Disentangling the Different Factors that Contribute to the Production of 3rd Person Singular Errors in Spanish

By Joseph D.S.B. Martin

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Department of Psychological Sciences

University of Liverpool

October 2019

UNIVERSITY OF LIVERPOOL

INSTITUTE OF LIFE AND HUMAN SCIENCES

PSYCHOLOGY (SCIENCE)

DECLARATION IN HIGHER DEGREE THESES

DECLARATION

This thesis is the result of my own work. The material contained in the thesis has not been presented, nor is currently being presented, either wholly or in part for any other degree or qualification.

Signed ..... (Candidate)

Signed .....31/10/2019.....(Candidate)

Acknowledgements

This thesis would not have been possible without the support of a number of people that I owe a great deal of gratitude to. Firstly, I would like to thank two supervisors, Dr. Colin Bannard and Professor Julian Pine. Their guidance and expertise over this process removed any anxiety I have had about my ability to complete this thesis. The ideas discussed in this thesis are the result of our long and thoughtful discussions. Colin, thank you for your patience during some very lengthy statistics sessions, I appreciate all your hard work over the past 4 years. Julian, thank you for always having an open door and for being willing to discuss anything that was on my mind on any given day, thank you again.

I would like to express my gratitude to the School of Psychology and the Department of Psychological Sciences for giving me this opportunity, employing me and funding my PhD. I am very grateful toward Maria Grimes, Karen Wayman, Pauline Hughes, Helen Hubbard, Sharon Clayton, Amy Hart, Elaine Fairhurst and Hayley Walsh who have looked after me over the past 4 years; their kindness and humour has made many stressful situations easier. I must also thank the people I have shared an office with over the past 4 years; Sam, Michelle, Sumitra, Jamie, Heather, Flo and Esmira. You have all helped me enormously.

I am extremely grateful to the schools, children who took part and the teachers welcomed me into their schools. Each school was very generous with their time and were always happy to help in any way they could. Special thanks must go to Mrs. Louise Young and Mr. Steve Lomas who were always tremendously helpful and did their utmost to make recruitment an easy process. I looked forward to every day I spent in your school. I must also thank Lindsay Bleazard, Nicola Yorke-Robinson and Fr. Brian Kealey who helped me to contact the schools. The final study would have been particularly difficult without your help.

Thanks to Stan Heaton and Keith Fairclough who are stalwarts of Lowerhouse Cricket Club. They have allowed me to complete my work in the West End Room and are always willing to discuss club affairs when I needed a break. I am very

appreciative of your kindness over the past 4 years. Thank you also, Chris Bleazard, Jonathan Finch, Francois Haasbroek, Ockert Erasmus, Cobus and Licinda Pienaar, Joe Beneduce, Charlie Cottam, Hayden Baxendale, Ainsley Brown and Nigel Brown. Each has always shown an interest in what I was doing and have always provided me with welcome distractions when I most needed them.

I must thank my family who have been with me on this process. To Paddy and Alex, who are always a source of fun in our house. I have been metaphorically absent over these past 4 years whilst I have been completing my PhD. I hope life will get back to normal soon. I must thank my Parents, Susan and Graham. Thank you, Dad, for always showing interest, stating the obvious, providing wonderful clichés, giving me support even after your heart attack and making me laugh when I needed to. Finally, Mum, my biggest supporter. You listened to my rants and help me reason through tasks I never thought I could. You never doubted me and always believed in me. For that I will be always be in your debt.

Thank you all.

Abstract

The objective of this thesis is to explain a type of error that is commonly observed in the speech of children in a number of languages – the tendency to frequently deploy a particular person-number form of a verb in contexts where another form is appropriate, a pattern that has been referred to as "defaulting". In particular we focus on the tendency of Spanish-learning children to "default" to the third person singular (3sg) form of verbs. For example, a child speaking Spanish frequently produces the 3sg form *-a* when the 3rd person plural (3pl) form *-an* is the target. We seek to understand what factors – e.g. the frequency with which the different forms occur, the way the different forms sound (phonology) or what the different forms mean (semantics) – cause children to produce this error. In particular we test an existing proposal that defaulting is the result of the relative frequency of the form in question.

We employ a novel training study approach designed to establish a causal link between linguistic experience and language development and disentangle different explanatory factors. This thesis begins with an overview of the previous literature, establishing the role that frequency, phonology and semantics play in morphological development. Chapter 3 investigates how much of an effect frequency has on the acquisition of present tense Spanish inflections by manipulating the input frequency in English-speaking monolingual adults (Experiment 1: Mean age - 19;4, Range - 18;3-25;1 – Experiment 2: Mean age – 19;6, Range – 18:3-24;8). Chapter 4 explores the role of phonology and semantics in the acquisition of the semantic 3sg form and the phonological *-a* form, by permuting forms across meanings using a sample of English-speaking monolingual adults (Mean age – 20;6 – Range – 18:4-35;2) . Chapter 5 explores the role of frequency in the acquisition of Spanish inflections using the same methodology as Chapter 3 but using an monolingual sample of English-speaking children (Experiment 4: Mean age – 9;2 - Range – 8;4-10;4 – Experiment 5: Mean age – 9;4 – Range – 8;4-10;8). Chapter 6 aims to improve on the methodology from Chapter 3 by introducing a physical "teacher" that the child can interact with. Again we tested monolingual English-speaking children (Experiment 6: Mean age – 9;4 –Range – 8:1-10:7 – Experiment 7: Mean age– 9;4 – Range – 8;1-10;1).

The results from these studies demonstrate that, while input frequency does affect the production of the 3sg form, its contribution differs from that previously proposed. We propose that the defaulting that has been found previously is at least partially a product of phonology and semantics. While input frequency does result in an overall increase in rate of 3sg production, it also contributes to increased overall performance (including on other forms) and in some cases, contrary to prior proposals, reduces rather than increases the rate of 3sg errors.

Contents

List of Figures

List of Tables

APPENDICIES

Abbreviations

1pl	First Person Plural
1sg	First Person Singular
2pl	Second Person Plural
2sg	Second Person Singular
3pl	Third Person Plural
3sg	Third Person Singular
ToM	Theory of Mind
ATOM	Agreement/Tense Omission Model
D-Feature	Determiner Feature
MOSAIC	Model of Syntax Acquisition in Children
OI	Optional Infinitive
RI	Root Infinitive
UG	Universal Grammar
VEKI	Very Early Knowledge of Inflection
VEPS	Very Early Parameter Setting
VLM	Variational Learning Model
UCC	Universal Checking Constraint

1 Introduction to the Thesis

1.1 Statement of the Problem

The objective of this thesis is to explain a type of error that is commonly observed in the speech of children in a number of languages – the tendency to frequently deploy a particular person-number form of a verb in contexts where another form is appropriate, a pattern that has been referred to as "defaulting". In particular we focus on the tendency of Spanish-learning children to "default" to the third person singular (3sg) form of verbs. For example, a child speaking Spanish will produce the 3sg form *-a* when the 3rd person plural (3pl) form *-an* is the target. We seek to understand what factors – e.g. the frequency with which the different forms occur, the way the different forms sound (phonology) or what the different forms mean (semantics) – cause children to produce this error.

1.2 Causes of 3rd Person Singular Errors

Wexler and colleagues (Harris & Wexler, 1996; Wexler, 1998) suggest that children make very few errors early in their language production and that they display an adult-like level of competence. This is demonstrated by children displaying low overall error rates (around 4% of productions) when producing inflected forms. However, Rubino and Pine (1998) and Aguado-Orea and Pine (2015) show that overall error rates are misleading, as they do not take into account the frequency with which the different inflections are heard in the input and the frequency with which they are produced in a child's output. They claim that the large number of high frequency forms that are produced correctly masks the high rate of low frequency forms that are produced incorrectly. Thus, when the error rates across each of the different inflections are isolated, some are produced much more successfully than others. For example, Aguado-Orea and Pine show that the error rate for the 3sg form

is 0.7% (for both of the children in their study) but the error rate for the 3pl form is over 30% for both children (Juan = 33.9%, Lucia = 46.4%). Critically, a large proportion of these errors involve the incorrect use of the 3sg.

Aguado-Orea and Pine (2015) argue that the high level of accuracy with which the 3sg form is produced is directly related to the frequency with which this form is heard in the input. They demonstrate that around 50% of the utterances heard in a child's input involve the use of the 3sg form. This input frequency is reflected in the child's output, with children producing this form more readily than the other available forms. The bias towards the 3sg form in the input, it is argued, pushes children to "default" to the 3sg form, due to it being the most frequent form, with children producing the 3sg form regardless of the target.

Ambridge, Kidd, Rowland and Theakston (2015) propose that frequency both protects and hinders language development. Children are able to successfully produce high frequency forms, such as the 3sg form, but frequency also hinders development, owing to children producing high frequency forms in place of low frequency forms when they are unsure of which inflection to use. However, frequency alone does not appear to completely explain differences in rates of correct production in a child's language development. Aguado-Orea and Pine note that frequency effects for different inflections are not always consistent. For example, the 2nd person singular (2sg) form is more frequent in a child's input than the 1st person singular (1sg) form, but Juan and Lucia made more errors in 2sg contexts despite the input frequency being lower for this form than for the 1sg form. Austin (2012) found

similar results, where the 2sg form was more frequent in the input, but 1sg and 2nd person plural (2pl) forms emerged in children's language before the 2sg form.

The frequency effects discussed above give us a plausible explanation of why children so readily "default" to the 3sg form and produce 3sg errors. However, frequency is not the only candidate explanation. Schwartz and Leonard (1982) propose that children select and omit certain verb forms as a function of their phonological properties. They suggest that children are more likely to produce phonologically simple forms rather than complex ones. Increasing the phonological complexity of an inflection increases the cognitive load for a child (Marshall & van der Lely, 2007). Barlow and Pruit-Lord (2014) propose that children will strive to produce the simplest form possible, in order to not overload their cognitive system. A child must be capable of producing the desired form, which means a child's early production is also determined at least in part by the child's phonological/articulatory capacity (Vihman & Croft, 2007). Teasing apart the effects of frequency and phonology is often difficult, since as Diessel (2007) points out, high frequency forms are often more phonologically simple (-*a*) than low frequency forms which are phonologically complex (-*amos).*

A third alternative is that it is meaning that drives the pattern of errors - there may be something special about the meaning of the 3sg form that increases the likelihood of that form being produced. A potential reason for the preference for the 3sg form, is that this form is semantically isolated when compared to the 1st and 2nd person forms. 1st and 2nd person forms require the child to have an understanding of perspective - which form is appropriate depends on who is speaking – them or their interlocutor. If

a speaker is describing an action that they alone are performing then a first person singular form is required, but if the other conversational participant is describing the same action then the second person form is required. By contrast, the 3sg form is the same regardless of who in the dyad is speaking – describing an action performed by someone other than the speaker or their interlocutor always requires the third person.

These three separate factors (i.e. frequency, phonology and semantics) are potentially important in both the successful and erroneous production of the 3sg form. However, these factors are intertwined within natural language, and so are difficult to separate. Our goal in this work was to develop a methodology that allowed us to isolate these different factors, but kept the properties of naturalistic Spanish, by maintaining the proportional frequency of different inflected forms in the input.

1.3 Experimental Methodology Used in the Thesis

A core aim of this thesis is to establish how much of an effect input frequency has on the successful and erroneous production of the 3sg form within the Spanish *-ar* conjugation. In order to examine this issue, participants (both adults and children) were randomly assigned to one of two conditions that directly manipulated the input frequency of the inflections that were taught. They were taught a series of verbs and their inflections over a period of 3 training days. At the end of each day the participant was then tested on their ability to produce the different verbs and inflections they had been taught.

The first condition, named the "skewed" condition, maintained the proportional frequency seen in naturalistic Spanish, where there is a bias toward the 3sg form (e.g. as seen in the work of Aguado-Orea, 2004 and Aguado-Orea & Pine, 2015). Other participants were assigned to the "uniform" condition. The proportional frequency of the taught inflections within the "uniform" condition were equal, where there was no bias towards any inflection. By comparing performance in these two conditions, we could therefore examine how much of an effect frequency has on production. This allowed us to ask the following questions: How much of an effect does frequency have on successful and erroneous production? And to what extent do other factors affect successful and erroneous production? We explore these questions in both children and adults.

There are other factors that may also contribute to the production of 3sg errors, notably phonology and semantics. In order to assess how much of an effect these factors have on production, we devised an experiment that disentangled phonology and semantics. We created a permuted inflectional paradigm that was unique to each participant. This allowed us to ask the following questions: How much of an effect does phonology have on successful and erroneous production? And how much of an effect does semantics have on successful and erroneous production? These questions were explored in the adult participants.

1.4 Structure of the Thesis

1.4.1 *Part One – Chapter 2 - Background Literature*

The thesis begins with an overview of the previous literature. We examine the arguments surrounding the production of errors, initially through the Optional

Infinitive (OI) hypothesis (Wexler, 1998), which suggests that children display adult-like competence in terms of use of inflections. We review alternative accounts, and consider the role of relative frequency is determining learning. We then explore literatures concerning the role of phonological and semantic factors in morphological development.

1.5 Part Two – Experimental Testing

1.5.1 *Chapter 3 - Assessing the Effect of Skewed Frequency Distributions on the Learning of Spanish Person and Number Morphology.*

Chapter 3 reports on experiments exploring the role of skewed frequency distributions in the learning of Spanish person and number morphology by English-speaking adults. In Experiment 1, participants were assigned to one of two conditions. The first was the skewed condition, where the input frequency was biased toward the 3sg form. The second was the uniform condition, where the input frequency was equal amongst the different inflections. This allowed us to look at whether, as has been claimed, it is relative frequency that drives 3sg errors. Experiment 2 compares the skewed condition to the uniform condition again, but to address a possible methodological confound, a change was made to the relative frequency of different trials.

1.5.2 *Chapter 4 - Assessing the Contribution of Phonology and Semantics Factors in Spanish Learners Successful and Erroneous Production.*

Chapter 4 aims to test how much of an effect phonology and semantics have on production, by isolating these two factors. Experiment 3 achieved this by creating a permuted inflectional paradigm for each participant who took part in the experiment. For example, in typical Spanish, the 3sg form equates to the -*a* form phonologically. However in the Randomisation condition, for one participant the 3sg form might equate to -*amos* phonologically and for another participant the 3sg form might equate to -*an* phonologically. By decoupling phonology and meaning in this way we are able to look separately at their effects.

1.5.3 *Chapter 5 - Assessing the role of a Skewed Frequency Distribution in the Learning of Spanish Person and Number forms in Children.*

Within Chapter 5, experiments 4 and 5 aim to use the same design described in Chapter 3 but in a child rather than an adult sample.

1.5.4 *Chapter 6 - Assessing the Role of a Skewed Frequency Distribution on the production of Spanish Person and Number combinations with the use of a new Methodology.*

In Chapter 6, we aim to improve on the methodology utilised in Chapter 5 by attempting to make the person and number features used to learn the different

inflections more salient for the children involved in the study. We did this by including a physical "teacher" for the children to interact with.

1.5.5 *Part Three – Chapter 7 – General Discussion*

Chapter 7 synthesises the information found within the previous experiments and the previous literature, and uses them to provide a unified account of children's 3sg errors.

Part One

2 Background Literature

This chapter reviews the existing literatures relevant to our understanding of children's 3sg errors and the different explanations of children's "defaulting" behaviour. The first part of this chapter provides a description of the Spanish morphological paradigm which is used in our experiments. We then provide an overview of the theoretical frameworks that have been applied to understanding our target phenomenon. We explain the Optional Infinitive stage of development and consider alternate explanations of why children produce Optional Infinitive errors (e.g. MOSAIC and the Variational Learning Model). The next section addresses the issue of error rates in children's early productions, where nativist theories suggest that the low overall error rates found show that children have an adult-like understanding of morphology. Alternatively, Constructivist theorists propose that these low overall error rates are misleading, due to fluctuating error rates for individual inflections. Finally, we discuss the different factors (i.e. frequency, phonology and semantics) that might cause children's "defaulting" behaviour.

2.1 A Brief Overview of Present Tense Spanish Verb Morphology

Inflectional morphology is the variation of the forms of words in ways that indicate grammatical features. An example is the marking of verbs for different "person" and "number" features. In order for children to be successful language users, they must learn to produce and interpret the different inflections of their target language.

Spanish is a morphologically rich language, where an adult speaker can produce verbs with over 40 possible different affixes. Spanish is also a "pro-drop" language where subjects are not overtly marked in language, thus allowing the subject to be omitted (Mann, 2012; Bedore & Leonard, 2001). While one is not necessarily a pre-

requisite for the other, as shown in languages such as Chinese which are pro-drop but not morphologically rich (Mann, 2012), in the case of Spanish, morphological richness and pro-drop go hand in hand.

The Spanish present tense has three verb conjugation classes that use 3 thematic vowels: [a], [e] and [i]. Each verb produced requires a suffix, as a stem cannot be produced as a free morpheme and therefore needs to be marked with an inflection. Verbs must agree with the subject in person and number. Table 2.1 outlines the different present tense person and number combinations a child can produce. The examples used in the Table are the Spanish verbs "bailar", "comer" and "vivir" which mean "to dance", "to eat" and "to live" respectively.

Table 2.1 – 3 conjugation classes displaying the different person and number combinations on Spanish present tense verbs.

	[-ar] – "bailar" – To Dance	*[-er]* – "comer" – To Eat	*[ir]* – "vivir" – To Live
1st Person Sing.	bailo (I dance)	como (I eat)	vivo (I live)
2nd Person Sing.	bailas (You dance)	comes (You eat)	vives (You live)
3rd Person Sing.	baila (He/She dances)	come (He/She eats)	vive (He/She lives)
1st Person Plural	bailamos (We dance)	comemos (We eat)	vivimos (We live)
2nd Person Plural	bailáis (You dance)	comèis (You eat)	vivís (You live)
3rd Person Plural	bailan (They dance)	comen (They eat)	viven (They live)

In the present tense, speakers have 6 different inflections to learn and they must understand different person and number combinations in order to be fully productive

and generate a correctly inflected form. This thesis will focus on the *"-ar"* conjugation class which has the highest type frequency (over 90% of the verbs belong to this class; Aguirre, 2003) and has regular stems and inflectional suffixes (Clahsen, Aveledo & Roca, 2002).

A number of researchers (e.g. Wexler, 1998; Hoeskra & Hyams, 1998) have claimed that despite the apparent cognitive complexity of acquiring an inflectional paradigm, children produce very few errors when doing so (at least in terms of the overall error rate). They argue that this low overall error rate indicates that children have adult-like understanding of inflection. However, Rubino and Pine (1998) have argued that this evidence is somewhat misleading, and a more fine-grained analysis is needed in order to show the true nature of young children's knowledge. Before exploring these arguments further, I will introduce the different theoretical frameworks that underlie them.

2.2 Theoretical Frameworks

There are two broad frameworks within which language researchers work – generativist and constructivist. I will describe each of these frameworks in turn.

Generativists suggest that a speaker's knowledge of grammar is built on a set of formal rules that operate on words (e.g. Nouns, Verbs, Determiners) or phrases (e.g. Noun Phrase or Verb Phrase) and these rules operate as categories (Guasti, 2002). For example, a morphological rule in English is adding the *-ed* past tense morpheme to verbs such as "push" to create the form "pushed". Other generative rules operate on word order – for example Phrase Structure Rules (Chomsky, 1957). These rules

allow a child to determine that (e.g. in English) the subject of a sentence is placed before the verb (e.g. “He is dancing” rather than “*Is dancing he”).

The challenge that language acquisition presents to generativist theory is explaining how children come to have this knowledge. One common solution is to assume that it is at least partially innate[1]. The nativist approach proposes that language development is biologically predetermined. The innate knowledge that children are claimed to possess is often referred to as Universal Grammar. The Universal Grammar (UG) approach proposes that all natural languages share a set of common structural properties e.g. categories (Verb, Noun), constraints and parameters (e.g. head direction) that are innate (Ambridge, Pine & Lieven, 2014), allowing the child to acquire languages despite a lack of rich input (Cowie, 1997; Chomsky, 1975, 1986). Chomsky (2005) and Yang (2004) suggests that, whilst it is possible to learn probabilistically from one’s linguistic environment, it would not be possible without having innate knowledge of both syntactic and phonological structure. Arguments for the UG approach point to a “Poverty of the Stimulus”, claiming that the language input children are exposed to is simply not sufficient to learn effectively, such that there must exist innate language universals (Kliesch, 2012; Dabrowska, 2015).

Wexler (1994, 1998) is a proponent of the nativist approach. He asserts that claims that children gradually learn the morphological properties of their language are fundamentally wrong. He proposes two hypotheses that outline the innate knowledge that children have before they begin to produce language. The first is “Very Early

[1] It is important to note (Ambridge & Lieven, 2011) that it is possible to be a Generativist without being a Nativist (where the rules, categories and phrases are acquired as the result of learning process rather than being innate) and a Nativist without being a Generativist (where innate knowledge refers to something other than grammar).

Parameter Settings" (VEPS) and the second is "Very Early Knowledge of Inflection" (VEKI). These two hypotheses are intertwined and can be difficult to tease apart.

VEPS refers to the basic parameters that are set from the earliest observable stage of language development. Wexler argues that children have discovered the functional properties of language before the two-word stage of language development, which enables them to apply these properties from the earliest observable stage. For example, the parameter for verb movement (i.e. head movement) is set before a child can combine verbs and arguments together.

VEKI refers to the child's early knowledge of inflection and is closely aligned to VEPS (Wexler, 1998). From the earliest observable stage, the child possesses and understands both the grammatical and phonological properties of varying inflectional elements. Poeppel and Wexler (1993) argue that young German-speaking children rarely make agreement errors (i.e. use verb forms that do not agree with the subject of the utterance). In a similar vein, Harris and Wexler (1996) show that English-speaking children rarely produce a 3rd person singular form when a 1st person form is required. The example they use is that children simply do not say "*I likes ice-cream" where children add "-*s*". However, "-*s*" occurs frequently when the subject is third singular (e.g. "She likes ice-cream"). Wexler (1998) makes the informal claim that children are "little inflection machines" suggesting that children have the ability to produce inflections almost instantaneously.

To summarise, Wexler proposes that children set the parameters of their language very early in development, have early knowledge of inflection and understand the

relationship between these two phenomena (i.e. VEPS and VEKI), where the relationship between VEPS and VEKI demonstrates that children have an adult-like understanding of inflection.

The second major theoretical approach to the development of grammar is Usage-based or Constructivist theory. This proposes that learning inflection is a gradual, piecemeal process of learning from experience. Tomasello (1992a; 1996; 2003) suggests that children are partially productive as opposed to fully competent from the first observable stage, opposing Wexler's (1998) view of children's inflectional production. Tomasello (2003) suggests that the development of inflection is a gradual process that begins with children rote learning utterances (called "Frozen Phrases" or "Holophrases" such as "She's kicking it", "She's pushing it") within cultural routines (e.g. learning "I'm eating it" at mealtime with a caregiver). The child then develops inflectional knowledge by generalising over the forms that have been rote learned.

2.3 The Optional Infinitive Stage

Wexler (1998) claims that evidence for both the VEPS and VEKI phenomena comes from the existence of an Optional Infinitive (OI) stage early in language development (Wexler, 1990, 1992, 1994). This stage is explained as a period in the child's development during which they use both the finite form (e.g. a verb that has an expressed subject – and can show person and number marking) and the infinitive form (a basic form of the verb which has no binding inflection) when an obligatory finite form is required. During this stage, children fail to understand that tense is obligatory in finite clauses. The child produces root infinitives in what they believe

is a grammatical sentence. For example, in English, the child might say “That go in there” rather than “That goes in there”. Despite children using these forms interchangeably and therefore producing errors, Wexler argues that children have all the necessary grammatical properties available such as verb movement and the ability to check for grammaticality.

According to Wexler’s (1998) model, during the OI stage of development, although all the parameters of the child’s language have been set, the child’s grammar also includes a “Unique Checking Constraint” which makes the child susceptible to certain types of grammatical errors. The UCC is a grammatical constraint which the child sometimes satisfies and sometimes violates in order to satisfy another competing constraint. When the child satisfies the UCC, only one D-feature is checked and OI errors are produced. In contrast, when the child violates the UCC both D-features are checked, and correct finite forms are produced. The child therefore uses finite and non-finite forms interchangeably until the UCC disappears from their grammar. Wexler (1998) argues that the UCC “withers away” as a result of maturation and that the child’s grammar becomes fully adult-like by the age of four years (Blom & Wijnen, 2013).

2.4 Model of Syntax Acquisition in Children – MOSAIC

An alternative explanation of the OI stage, comes from Freudenthal and colleagues (2005, 2006, 2007, 2009, 2010, 2015), who use a computational model (MOSAIC – Model of Syntax Acquisition in Children) to explain why OI errors are produced in languages such as English and German. The MOSAIC model (developed from CHREST; Gobet, 1998) uses corpus data from parent and child interactions to show

that OI errors can be explained in terms of input-driven learning. OI errors are simulated as 'truncated compound finites', which are auxiliary and infinitive constructions with a missing modal or auxiliary (Freudenthal, Pine, Gobet, 2006). OI errors are a product of children learning truncated verb forms from compound finite structures in the input. This process reflects the restrictions a child has on their language learning. MOSAIC has no in-built knowledge of morphology or syntax. Instead, the model rote learns strings of words from the input and thereby develops incrementally. The model learns from the left and right edge of the input. For example, if a child hears the phrase "Chris wants to play cricket", they will initially learn from the left and right edge of the phrase to produce "Chris play cricket" – an OI error. These errors will cease as the model's learning capabilities increase as a function of input exposure.

The input to MOSAIC is created by taking the child-directed speech from transcribed naturalistic recordings of caregiver-child interactions. The child-directed speech is then fed through the model several times. Output is generated after each round of exposure. MOSAIC learns in a slow, probabilistic manner, where the length of utterance is dependent upon the amount of input to which the model has been exposed. MOSAIC simulates OI errors by learning them from compound finites on the far right and left of the input (for example, "He can go" is learned as "he go"). The model will therefore produce truncated (shortened) compound finite utterances.

MOSAIC simulates development as it progressively learns to produce longer utterances by increasing the amount of input the model is exposed to. As the input increases the number of OI errors produced decreases over time. Freudenthal et al.

(2010) explain that this process reflects the pattern of compound finites in various languages such as German and Dutch, where finite modals and auxiliaries appear before the infinitive form. Early utterances learned in MOSAIC are learned from the right edge, leading to infinitives being produced as single word utterances. As exposure and input increases, finite modals and auxiliaries begin to appear as more of the input is retained and OI errors begin to be replaced by compound finites.

MOSAIC is very good at simulating the developmental pattern of OI errors across various languages. Freudenthal, Pine and Gobet (2006) showed that MOSAIC could simulate developmental pattern of OI errors across two OI languages; English and Dutch. Freudenthal, Pine, Aguado-Orea and Gobet (2007) developed the model further by simulating the developmental pattern in 2 additional languages (an OI language: German and a Non-OI language: Spanish).

Freudenthal et al. (2007) successfully simulate the low rate of OI errors in Spanish when compared to the other languages input to the model (i.e.. English, German and Dutch). Freudenthal et al. show that, despite Spanish having the same level of compound finites as German and Dutch, MOSAIC still produces the low level of OI errors found in naturalistic speech. These results are found due to the rate of non-finite forms appearing in utterance final position being very low (26%) when compared to other languages such as German (65%) and Dutch (85%), thus explaining the absence of a pronounced OI stage in Spanish.

It also must be acknowledged that, while MOSAIC does successfully simulate difrerences in the rate of OI errors across several languages (Dutch, German, French

and Spanish), the model does not simulate the very high rate of OI errors found in English. The rate of OI errors in English is around 85% during the early stages, but the rate at which English OI errors are produced in MOSAIC is closer to 60%. Freudenthal et al. (2010) suggest that these results are a product of two distinct processes: one which produces infinitive forms and one that produces bare stem errors. They propose that OI errors can reflect the learning of non-finite forms from compound finite structures, but OI errors in simple finite contexts can be explained by the process of defaulting to the most frequent or phonologically simple verb form (see Rasanen, Ambridge and Pine (2014) for an elicitation study that provides empirical support for this view).

2.5 The Variational Learning Model

An alternative explanation of OI errors, this time in a generativist tradition, is the Variational Learning approach (Yang, 2002, 2004; Legate & Yang, 2007). Legate and Yang's (2007) Variational Learning Model (VLM) suggests that each child has a range of possible grammars (or parameters) that are associated with probabilities acquired from the input. These probabilities can change depending on the child's linguistic environment. Legate and Yang propose that the period of time a child spends in the OI stage is directly proportional to the level of overt tense and agreement marking in that language. For example, children learning to speak English (which has an impoverished inflectional paradigm) will spend more time in the OI stage than children learning to speak Spanish as a result of the high degree of overt tense and agreement marking in Spanish.

Legate and Yang suggest that when a particular grammar is used, it is rewarded by utterances that are consistent with that particular grammar. Alternatively, the grammar can be "punished" if the utterance is not consistent with the grammar (or parameter) – for example, when a +Tense grammar is used to understand an utterance with no overt tense marking. The VLM integrates a probabilistic learning mechanism into the model, where the child initially has access to both the positive and negative values of the parameter, until the environment sets the parameter to a particular value.

For the VLM approach, OI errors reflect the fact that the child's grammar is not set immediately in response to a single utterance. Instead, parameters are set gradually and grammars are eventually abandoned when it has become clear that they are inconsistent with the input, gradually reducing the number of OI errors produced. The rate of reduction in OI errors is dependent on the amount of morphological evidence in the input. For example, Legate and Yang use the examples of Spanish, French and English to explain the varying rates of OI errors across languages (English has a high rate of OI error, French a moderately high rate and Spanish a low rate). Legate and Yang show that Spanish input is more likely to reward the (+*tense)* grammar than French and even more so than English, thus explaining the differences in the OI rates across the different languages.

As previously discussed, different languages have different rates of OI errors. For example, languages such as English, Dutch and German have much higher rates of OI error than languages such as Spanish or Italian. Wexler (1998) argues that since the only difference between child and adult grammars is that children's grammars

include a Unique Checking Constraint, children learning language like Spanish and Italian in which the UCC has little effect, should make very few errors of any kind in their use of verb inflection.

It is clear that explaining why errors occur in children's production of different inflections is important for understanding production. Generativists suggest that, OI errors aside, children make few errors when producing different inflectional forms. However, this assumption is disputed by constructivist theorists. Constructivists suggest that using overall error rates to explain production is problematic. The next section will explain this argument in more detail.

An important feature of the OI hypothesis, is that children learning null-subject languages (like Spanish and Italian) do not go through the OI stage. Therefore, in some Romance languages (including Spanish and Italian, but excluding French) Wexler suggests that there will be no OI errors due to these languages being pro-drop or null-subject languages, where there is no overt subject and hence no requirement to check agreement (For example, the *-a* in the Spanish verb "*baila*" carries the semantic information that "he/she is performing an action, whilst 1sg *"-o"* means "I am performing an action") and therefore does not need to be checked against the subject of ther utterance. This means that errors should not be made on finite verb forms and consequently children should not produce either incorrect finite inflections or incorrect bare infinitives.

2.6 What Errors do Children Make?

Wexler (1998) claims that in certain languages (such as Spanish and Italian) children's use of verb inflection is error-free and adult-like from its first emergence. This, it is claimed, is demonstrated by children having surprisingly low overall error rates in their use of inflected forms. For example, Hoeskra and Hyams (1998) found that in various languages (including Spanish and Italian), overall error rates were consistently low at around 4%, suggesting that children do have an adult-like inflectional system.

Wexler proposes that these low overall error rates are evidence that children have Very Early Knowledge of Inflection (VEKI). At first glance, this claim might appear uncontroversial. However, the use of overall error rates can be misleading. High-frequency forms are usually produced correctly. Forms that are infrequent in the input are less likely to be produced in children's speech and are more likely to be produced incorrectly if or when they are produced. Consequently, low overall error rates can hide pockets of high error on low frequency inflections and verbs and result in a simplistic view of children's inflectional usage.

That the use of low overall error rates to show competence might be misleading was pointed out by Rubino and Pine (1998) who used Brazilian Portuguese (a language which has a similar inflectional system to Spanish) to show that using overall error rates as a measure of competence is problematic. In Rubino and Pine's study, overall error rates were low (at around 4%). However, by differentiating between the error rates for verbs in different person-number contexts they discovered that the error rates varied substantially depending on the target. For example, when producing 3rd

person plural forms (3pl) the child's error rate was 43%. This was high compared to the other forms produced. For example, the error rate for 1st person singular and 1st person plural forms were 8.4% and 23.5% respectively. The error rate for the highest frequency 3sg form was only 0.5%.

Aguado-Orea and Pine (2015) found similar results to Rubino and Pine in a study that assessed the production of different verb inflections in children's Spanish. Two children took part in the study (Juan and Lucia) and again, by-inflection error rates varied for both children, despite their having low overall error rates (Juan's overall error rate was 4.6% and Lucia's overall error rate was 3.9%). For both children involved in the study, the error rate for the 3sg form was low at 0.7%. However, the error rate for the 3pl form was much higher at 33.9% (Juan) and 46.4% (Lucia). This discrepancy creates an issue for the full competence model in view of children appearing to have varying levels of competence for different forms.

If constructivist claims that early language use involves the rote learning of phrases are correct, then we expect that inflected forms would only be used in a fixed manner at the start of development (e.g. "I'm kicking it" or "I'm pushing it"). For example, children may simply produce one or two fixed slot and frame constructions early on, prior to their becoming productive and the child's inflectional paradigm developing. Veneziano and Parisse (2010) propose that, for children's early production of French, this appears to be the case. They report that for early inflectional productions, children consistently produce verbs in one inflection or construction. This result has been found for a variety of languages, including Spanish (Gathercoe, Sebastian & Soto, 1999), Italian (Pizzuto & Casselli, 1992) and

English (Tomasello, 1992a). These findings suggest that children do not have an adult-like use of inflection like Wexler suggests.

Supporting these claims, Aguado-Orea (2004) provides an example of Spanish-speaking children's use of frozen forms. In his study, the 1sg form had a low error rate (4.9% for Juan and 3% for Lucia). However, when this inflection was produced, it was produced most frequently using two verbs – "*quiero*" ("I want") and "*puedo*" ("I can"). This suggests that for certain inflections, children appear to learn whole forms and frequent verbs as opposed to specific morphological properties. For example, when using the 1sg form, the error rate for the highest frequency verbs was low at 1.6%. However, for less frequent verbs using the 1sg inflection, the error rate rose to 23.4%.

It would be logical to assume that if children really did have an adult-like understanding of inflection, they again should be able to produce all of the possible verb inflections that an adult can. The fact that they do not suggests that their knowledge is incomplete. A great deal of other research into the order of acquisition of individual verb inflections outlines a difference in onset for the production of individual inflectional forms. Ezeizabarrena (1996) studied 2 children learning Basque and Spanish. Ezeizabarrena reports that in both Basque and Spanish, 3sg verbs are produced earlier than other inflections. The proposal that inflections appear early rather than late in development is fundamental to Wexler's arguments and the differences in use of particular verb forms suggests that children learn different forms at different times rather than having full competence at an early stage of development.

Pizzuto and Casselli (1992) suggest that while certain inflections do appear early in development (such as the 3sg form), the full, complex inflectional paradigm emerges late in development and its emergence is a slow process. They propose that only the 3sg form can actually be produced early in development and produced correctly. This provides evidence that, whilst some inflectional forms are available early in development, not all inflectional forms are productive from an early stage.

Children learning certain languages, such as Spanish, do not go through a pronounced OI stage. According to Wexler, this is because the unique checking constraint does not have the same effects in null subject and obligatory subject languages.

2.7 Optional Infinitive Errors in Pro-Drop Languages

An alternative view of children's knowledge of verb inflection suggests a mirroring of the OI stage in null subject languages such that young children use a form of the verb other than the infinitive as a Root Infinitive (RI) analogue. Salustri and Hyams (2003) argue for a stage in early Italian in which children use the imperative rather than the infinitive as an RI analogue. They argue that RI analogues can be identified on the basis of the following characterisitics. First, the RI analogue will occur significantly more in child speech than in adult speech and second, it will occur significantly more in pro-drop languages than in OI languages.

Salustri and Hyams (2003) compare German (an OI language) and Italian (a pro-drop language). In Italian, imperatives are used early in language, with the first 4

verbs produced being imperatives, and these forms are overused at a similar stage to the OI stage of development. In German the frequency of the imperative form in the input is similar to Italian (around 36%). However, in German, the use of imperatives by the child is infrequent when compared to Italian. In a similar vein, Grinstead, De La Mora, Vega-Mendoza and Flores (2009) suggest that in Spanish, the 3sg present tense form is an RI analogue, thus potentially explaining the overuse of 3sg forms in Spanish.

Despite RI analogues being an elegant addition to the OI hypothesis, the theory does encounter some problems. Firstly, Tatsumi and Pine (2016) suggest that RI analogue theorists assume "inflectional imperialism" (Slobin, 1973,1985), where children will use a single affix for each stem, thus producing errors that involve defaulting to a single dominant pattern. Tatsumi and Pine highlight that, despite this pattern being shown in languages such as Spanish and Italian, it is not the case for all languages. Languages such as Turkish and Hungarian do not follow such a default pattern of use. Secondly, the RI analogue approach does not extend far enough to explain what RIs can account for and what input frequency can account for. Distinguishing between the two approaches (RIA approach and input frequency, of which the latter will be discussed in the next section) is difficult since RI analogues tend to be high frequency forms. For example, Grinstead et al. (2009) propose that in Spanish the 3sg form should be treated as an RI analogue. However, the 3sg form also happens to be the most frequent form in both the child's input and in the child's speech. The situation is similar in Italian, where Salustri and Hyams (2003) state that the imperative is an RI analogue but the imperative is also a high frequency form which is homophonous with the 3sg present tense form. It is therefore difficult to ascertain

what is actually causing the error – the fact that the form has some special linguistic status for the child or the fact that it is the highest frequency form in the input.

2.8 The Effect of Frequency on Morphological Acquisition

An important implication of Wexler's (1994, 1998) approach is that the errors produced by children are not a result of frequency effects in child directed speech. Wexler suggests that frequency is merely a triggering mechanism for other elements in language to develop rather than being fundamental to the learning process. Early work by Brown (1973) found no correlation between frequency and age of acquisition, suggesting that frequency has no role in the development of a child's inflectional system (see also De Villiers, 1985; Newport, Gleitmann & Gleitmann, 1977 for similar claims). Despite these early claims, more recent evidence suggests that frequency is an important factor (Rasenan, Pine & Ambridge, 2014; Aguado-Orea and Pine, 2015).

Ambridge, Kidd, Rowland and Theakston (2015) suggest that frequency plays an important role in the production of both correct and incorrect forms. Frequency appears to protect children from error and result in particular kinds of errors, by supporting the correct production of highly frequent forms learned from the input (for example, the high frequency of 3sg forms in Spanish input appears to result in virtually error-free performance in 3sg contexts on the part of the child). Ambridge et al. call this the "Prevent Error Thesis". At the same time, high frequency forms learned from the input interfere with the learning of low frequency forms and result in errors in which these forms are used incorrectly when a low frequency form is the

target. Ambridge et al. call this the "Cause Error Thesis" (for example, using the 3sg form in place of the 3pl form).

The two theses explained above are integral to the acquisition of different inflectional forms across various languages. "Defaulting" is the process where children consistently use a particular form (e.g. the most frequent or prototypical form) in contexts in which other forms are required. For example, in Spanish, the most frequent form is the 3sg present tense form *"-a"*. This form is often produced when another form is required (e.g. when children produce the 3sg form, ending in *"-a",* instead of the correct 3pl present tense form, ending in *'-an'*). As a result of "defaulting", the error rate for the "defaulting" form (i.e. the most frequent form) will be low owing to both learning when to use the form and using the form when the child is unsure of which inflection to use. However, the error rates for the other forms will be high because of the use of the "default" form where those forms are required, at least until the child has adult-like knowledge of the low frequency forms.

Some forms are more frequent in the input that children hear than others (e.g. the 3sg form *"-a")*. As discussed previously (Aguado-Orea & Pine, 2015), the error rate for the 3sg form is low (0.7% for both of the children in the study). However, the error rates varied across the other forms. An interesting finding from this study was that for the different errors produced, 80% of the time, the errors involved using the 3sg form (83.6% for Juan and 92% for Lucia) in a non-3sg context (for example, using the 3sg form *-a,* "**Baila*" when the 3rd person plural form (3pl) *-an* was required, "*Bailan*"). Leonard, Caselli and Devescovi (2002) show similar results in Italian where children show high accuracy levels with the 3sg form and yet struggle with

3pl forms. They also note that the 3sg form is often used when a 3pl form is required.

Comparable results were found by Räsänen, Ambridge and Pine (2016) and Tatsumi, Ambridge and Pine (2017). Räsänen et al demonstrated that when using an elicitation production paradigm, children in Finnish are more likely to produce errors when a low frequency form is required. The rate of error for non-3sg inflected forms ranged from 10% - 32%, significantly higher than the 3sg form error rate (0.46%). These trends are consistent across other languages such as English and Spanish. The results also show that Finnish children are more likely to replace low frequency tense/agreement forms with high frequency tense/agreement forms when a low frequency form is required.

Tatsumi, Ambridge and Pine (2018) found results akin to those found by Räsänen et al. Tatsumi et al. used an elicitation paradigm to test 30 Japanese-speaking children on their ability to produce both simple and complex verb forms. They found that the children were more likely to use a high frequency form when a low frequency form was required, even after controlling for morphological complexity. Again, this demonstrates that children are more likely to produce forms that are frequent in the input as opposed to low frequency forms. These results therefore suggest that children are "defaulting" to the form with highest frequency.

Räsänan, Ambridge and Pine (2014) studied 'defaulting' in mono-lingual English-speaking children. In English, the most frequent form produced is usually the bare form, which is indistinguishable from the infinitive form. Rasanen et al. used an

elicited production paradigm to assess whether children's errors were a result of "defaulting" to the most frequent form, therefore potentially explaining why English-speaking children appear to have a particularly high rate of OI errors. The results from this study showed the proportion of bare stem forms vs 3sg forms in the input negatively predicted the rate of correct 3sg production vs OI errors.

It appears then that input frequency is a key factor in the production of different inflected forms. Similar results have been found for Polish noun marking. Dabrowska and Szczerbinski (2006) found that children aged 2;7 were highly productive with high frequency forms learned from the input and produced few errors. Alternatively, children as old as 4;5 demonstrated poor performance on low frequency forms learned from the input, again highlighting the importance of input frequency.

While there is much evidence that children "default" to the most frequent form in the input, it is worth noting that frequency cannot explain all aspects of children's competence with or selection of different forms. Austin (2012) found that children learning Spanish produced the 3sg form first and more frequently than other forms, mirroring the high frequency with which this form is produced in the input. However, the order of emergence of the different inflectional forms does not match their input frequency. For example, in Spanish the 2sg form is the second most frequent form produced by adult Spanish speakers. However, the 1sg form emerges earlier in Spanish children's speech.

A related trend is reported in the same paper for Basque, where for ergative subjects, 2sg forms are produced most frequently in the input, but children produce 2sg forms less frequently than both 3sg forms and 1sg forms. Further to Austin's study, Aguado-Orea and Pine (2015) demonstrated that the pattern of errors produced by children learning Spanish (Juan and Lucia), again did not completely match the input frequency. The two children involved in this study produced a higher proportion of errors in 2sg contexts compared to 1sg contexts, despite the 1sg forms being less common in their caregiver's speech.

Aguado-Orea (2004) provides a more in-depth analysis of Juan and Lucia's speech production. The results from this study clearly highlight the role of frequency in the successful production of the 3sg form (the highest frequency form) and how defaulting to this form causes error through the production of the 3sg form in the wrong context (e.g. Juan and Lucia produced a high amount of errors where they used the 3sg form when the 3pl form was the target form).

Kueser, Leonard and Deevy (2018) conducted a study assessing both children with Developmental Language Disorder (DLD) and typically developing children's 3sg production from an American-English corpus of child-direct speech. The high proportion of 3sg forms produced in the input were associated with 3sg use in children with DLD and typically developing children. As expected, Kueser et al. found that, while children with DLD produced the 3sg form less frequently than typically developing children, they also showed that frequency adjusted input increased the 3sg production for children with DLD, suggesting that children with

DLD are sensitive to the statistical properties of the input when producing certain forms.

As stated previously, the 3sg form in Spanish (and other languages) appears to dominate both children's input and their own early productions (see Aguado-Orea, 2004; Aguado-Orea and Pine, 2015). We established that frequency plays a prominent role in the acquisition of different inflectional forms. However, input frequency cannot explain the complete pattern of both correct production and errors. It is important to note that we see that there are other processes that potentially interact with frequency to create the pattern of acquisition we see in many different languages including Spanish.

In the next section, we discuss various other factors that might contribute to the pattern of production we see across languages, but in particular, of the pattern that we see in Spanish. We identify 3 factors that affect learning, which are; morphophonological complexity, lexical neighbourhood density and semantics. We will establish the role these factors play in the acquisition of different inflectional forms.

2.9 Beyond Frequency

2.9.1 The Role of Phonology and Articulatory Complexity

It is difficult to discuss children's morphological development without discussing phonology. Phonological processing of both simple and complex forms contributes to when and how a child produces those forms, meaning that phonology places some constraints on inflectional development. For a child learning a complex inflectional

system, producing different inflections is a difficult learning process. Barlow and Pruit-Lord (2014) propose that children initially strive to produce the simplest form possible. As such, children are more likely to produce lexical items that have a syllable that ends in a final vowel as opposed to a consonant cluster. A form with a final consonant cluster such as "*drinks"* is more complex to produce than a form such as "*plays"*, therefore increasing the likelihood of a child producing the latter.

The role of phonology in morphological development is particularly clear in children with DLD. It is well documented that children with DLD encounter problems with some aspects of inflectional morphology (Kunnari et al., 2011; Bedore & Leonard, 2001, 2005). In terms of production, Children with DLD are more likely to omit an obligatory morpheme due to the phonological complexity of the form that is required, thus hindering development and usage. Marshall and van der Lely (2007) show that phonological complexity affects suffixation as DLD children are more likely to omit an inflected form that ends in a complex consonant cluster as opposed to a vowel.

A child learning an inflectional system might look to reduce cognitive load by producing the most phonologically simple form available. Children can select and omit certain verb forms as a function of their phonological properties (Ferguson & Farwell, 1975; Schwartz & Leonard, 1982). This process has been described as "lexical selection" and "lexical avoidance", where children will either select words or avoid words based on their syllabic shape and structure. Schwartz and Leonard propose that children will more readily produce inflections that are phonologically simpler than inflections that are more complex. Davis, Chenu and Yi (2018) report

on two French-speaking children who appeared to select words that matched their own capacity. In the case of Spanish, it could be argued that producing the 3sg form "-*a*" is easier for a child than producing the 3pl form "-*an*".

Penny (2000) points out that, in Spanish, the 2sg morpheme "-*as*" is more complex than the 3sg morpheme "-*a*". This can potentially explain why the 3sg form is more likely to be produced than the 2sg form. Aguillar-Mediavilla, Sanz-Torrent and Serra-Raventos (2007) suggest that when children overload their processing system, they are more likely to produce automatic patterns that are easier to both process and produce. Children are therefore likely to omit phonemes to reduce cognitive load and produce phonetically simple forms.

Expanding on these claims, a study of a single Spanish child's inflectional development demonstrated that frequency and phonology are intertwined when children are learning the different inflections (Aguirre, 2003). In this study, Aguirre suggests that an inflected form is more likely to be produced if it is 1) the most frequent form and/or 2) the most phonologically simple. This is because when the form is both frequent and simple to produce, the form will become more salient in the lexicon. In the case of the 3sg form (which is produced more readily that other forms), this form is phonologically simple because it consists of a single vowel (-*a* compared to 2sg -as and 3pl -*an).* Ettlinger and Zapf (2011) suggest that for English-speaking children, the production of noun plurals is severely limited because of the complex phonological properties of plural nouns. Roark and Demuth (2000) also suggest that children's early productions reflect the frequency with which syllables and word structures appear in the input.

Early childhood productions of different inflected forms can also be explained through a child's phonotactic constraints. Demuth and Tremblay (2008), in a study of two French-speaking children's production of determiners, found that children produced phonologically simple monosyllabic forms before disyllabic and trisyllabic forms. This suggests that children will produce simple forms earlier than more complex forms as a function of the phonotactic constraints that operate on a child's early productions. These constraints can inhibit the production of certain forms and can increase the use of other forms. Song, Sundra and Demuth (2009) propose that a child's language development is somewhat opportunistic. Song et al. suggest that in order for children to be successful, children often default to the simplest form they can produce in order to increase the chance of success. For children producing inflected forms in Spanish, the simplest form to produce is the 3sg form *"-a"*. Song et al. propose that the 3sg form is particularly influenced by complexity and therefore is produced in phonologically simple contexts.

Demuth and McCullough (2009) suggest that this process is fundamental to children's language development. They propose a "Prosodic Licencing Hypothesis" where learners are more likely to produce inflections in prosodically licensed contexts which are phonologically simple contexts, noting that the 3sg form "-*s*" in English will be produced more readily in simple contexts. During the early stages of development, children will use these prosodic contexts to produce simple forms more readily than complex forms. As learners' abilities increase, children will be able to produce progressively more phonologically complex grammatical forms.

2.9.2 The Role of Phonological Neighbourhood Density

Another aspect of word structure that has been shown to be important in lexical acquisition is phonological neighbourhood density (PND). A word's phonological neighbourhood density is defined as the number of other words in the language that differ by a single phoneme (Munson & Solomon, 2004; Hansen, 2017; Chan & Vitevitch, 2010). Munson and Solomon (2004) use the word "cat" as an example of a word with a rich phonological neighbourhood density with neighbours including forms such as "at", "sat", "cap" and coat". Alternatively, a word like "choice" has relatively few neighbours e.g. "voice" and "chase" (these lists are illustrative rather than exhaustive). Words such as "cat" that have many phonological neighbours are described as having a dense neighbourhood. Alternatively, words like "choice" with relatively few neighbours are described as having a sparse neighbourhood (Vitevitch & Stamer, 2006).

Luce and Pisoni (1998) and Vitevitch and Luce (1998; 1999) demonstrated that for adults, words with sparse neighbourhoods are recognised more quickly than words with dense neighbourhoods. This suggests that words with dense neighbourhoods trigger various other words among their neighbours creating competition between words, thus increasing the time to respond. Vitevitch and Rodriguez (2005) examined the role of PND in native Spanish speakers using an auditory lexical decision task. They predicted that there would be a competition effect (similar to results found in English), where words with sparse neighbourhoods will be responded to more quickly than words with dense neighbourhoods. However, in terms of production, dense neighbourhoods have a facilitatory effect, where words

from dense neighbourhoods are more likely to be produced by adults (Vitevitch, 1997).

In terms of recognition, Garlock, Walley and Metsala (2001) and Metsala (1997) found that words from dense neighbourhoods require more phonetic information in order for words to be recognised when compared to words from sparse neighbourhoods. However, with regard to production, children appear to mirror the trends displayed by adults. When learning is measured longitudinally, children acquire words from dense neighbourhoods at a faster rate than sparse neighbourhoods (Storkel, 2001; Storkel & Rogers, 2000). Extending this finding, Storkel (2004) suggests that early in development, children are more likely to acquire words that exist in dense neighbourhoods than sparse neighbourhoods.

Hogan, Bowles, Catts and Storkel (2011) measured the effect of neighbourhood density on phoneme accuracy in 2nd and 4th grade children. They predicted that phoneme awareness would increase in dense neighbourhoods. This prediction was confirmed, as children were more likely to produce an accurate word in a dense neighbourhood than a sparse neighbourhood. Hogan et al. also examined the effect of word frequency on phoneme accuracy, finding that children would also be more accurate producing high frequency words than low frequency words. Further to this, they found an interaction between frequency and neighbourhood density; as the frequency increased the effect of neighbourhood density decreased and vice versa. Hogan et al. suggest that there is a threshold effect, where both frequency and neighbourhood density are factors in children's early productions, but the effect of

neighbourhood density diminishes when word frequency is high or inversely the effect of word frequency diminishes when neighbourhood density is high.

The results of this study highlight that there is some relationship between frequency and phonology in terms of having awareness of and producing different phonemes. Hogan et al. propose a threshold effect, where high phoneme awareness is a product of either high neighbourhood density or high word frequency, but not both.

2.9.3 The Effect of Meaning on Inflectional Acquisition

Much of the work outlined above has shown an advantage for particular forms of words, such as the 3sg. We considered frequency and phonological explanations for this. However, an alternative explanation for children's early production of the 3sg form, is that there may be something special about the meaning of the 3sg form that increases the likelihood of the form being produced. Pinker and Prince (1988) argue that phonology alone cannot completely explain children's early inflectional use. The example they use is the verbs "*Break*" and "*Brake*" which are phonologically the same but semantically different. Therefore, children must distinguish between the two forms based on semantics rather than other lexical features. This suggests that children have a more abstract representation of different forms rather than just their phonology and/or their input frequency.

A potential reason for the 3sg preference in a child's language is that the other forms (e.g. 1st and 2nd person pronouns) require a child to understand that which form is appropriate depends on who is speaking – them or their interlocutor. If a speaker is describing an action that they alone are performing then a first person singular form

is required, but if the other conversational participant is describing the same action then the second person form is required. By contrast, the 3sg form is the same regardless of who in the dyad is speaking – describing an action performed by someone other than the speaker or their interlocutor always requires the third person.

Wechsler (2010) suggests that 3sg forms are semantically isolated when compared to 1st and 2nd person forms. This is as a result of 1st and 2nd person forms being contextually related to either the speaker (i.e. the person producing the utterance – "*I am playing*") or the addressee (i.e. the person receiving the utterance – "*you are playing*"). In contrast, 3rd person forms exclude both the speaker and addressee; instead they make reference to either "*he*" or "*she*", the most salient person who is not being addressed (e.g. He is playing). Thus, the difference between 1st, 2nd and 3rd person forms is whether the communicative intent is related to a person either producing or hearing an utterance or the utterance is about someone outside of the conversation.

It has been pointed out that personal pronouns are often omitted from early speech productions in favour of 3rd person forms or proper names, yet when they are produced, they are often used erroneously (Brown, 1973; Evans & Demuth, 2012). Lewis and Ramsay (2004) suggest that these trends can potentially be explained by children not yet having the capacity to understand that others have either similar or different viewpoints that can be expressed. Therefore, children must be able to demonstrate a certain level of social understanding (e.g. self-recognition, understanding perspective and speech roles) to correctly produce certain forms (Lee, Hobson & Chiat, 1994; Lewis & Ramsay, 2004).

Wechsler (2010) argues that forms that take into account perspective and speech roles (e.g. 1st and 2nd Person forms) require the skills of self-ascription and Theory of Mind (ToM) in order to be successfully produced. Weschler describes "self-ascription" as the presentation of an utterance that contains a first person indexical. Weschler suggests that the first-person form is unique as it carries information about the speaker. Theory of Mind is described as the ability to attribute mental states to oneself and others, allowing an individual to make inferences about what others are thinking (Premack & Woodruff, 1978; Wimmer & Perner, 1983; Astington, 1994; Schlinger, 2009). ToM allows the addressee to understand the mental state of the speaker and infer meaning (e.g. an addressee understanding that a 1st person utterance is in relation to the speaker). Weschler suggests that self-ascription allows the speaker to correctly use 1st person pronouns and the addressee to correctly interpret 2nd person pronouns. ToM allows the addressee to correctly interpret 1st person pronouns and the speaker to correctly use 2nd person pronouns.

By contrast, children do not require either ToM or self-ascription to produce third person forms (i.e. the speaker does not have to understand the mental state of the addressee nor be able to ascribe a role to themselves). Therefore, this could explain children defaulting to the 3sg form early in their language learning as the use of 3sg form does not require high order understanding. Wechsler (2010) proposes that self-ascription allows for children to omit 1st or 2nd person forms and replace them with 3rd person forms if they cannot produce a 1st or 2nd person form.

Mazzaggio (2016) suggests that children learning to produce language must take into account two factors; the first is the intrinsic grammatical perspective (e.g. Person) and the second is the speech roles (e.g. Speaker and Addressee). Mazzagio tested children aged between 38 and 70 months to assess to what extent children use ToM to understand 1st and 2nd Person pronouns. Mazzagio used a battery of ToM tests (using Wellman & Liu's (2004) seven task methodology) and a series of pronoun tasks (i.e. Use of pronouns and verbs, pronouns in isolation and a picture selection task to measure understanding of pronouns). The results showed there was no relationship between ToM and 1st person pronoun production, but there was a correlation between ToM and 2nd person pronoun production.

Similar results were found by Markova and Smolík (2014) who studied native speakers of Czech, a morphologically rich language that allowed the researchers to assess the use of personal pronouns and different verb conjugations. In particular, they looked at the use of mental state language (i.e. use of 1sg and 2nd person forms) and its relationship to pronoun use. Markova and Smolík found that use of 2nd person pronouns (but not 1st person pronouns) was related to a child's ability to understand the mental state of another (i.e. understand another's thought processes). For example, a child at a certain stage of development can understand 1st person pronouns as they require an understanding of an internal mental state, whereas 2nd person pronouns require their understanding to be pushed further as they have to be able to appreciate another's mental state.

These results support the claims of Wechsler (2010), who suggested that for a speaker to produce 2nd person pronouns successfully, they must have ToM – they

must understand the mental state of the addressee. Mazzaggio goes further, suggesting that this relationship leads to children producing errors on the basis of "Pronoun Reversal". Pronoun Reversal involves children substituting the 1st person pronoun "I" for the 2nd person pronoun "you" and vice versa (Chiat, 1986). The confusability between the 1st and 2nd person forms has implications for children using the 3sg form. The 3sg form is non-ambiguous and contextually non-variant as using this form does not involve having to understanding another person's mental state. Under this proposal, the 3sg form is also semantically separate from the 1st and 2nd person forms as it does not involve the speaker or the addressee.

2.10 Theoretical Predictions

The aim of the studies in the current thesis is not to test the predictions made by particular models of early morpho-syntactic development but rather to investigate whether non-Spanish-speaking adult and child participants learning a simplified version of the Spanish present tense paradigm show a similar pattern of performance to young Spanish-speaking children and, to the extent that they do, to determine what factors are responsible for this pattern. However, since the results of these studies have the potential to tell us why learners show the patterns of performance shown by young Spanish-speaking children, they also have the potential to provide evidence for and against the theoretical positions presented above. Thus, the constructivist theories described above assume that the pattern of performance shown by young Spanish-speaking children reflects the interaction between domain-general learning mechanisms and the semantic-distributional properties of the language to which they are exposed. It follows that, if our results show that older second-language learners show the same pattern as Spanish-speaking two-year-olds

and that this pattern can be explained in terms of the semantic-distributional properties of the training set (such as differences in input frequency), then these results would provide support for a constructivist position. The generativist theories, on the other hand, assume that the pattern of performance shown by young Spanish-speaking children reflects the tendency to use a particular underspecified form of the verb (in this case the 3sg) in an inappropriate finite context because of a maturationally controlled difference between the child and the adult grammar. It follows that, if our results show a similar pattern of performance in language-learning adults to that found in language-learning children, they would raise doubts about the idea that this pattern reflects an underlying difference between the child and the adult grammar.

2.11 Conclusion

We established that three separate factors might be expected to contribute to children's "defaulting" behaviour. However, it is clear that these processes are not mutually exclusive but are intertwined. The aim of the following set of studies is to disentangle the roles that frequency, phonology and semantics play in both the acquisition and production of inflection. In order to do this, both children and adults took part in a set of training studies, where they learned different Spanish verbs and their inflections to establish how much of an effect these three factors have.

Part Two

Chapter Three

3 Assessing the Effect of a Skewed Distribution on the Learning of Spanish Person and Number Morphology

3.1 Introduction

The aim of the current study is to test the claim that the pattern of errors seen in Spanish two year olds reflects the statistical properties of the Spanish language. In particular, we are interested in the claim that one particular common error type - the use of the 3sg form where other forms are contextually appropriate - reflects the speaker's defaulting to the most frequent form learned from the input. The study aimed to measure how much of an effect input frequency has on monolingual English-speaking adults learning different inflected forms in Spanish. We did this by directly manipulating the input frequency during learning so that in one condition the relative frequency matched that of real child-directed speech and in the other condition the different verb forms occurred an equal number of times (thereby removing the skew observed in real Spanish use). Participants were taught twelve different verbs and five inflections using animations depicting the relevant actions. Participants were tested on three separate days.

By directly manipulating the frequency of the forms that participants hear, we can test whether the speaker bias towards the 3sg form is caused by skewed input during learning. This will allow us to determine how much of an effect frequency has on learning and whether error rates are a direct result of the greater frequency. We hypothesise (on the basis of the idea that frequency is what drives the special status of the 3sg form) that the 3sg form will be produced more successfully than the other forms, but only in the skewed condition, and that more 3sg errors will be made in the skewed condition than in the other conditions.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Participants

45 adult, monolingual speakers of English took part in the study. Twenty participants were assigned to the skewed condition and 25 participants were assigned to the uniform condition. The sample consisted of 37 females and 8 males, with a mean age of 19;4 (range 18;3 – 25;1). All participants were first year students from the University of Liverpool recruited using the Experiment Participant Requirement system, where each participant gained course credit for taking part in the study.

3.2.1.2 Design and Materials

The study employed a mixed factorial design. The between-subjects factor was the condition assigned to each participant (skewed or uniform); the within-subjects factor was target verb ending *(-a, -amos, -an, -as, -o*). In order to maintain the natural properties of Spanish, 1pl/-*amos* remained as a stressed inflection within our learning paradigm whereas the 1sg/-*o and* 3sg/-*a* forms remained unstressed or reduced, mirroring the properties of spoken Spanish. Any preference for 1pl/-*amos* over other inflections, in our data as in real usage, could be explained by the 1pl form being more salient as a result of being stressed. Participants' production of verb endings was the outcome of interest, and we derived two binary trial-by-trial dependent variables from this, both of which we explore below – whether the participant produced the correct verb ending, and, for the trials on which an error was made, whether the errorfully-produced verb ending was the 3sg form.

Participants viewed computer animations of people performing different actions. The animations were created using Anime (Smith Micro Software, 2015), and presented using a purpose-made app created using the Processing language and environment (Reas & Fry, 2007). The animations depicted actions that could be described using the different person and number combinations that the participants were to learn. There were five different forms that they learned - 3rd person singular (-*a*), 2nd person singular (-*as*), 1st person singular (-*o*), 1st person plural (-*amos*) and 3rd person plural (-*an*). 2nd person plural was not included in the experiment, as this form is so rare in the input, and so rarely attested in children's productions.

During a training phase, participants viewed 80 trials. Each trial consisted of 2 animations – a first animation that was paired with a description of the action, and a second animation that was not. During the second animation the participant was asked to imitate the description that they had heard accompanying the first video. The relative frequencies of the different endings included varied with condition, as shown in Tables 3.1 and 3.2. Table 3.1 shows the proportional frequency of the different person and number combination for the skewed condition. In the skewed condition, there is a strong bias towards the 3sg form, and a slight skew toward the 2sg form, reflecting the biases seen in naturally occurring Spanish. Table 3.2 shows the proportional frequency of the different person and number combinations for the uniform condition. In this condition, there is no bias towards any form and the frequency with which participants hear the different forms is equal.

In both conditions, some combinations of verb stems and endings were unseen during the training phase, as indicated by the empty cells in the Table. This means

that some stem and ending combinations were presented for the first time during the testing phase, allowing us to look at generalisation during this phase of the experiment.

Table 3.1 - Frequency distribution of the verb/ending combinations in the Skewed Condition on each training day.

	3sg (a)	2sg (as)	1pl (amos)	3pl (an)	1sg (o)
Rasg-			1	1	1
Escal-			1	1	1
Dibuj-		2	1	1	1
Salt-		2	1	1	1
And-	5		1	1	1
Empuj-	5	2	1		
Golpe-	5	2		1	
Pate-	5	2			1
Dispar-	5		1	1	1
Bail-	5	2	1		
Cant-	5	2		1	
Tir-	5	2			1
FREQUENCY OF PERSON/NUMBER FORM	40	16	8	8	8
PROPORTION OF TOTAL TRIALS	0.5	0.2	0.1	0.1	0.1

Table 3.2 - Frequency distribution of the verb/ending combinations in the Uniform Condition on each training day.

	3sg (a)	2sg (as)	1pl (amos)	3pl (an)	1sg (o)
Rasg-		2			2
Escal-			2	2	
Dibuj-		2	2		
Salt-				2	2
And-	2		2	2	2
Empuj-	2		2	2	2
Golpe-	2	2		2	2
Pate-	2	2		2	2
Dispar-	2	2	2		2
Bail-	2	2	2		2
Cant-	2	2	2	2	
Tir-	2	2	2	2	
FREQUENCY OF PERSON/NUMBER FORM	16	16	16	16	16
PROPORTION OF TOTAL TRIALS	0.2	0.2	0.2	0.2	0.2

During the testing phase, learners viewed the animations once each without any description, and were asked to describe the action they saw. Table 3.3 shows the frequency of the different person and number combinations for the test phase. During the testing phase, the unseen verb forms were presented to test the participant's ability to generalize their knowledge to novel verb-ending combinations. Each participant viewed 40 trials during the testing phase, with 8 verbs and 5 endings. They viewed the animations and were asked to produce the correct form.

Table 3.3 - Frequency distribution of the verb/ending combinations in the test phase (40 test trials per day) .

	3sg (a)	2sg (as)	1pl (amos)	3pl (an)	1sg (o)
And-	1	1	1	1	1
Empuj-	1	1	1	1	1
Golpe-	1	1	1	1	1
Pate-	1	1	1	1	1
Dispar-	1	1	1	1	1
Bail-	1	1	1	1	1
Cant-	1	1	1	1	1
Tir-	1	1	1	1	1
FREQUENCY OF PERSON/NUMBER FORM	8	8	8	8	8

3.2.1.3 Procedure

The participants were tested in a quiet room at the University of Liverpool. They were seated directly in front of a desktop computer. There was no time limit imposed.

To explain how the animations worked, we will use the Spanish verb "bailar" meaning "to dance". The first animation that the participants viewed showed "Juan" pointing towards an action and then describing that action, so that, for the 3rd person singular form of the verb "to dance", Juan pointed at a character dancing on the screen and say "baila". The animation played again, but with Juan no longer on the screen, and the participant was asked to repeat the form produced by Juan, so that, for a video that featured a third actor (neither Juan nor the participant themselves), the participant would then attempt to say "baila". The same format would apply for

the 3rd person plural form. However, this time “Juan” would point at 2 people dancing and produce the form “bailan”. The participant would then see the video again but without Juan being shown and would repeat the form.

To illustrate first and second person forms, two photographs were taken - one of the experimenter and one of the participant. These pictures were then superimposed on the head of one of the characters. The experimenter would demonstrate how the experiment worked, first in English and then in Spanish. The examples involved the use of 5 different forms; 3rd person singular, 1st person singular, 2nd person singular, 3rd person plural and 1st person plural.

The English examples used the verb “climb”. The animation would play and Juan would describe the action. For the 3rd person singular form, Juan would say “He’s climbing”. The animation would show Juan pointing towards another character performing the climbing action. The animation would play again, but this time without a description and the experimenter would repeat the form. For the 1st person singular form, Juan would point towards an animation of himself performing the climbing action. He would say “I’m climbing”. The animation would play again but with an image of the experimenter superimposed on the animated character. The experimenter would then repeat the form. For the 3rd person plural form, Juan would point towards two different characters performing the climbing action and say, “They’re climbing”. This animation did not use either Juan or the experimenter. The animation would finish and the same two characters would appear again. The experimenter would repeat the form. The 1st person plural form again had Juan pointing towards two characters - one character being Juan and the other being the

experimenter. Juan would observe aloud that “We’re climbing”. The same animation would play again, and the experimenter would repeat the phrase. A last animation was the 2nd person singular form. Juan would point at the character in the animation. The character here had the experimenter’s face. Juan would say the phrase “You’re climbing”. The animation would play again, with Juan performing the climbing action. The experimenter would then repeat the phrase “You’re climbing”. The English examples would finish here.

The experimenter would then perform the same task, this time using the Spanish verb “cortar”, meaning “to cut”. The same format of animations would play. The 3rd person singular animation used a male character for both animations. For the 1st person singular form, Juan would perform the action for the demonstration and the experimenter’s face would be used when the experimenter was repeating the phrase. The 2nd person singular form used the experimenter’s face for Juan’s demonstration and then Juan would replace the experimenter’s character when the 2nd animation was shown. The 1st person plural form used both Juan and the experimenter for both the demonstration and the repetition phase. For the 3rd person plural form, 2 separate characters were used for the demonstration and repetition phases. Again, each of the 5 different verb endings was used; 3sg (corta), 2sg (cortas), 1sg (corto), 3pl (cortan) 1pl (cortamos). The experimenter would see and hear Juan explaining the action. The same animation was played again without the description. The experimenter would then have to repeat the Spanish verb and ending combination (e.g. the 3sg form “corta”.) On the first testing day, the experimenter would demonstrate both the English and Spanish forms. However, on the second and third day, only the Spanish forms were demonstrated.

After the experimenter had completed these examples, the experiment would begin. The Spanish process was repeated for 80 trials per day for both the uniform and the skewed conditions. The experimenter would click after each animation in order to move from the demonstration animation to the repetition animation. If the participants did not hear a verb during the demonstration phase, the experimenter would move on to the next animation. For the 1st person singular form, the first animation would show "Juan" pointing at "Juan" dancing and would produce the form "bailo". The second animation would show the participant's head on another character performing the same action. The participant would repeat this form. For the 2nd person singular form, the first animation would show the participant dancing, as Juan produced the form "bailas". The second animation would show Juan performing the same action, and the participant would repeat the form "bailas". For the 1st person plural form, both Juan and the participant would be dancing, with Juan pointing and saying "bailamos". The same animation would play again, and the participant would repeat the spoken verb.

After the training stage, participants took part in a 40-trial test phase. During this phase, the verb/ending combination would not be given during the animation. The participants were required to remember the different verb/ending combinations. The experimental procedure was repeated on each of three training days, with a slight difference to demonstrations on day 2 and 3, when the participants did not experience the animations in English. Each training day was audio-recorded.

3.2.1.4 Transcription and Coding

The audio-recorded Spanish responses were transcribed by the researcher. Correct verb ending responses were coded as 1 and incorrect verb ending responses were coded as 0, regardless of whether the stem produced was correct. When no answer was given, the trials were omitted from the analysis. Both the verb stem and the ending were coded. Responses were only included if a full verb was produced (i.e. both a verb stem from the set of verb stems seen in training and an ending was produced). If either was missing, the trial was omitted. There were 134 trials excluded, out of a total of 1840. The reasons for exclusion were as follows - not producing a full verb (i.e. not producing an eligible stem or ending) (44), not providing an answer (63) and not producing an audible response (27). 20% of the participants from each condition were second coded by a research assistant who was blind to condition. Inter-rater reliability, measured by Cohen's Kappa, showed a high level of agreement between coders (agreement = 91.5%; kappa = .869).

3.3 Results

For this study, 2 analysis procedures were performed. We built 2 sets of mixed effects logistic regression models: one set to assess the effect condition had on the accuracy of the production of the different verb endings and one set to assess the effect condition had on the rate of 3rd singular errors produced. Table 3.4 shows the proportion of total responses made up by each ending. Table 3.5 shows the proportion of trials on which participants produced a response for each ending.

Table 3.4 - Proportion of coded trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.37(±.15)	0.15(±.09)	0.15(±.09)	0.21(±.11)	0.12(±.06)
Uniform	0.33(±.18)	0.18(±.07)	0.18(±.09)	0.16(±.10)	0.15(±.08)

Table 3.5 - Proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.36(±.15)	0.14(±.09)	0.14(±.09)	0.21(±.11)	0.12(±.06)
Uniform	0.29(±.14)	0.17(±.08)	0.16(±.08)	0.14(±.08)	0.13(±.07)

3.3.1 Correct Verb Ending on Day 3

The main aim of this analysis was to assess what effect condition and context (ending targeted by the animation) had on the correct production of verb endings. For this analysis, 1706 responses were analysed. We started with a mixed effects logistic regression model that included Condition and the Target Ending as predictors, along with an interaction between them. We took a leave-one-out approach to model selection. Likelihood ratio tests were used to compare models and determine whether each predictor term had explanatory value – whether removing these terms damaged the fit of the model to the data. If removing a term produced a model with at least as good a fit, then that term was excluded. Participant was included as a random effect on the intercept and the Target Ending slope. A target

stem varying intercept was not included in the model due to its providing a singular fit (as recommended by Barr et al., 2013).

The experimental design allowed a test of whether the participants could generalise to new verb-ending combinations. It is important to note that the inclusion or non-inclusion of a particular verb and ending combination at training is confounded with ending type and with condition. We therefore need to ensure that these two predictors account for variance over and above whether a form is seen or not. Once the value of these predictors has been established, we can then look at whether each verb's having been seen or unseen at training explains independent variance or merely a subset of that explained by ending and condition. To test the effect that exposure to a particular verb-ending combination had on the data, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that the model with the seen/unseen status of trials (coded as 0 for seen and 1 for unseen; hereafter referred to as simply "unseen") as a predictor variable provided a significantly better fit to the data than a null model ($X^2(1) = 23.76$, $p<.001$).

A full model where fixed effects of Target Ending, Condition and an interaction between the two were included along with exposure status was built (Target Ending*Condition + Unseen). This model was compared to a model where the interaction had been removed (Target Ending + Condition + Unseen). The full model with the interaction variable included did not provide a significantly better fit to the data than a model with the interaction removed ($X^2(4)=2.82$, $p=.587$). This indicates that the interaction does not usefully explain variance. The model with the

interaction removed was compared to a model with condition removed as a fixed effect (e.g. Target Ending + Unseen). The model with Target Ending, Condition and the Unseen trials included as fixed effects provided a significantly better fit to the data than when Condition was removed ($X^2(1)=34.94$, $p=.026$). The model with Target Ending, Condition and the Unseen trials included as fixed effects (Target Ending + Condition + Unseen) was compared to a model with Condition and the Unseen trials (Condition + Unseen) as fixed effects. The model that included Target Ending, Condition and the Unseen trials as predictor variables provided a significantly better fit to the data as opposed to the model with only Target Ending included ($X^2(4)=34.69$, $p<.001$).

The model with Target Ending, Condition and the Unseen trials included as fixed effects was compared to a model with only 'unseen' included as a fixed effect (due to the unseen model providing a significantly better fit to the data than a null model). The model with Target Ending, Condition and the Unseen trials included as fixed effects provided a significantly better fit to the data than the model with only 'unseen' included ($X^2(5)=40.31$, $p<.001$). This model (Target Ending + Condition + Unseen) was compared to a model with the Unseen trials removed as a fixed effect (Target Ending + Condition). This was to examine how much predictive value the Unseen factor had when included with a reduced set of selected model terms. The model with the Unseen trials included as a fixed effect provided a significantly better fit to the data than when the Unseen trials had been removed ($X^2(1)=4.77$, $p=.029$). It is thus concluded that a model with Target Ending, Condition and the Unseen trials as fixed effects provided the best fit to the data.

Figure 3.1 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015) in order to visualise our model. The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional, we treat as its estimate. Where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed). In order to aid the interpretation of our plots throughout this thesis we chose to exclude the terms that were suggested by the above model comparison to not improve model fit. We used treatment coding for all models, where the -a/3sg form in the uniform condition was treated as the reference class (and thus corresponds to the intercept in the model). To generate the log odds for Figure 3.1, we combine model coefficients to give unique estimates for each combination of form and condition. A table displaying output from the full Bayesian Model with predictor variables Target Ending, Condition and the interaction between the two variables is included in Appendix 1 – Table 1.

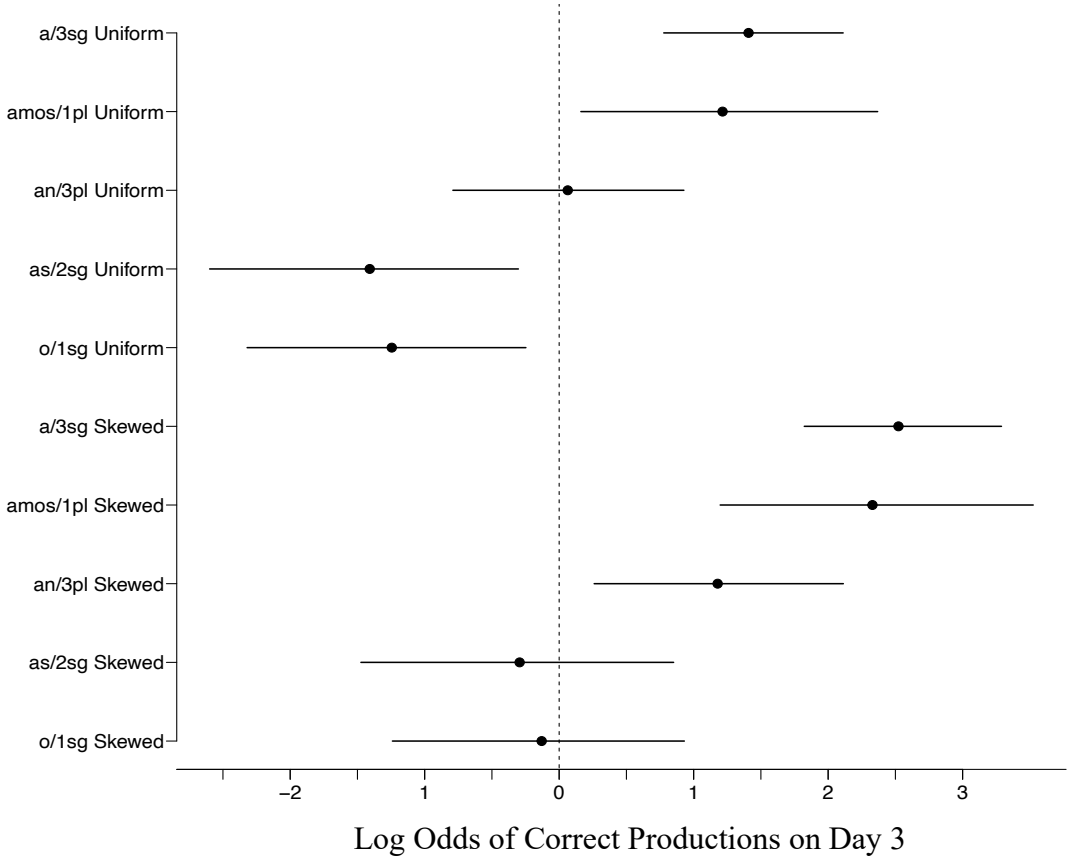

Figure 3.1 - The log odds and confidence intervals (95%) for correct productions on Day 3 (Unseen - β = -0.59, CI = -1.05 – -0.16).

In both conditions, participants were significantly more likely to produce *"-a"*. correctly than *"-an", "-as"* and "-*o*".

3.3.2 3rd Person Singular Error on Day 3

The main aim of this analysis was to assess what effect condition and target ending have on the production of 3rd person singular errors. For this analysis, all 679 responses were analysed where the 3sg form was not the target. This type of error is where "a" is produced instead of the target form - for example, producing "**baila*" instead of "*bailamos*" when "amos" was the target ending. We began with a full

model that included an interaction between condition and the target ending variable. We took a leave-one-out approach to model selection. Likelihood ratio tests were used to assess the fit of the model and whether each predictor term had explanatory value – whether removing these terms damaged the fit of the model to the data. A random effect of participant on the intercept was included in our model. A model that additionally included a random effect of participant on the Target Ending slope gave a singular fit suggesting that the model was overparameterised. A random effect of target stem on the intercept was considered but not included in the model due to not improving the fit of the data when added as a random slope and again providing a singular fit. Both terms were therefore omitted in line with Barr et al. (2013).

A full model in which Condition, Target Ending and a Condition by Target Ending interaction were included as fixed effects was compared to a model where the interaction was removed (Condition + Target Ending only). The full model provided a significantly better fit to the data than the reduced model ($X^2(3) = 11.75$, $p=.008$). The full model was also compared to a model where Condition had been removed. The model with the interaction included provided a significantly better fit than a model with only Target Ending included ($X^2(4) = 12.10$, $p=.016$). The full model was compared to a model with only Condition as a predictor variable. Again, the full model provided a significantly better fit than the model with only Condition in it ($X^2(6) = 85.13$, $p<.001$). Finally, the full model provided a significantly better fit than a null model where no fixed effects were present ($X^2(7) = 85.42$, $p<.001$). It was thus concluded that a full model (with Condition, Target Ending and an interaction between them) gives the best fit to the data.

Figure 3.2 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed. We used treatment coding for all models, where the -a/3sg form in the uniform condition was treated as the reference class.

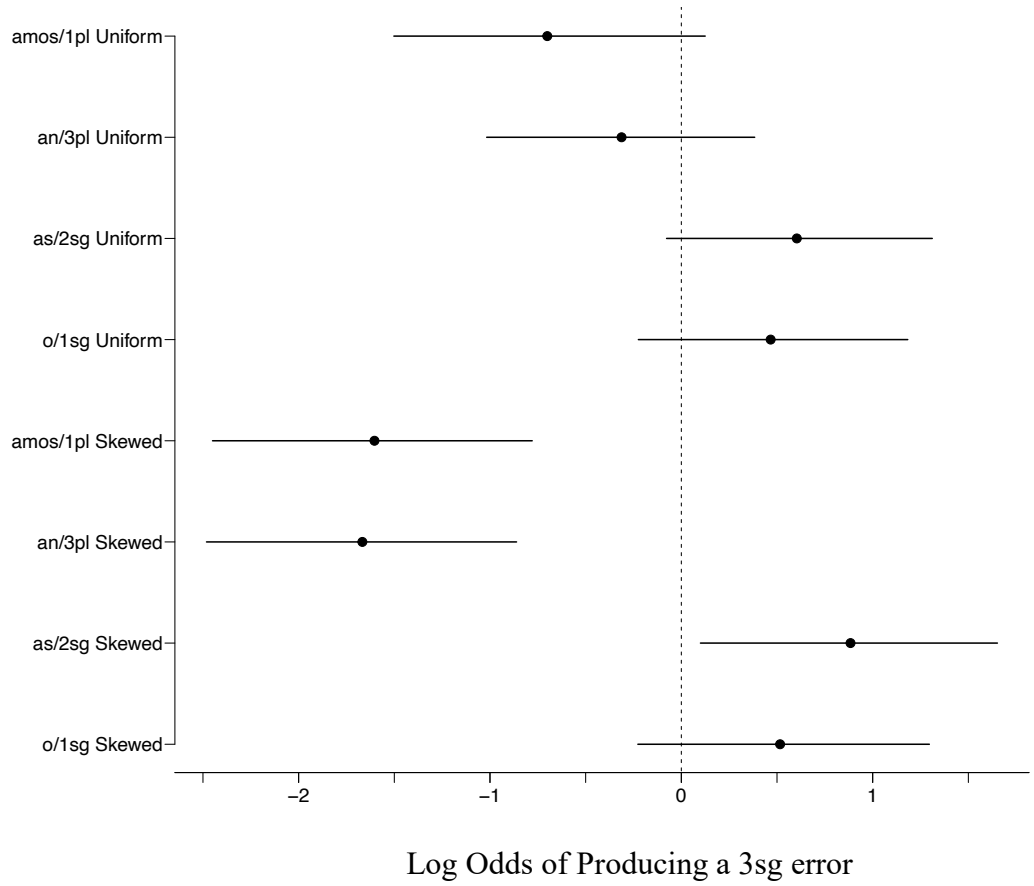

Figure 3.2 - The log odds and confidence intervals (95%) for 3rd Person Singular Errors on Day 3

Figure 3.2 shows that, in both conditions, participants were significantly less likely to produce a 3rd person singular error when the target was "-*amos*" or "-*an*" (plural targets) than when it was "-*as*" or "-*o*" (singular targets). There were more 3rd person singular errors overall in the uniform than in the skewed condition. The interaction between Target Ending and Condition is reflected in that fact that this pattern is driven largely by "-*amos*" and "-*an*".

3.4 Interim Discussion

The current study aimed to measure the effect that frequency had on learning different morphological forms. Participants were split into 2 conditions - one

condition skewed the input frequency towards the 3sg form, in line with the naturally occurring distribution reported for Spanish, and the other condition kept the input frequency equal across the different inflected forms. We predicted that the 3sg form would be produced more successfully than the other forms, but only in the skewed condition. We also predicted that there would be more 3sg errors in the skewed condition when compared to the uniform condition. The results do not support the first hypothesis because, while there are more correct productions of the 3sg form than of the other endings, this applies in both conditions, rather than just the uniform condition. This suggests that frequency cannot be the only driver of accuracy with the 3sg. There are also more correct productions overall in the skewed condition than the uniform condition, but there was no interaction, indicating that the skew improves performance across all endings. The second hypothesis was also not supported. The rate of 3sg errors was higher in the uniform than the skewed condition. An important pattern is that participants were more likely to produce 3sg errors when another singular form was the target, and this difference between plurals and singulars was greater in the skewed condition than the uniform condition.

Before we can draw conclusions from these results there is one methodological concern that must be addressed. In both conditions, participants took part in 80 training trials. However, keeping the number of training trials constant across conditions required the introduction of a potential confound. Achieving the desired proportions in the skewed condition required that each participant actually heard 3 of the different forms less in the skewed than in the uniform condition (i.e. *-amos, -an,* and *-o)* .

In order to explore this possible confound, we ran a second version of the uniform condition which had 40 training trials rather than the 80 we used previously. This was then compared to the same skewed data as in Experiment 1. This allowed us to have the same relative frequency for each of the three non-privileged endings across the two conditions.

3.5 Experiment 2

3.5.1 Method

3.5.1.1 Participants

44 adult, monolingual speakers of English took part in the study. Twenty-four participants were assigned to the uniform 40 trial condition and this was combined with the 20 participants collected in the skewed condition from Experiment 1. There were 8 male participants and 36 female participants, with a mean age of 19;6 (range 18;3-24;8). All participants were 1st year undergraduate students from the University of Liverpool. The participants were recruited using the Experiment Participant Requirement system where students gained course credit for taking part in the study.

3.5.1.2 Design and Materials

The design and materials were the same as for Experiment 1 with the exception of the change in the relative frequency of items in the uniform condition. Table 3.5 shows the proportional and relative frequencies are equal across the different forms. However, the frequency is lower than in Experiment 1 (8 per inflection rather than 16 per inflection).

Table 3.6 – The relative and proportional frequency of the different verb/ending combinations in the uniform condition with 40 training trials received on each training day.

	3sg (a)	2sg (as)	1pl (amos)	3pl (an)	1sg (o)
Rasg-		1			1
Escal-			1	1	
Dibuj-		1	1		
Salt-				1	1
And-	1		1	1	1
Empuj-	1		1	1	1
Golpe-	1	1		1	1
Pate-	1	1		1	1
Dispar-	1	1	1		1
Bail-	1	1	1		1
Cant-	1	1	1	1	
Tir-	1	1	1	1	
FREQUENCY OF PERSON/NUMBER FORM	8	8	8	8	8
PROPORTION OF TOTAL TRIALS	0.2	0.2	0.2	0.2	0.2

3.5.1.3 Procedure

This was the same as for Experiment 1.

3.5.1.4 Transcription and Coding

This was the same as for Experiment 1. 1800 trials were analysed and there were 181 trials excluded based on; not producing an eligible stem (14), not producing an eligible ending (1), not providing an answer (133) and not producing an audible response (33). Table 6 shows the response rate for each verb ending. The data for the uniform condition indicate that there is a much lower response rate for each verb during the testing phase. This could potentially be explained by the participants in this condition receiving fewer trials (40 trials) than the skewed condition (80 trials).

Inter-rater reliability, measured by Cohen's Kappa, showed a high level of agreement between coders (Agreement = 88.8%; Kappa = .859).

3.6 Results

The analyses performed were the same as for Experiment 1. The skewed data reported here are those collected in Experiment 1. The uniform data, however, are new. Table 3.6 shows the response rates for the proportion of trials in which each response was produced across the two conditions. Table 3.8 shows the proportion of trials in which participants produced a response for each ending.

Table 3.7 - The proportion of coded trials (±SD) on Day 3 in which each verb ending was produced, separated by condition, where the uniform condition featured 40 Training Trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.37(±.15)	0.15(±.09)	0.15(±.09)	0.21(±.11)	0.12(±.06)
Uniform	0.39(±.14)	0.17(±.09)	0.13(±.09)	0.17(±.15)	0.15(±.09)

Table 3.8 – The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced, separated by condition, where the uniform condition featured 40 Training Trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.36(±.14)	0.14(±.08)	0.14(±.09)	0.21(±.11)	0.12(±.06)
Uniform	0.32(±.13)	0.14(±.08)	0.11(±.07)	0.14(±.10)	0.13(±.08)

3.6.1 Correct Productions on Day 3

In this analysis, 1619 trials were analysed using mixed effects logistic regression models. The random effects structure and the model comparison process was the same as for Experiment 1. Again, a random effect of target stem on the intercept was considered but not included in the model due to its not improving the fit to the data when added ($X^2(1) = 0.03$, $p=.837$).

The experimental design contained a test of whether the participants could generalise to new verb-ending combinations. To examine the effect that exposure to a particular verb-ending combination had on the data, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that the model with the "unseen" trials as a predictor variable provided a significantly better fit to the data than a null model ($X^2(1) = 12.13$, $p<.001$).

A full model which included target ending, condition, and an interaction between target ending and condition as fixed effects in addition to the 'unseen' variable as a fixed effect was compared to a model where the interaction was removed as a fixed effect (Target ending + Condition + Unseen). The full model did not provide a significantly better fit to the data than the reduced model ($X^2(4)=5.31$, $p=.257$), indicating. The model which included Target Ending, Condition and the Unseen Trials as fixed effects was then compared to a model with condition removed (Target Ending + Unseen). The model that included Target Ending, Condition and the Unseen variable as a fixed effect provided a significantly better fit to the data than a model with only Target Ending and the Unseen variable included as a fixed effect ($X^2(1)=12.21$, $p<.001$). The model which included Target Ending, Condition and the

Unseen variable as fixed effects was compared to a model with Condition and the Unseen trials included as fixed effects (Condition + Unseen). The model which included Target Ending, Condition and the Unseen variable as fixed effects provided a significantly better fit to the data than the model with target ending and the 'unseen' trials included as fixed effect ($X^2(4)=34.98$, $p<.001$). Finally, the model which included Target Ending , Condition and the Unseen variables as fixed effects was compared to a model which included the Unseen variable as a fixed effect (due to a model with the Unseen variable as a fixed effect provided a significantly better fit to the data than a null model with no fixed effects included). The model which included Target Ending, Condition and the Unseen trials as fixed effects provided a significantly better fit to the data than a model which included only the Unseen trials as a fixed effect, ($X^2(4)=46.75$, $p<.001$).

We need to assess whether the "unseen" fixed effect can be removed from the model. The model which included Target Ending, Condition and the Unseen trials as fixed effects was compared to a model with Target Ending and Condition as fixed effects, but with the Unseen trial removed. The comparison revealed that a model which included Target Ending, Condition and the 'Unseen' trials as fixed effects provided a better fit to the data than a model with only Target Ending and Condition included ($X^2(1)=10.98$, $p<.001$). This means that having the 'Unseen' trials included in the model explains more of the data than when it was removed.

Figure 3.3 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible

intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at α = 0.05 (two-tailed). In order to aid the interpretation of the plot we chose to exclude the non-significant items. We used treatment coding for all models, where the -a/3sg form in the uniform condition was treated as the reference class. (A table displaying the Bayesian Model including predictor variables Target Ending, Condition and the interaction between the two is included in Appendix 1 – Table 3).

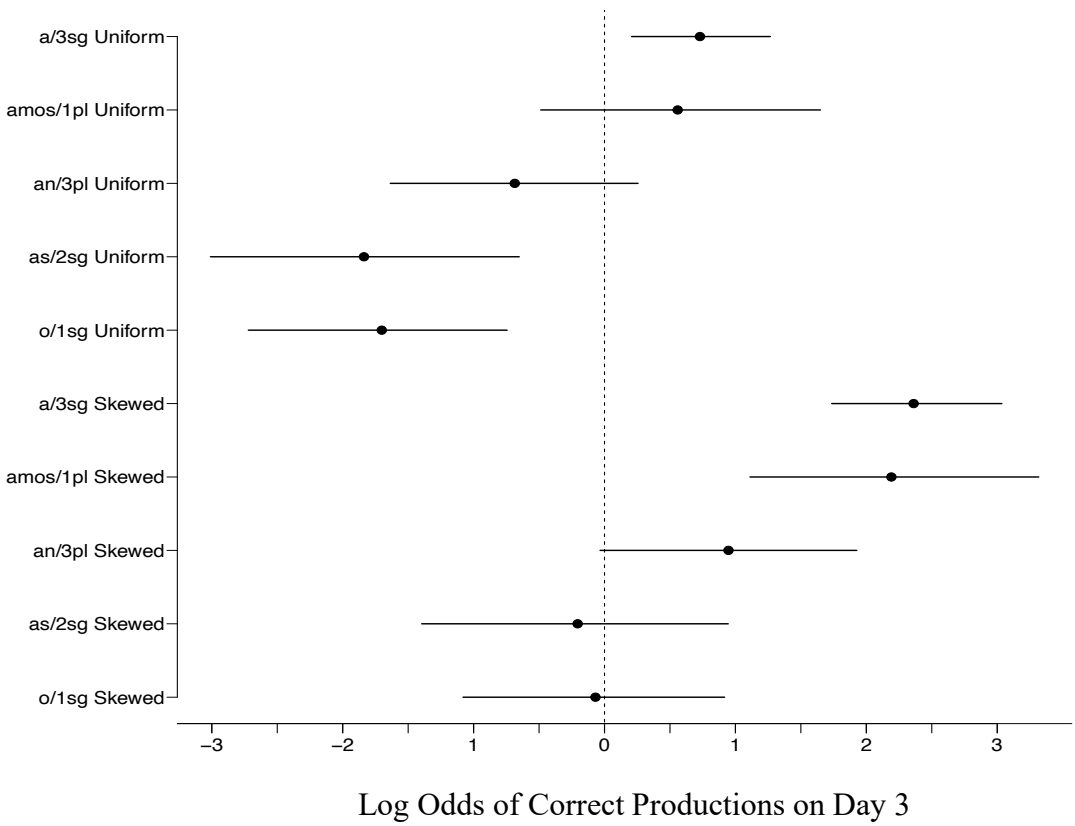

Figure 3.3 - The log odds and confidence intervals (95%) for correct verb ending production on Day 3 (Unseen - β = -0.65, CI = -0.99 – -0.32).

Figure 3.3 shows that in the uniform condition, participants were more likely to produce "-*a*" succesfully than "-*an*", "-*as*" and "-*o*". In the skewed condition,

participants were also more likely to produce "-*a*" succesfully than "-*an*", "-*as*" and "-*o*". Across both condition, there was no difference in successful production between "-*a" and "-amos*".

3.6.2 3 Person Singular Error Production on Day 3

The main aim of this analysis was to assess what effect condition and target ending have on the production of 3rd person singular errors. This type of error is where "*a"* is produced instead of the target form, for example, producing "**baila*" instead of "*bailamos*" when "-*amos*" was the target ending. 704 responses on Day 3 (all responses for non-3sg trials) were analysed using mixed effects logistic regression models. A random effect of participant on the intercept was included in our model. A model that additionally included a random effect of participant on the Target Ending slope gave a singular fit suggesting that the model was overparameterised. A random effect of target stem on the intercept was considered but not included in the model due to not improving the fit of the data when added as a random slope and again providing a singular fit. These terms were therefore omitted in line with Barr et al. (2013).

A full model where Condition, Target Ending and a Condition by Target Ending interaction were included as fixed effects was compared to a model where the interaction was removed (Condition + Target Ending only). The full model did not provide a significantly better fit to the data than the reduced model ($X^2(3) = 4.39$, $p=.222$). The reduced model with Target Ending and Condition included as fixed effects was compared to a model that included only Target Ending as a fixed effect (i.e. Condition was removed as a fixed effect). The model with Target Ending and

Condition included did not provide a significantly better fit to the data than a model with Target Ending included as fixed effect ($X^2(1) = 2.65$, $p=.103$). Finally, the model that included only Target Ending was compared to a Null Model. The model with only Target Ending included provided a significantly better fit than a null model where no fixed effects were present ($X^2(3) = 116.34$, $p<.001$). It was thus concluded that a model which included Target Ending as a single fixed effect provided the best fit to the data.

Figure 3.4 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional, we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

In order to aid the interpretation of the plot we chose to exclude the non-significant items (A table displaying the Bayesian Model including predictor variables Target Ending, Condition and the interaction between the two is included in Appendix 1 – Table 4).

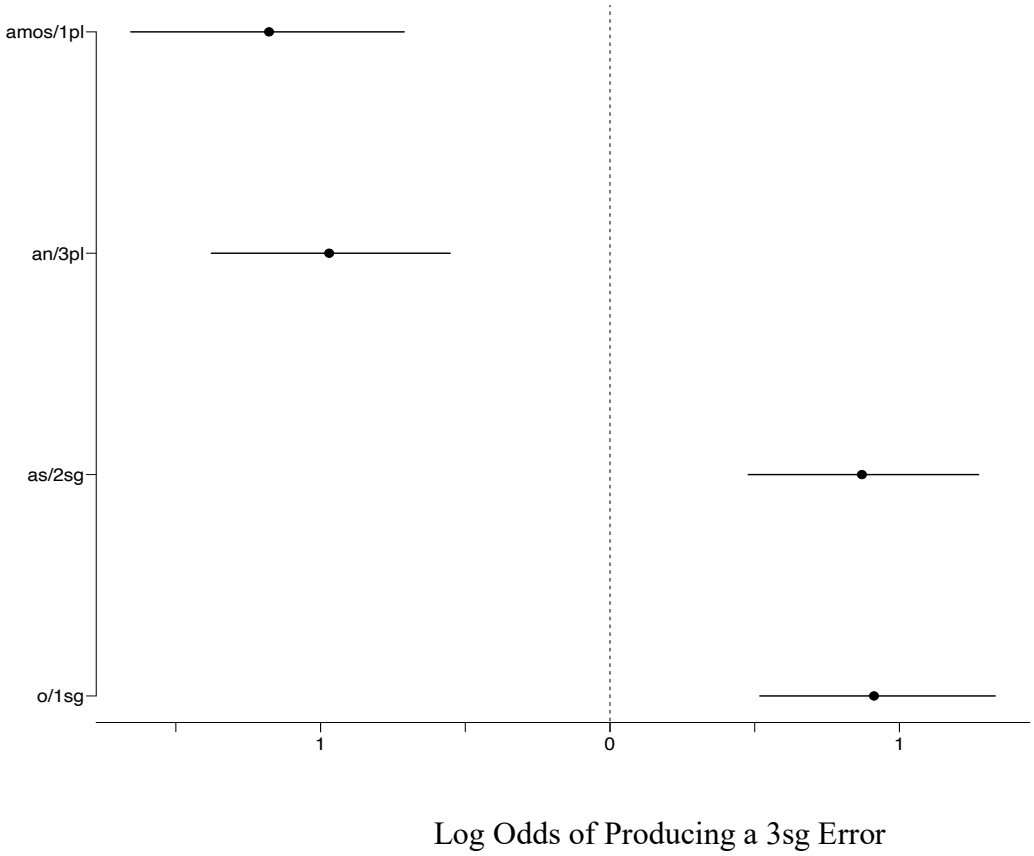

Figure 3.4 - The log odds and confidence intervals (95%) for 3rd Person Singular Errors Produced on Day 3

Figure 3.4 shows participants were significantly more likely to produce a 3rd person singular error when the target was singular ("-*as*" or "-*o*") than when it was plural ("-*amos"* or *"-an"*), regardless of condition.

3.7 How do participants make use of the cues?

Studies 1 and 2 suggest that participants were likely to produce a 3rd person singular error when the target was another singular form (e.g. producing the 3sg "-*a*" when the 2sg "-*as*" was required). In order to further understand this, we performed an

additional set of analyses to determine which cues the participants used to decide which verb inflection to produce.

Participants in our studies were tasked with learning to map different inflections to cues in a series of animated videos. These cues took the form of the different "person" and "number" features of the different inflections. For example, the verb "*Bailan*" is the 3pl form, so in the video two characters, neither the participant nor the "teacher", danced together. The verb "*Bailo*" is the 1sg form, so participants would see their own character dancing alone.

In order to assess how the participants were using the person and number cues, we fitted Bayesian Multinomial Regressions to the data from the three conditions - the uniform condition with 40 trials, the uniform condition with 80 trials and the skewed condition. The models were created using an R tool called *brms* (Burkner, 2017) which in turn makes use of *STAN*. Default weakly informative priors were used for all models. The 3sg form was set to be the reference class. This means that the model coefficients indicate the effect of the cues on participants' preference for this morphological form over the others (e.g. 1pl, 3pl, 2sg, 1sg).

3.8 Results

Using the data from Day 3 of testing we fitted three Bayesian multinomial regression models that included Person and Number features as individual predictor variables. Participant identity was included as a random effect on the intercept. We did not include by-participant random slopes because we are not seeking to generalise to the wider population. Instead, we are looking to determine whether the participants

within our experiments could use the different morphological features, and, if they could, which did they use, in order to understand the results reported above. We include the random intercepts to account for differences between participants that are not related to our question. This was done separately for each of the three conditions.

Leave-one-out cross validation (LOOIC; Gefland, Dey & Chang, 1992; Gefland, 1996) was used to assess the fit of the models. Removing the different terms and comparing the reduced model with the full model allowed us to measure how much explanatory value the different terms provided. If removing a term produced a model with at least as good a fit, then that term was excluded from our final model. Three separate models were constructed in order to assess the effect of Person and Number features within each condition rather than comparing the conditions.

3.8.1 Uniform Condition (80 Training Trials)

A model which included both Person and Number as fixed effects (LOOIC = 2099.73, SE = 56.28) provided a better fit to the data than a model with only Person included (LOOIC = 2426.78, SE = 40.27), a model with only Number included (LOOIC = 2391.43, SE = 39.19) and a null model where no fixed effects were present (LOOIC = 2730.11, SE = 24.14). The best model thus had both Person and Number included as fixed effects. This suggests that participants use both the person and number features within the Uniform 80 condition to learn the different morphological forms.

3.8.2 Uniform Condition (40 Training trials)

A model which included both Person and Number (LOOIC = 1840.68, SE = 50.80) provided a better fit to the data than a model with only Person included (LOOIC = 2180.13, SE = 36.66), a model with only Number included (LOOIC = 1980.81, SE = 42.44), and a null model where no fixed effects were present (LOOIC = 2333.45, SE = 29.56). The best model thus had both Person and Number included as fixed effects. This suggests that participants use both the person and number features within the Uniform 40 condition to learn the different morphological forms.

3.8.3 Skewed Condition

A model which included both Person and Number (LOOIC = 1732.14, SE = 58.29) provided a better fit to the data than a model with only Person included (LOOIC = 2144.05, SE = 41.23), a model with only Number included (LOOIC = 1983.76, SE = 42.31), and a null model where no fixed effects were present (LOOIC = 2410.56, SE = 28.16). The best model thus had both Person and Number included as fixed effects. This suggests that participants use both the person and number features within the Skewed condition to learn the different morphological forms.

The three multinomial analyses demonstrate that participants in the adult study effectively use the different morphological features. This tells us that the accuracy reported in our prior analysis is the result of their use of both cues. However, it is also useful to look at how their knowledge of cues is reflected in their errors. This is revealed by Figures 3.5 – 3.9 which show the rate at which each of the five different inflections were errorfully produced in each of the four other contexts across all three conditions in Experiment 1 and Experiment 2. We can see that in many cases

participants appear to make errors by classifying on the basis of a single feature, dimension, typically the number cue, and ignoring the other . For example, we can see in Figure 3.5 that participants erroneously use the 3sg form when the target is another singular form. This is also found in Figure 3.9, where participants are more likely to produce a 1st person singular error when the target is also a singular form. Similarly, for 3rd person plural errors and 1st person plural errors, there was more confusability when the target form was another plural form.

We previously discussed the stressed/unstressed inflections within the Spanish inflectional paradigm, where 1pl/-*amos* is a stressed form and the 3sg/-*a* and 1sg/-o are unstressed. The participants across the three conditions were more likely to confuse the 3sg/-*a* form with the 1sg/-*o* (and vice versa) compared to 1pl/-*amos*. It could be suggested that the fact that -amos is stressed made 1pl-amos forms less confusable with forms marked with the unstressed -a and -o morphemes.

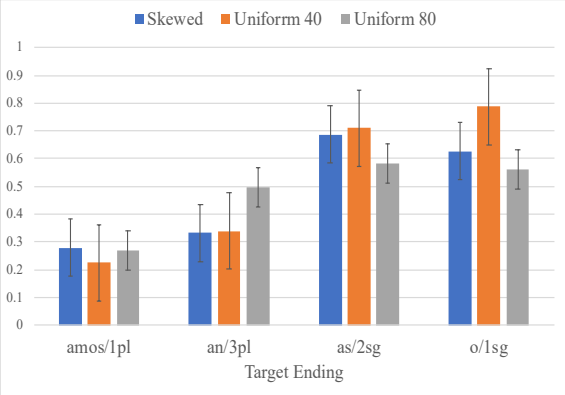

Figure 3.5 – Proportion of 3sg Errors produced on Day 3.

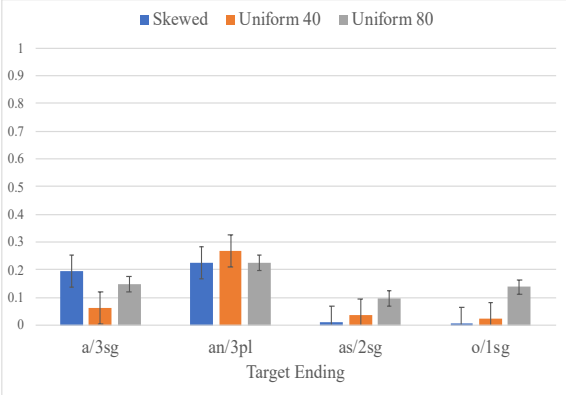

Figure 3.6 – Proportion of 1pl Errors produced on Day 3.

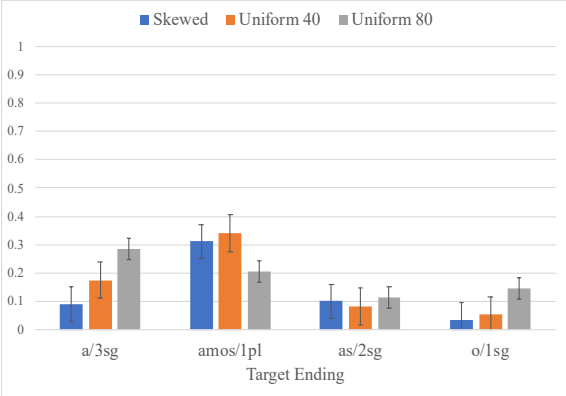

Figure 3.7 – Proportion of 3pl Errors produced on Day 3.

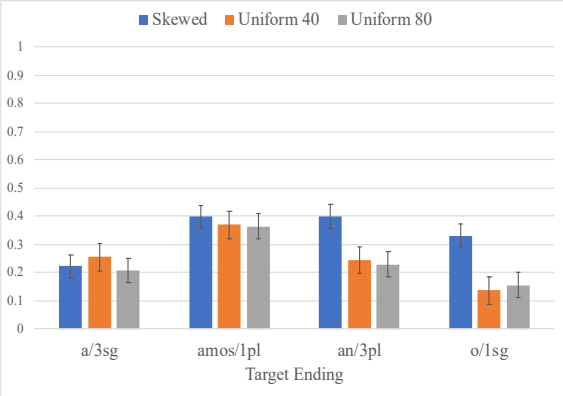

Figure 3.8 – Proportion of 2sg Errors produced on Day 3.

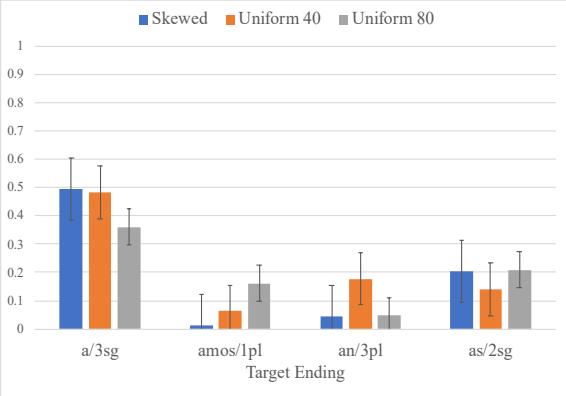

Figure 3.9 – Proportion of 1sg Errors produced on Day 3.

Figure 3.5 - 3.9 - Plots demonstrating where errors are made across each target inflection on Day 3.

3.9 Discussion

The aim of the studies reported in this chapter was to understand the effect frequency has on the production of person and number variants of Spanish verbs. In order to assess this relationship, participants were assigned to either a skewed condition, where the input was biased towards the 3sg form or a uniform condition, where the input was equal across the different verb forms. In Experiment 1, participants in both conditions heard the same number of training trials (80 Trials). In Experiment 2, participants in the skewed condition took part in 80 training trials but in order to check for a possible confound in Experiment 1 (despite the bias towards the 3sg form in the skewed condition, participant in the uniform condition heard the 3 lowest frequency forms more than in the skewed condition) participants in the uniform condition took part in 40 training trials. For both experiments, we hypothesised that the 3sg form would be produced more successfully than the other endings, but only in the skewed condition. We also hypothesised that there would be significantly more 3sg errors produced in the skewed condition when compared to the uniform condition.

For both experiments 1 and 2, the first hypothesis was not supported, with participants producing the 3sg form more successfully than the other endings but doing so regardless of condition. The second hypothesis was also not supported. There was an effect of condition on the production of 3sg errors in the opposite direction to that we predicted - the rate of 3sg errors was higher in the uniform than in the skewed condition. There was also an unpredicted effect of target ending, reflecting the fact that fewer errors were made for the plural forms than the singular forms, and an interaction between target ending and condition, reflecting the fact

that this difference was greater in the skewed than in the uniform condition. This suggests that participants are differentiating well between singular forms and plural forms, but failing to distinguish as well on the basis of person. In the skewed condition performance improves because participants seem to learn to distinguish the 3sg form better from the plural forms, but they do not show any better ability to do so based on person.

The model comparisons also showed us that in both experiments 1 and 2, trials that were Unseen or Seen during the training phase, had a significant impact on the production of correct forms, and thus had more predictive value. This demonstrates that participants do worse when the verb endings are produced with forms they have not heard before. This suggests that their knowledge is not completely generalised and abstract. Thus, whether the verb and ending combination they are tested on has been seen before is important.

Our initial analyses allowed us to assess where participants correctly produced and made errors. However, we also looked to establish whether the participants in the study used the different Person and Number cues to learn the morphological features of the different inflections. In order to answer this, we fitted Bayesian Multinomial Regression models to the data from the three different conditions. The models showed that across the three separate conditions (Uniform condition with 80 trials, Uniform condition with 40 trials and the skewed condition) the adults in our study could effectively use the different person and number cues embedded in the testing paradigm to produce the different present tense Spanish inflections, confirming the accuracy levels reported in Experiment 1 and Experiment 2. However, when

analysing the errors made across the three conditions, it could be suggested that the confusability between 3sg and 1sg forms could be explained by these inflections being unstressed in the input. The stressed/unstressed distinction could also explain why there is little confusability between 3sg and 1pl forms. However, the stressed/unstressed distinction does not explain why there is little confusability between 3sg forms and 3pl forms, both of which are unstressed. It is more likely that the errors made are the result of classifying inflections on the basis of a single feature (e.g. number).

The results support the previous claims of Aguado-Orea and Pine (2015) that participants show lower error rates on the 3sg form, and that children's errors often involve the use of the 3sg form in contexts where other forms are required. However the results do not provide support for the claim that this is due to the greater frequency of the 3sg form. An important, and unexpected trend from both studies, was the bias toward the 3sg form within the uniform condition, where there was no difference in the frequency of the different endings. Within the uniform condition, the 3sg form was produced more successfully than the 3pl form, the 2sg form and the 1sg form. It is therefore clear that whilst frequency affected performance in a number of other ways, it did not drive the 3sg advantage.

We must then look to other factors to explain the observed advantage for 3sg. The additional factors we are considering in this thesis are phonology and semantics. In order to explore the effects of phonology and semantics, we created a testing paradigm that separated the different phonological and semantic factors involved in learning different inflections. Dissociating phonology and semantics in this way

allows to determine how much of an effect each of them has on the learning of inflection. It is this study that we turn to in the next chapter.

Chapter Four

4 Assessing the Effect of Phonology and Semantics

4.1 Introduction

The aim of the current study was to explore the effect of phonology and semantics on the learning of the Spanish present tense. In studies 1 and 2, we found that the relative frequency of the different endings in an inflectional paradigm had an effect on learning, but could not explain the pattern of defaulting.

Our goal in this third study was to see whether the bias towards the 3sg form might be explicable in terms of two other properties – phonology and semantics. Aguado-Orea and Pine (2015) demonstrate that the 3sg form is the most frequent in the input (over 50% of the input produced is in the 3sg form) and this is then replicated in the child's output. However, it is important to acknowledge that the 3sg inflection ("-*a*") is also the most phonologically simple to produce. Moreover, as we will explain, it also has the most phonological neighbours (i.e *"-as" and "-an"*) within the paradigm. A further possibility is that there may be something semantically special about the 3sg form that makes it easier or more salient than the other inflections, thus making it more distinctive when a child is acquiring language.

In order to test how much of an effect these two factors have on language acquisition, we designed an experiment where we permuted the inflections across meaning for the Spanish inflections learned in the previous experiments. By permuting the endings, we are able to dissociate phonology and semantics and thereby tease apart the effects they have on inflectional learning.

4.2 Experiment 3

4.2.1 Method

4.2.1.1 Participants

20 new monolingual English-speaking participants were recruited for this study and all 20 were assigned to the Randomisation group. The sample consisted of 15 females and 5 males with a mean age of 20;6 (range 18;4 - 35;2). The participants were all first-year undergraduate students at the University of Liverpool and were recruited using the Experiment Participant Recruitment (EPR) Scheme where they gained course credit for taking part in the study. This group was compared to the 25 participants in the uniform condition (with 80 training trials; see Experiment 2 for details), which for clarity we will here call the "canonical" condition since it follows the canonical mapping of form to meaning.

4.2.1.2 Design and Materials

Each participant in the randomization condition was assigned a unique mapping of form to meaning. Table 4.1 shows the mapping for 4 example participants (A-D) in this condition, alongside the canonical mapping for comparison.

Table 4.1 - The randomised phonological and semantic verb endings for four hypothetical participants.

Participant	3sg	1pl	3pl	2sg	1sg
Canonical Participant	a	amos	an	as	O
Participant A	amos	An	as	o	a
Participant B	an	As	o	a	amos
Participant C	as	O	a	amos	an
Participant D	o	As	amos	an	a

The study employed a mixed-factorial design for two separate sets of analyses. The between subjects factor for the first set of analyses was the condition to which each participant was assigned (Randomised or Canonical; 80 Training Trials); the within subjects factor was target phonological verb ending (*-a, -amos, -an, -as, -o*; based on each participant's unique mapping from forms to meanings). Participants' production of phonological verb endings was the outcome variable and we derive two binary trial-by-trial dependent variables from this - whether the phonological verb ending produced was correct, and, when an error was made, whether the errorfully produced ending was the "-*a*" form.

For the second set of analyses, we again employed a mixed-factorial design where the between subjects factor was again condition (Randomisation or Canonical) but the within-subjects factor was target meaning/context (3sg, 1pl, 3pl, 2sg, 1sg) rather than phonological form. The meaning of the participant's produced ending based on the participant-specific form-meaning mapping was the coded behaviour (so, for example if the imagined participant A as described above were to produce the ending

-*o* then the outcome would be 2sg, whereas if participant B were to produce -*o* then the outcome would be 3pl), and we created two binary trial-by-trial dependent variables from this – whether the correct semantic verb ending was produced and, when an error was made, whether the errorfully produced semantic ending was that participant's "3sg" form.

During the training phase in the Randomisation condition, the participants were exposed to 80 training trials. Each trial consisted of 2 animations; the first animation paired the action and the description of action; the second animation displayed only the action. For the second animation, the participant was expected to repeat the verb. Table 4.2 shows the relative and proportional frequency of the different person and number combinations in the Randomisation condition. Within this condition, there was no bias towards any semantic or phonological form.

Table 4.2 - The frequency distribution of the person/number combinations in the Randomisation condition received on each training day.

	3sg	2sg	1pl	3pl	1sg
Rasg-		2			2
Escal-			2	2	
Dibuj-		2	2		
Salt-				2	2
And-	2		2	2	2
Empuj-	2		2	2	2
Golpe-	2	2		2	2
Pate-	2	2		2	2
Dispar-	2	2	2		2
Bail-	2	2	2		2
Cant-	2	2	2	2	
Tir-	2	2	2	2	
FREQUENCY OF PERSON/NUMBER FORM	16	16	16	16	16
PROPORTION OF TOTAL TRIALS	0.2	0.2	0.2	0.2	0.2

During the testing phase, participants viewed 40 trials with each of the unique videos being seen once. However, during this phase they only viewed the animations once without a description and would attempt to describe what they saw (rather than seeing a learning animation and then an imitation animation). Table 4.3 shows the relative frequency of the forms presented during the testing phase of the randomisation condition. During the testing phase, the unseen verb-ending combinations were presented to require the participants to generalize their knowledge to novel verb-ending combinations. This process was repeated 3 times on consecutive days.

Table 4.3 - The frequency distribution of the verb/ending combinations in the test phase (40 test trials per day).

	3sg	2sg	1pl	3pl	1sg
And-	1	1	1	1	1
Empuj-	1	1	1	1	1
Golpe-	1	1	1	1	1
Pate-	1	1	1	1	1
Dispar-	1	1	1	1	1
Bail-	1	1	1	1	1
Cant-	1	1	1	1	1
Tir-	1	1	1	1	1
FREQUENCY OF PERSON/NUMBER FORM	8	8	8	8	8

4.2.1.3 Procedure

The procedure was the same as for Experiment 1 and 2.

4.2.1.4 Transcription and Coding

Spanish responses were audio-recorded and then transcribed by the researcher. Responses were transcribed fully first and then separated into the stem and ending.

Correct responses were codes as 1 and incorrect responses were codes as 0. Responses were only included if a full verb was produced (i.e. both a stem and an ending from the set of endings seen in training, not simply a stem or an ending). If either the stem or ending was missing from the response, the trial was omitted from the analysis. 250 trials were excluded, out of 1799 trials, as follows; not producing an eligible stem (53), not producing an eligible inflection (5), not providing an answer (166) and not producing an audible response (24). 20% of the participants within each condition were coded by a second coder who was blind to the condition. High agreement was obtained (agreement = 92.5%, kappa = .908).

4.2.2 Results

For this study, we performed 4 analyses. We built 4 mixed effects logistic regression models to assess: (1) the effect of our predictors on the correct production of different verb endings across different meanings/contexts of use and (2) the effect of our predictors on the correct production across different phonological target forms. For the second part of the analyses, we assessed (3) the effect of our predictors on the rate of 3rd person singular errors produced (when a participant produced the form mapped to 3sg for them at training – "-*a*" for canonical condition participants but most often other forms in the randomised condition - when another form is required) and 4) the effect of our predictors on the rate of "-*a*" errors produced (when a participant produced an "-*a*" form when another was required). Tables 4.4 and 4.5 show the rates of production for the different person and number combinations within the Randomisation study. The Table shows the rate for both the semantic forms and the phonological forms in order to demonstrate the differences between the factors we are attempting to disentangle. Table 4.4 displays the rate at which

each form was produced for the canonical condition (e.g. the uniform condition from Experiment 1 and 2) and then the rate at which each form was produced when defined separately by phonological form and by semantics.

Table 4.4 - The proportion of coded trials (±SD) on Day 3 in which each verb ending was produced.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Canonical	0.33(±.18)	0.18(±.07)	0.18(±.09)	0.16(±.10)	0.15(±.08)
Phonology	0.37(±.14)	0.17(±.11)	0.07(±.08)	0.19(±.10)	0.19(±.09)
Semantics	0.18(±.17)	0.19(±.10)	0.18(±.14)	0.25(±.18)	0.20(±.10)

Table 4.5 - The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Canonical	0.29(±.14)	0.17(±.08)	0.16(±.08)	0.14(±.08)	0.13(±.07)
Phonology	0.30(±.11)	0.14(±.10)	0.07(±.08)	0.15(±.07)	0.16(±10)
Semantics	0.15(±.14)	0.15(±.12)	0.16(±.08)	0.21(±.15)	0.17(±.10)

4.2.3 Phonological Effects – Correct Production on Day 3

The main aim of this analysis was to establish what effect Condition (Randomisation or Canonical) and Phonological Target Ending had on the correct production of verb endings. For this analysis, 1548 responses were analysed. We fitted a mixed effects logistic regression model that included Condition and the Target Ending as predictors, along with an interaction between them. Participant was included as a

random effect on the intercept and Target Ending slope. A target-stem-varying intercept was not included in the model due to its resulting in a singular fit (in line with Barr et al., 2013). Likelihood ratio tests were used to assess the fit of the model and whether each predictor term had explanatory value – whether removing these terms damaged the fit of the model to the data. Removing the different terms allowed us to measure how much explanatory value the different terms provided. If removing a term produced a model with at least as good a fit, then the term was excluded from out final model.

The experimental design contained a test of whether the participants could generalise to new verb-ending combinations. To test the effect that exposure to a particular verb-ending combination had on the data, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that the model with the "Unseen" trials included as a fixed effect did not provide a significantly better fit to the data than a Null model which included no fixed effects ($X^2(1) = 1.40$, $p=.237$). Thus, the "Unseen" variable was omitted from the models.

A full model including fixed effects of Target Ending, Condition and an interaction between the two was built (Target Ending * Condition). This model was compared to a model where the interaction between Condition and the Target Ending Variable was removed (Target Ending + Condition). The full model provided a significantly better fit to the data than the reduced model ($X^2(4) = 17.27$, $p=.002$). This suggests that the interaction usefully explains variance in the model. The full model (with an interaction) was then compared to a model with only Target Ending included as a

fixed effect (i.e. Condition had been removed). The full model again provided a significantly better fit than a model where only Target Ending was included ($X^2(5) = 35.59$, $p<.001$). The model with an interaction was also compared to a model where only Condition was included (i.e. Target Ending removed). Again, the model with the interaction included provided a significantly better fit to the data than a model that solely included Condition ($X^2(8) = 53.17$, $p<.001$). Finally, the full model was compared to a null model where Condition and Target Ending had been removed. The full model provided a significantly better fit to the data than the null model, ($X^2(9) = 72.43$, $p<.001$). It was therefore concluded that a model which included Condition, Target Ending and an Interaction between them, gives the best fit for the data.

Figure 4.1 shows the parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "Rethinking" package (McElreath, 2015). The error bars show the credible intervals for each parameter and show its mean value which, as is conventional we treat as its estimate. Where a mean parameter values lies outside the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

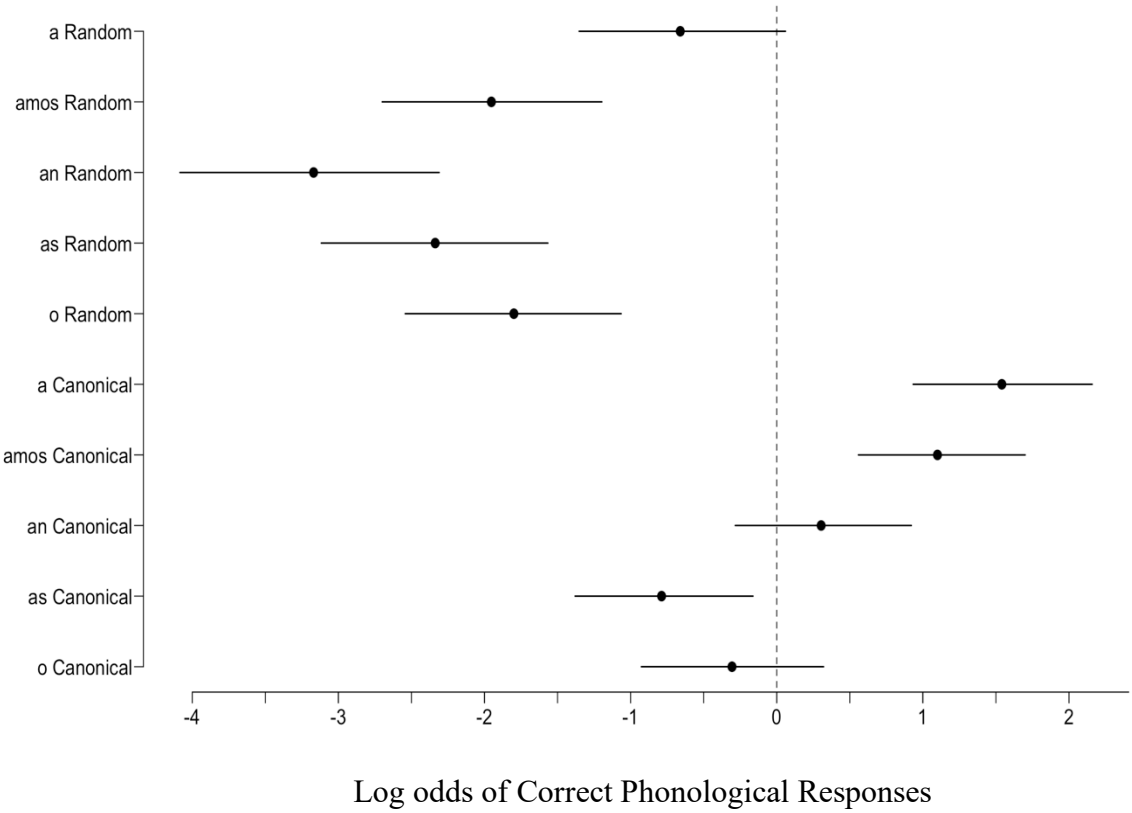

Figure 4.1 - The log odds and confidence intervals (95%) for the correct phonological verb endings produced on Day 3 (where 0 corresponds to 50% accuracy).

Figure 4.1 shows that the best performance in both conditions was for the "-a" ending. The novel finding here is that this occurs in the randomisation condition, where participants were significantly less likely to produce "-*amos*", "-*an*", "-*as*" and "-*o*" correctly than "-*a*". For example, the rate of correct production for the *"-a"* form was 39% compared to that for "-*amos*" at 16% and "-*an*" at 17%. The model also revealed that performance overall was better in the canonical condition although there was a significant interaction between Condition and Target Ending, indicating that the strength of this effect varies somewhat by ending.

4.2.4 Phonological Effects – "a" error on Day 3

The aim of this analysis was to assess the effect of Condition and Target ending on the production of "-*a*" errors. We defined an "-*a*" error as one where "-*a*" is produced instead of the target form, for example, producing *"canta"* rather than *"cantamos",* where "-*amos*" was the target ending phonologically. In the randomization condition this means forms produced for different target videos/meanings for different participants. 797 of all Day 3 productions where "-*a*" was not the target were analysed using mixed effects logistic regression models. "-*a"* errors were coded as 1 and all other responses were coded as 0. The random effects structure included a random effect of participant on the intercept and the target ending slope and an effect of target stem on the intercept.

A full model where Condition, Target Ending and an interaction between the two were included as fixed effects (Target Ending * Condition) was compared to a model where the interaction was removed (Target Ending + Condition). The full model did not provide a significantly better fit to the data than the model with the interaction removed ($X^2(3) = 3.23$, p=.358). The reduced model (Target Ending + Condition) was then compared to a model where only Condition was included as a predictor variable. The model which included Target Ending and Condition as fixed effects provided a significantly better fit to the data than a model with only Condition included ($X^2(3) = 10.31$, p=.016). The model which included Target Ending and Condition was also compared to a model with only Target Ending included as a Predictor Variable. The model with both Target Ending and Condition did not provide a significantly better fit to the data than a model with only Target Ending included ($X^2(1) = 0.14$, p=.712). Finally, the model which included only Target

Ending as a fixed effect was compared to a null model where both Target Ending and Condition had been removed as fixed effects. The model which included Target Ending as a fixed effect provided a significantly better fit to the data than a model with both predictor variables removed ($X^2(2) = 10.34$, $p=.005$). It can therefore be concluded that a model with Target Ending included provided the best possible fit to the data.

Figure 4.2 shows the parameters derived from our chosen model, generated from the model comparison. We again ran a Bayesian version of the model using STAN via the "rethinking" package. The error bars show the 95% credible intervals for each parameter and points show the mean value which we will treat as its estimate. Where a mean parameter value lies outside of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

In order to aid the interpretation of the plot we chose to exclude the non-significant items (A table displaying the Bayesian Model including predictor variables Target Ending, Condition and the interaction between the two is included in Appendix 1 – Table 6).

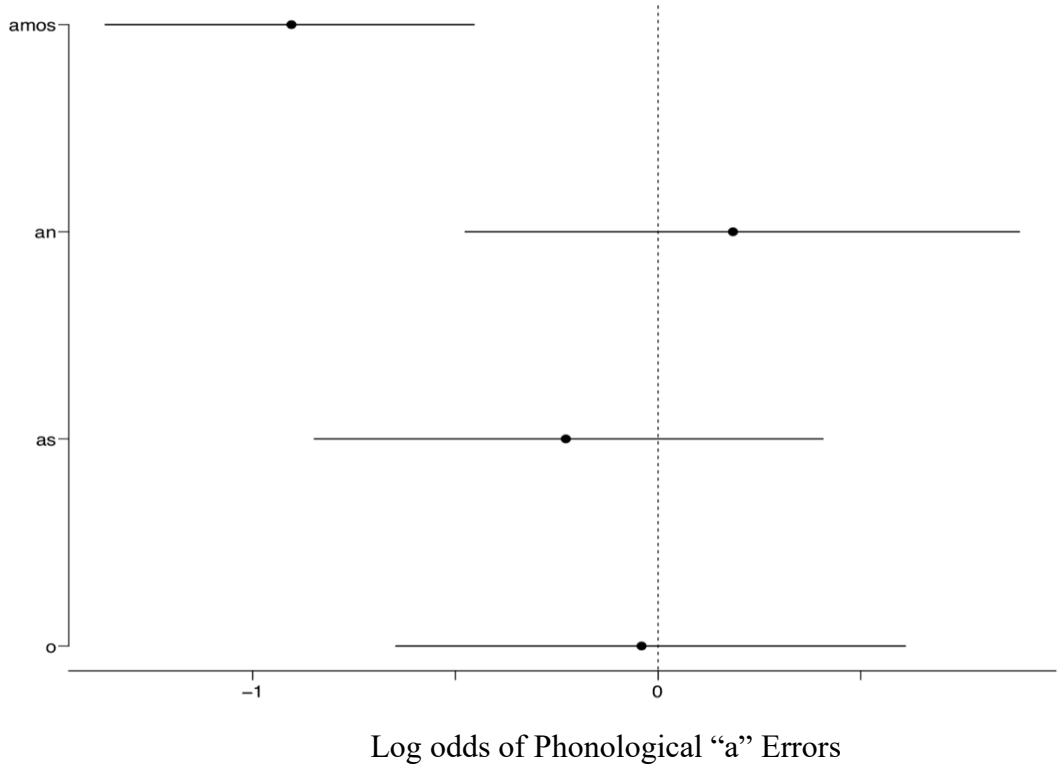

Figure 4.2 - The log odds and confidence intervals (95%) for "-a" errors produced on Day 3 (where 0 corresponds to 50% accuracy)

Figure 4.2 demonstrates that within both condition participants were significantly less likely to make a phonological “-*a*” error when “-*amos*” was the target than they were when any other ending was the target. The largest number of “-a” errors occurs when “-an” is the target.

4.2.5 Semantic Effects – Correct Production on Day 3

The permutation of person and number combinations within the Randomisation Condition (disentangling the phonological and semantic effects) allowed us to establish what effect these two linguistic factors have on language acquisition. We therefore analysed the effect of Condition and Target Ending on the correct semantic production of verb inflections. For this analysis, 1547 responses were analysed. We

fitted a mixed effects logistic regression model that included an interaction between Condition and the Semantic Target Ending variable.

Likelihood ratio tests were used to assess the fit of the model and whether each predictor term had explanatory value – whether removing these terms damaged the fit of the model to the data. Removing the different terms allowed us to measure how much explanatory value the different terms provided. If removing a term produced a model with at least as good a fit, then that term was excluded from out final model. Participant identity was included as a random effect on the intercept and on the the slope for Target Ending. A Target Stem varying intercept was omitted due to providing a singular fit (in line with Barr et al., 2013). The semantic target ending variable here was the semantic target for each participant (3rd Person Singular, 1st Person Plural, 3rd Person Plural, 2nd Person Singular and 1st Person Singular) as distinct from the phonological endings we analysed above. In the canonical condition, phonological form and meaning have a constant relationship, whereas in the randomization condition they are dissociated.

The experimental design allowed a test of whether the participants could generalise to new verb-ending combinations. To examine the effect exposure to a particular verb-ending combination had on the data, we compared a model with exposure status to a null model with no fixed effects. The results showed that a model with the "Unseen" trials as a fixed effect did not provide a significantly better fit to the data than a null model, ($X^2(1) = 1.69$, $p=.193$). Thus, the Unseen variable was excluded from subsequent models.

A full model where fixed effects of Target Ending (semantically defined), Condition and an interaction between the two was built first (Target Ending * Condition). This model was compared to a model where the interaction was removed from the model (Target Ending + Condition). The full model provided a significantly better fit to the data than the model with the interaction removed ($X^2(4) = 12.48$, $p = .014$). The model with the interaction included was then compared to a model with only Condition included. The full model again provided a significantly better fit to the data than a model with only Condition included ($X^2(8) = 28.29$, $p < .001$). The full model was compared to a model with only Target Ending (Semantic) included. The full model provided a significantly better fit to the data than a model with only Target Ending included ($X^2(5) = 32.05$, $p < .001$). Finally, the full model (with an interaction included) was compared to a Null model with the interaction and predictor variables removed. The full model again provided a significantly better fit for the data than a Null model ($X^2(9) = 45.90$, $p < .001$). We therefore concluded that a model which included Target Ending (semantic), Condition and an interaction between the two gives the best fit to the data.

Figure 4.3 shows the parameters derived from the chosen model. For the purposes of reporting we ran a Bayesian Version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter. The points show the mean values which we treat as its estimates; where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

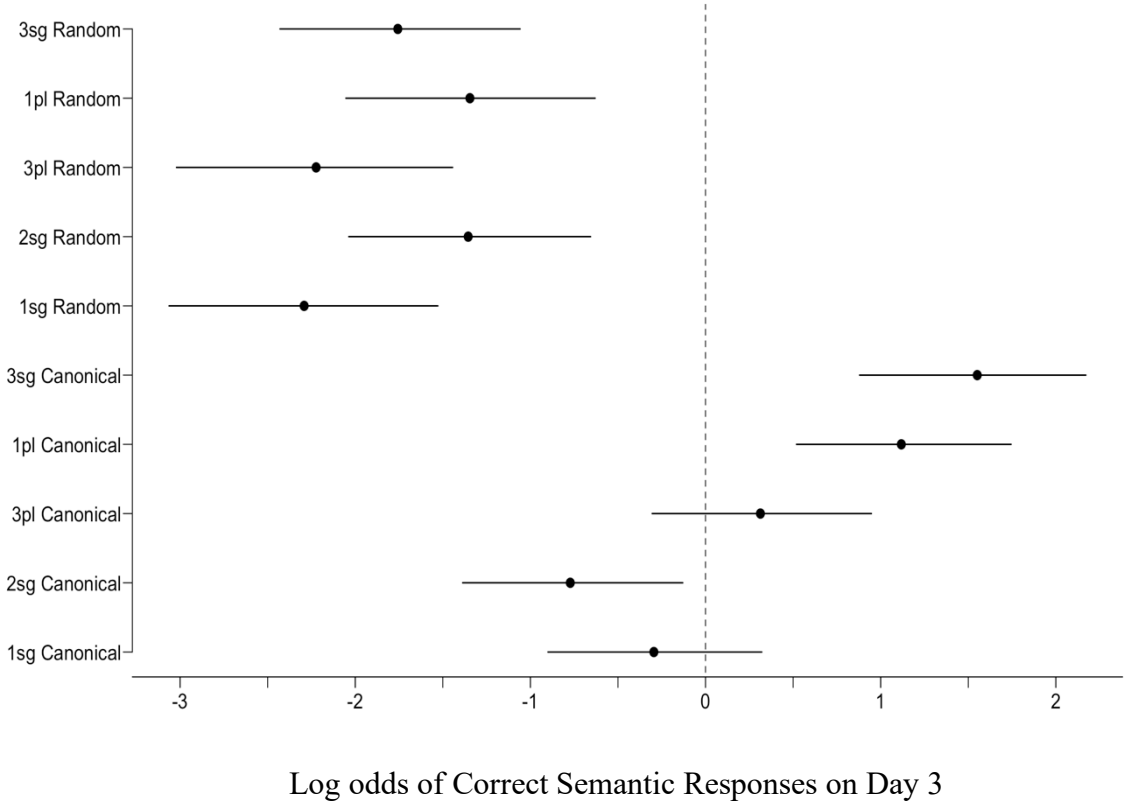

Figure 4.3 - The log odds and confidence intervals (95%) for the correct semantic verb endings produced on Day 3 (where 0 corresponds to 50% accuracy).

Figure 4.3 demonstrates that within the Randomisation condition, participants were no more likely to correctly produce any of the forms (including the 3sg form) than any of the other forms. In the canonical condition, participants were more accurate than in the randomisation condition for all forms and more likely to correctly produce a 3sg form than any of the others.

4.2.6 Semantic Effects – 3rd Person Singular Errors Produced on Day 3

The second part of the semantic analysis allowed us to assess what effect Condition and Target Ending (Semantic) has on the production of 3sg errors. This type of error is produced in our study when participants produced a 3sg form when another

semantic form was the target. For example, producing“bail-3sg” instead of “bail-1pl” when a “1pl” form is required. Because each participant in the randomization condition had a unique mapping of form to meaning, the phonological form that corresponds to the 3sg meaning differs across participants. 767 of all non-3sg trials on Day 3 were analysed using mixed effect logistic regression models. The random effects structure included a random effect of participant on the intercept. A model that additionally included a random effect of participant on the Target Ending slope gave a singular fit, suggesting that the model was overparameterised. A random effect of target stem on the intercept was considered but not included in the model due to not improving the fit to the data when added as a random slope and again providing a singular fit. These terms were therefore omitted in line with Barr et al. (2013). The model comparison was the same as for the previous analyses.

A full model which included target ending, condition and an interaction between the two (Target Ending * Condition) was compared to a model with the interaction removed (Target Ending + Condition). The full model provided a significantly better fit to the data than a model with the target removed ($X^2(3) = 15.60$, p=.001). The full model was then compared to a model with only condition included as a predictor variable. The full model again provided a significantly better fit to the data than the model with only condition included ($X^2(6) = 18.88$, p=.004). The full model with an interaction included was compared to a model with only Target Ending included as a predictor variable. Again, the full model provided a significantly better fit to the data ($X^2(4) = 31.76$, p=.001). Finally, the full model was compared to a Null model where the predictor variables had been removed. The full model provided a significantly better fit to the data than a Null model ($X^2(7) = 35.52$, p=.001). Again,

the full model with Target Ending, Condition and an interaction between the two fixed effects was the chosen model.

Figure 4.4 shows the parameters derived from our final chosen model. For the purposes of reporting, we again ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional, we treat as its estimate. Where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say the former is different from the latter at $\alpha = 0.05$ (two-tailed).

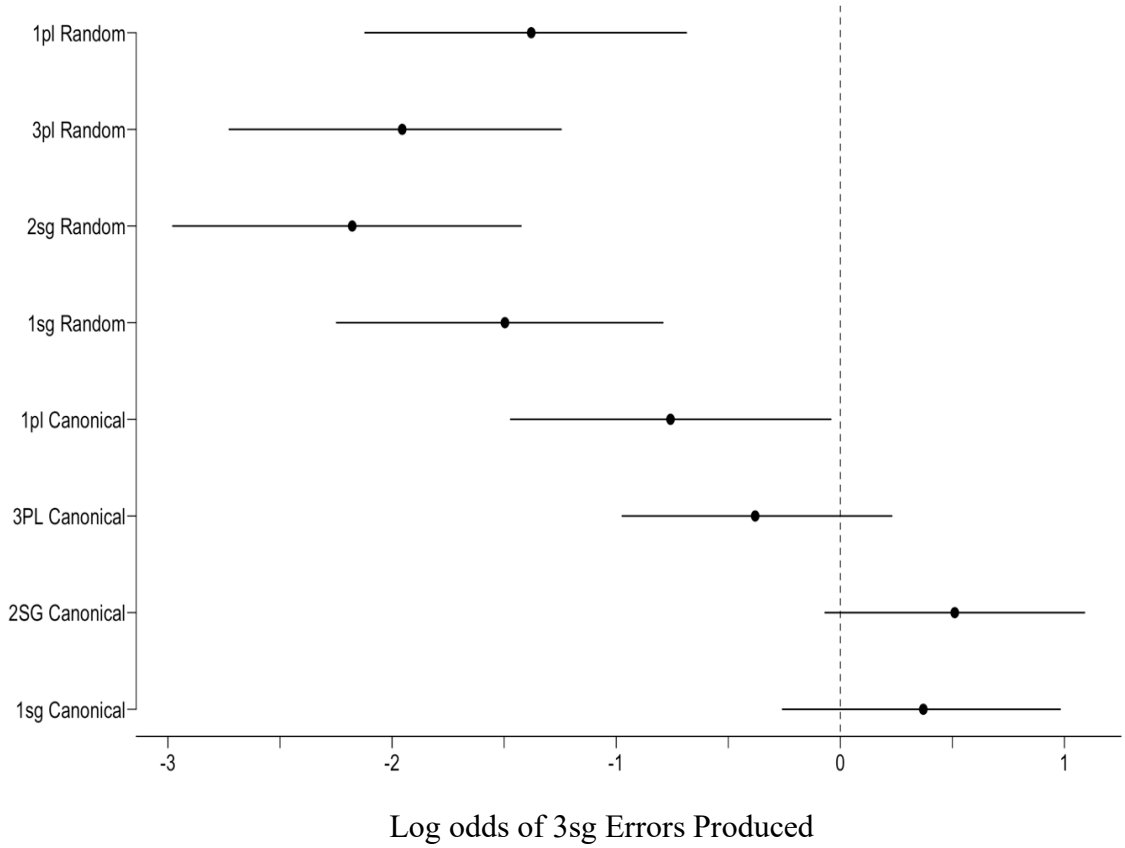

Figure 4.4 - The log odds and confidence intervals (95%) for the 3rd person singular error on Day 3.

Figure 4.4 shows that in the randomisation condition, the rate of 3sg errors was generally low (at or below 20% production for all endings) and did not differ much across the different targets. However, in the canonical condition, the rate of 3sg errors was much higher, and participants were more likely to produce a 3sg error when the target was a singular form. The 3sg error rate when the 2sg form was the target was 56% and the 3sg error rate when the 1sg form was the target was 52%. These results, specifically the absence of a tendency to make more 3sg errors on singular forms in the randomisation condition, suggest that semantics on its own cannot explain the high rate of 3sg errors in Spanish.

4.2.7 Summary of the Phonological and Semantic Effects

To summarise, we see an effect of phonology but not of semantics in the randomization condition. It is interesting then to look at the rates of production of the different forms regardless of context. These were reported above but we extract the important information below for convenience of reference. When we do so, we see a bias in production towards the "-*a*"/3sg form in the canonical condition and the "-*a*" form in the randomisation condition, but no bias in production towards the 3sg form. Table 4.6 shows the phonological rates of production for both the canonical and the randomisation condition. The Table demonstrates that the rate of production for the "-*a*" form is higher than the other forms suggesting that participants are more likely to produce that form than the other available inflections.

Table 4.6 – The proportion of coded trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a	amos	An	as	o
Canonical	0.33(±.18)	0.18(±.07)	0.18(±.09)	0.16(±.10)	0.15(±.08)
Phonology	0.37(±.14)	0.17(±.11)	0.07(±.08)	0.19(±.10)	0.19(±.09)

Table 4.7 shows the semantic rates of production for the canonical and randomisation conditions. The rates of production for the randomisation condition suggest that there is no real preference for any semantic form.

Table 4.7 - The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced.

	3sg	1pl	3pl	2sg	1sg
Canonical	0.33(±.18)	0.18(±.07)	0.18(±.09)	0.16(±.10)	0.15(±.08)
Semantics	0.18(±.17)	0.19(±.10)	0.18(±.14)	0.25(±.18)	0.20(±.10)

4.3 How do participants make use of the cues?

In order to assess how the participants are using the person and number cues, we fitted Bayesian Multinomial Regressions to the data in the Randomisation study. The models are created using an R tool called *brms* (Burker, 2017), which, in turn, makes use of STAN. Default weakly-informative priors were used in all models. The 3sg form was set to be the reference class.

4.4 Results

Using the data from Day 3 of testing, we fitted Bayesian multinomial regression models that included Person and Number features as individual predictor variables. The outcome in these models was the meaning/context that corresponded to the form produced for the participant in question (so that if a participant produced "-*a*" and for that participant "-*a*" was mapped to 1pl then 1pl was the outcome for that trial). Participant identity was included as a random effect on the intercept.

Leave-one-out cross validation (LOOIC; Gefland, Dey & Chang, 1992; Gefland, 1996) was used to assess the fit of the model and whether each predictor term had explanatory value. Removing the different terms allowed us to measure how much explanatory value the different terms provided. If removing a term produced a model

with at least as good a fit, then that term was excluded from our final model. Three separate models were constructed in order to assess the effect of Person and Number features within the randomisation condition.

4.4.1 Randomisation Condition

A model which included both Person and Number as fixed effects (LOOIC = 4265.88, SE = 55.02) provided a better fit to the data than a model with only Person included (LOOIC = 4497.49, SE = 44.44), a model with only Number included (LOOIC = 4402.80, SE = 46.88) and a null model, where no fixed effects were present (LOOIC = 4641.29, SE = 36.93). The selected model was thus a model with both Person and Number included as fixed effects. This suggests that participants use both the person and number features within the randomisation condition to learn the different forms.

The multinomial regression analysis demonstrates that participants in the Randomisation study could use the different morphological features to learn the semantic elements of the different inflections. Nonetheless the participants in the Randomisation condition produced a low overall accuracy rate (19% overall accuracy rate – no different from chance). It is thus particularly informative to see how their knowledge of cues is reflected in their errors. This is revealed in Figures 4.5-4.9, which show the rate with which each of the five different inflections were erroneously produced in each of the four other contexts. For illustrative purposes, the randomisation condition is compared to the canonical condition.

Figures 4.5 – 4.9 demonstrate that the rate of errors is relatively even across the different inflections. However, for some inflections, the results show that participants do make errors by classifying on the basis of a single feature – the number cue. For example, Figure 4.6 shows that participants are more likely to produce a 1st person plural error when the target is another plural form.

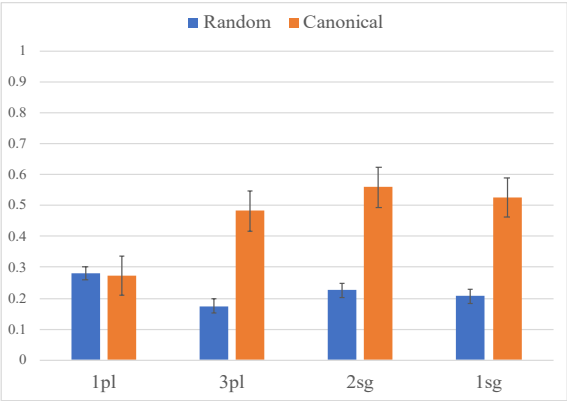

Figure 4.5 – Proportion of 3sg Errors produced on Day 3.

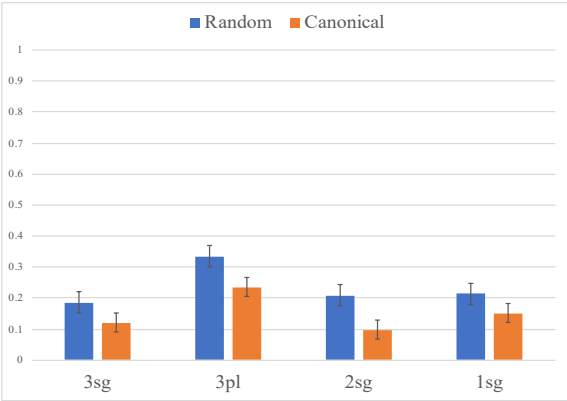

Figure 4.6 – Proportion of 1pl Errors produced on Day 3.

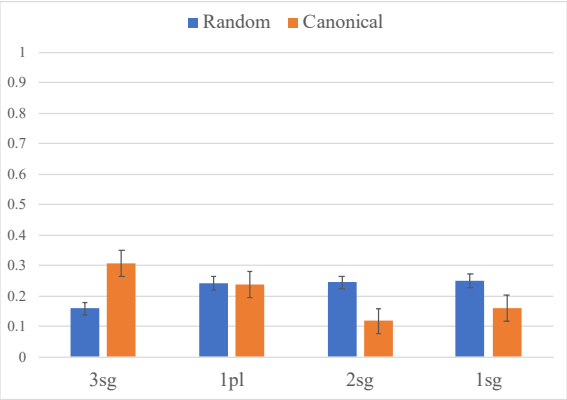

Figure 4.7 – Proportion of 3pl Errors produced on Day 3.

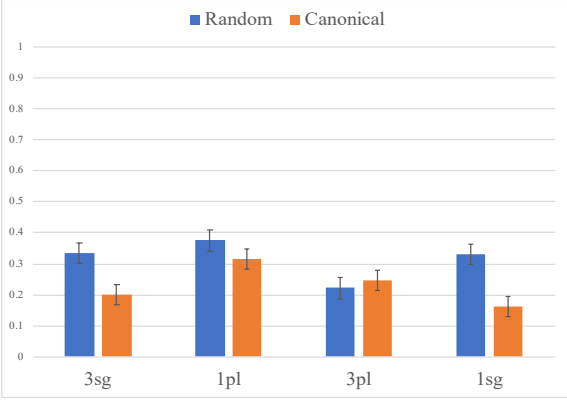

Figure 4.8 – Proportion of 2sg Errors produced on Day 3.

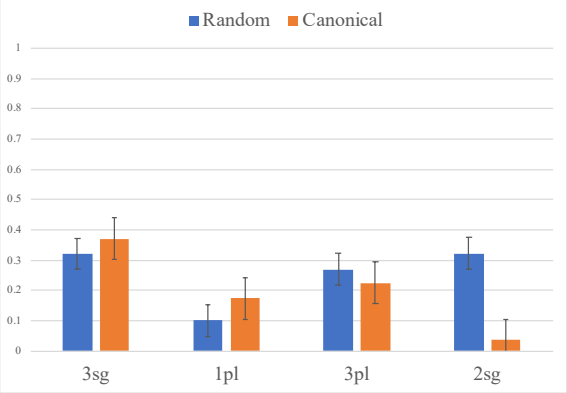

Figure 4.9 – Proportion of 1sg Errors produced on Day 3.

Figure 4.5 - 4.9 - Plots demonstrating where errors are made across each inflection learned on Day 3.

4.5 Discussion

This study aimed to assess to what extent phonology and semantics affect the acquisition of different Spanish present tense inflected forms. The goal was to shed light on the effects observed in experiments 1 and 2, where participants were seen to "default" but where we saw that frequency could not explain the pattern (the differences between conditions were not as predicted). In order to explore these effects, we devised a study where we permuted the different person and number combinations for each participant. This meant that each participant was taught a unique morphological paradigm, where the person and number combinations were different for each participant. Based on the proposal that it is the phonology of the different endings that drives the 3sg advantage and defaulting, we hypothesised that participants would be significantly more successful in producing the phonological "-*a*" when compared to the production of other inflected forms and also produce more "-*a*" errors (producing an "-*a*" form when another phonological target form was required). Based on the proposal that it is the meaning of the different endings that drives the 3sg advantage and defaulting, we also predicted that participants would be more successful at producing 3sg forms and also produce more 3sg errors (producing a 3sg form when another semantic form is required)

In Experiment 3, we conducted 4 analyses. Experiment 3 focused on phonological and semantic effects on production. In terms of successful phonological production participants were significantly more likely to correctly produce "-*a*" than the other inflected forms taught during the study. This suggests that participants are significantly more likely to produce the most phonologically simple form (in Spanish this is the "-*a*" form). However, within the canonical condition participants were

significantly more likely to produce "-*a*" when compared to the Randomisation condition.

With regards to semantic accuracy, the analysis regarding successful semantic production demonstrates that within the randomisation condition, there is no semantically driven difference in the preference for different forms. Within the randomisation condition, there is also no preference for any semantic form in terms of the production rate. However, within the canonical condition, there is a clear preference for the form that corresponds to the 3sg - participants were more likely to produce a 3sg form than 3 out of the 4 other possible forms that were learned. In the randomisation condition, we found no preference for using a 3sg form when another semantic form was the target (e.g. producing a 3sg form when 1pl was the target form). Within the canonical condition, participants were more likely to produce a 3sg error when the target form was a singular (e.g. 1st or 2nd person).

In terms of stressed and unstressed inflections, the lack of confusability between -*a* and -*amos* could suggest that the participants in our study were sensitive to the difference between the stressed and unstressed forms (i.e. -*amos* being stressed). However, participants were more likely to confuse "-*a*" with "-*an*" rather than "-*o"*. Instead, these results suggest that "-*a*" errors are driven by phonological simplicity and similarity. The latter effect (of similarity) is demonstrated by participants being more likely to produce an "-*a*" error when "-*an*" was the target form. This is due to "-*an*" being the most phonologically confusable form (see Figure 4.2). The reason that this might be so is shown in Figure 4.10.

Figure 4.10 is a plot of phonological space which shows the different inflections in the *-ar* conjugation in Spanish. This is built using a Levenshtein distance measure. The Levenshtein measure is a metric that measures the distance between any two sequences of symbols. The distance between two strings is derived from the number and type of operations needed to turn one string into the other. The different operations are three-fold; deletions, insertions and substitutions. Following Kessler (1995), deletions and insertions had a value of 1, where the distance between characters is the number of operations performed. For example, the small distance between "*-a*" and "*-as*" is 1 due to having 1 insertion. The larger distance between "*-a*" and "*-amos*" is 3 as a result of having 3 insertions. However, substitutions have a value of 2 due to the operation requiring both a deletion and an insertion. For example, the distance between "*-a*" and "*-o*" is 2. We then used principal component analysis to reduce these factors to two dimensions for visualization purposes, resulting in the plot seen in Figure 4.10.

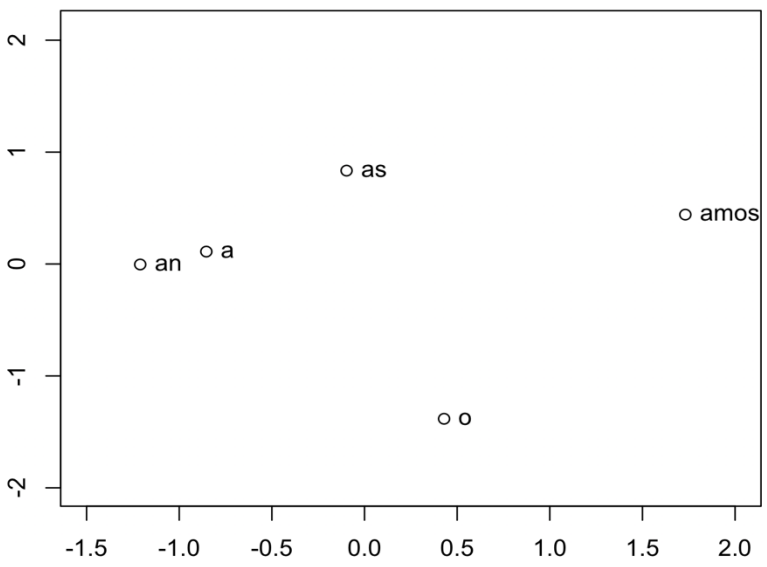

Figure 4.10 - The phonological distance between forms using the Levenshtein measure.

Figure 4.10 shows that "-*a*" and "-*an*" have the smallest pairwise phonological distance. This can explain the results found in the phonological analyses (Figure 4.2), where the most "-*a*" errors are produced when "-*an*" is the target as only 1 operation is required (1 insertion). Similarly, "-*a*" and "-*as*" are also close together due to also only requiring 1 insertion. The distance between "-*a*" and "-*o*" is larger than that of both "-*an*" and "-*as*" due to it requiring a substitution, which is a more complex operation that requires both a deletion and insertion. Phonological distance can also explain why few "-*a*" errors are made when "-*amos*" was the target form due to participants needing to perform more operations (3 insertions).

Our Bayesian Multinomial Regression analysis demonstrated that the participants in the Randomisation Condition did use the Person and Number features of language. However the semantic information they have learned does not result in overall good performance as demonstrated by Figures 4.5 – 4.9.

On the basis of the above phonological analysis we propose a phonological centrality effect, where "-*a*" is the most phonologically central item within our Spanish morphological paradigm. In relation to the other forms learned, the "-*a*" form is the central form in phonological space. Therefore, when we make errors we move to the most similar form. This explains why "-*a*" is used when other forms are required.

A related claim to centrality can be made based on meaning. For inflected forms in Spanish, the 2nd person plural form is extremely rare in the input. Thus, we can treat it as effectively missing from the paradigm. As a result, the forms that have the greatest feature overlap with other forms in the paradigm are the first and third

person singulars. This is due to 3sg and 1sg forms having the most semantic overlap due to 3sg overlapping in at least 1 feature with 1sg, 2sg and 3pl forms and 1sg forms overlapping in at least one feature with 3sg, 2sg and 1pl forms. However, 2sg forms only overlap with 1sg and 3sg forms due to 2pl being rare (or absent from our paradigm) in the input. Taking these phonological and semantic factors together we propose that speakers' defaulting errors are the result of defaulting not to the most frequent form, but to the phonologically and semantically most typical.

That we are assigning a role to semantics here, when it seemed to have no effect in the randomization condition is down to the fact that performance in the Canonical condition is superior to performance in the Randomisation condition, even when analysed by phonology. Thus, phonology alone cannot explain the advantage in production for the 3sg/"-*a*" form. The only difference between the Canonical condition and the Randomisation condition is the "meaning" of the phonological forms (i.e. the phonological forms are constant between the two conditions, but the "meaning" of each phonological form is different). Thus, "meaning/semantics" while it has no evident effect on its own, increases accuracy when combined with the effect of phonology. Answering the question as to why the combination of the two properties should be more than the sum of their parts in this study will require further research.

Chapter Five

5 Assessing the Effect of a Skewed Frequency Distribution on the Learning of Spanish Person and Number Forms in Children

5.1 Introduction

This goal of this thesis is to understand a pattern observed in morphological errors made by preschool children. In order to look at this we designed a novel training-study methodology. In previous chapters we pursued this approach with adults. In this chapter we explore this paradigm with child learners. Conducting a training study with an unfamiliar morphology with the age that we are most interested in (2 and 3 years olds) is not possible. Therefore, as a way of more closely approximating the behavior of children learning their first language, we run a version of our study with school age children – specifically children aged 8 and 10 years - using the same testing paradigm. Children were taught the same Spanish verbs as the adults from experiments 1 and 2. Like for the adults, these children were taught the verbs during three training sessions and then tested on their ability to recogni
se the different forms after each testing session.

5.2 Experiment 4

5.2.1 Method

5.2.1.1 Participants

36 monolingual English-speaking children took part in the study. Seventeen children were assigned to the skewed condition and 19 were assigned to the uniform condition. The mean age of the children in this study was 9;2 (range 8;4-10;8). The children were recruited from primary schools in the Liverpool and Burnley areas of North West England. The 19 children assigned to the uniform condition were from the Burnley school. The 17 children assigned to the skewed condition were from the Liverpool schools.

5.2.1.2 Design and Materials

This was the same as for experiments 1.

5.2.1.3 Procedure

The children were tested in a quiet room at their school. They were seated directly in front of a laptop computer. The procedure was the same as Experiment 1 and 2.

5.2.1.4 Transcription and Coding

Spanish responses were audio-recorded and then transcribed by the researcher. Correct responses were coded as 1 and incorrect responses were coded as 0. When no answer was given, the cells were left blank and then removed for the analysis. Both the verb stem and the ending were coded. Responses were only included if a full verb was produced (i.e. both a verb stem from the set of verb stems heard at training and an ending were produced). If either were missing, this was not counted as a complete trial. There were 412 trials excluded, out of a total of 1394 trials, for the following reasons - not producing a full verb (i.e. not producing either an eligible stem or ending) (47), not producing an eligible ending (18), not providing an answer (360) and not producing an audible response (5). 20% of the participants within each condition were coded by a second coder who was blind to the condition. Inter-rater reliability, measured by Cohen's Kappa, demonstrated a high level of agreement between coders (agreement = 92.5%; Kappa = .875).

5.3 Results

For this study, 2 analyses were performed. We built a series of mixed effects logistic regression models to assess the effect condition had on the correct production of the

different verb endings. The second analysis assessed the effect condition had on the rate of 3rd singular errors produced. Table 5.1 shows the proportion of trials on which each response was given across the two conditions. Table 5.2 shows the proportion of trials on which participants produced an accepted response when each ending was the target, separated by condition.

Table 5.1 - The mean(±SD) response rate for each verb ending in the Skewed and Uniform Condition with 80 Training Trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.50(±.16)	0.04(±.04)	0.08(±.11)	0.29(±.19)	0.09(±.09)
Uniform	0.36(±.21)	0.04(±.07)	0.14(±.14)	0.28(±.28)	0.18(±.17)

Table 5.2 - The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.40(±.18)	0.04(±.06)	0.06(±.08)	0.23(±.17)	0.06(±.05)
Uniform	0.23(±.15)	0.03(±.05)	0.08(±.09)	0.15(±.16)	0.11(±.12)

5.3.1 Correct Verb Ending on Day 3

The aim of this analysis was to examine what effect target ending and condition had on the accuracy of production of verb endings. For the analysis of the children's data, 982 responses were analysed. Likelihood ratio tests were used to assess the fit of the models and whether each predictor term had explanatory value - whether removing these terms damaged the fit of the model to the data. If removing a term

provided a model with at least as good a fit, then that term was excluded from our final model. We produced a mixed effects logistic regression model that included an interaction between Condition and the Target Endings variable. Child identity was included as a random effect on the intercept. A random effect of participant on the slope for target ending was considered but excluded as it gave a singular fit. A random effect of target stem on the intercept was also considered but resulted in a singular fit, and so was omitted in line with Barr et al. (2013).

As in previous studies the experimental design included a test of whether the children could generalise to new verb-ending combinations. To test the effect that exposure to a particular verb-ending combination had on the data, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that the model with the 'unseen' trials as a predictor variable did not provide a significantly better fit to the data than a null model ($X^2(1) = 1.80$, $p=.179$).

A full model with the fixed effects of Target Ending, Condition and an interaction between the two (Target Ending*Condition) was compared to a model where the interaction had been removed (Target Ending + Condition). The model with the interaction included provided a significantly better fit to the data than the model which excluded the interaction ($X^2(4) = 28.58$, $p<.001$). The full model with the interaction included was compared to a model that only included target ending. Again, the full model provided a significantly better fit to the data than a model with only Target Ending included, ($X^2(5) = 29.18$, $p<.001$). The full model was compared to a model that included condition as a fixed effect. The full model again

provided a significantly better fit to the data than a model with only condition included, ($X^2(8) = 174.20$, $p<.001$). Finally, the full model was compared to a null model. The full model provided a significantly better fit to the data than the null model, ($X^2(9) = 174.88$, $p<.001$).

Figure 5.1 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

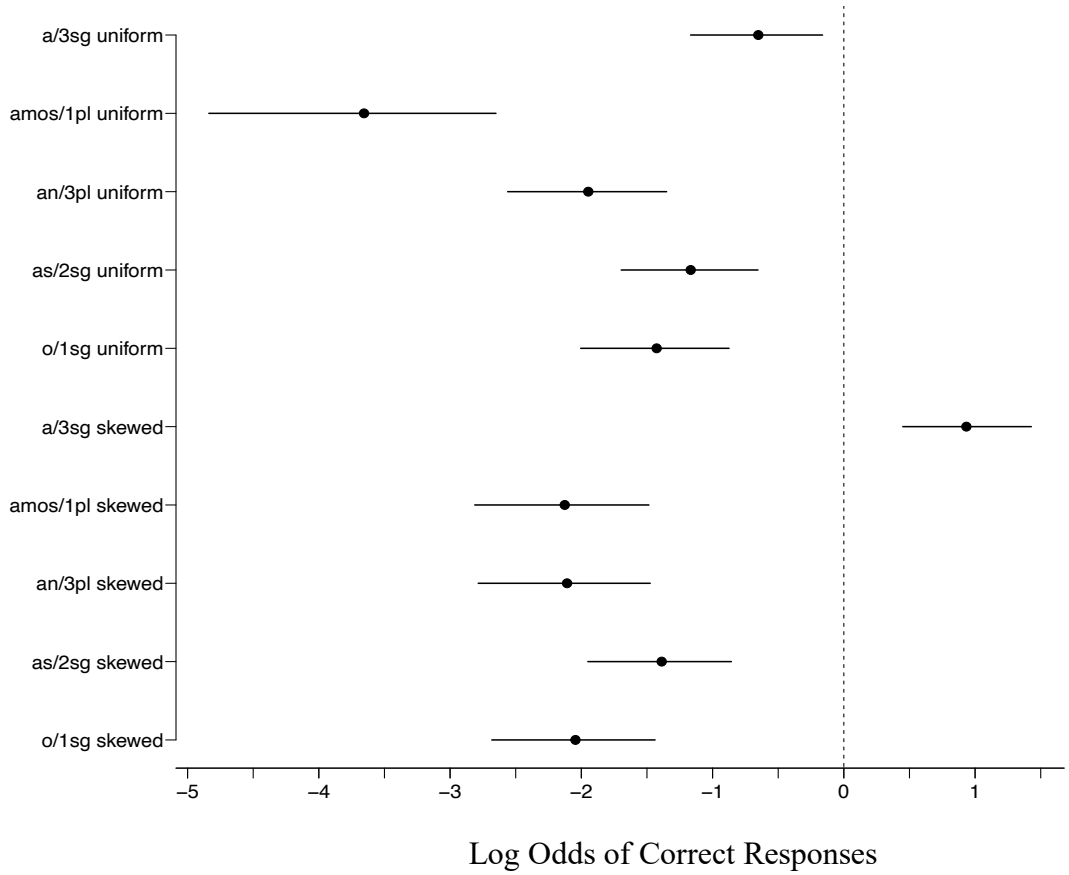

Figure 5.1 - The log odds and confidence intervals (95%) for the correct verb endings produced on Day 3.

Figure 5.1 shows that in both conditions, children were significantly less likely to produce "-*amos*", "-*an*" and "-*o*" correctly than "-*a*. The model comparison revealed that there was a significant interaction between target ending and condition - children were even more likely to produce the "-*a*" form successfully in the skewed condition, and there was significantly lower accuracy for "-*amos*" than the other forms in the uniform condition only.

5.3.2 3rd Person Singular Error on Day 3

The main aim of this analysis was to assess what effect condition and target ending have on the production of 3rd person singular errors. This type of error is where "-*a*" is produced instead of the target form, for example, producing "**baila*" instead of

"*bailamos*" when "-*amos*" was the target ending. 704 responses on Day 3 (all responses for non-3sg trials) were analysed using mixed effects logistic regression models. The random effects structure included a random effect of participant on the intercept. A model that additionally included a random effect of participant on the Target Ending slope gave a singular fit, suggesting that the model was overparameterised. A random effect of target stem on the intercept was considered but not included in the model due to not improving the fit to the data when added as a random slope and again providing a singular fit. Both these terms were therefore omitted in line with Barr et al. (2013). The model comparison was the same as for the previous analyses.

A full model which included target ending, condition and an interaction between the two (Target Ending*Condition) was compared to a model with the interaction removed (Target Ending + Condition). The full model provided a significantly better fit to the data, ($X^2(3) = 8.27$, $p=.041$). The full model was then compared to a model which only included target ending as a fixed effect. The full model again provided a significantly better fit to the data than a model which only included target ending as a predictor ($X^2(4) = 10.59$, $p=.032$). The full model was then compared to a model with only condition included as a fixed effect. The full model again provided a significantly better fit to the data than a model with only condition included, ($X^2(6) = 30.45$, $p<.001$). Finally, the full model was compared to a null model. Again, the full model provided a significantly better fit to the data than a null model, ($X^2(7) = 33.35$, $p<.001$).

Figure 5.2 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional, we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

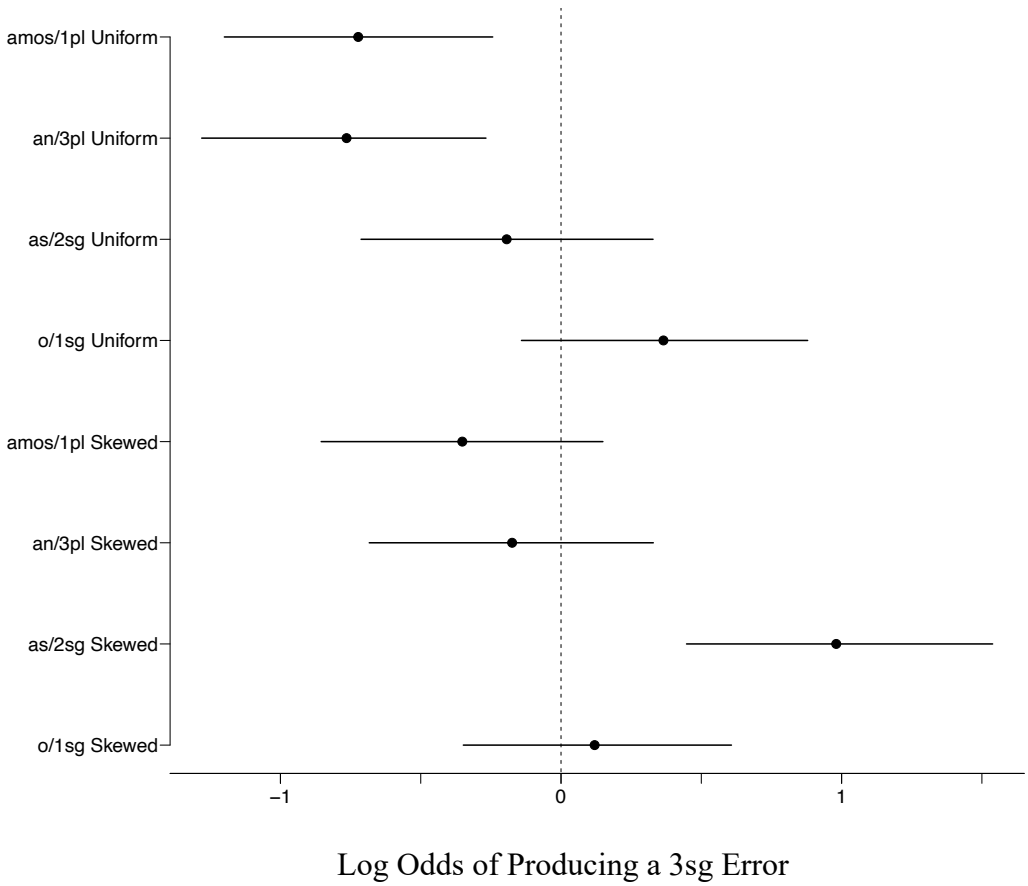

Figure 5.2 - The log odds and confidence intervals (95%) for 3rd person singular errors on Day 3.

Figure 5.2 shows that, unlike in the adult learners, but in line with our initial predictions, more 3sg errors were produced in the skewed condition when compared to the uniform condition. The Figure also shows that more 3sg errors were produced

on the singular forms when compared to the plural forms. The model also showed that there was a significant interaction between target ending and condition, reflecting the increased rate of 3sg errors when "-*as*" was the target form in the skewed condition only.

5.4 Interim Discussion

Study 4 used the same methodology as study 1 to assess the effect a frequency bias had on a child's ability to learn the different Spanish verb endings. Children were split into 2 conditions; "skewed" and "uniform" with 80 training trials. Each child was taught a series of Spanish verbs and then tested on their ability to produce the verbs again. We predicted that there would be more correct responses for 3sg forms than the other targets but only in the skewed condition. We also predicted that there would be more 3sg errors in the skewed condition when compared to the uniform condition. This is expected as a result of the bias in the input towards the 3sg form. The results partially support the first hypothesis, unlike with the adult participants; since, for the children, the extent to which 3sg performance was higher that for the other forms was greater in the skewed than the uniform condition – unlike with the adult learners, there is a significant interaction between condition and target ending. However, not consistent with our predictions, performance on the 3sg is significantly better than most other targets in the uniform as well as the skewed condition, just with a smaller difference.

The base rates of production (Table 5.1) demonstrates that, despite there being no bias toward the 3sg form in the uniform condition in the input, participants in the uniform condition are more likely to produce the 3sg form regardless of whether this

is the correct form to produce. This suggests that there is something in addition to frequency that is driving production.

For the 3sg error data, the results indicate that more 3sg errors are made in the skewed condition than in the uniform condition. Figure 5.2 also demonstrates that within both the skewed and uniform conditions, more 3sg errors were made for singular forms than for plural forms. This supports the hypothesis as more 3sg errors were made in the skewed condition when compared to the uniform condition. This is the opposite pattern to that observed in the adults. The model also shows that there is a significant interaction between Target Ending and Condition. This effect is driven by the increased likelihood of children producing 3sg errors when the target form was "-*as*" in the skewed condition, when compared to the uniform condition. This suggests that for children, the bias toward the 3sg form in the input increases the likelihood of producing a 3sg form in the wrong context, the opposite to what we saw in Chapter 3 where the bias toward the 3sg form for the adult participants, reduced the likelihood of producing a 3sg form in the wrong context.

This study has the same possible confound as Experiment 1. Achieving the desired proportions in the skewed condition required that each participant actually heard 3 of the different forms less in the skewed than in the uniform condition (i.e. -*amos,* -*an,* and -*o)*. In order to combat this, we ran the uniform condition with 40 trials rather than the 80 conditions. This was then compared to the same skewed data as in Experiment 4. This allowed us to have the same relative frequency for each of the three non-privileged endings across the two conditions.

5.5 Experiment 5

5.5.1 Method

5.5.1.1 Participants

34 monolingual English-speaking children took part in the study. Eighteen children were assigned to the uniform condition and 17 children were assigned to the skewed condition, with a mean age of 9;4 (range 8;4-10;8). The children in the skewed condition were the same as in study 3 and recruited from the Merseyside area. The children recruited for the new uniform condition (40 trials) were also recruited from various Liverpool primary schools.

5.5.1.2 Design and Materials

The design and materials were the same as for Experiment 2.

5.5.1.3 Procedure

This was the same as for Experiment 3.

5.5.1.4 Transcription and Coding

This was the same as for Experiment 1. 1400 trials were analysed and there were 504 trials excluded for the following reasons - not producing an eligible stem (49), not producing an eligible ending (33), not providing an answer (359) and not producing an audible response (63). 20% of the children from each condition were second coded by a coder who was blind to the condition. Inter-rater reliability, measured by Cohen's Kappa, demonstrated a high level of agreement between coders, (agreement = 88.4%; kappa = .827).

5.6 Results

5.6.1 Correct Productions on Day 3

In this analysis, 896 trials were analysed using mixed effects logistic regression models. The random effects structure and the model comparison process was the same as for Experiment 3. The analyses performed were the same as for Experiment 3. The skewed data reported here are those collected in Experiment 3. The uniform data, however, are new. Table 5.2 shows the proportion of trials on which each response was given across the two conditions. Table 5.3 shows the proportion of trials on which participants produced a response for each ending in each condition.

Table 5.3 – The proportion of trials (±SD) on Day 3 in which each verb ending was produced, separated by condition, where the uniform condition featured 40 Training Trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.50(±.16)	0.04(±.06)	0.08(±.11)	0.29(±.19)	0.09(±.09)
Uniform	0.26(±.21)	0.03(±.06)	0.15(±.17)	0.26(±.18)	0.29(±.20)

Table 5.4 - The proportion (±SD) of trials on Day 3 in which any response was given when each ending was the target, separated by condition, where the uniform condition featured 40 Training Trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.40(±.18)	0.04(±.06)	0.06(±.08)	0.23(±.17)	0.06(±.05)
Uniform	0.14(±.14)	0.02(±.05)	0.07(±.09)	0.11(±.09)	0.14(±.11)

The experimental design included a test of whether the children could generalise to new verb-ending combinations. To examine the effect that exposure to a particular verb-ending combination had on responses, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that the model with the 'unseen' trials as a predictor variable did not provide a significantly better fit for the data than a null model ($X^2(1) = 3e\text{-}04$, $p=.987$).

A full model which included target ending, condition and an interaction between the two (Target Ending*Condition) was compared to a model where the interaction was removed (Target Ending + Condition). The model comparison showed that the full model provided a significantly better fit to the data than the second model with the interaction removed ($X^2(5) = 28.27$, $p<.001$). The full model was then compared to a model with only target ending included as a fixed effect. The full model again provided a significantly better fit to the data than a model with only target ending included, ($X^2(5)= 28.65$, $p<.001$). The full model was then compared to a model with only condition included as a fixed effect. The full model again provided a significantly better fit to the data than a model with only condition included, ($X^2(8)$

= 181.61, p<.001). The full model was finally compared to a null model. Again, the full model provided a significantly better fit to the data, ($X^2(9) = 182.01$, p<.001).

Figure 5.3 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

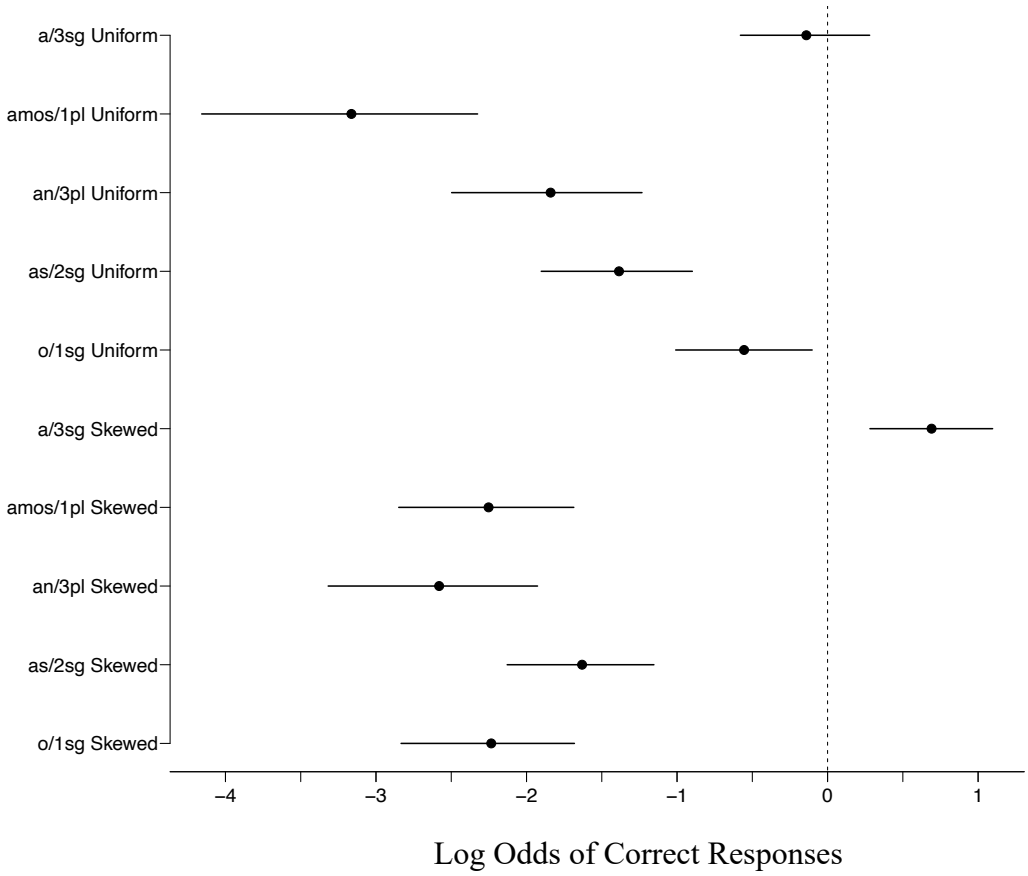

Figure 5.3 - The log odds and confidence intervals (95%) for the correct verb endings produced on Day 3.

Figure 5.3 shows that children were less likely to produce "-*amos*", "-*an*" and "-*as*" correctly when compared to the 3sg "-*a*" form in both conditions. The model indicated that that there is a significant interaction between target ending and condition. This reflects the fact that there is a stronger "-*a*" advantage in the skewed than the uniform condition, and there is significant variation between accuracy rates for the non-3sg targets in the uniform but not the skewed condition.

5.6.2 3rd Person Singular Error on Day 3

The main aim of this analysis was to assess what effect condition and target ending have on the production of 3rd person singular errors. This type of error is where "-*a"* is produced instead of the target form, for example, producing "**baila*" instead of "*bailamos*" when "-*amos*" was the target ending. 704 responses on Day 3 (all responses for non-3sg trials) were analysed using mixed effects logistic regression models. The random effects structure and the model comparison process was the same as for Experiment 4.

A full model where target ending, condition and an interaction between the two were included (Target Ending*Condition) was a compared to a model with the interaction removed (Target Ending + Condition). The model comparison demonstrated that the full model did not provide a significantly better fit to the data, ($X^2(3) = 5.78$, p=123). The model with the interaction removed (Target Ending + Condition) was then compared to a model with only target ending included as a fixed effect. The model comparison showed that the model with target ending and condition included (Target Ending + Condition) provided a significantly better fit to that data than a model with only target ending included, ($X^2(1) = 7.86$, p=.005). The model with the interaction

removed (Target Ending + Condition) was then compared to a model with only condition included as a fixed effect. The model comparison revealed that again, the model with the interaction removed (Target Ending + Condition) provided a significantly better fit to the data, ($X^2(3) = 21.58$, $p<.001$). Finally, the model with the interaction removed (Target Ending + Condition) was compared to a null model. The model with target ending and condition included as fixed effects provided a significantly better fit to the data than a null model, ($X^2(4) = 29.92$, $p<.001$).

Figure 5.4 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean value of a parameter lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

In order to aid the interpretation of the plot we chose to exclude the non-significant items (A table displaying the Bayesian Model including predictor variables Target Ending, Condition and the interaction between the two is included in Appendix 1 – Table 12).

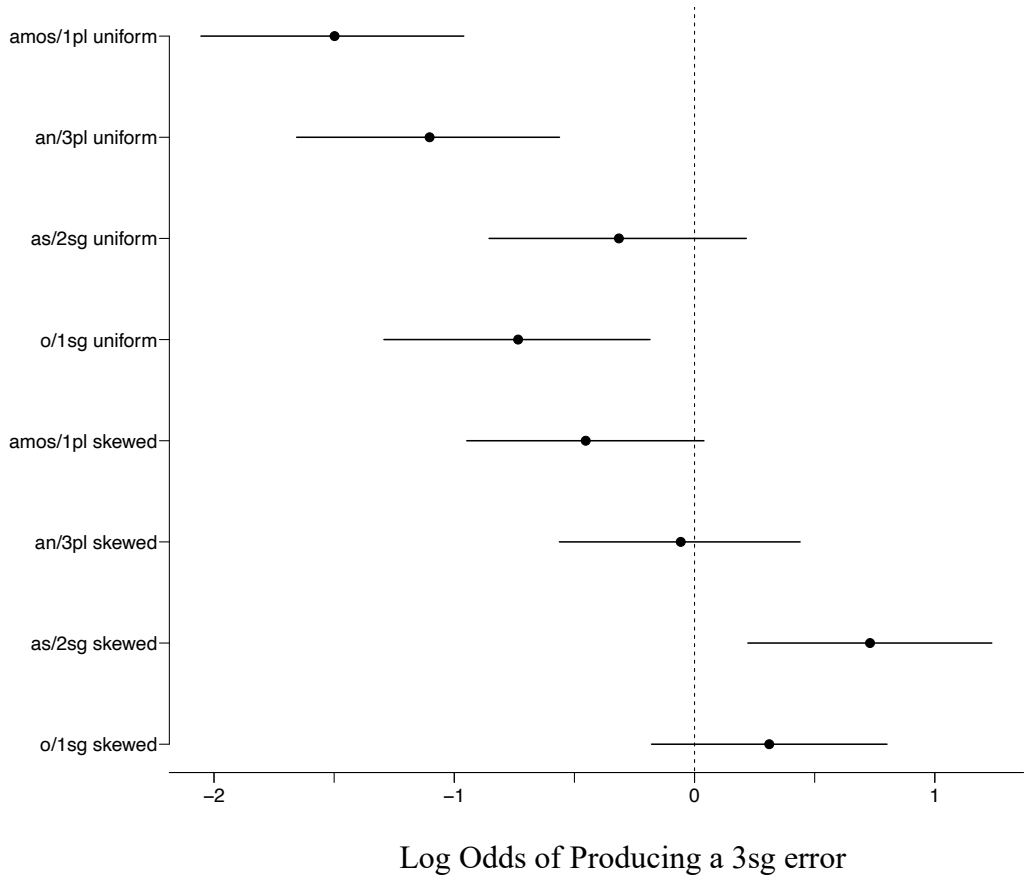

Figure 5.4 - The log odds and confidence intervals (95%) for the correct verb endings produced on Day 3.

Figure 5.4 shows that children were most likely to make 3sg errors when the singular forms ("-*as*" and "-*o*") were the targets in both conditions, and more likely to produce a 3sg error in the skewed condition compared to the uniform condition overall.

5.7 How do children make use of the cues?

Experiments 3 and 4 suggest that children were more likely to correctly produce 3rd person singular forms than they were the other forms, and more likely to replace other forms with the 3rd person singular (i.e. make 3rd person singular errors) in the

skewed condition than in the two uniform conditions (with 80 trials and 40 trials). In both experiments, children were most likely to produce a 3sg error when another singular form was required, mirroring the results of study 1 and 2, albeit with a smaller effect.

In order to assess how the participants are using the person and number cues, we fitted Bayesian Multinomial Regressions to the data from the three conditions – the uniform condition with 40 trials, the uniform conditions with 80 trials and the skewed condition. The models are created using an R tool called *brms* (Burkner, 2017) which in turn makes use of STAN. Default weakly informative priors were used in all models. The 3sg form was set to be the reference class.

5.8 Results

Using the data from Day 3, we fitted three Bayesian multinomial regression models that included Person and Number features as individual predictor variables. Participant identity was included as a random effect on the intercept. This was done separately for each of the three conditions.

Leave-one-out cross validation (LOOIC; Gefland, Dey & Chang, 1992; Gelfand, 1996) was used to assess the fit of the model and whether each predictor term had explanatory value. Removing the different terms allowed us to measure how much explanatory value the different terms provided. If removing a term produced a model with at least as good a fit, then that term was excluded from our final model. Three separate models were constructed in order to assess the effect of Person and Number features within each condition rather than comparing the conditions.

5.8.1 Uniform Condition (80 Training Trials)

A null model (which included no fixed effects) provided a significantly better fit for the data (LOOIC = 1114.09 SE = 28.79) than a model with both Person and Number included (LOOIC = 1129.13, SE = 29.67), a model with only Person included (LOOIC = 1126.84, SE = 29.72) and a model with only Number included as a predictor variable (LOOIC = 1116.46, SE = 28.79). The selected model did not include either Person or Number as fixed effects, suggesting that children within this condition did not use either of these cues to learn the different morphological forms.

5.8.2 Uniform Condition (40 Training Trials)

A model with only Number included as a fixed effect (LOOIC = 957.98, SE = 26.31) provided a significantly better fit for the data than a model with both Person and Number included (LOOIC = 958.62, SE = 27.42), a model with just person included (LOOIC = 962.70, SE = 25.91) and a Null model with no fixed effects included (LOOIC = 960.15, SE = 25.21). The selected model was a model with only Number included as a fixed effect. This suggests that in this condition, children were able to use the Number cue to learn the different forms but did not use the differences in person. This would potentially explain the results from experiments 3 and 4, where children are able to distinguish between singular forms and plural forms but are unable to distinguish between the different Person forms (e.g. 1st, 2nd and 3rd Person).

5.8.3 Skewed Condition

A model with only Number included as fixed effects (LOOIC = 1132.10, SE = 36.47) provided a significantly better fit for the data than a model with both Person

and Number included (LOOIC = 1137.37, SE = 38.38), a model with only Person included (LOOIC = 1146.37, SE = 37.67) and a Null model with no fixed effects (LOOIC = 1158.48, SE = 35.76). This suggests that within the skewed condition, children could effectively use the number cue to learn the different morphological forms.

The three multinomial analyses demonstrate that the condition the child was allocated to had a significant effect on whether they could use the different Person and Number cues to produce the different morphological forms. The results suggest that within the uniform condition (with 80 training trials) children failed to use either the Person or Number cues, whilst the children in the uniform condition (with 40 training trials) effectively used the Number cue to learn the difference between singular and plural forms, but could not discriminate between the different Person cues (e.g. identifying when a 1st person form is required or a 3rd person form is required). In the Skewed condition, children could effectively use the Number cue to learn the different inflections, where children could discriminate between different number cues, but not between different person cues.

It is also interesting to note how the children's knowledge of cues is reflected in their errors. Figures 5.5-5.9 show the rate at which each of the five different inflections were erroneously produced in each of the four other contexts across all three conditions in experiments 4 and 5. The Figures show us that there is some confusability among all the five different inflections. In some cases, participants appear to make errors by classifying on the basis of a single feature (e.g. the number feature). This is demonstrated in Figure 5.5, where children's erroneous use of the

3sg form is usually when another singular form is the target. This is also shown in Figure 5.9, where children are more likely to produce a 1sg error when a singular form is the target. The confusability between the different number features is less clear when a plural form is the target, where there are differences between the conditions in terms of erroneous use. These results potentially explain the low accuracy rates across the 3 conditions (e.g. Skewed = 28% correct, Uniform (40 trials) = 26% correct and Uniform (80 trials) = 24% correct; where chance accuracy = 20%).

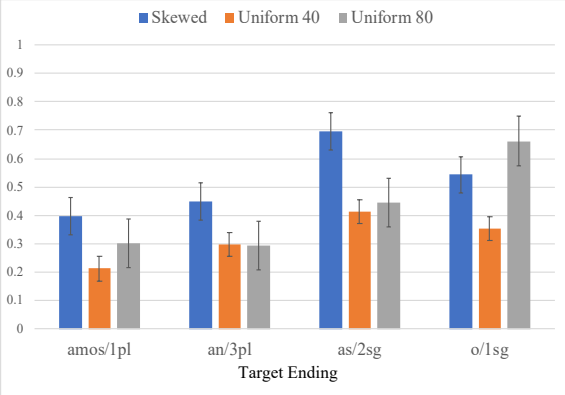

Figure 5.5 – Proportion of 3sg Errors produced on Day 3.

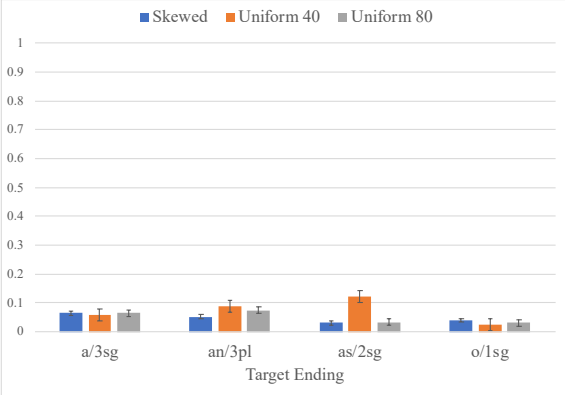

Figure 5.6 – Proportion of 1pl Errors produced on Day 3.

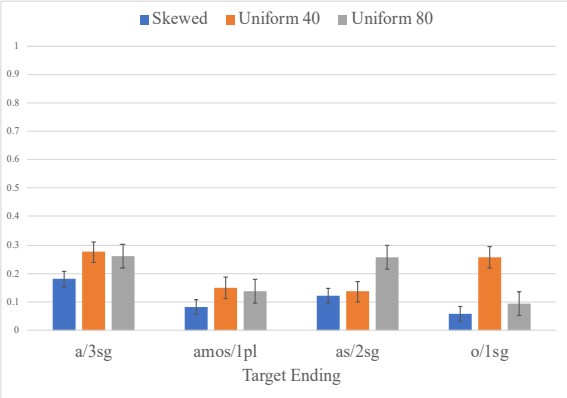

Figure 5.7 – Proportion of 3pl Errors produced on Day 3.

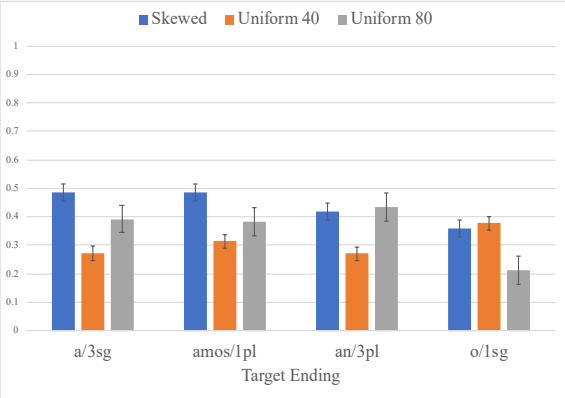

Figure 5.8 – Proportion of 2sg Errors produced on Day 3.

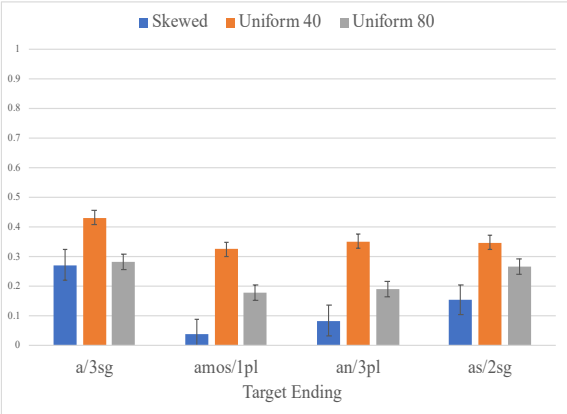

Figure 5.9 – Proportion of 1sg Errors produced on Day 3.

Figure 5.5 - 5.9 - Plots demonstrating where errors are made across each inflection learned on Day 3.

5.9 Discussion

The current study aimed to assess how children acquire different Spanish present tense inflected forms. In line with experiments 1 and 2, we hypothesised that there would be more successful productions of the 3sg than the other forms in the skewed condition but not the two uniform conditions. We also hypothesized that there would be significantly more 3sg errors produced in the skewed condition, when compared to the uniform conditions.

In Experiment 4, there was a significant interaction between target ending and condition, where the 3sg form was produced significantly more successfully than the other forms, and this effect was greater in the skewed condition than in the uniform condition (80 training trials). This partially supports our hypotheses in that frequency appears to affect 3sg production. However, the results in Experiment 4 also partially mirror the results from study 1 and 2, where the 3sg form within the uniform condition was produced significantly more successfully than 3 out of the 4 inflections. These results suggest that, while frequency is important in the acquisition of different inflections (there is, in fact, greater evidence of an effect of frequency here than for the adults), it cannot completely explain the patterns of use.

In terms of the production of 3sg errors, the model showed that there was a significant overall effect of condition, where more 3sg errors were made in the skewed condition when compared to the uniform condition. This suggests that the bias in the input increases the chance of producing a 3sg error, further suggesting that frequency is an important factor in both the correct production of the 3sg form and in the production of 3sg errors. It is important to note, however, that this was the

opposite pattern from that seen in the adults, where the rate of 3sg errors was higher overall in the uniform condition.

In Experiment 5 there was again a significant interaction between target ending and condition, where the 3sg form was produced significantly more successfully than the other forms, and this effect was greater in the skewed condition than in the uniform condition. As in Experiment 4, the 3sg form was produced more successfully than the other inflections (i.e. 1sg, 2sg, 1pl) in the uniform as well as the skewed condition.

The 3sg error analysis for Experiment 5 found effects of condition and target ending but no significant interaction. The hypothesis is supported, as participants were significantly more likely to produce a 3sg error in the skewed condition than the uniform condition.

In order to assess whether children use the Person and Number cues to learn the morphological features of the five different inflections, we fitted Bayesian Multinomial Regression models to the data from the three different conditions. The analysis showed that the ability to use the different person and number cues to learn the different morphological features varied by condition. Children in the skewed condition and uniform condition (with 40 training trials) could effectively use the Number cue to learn this morphological feature. Children in the Uniform condition (with 80 training trials) did not effectively use the Person and Number cues to learn the different morphological features. What is clear from experiments 4 and 5 is that there is a lot of confusability among the different forms, which reflects the low

accuracy rates found across the three conditions. In the skewed and uniform condition with 40 training trials and with 80 training trials the accuracy rate was just above chance at 28%, 26% and 24% respectively.

With regards to the stressed/unstressed distinction between inflections, it is clear that again, there is little confusability between 3sg and 1pl, potentially suggesting that children can use this distinction to discriminate between forms. However, the lack of confusability also reflects the lack of production of 1pl/-*amos* (i.e. participants do not produce the 1pl form and in the rare circumstances that they do produce 1pl, they do not use it correctly). There is confusability between 3sg and 1sg forms, which could be due to both forms being unstressed when spoken. However, there is also confusability between 3sg and the other inflections, leading us to suggest that the stressed/unstressed distinction is not an important factor in our studies.

The results from these studies are informative in various ways. It is clear that frequency has an effect on children production of correct forms and on their defaulting. In fact there is more evidence for the former in the children than in the adults, with interactions between condition and target ending (absent in the adults) being observed. However, as in experiments 1 and 2, in the uniform condition there still appears to be a preference for the use of the 3sg form, despite there being no bias in the input.

It is also important to acknowledge that the accuracy rates of the children in this study were not greatly above chance. They seem to struggle most notably in the use of the person cue - they often appear to utilize only one cue (i.e. the number cue) to

distinguish between inflections. Chapter 6 aims to address this issue by making the Person cues more salient to the child by adding a physically co-present “teacher” to the learning paradigm.

Chapter Six

6 Attempting to Improve Children's Learning with the Use of a New Methodology

6.1 Introduction

While studies 4 and 5 were informative regarding our main hypotheses, it is important to note that the accuracy rates for the children in this study were only marginally above chance. Our analysis revealed that children in the skewed condition and the children assigned to the uniform condition (with 40 training trials) only used the number cue to learn the different inflections and children in the uniform condition (with 80 training trials) did not effectively use either the person or number cues to learn the different inflections.

This pattern differs from the adult studies, where performance was much better. Clearly, we would be in a stronger position to draw conclusions from the child studies, were the effects observed in a group where a better rate of learning was seen. In order to try to improve learning, we therefore ran a new version of the study in which we increased the salience of the person cue in particular (i.e. the cue with which the children seemed to struggle most) in order to make it clearer for the children. To do this, we included a physical "teacher" in the training study in the form of a physically-present stuffed animal in place of the previous "teacher" who was embedded in the animations. By adding the physical "teacher", we hoped to make the interpersonal contexts that define different person contexts more salient.

We again tested children aged between 8 years old and 10 years old learning Spanish for the first time as a way of approximating the learning behaviour of Spanish-speaking children. The children were again taught the different verbs and inflections

over 3 training and testing days, in order to test their ability to produce the different forms correctly after each testing session.

6.2 Experiment 6

6.2.1 Methods Section

6.2.1.1 Participants

73 monolingual English-speaking children took part in the study. The target sample is based upon a bootstrap power analysis (Efron & Tibishani, 1994) using a sample of 36 participants (from Experiment 4) collected using an identical stimulus set and a very similar procedure to the skewed and uniform condition with 80 training trials. The model used was a multilevel logistic regression model using lme4 (Bates, et al. 2013). By-participants random slopes were included in all models. Ten thousand samples were drawn in the bootstrap and, for each sample, the target model and nested models were fitted and AIC values calculated. 37 participants per condition was the smallest sample size for which the target model provided a better fit than all nested models with a probability of >=.8, to give 80% power.

Thirty-seven children were assigned to the skewed condition; Thirty-six children were assigned to the uniform condition (with 80 training trials) with a mean age of 9;4 (range 8;1-10;7). Children were recruited from primary schools in the North West area of England. The children were randomly assigned to their group. This study was preregistered on the Open Science Framework (Martin, Bannard & Pine, 2018).

6.2.1.2 Design and Materials

This was the same as for Experiment 4 and 5, apart from the use of the stuffed animal, which was used as the "teacher" within the testing paradigm.

6.2.1.3 Procedure

The children were tested in a quiet room at their school. They were seated directly in front of the laptop computer. There was no time limit for the children to complete the different learning tasks. At the beginning of the experiment, the child was introduced to the stuffed animal who would act as the "teacher" in this experiment. The "teacher" was seated next to the laptop, with the head of the "teacher" facing toward the screen. The child could interact with the "teacher", whilst understanding that the "teacher" would be teaching them the different Spanish verbs and inflections.

In order to explain how the animations worked, we will use the Spanish verb "*patear*", which means "to kick". The first animation the children viewed showed an action and an audio description of what the animation is. For example, for the third person singular form, a character would kick a ball on screen and be accompanied by a description of the action "*patea*". The animation would then play again, but no description would be given. The child would then be asked to replicate the form produced by the "teacher" in the previous animation by saying "*patea*". For a video that featured a third person form, a different character was used (neither the teacher nor the child was included). The same format would apply for the 3rd person plural form, but this time two people would each kick a ball and the teacher would produce

the form "*patean*". The child would then see the video again but without the description being shown and would repeat the form.

To illustrate first and second person forms, three photographs were taken, the first was of the "teacher", the second was of the experimenter (for the demonstrations at the start of the study) and the third was of the child. Each picture was superimposed on the head of the appropriate character. The experimenter would demonstrate how the experiment worked, first in English and then in Spanish. The examples involved the use of 5 different forms; 3rd person singular, 1st person singular, 2nd person singular, 3rd person plural and 1st person plural. The format of the practice trials (i.e. before the training trials took place, in order to demonstrate the learning paradigm) involved watching an animation that were defined by the "teacher" who described the actions depicted in the animations.

The English examples used the verb "climb". The animation would play, and the teacher would describe the action. For the 3rd person singular form, the animation would show the character (who was not the experimenter or the teacher) climbing up a rock face and the teacher would describe the action by saying "He's climbing". The same animation would play again, but no description would be given. The experimenter would then repeat the form.

For the 3rd person plural form, two different characters would perform the climbing action. The "teacher" was again seated next to the laptop and would describe the action by saying "They're climbing". The animation did not feature the "teacher" or the "experimenter". The animation would finish, and the same two characters would

appear again. The experimenter would repeat the form. The 1st person plural form had two characters: the "teacher" and the "experimenter". The "teacher" would state the phrase "We're climbing". The animation would play again (with the same characters) and the experimenter would repeat the phrase. The final animation was the 2nd person singular form. The animation displayed the "experimenter's" character and the "teacher" would say the phrase "You're climbing". The animation would play again, with the "teacher" performing the action. The experimenter would then repeat the phrase. The English examples would then finish.

The experimenter would then perform the same task, but using the Spanish verb "*cortar*", meaning "to cut". The animations would play in the same format. The 3rd person singular animation used a male character for both animations. For the 1st person singular form, the "teacher" would perform the action for the demonstration and the experimenter's face would be used when the experimenter was imitating the phrase. The 2nd person singular form used the experimenter's face for the "teacher's" demonstration and then the "teacher" would replace the experimenter's face for the imitation phase. The 1st person plural form used both the "teacher's" and the "experimenter's" face for both the demonstration and repetition phases. For the 3rd person plural form, two separate characters were used for the demonstration and the imitation phases. Each of the five different verb endings was used: 3sg (*corta*), 2sg (*cortas*), 1sg (*corto*), 3pl (*cortan*) 1pl (*cortamos*). The experimenter would see and hear the "teacher" explaining the different actions. The same animations were played again without the description. The experimenter would then repeat the Spanish verb and ending combinations (e.g. the 3sg form "*corta*"). On the first testing day, the

experimenter would demonstrate both the English and Spanish forms. However, on the second and third day, only the Spanish forms were demonstrated.

After the experimenter had completed these examples, the experiment would begin. The Spanish process was repeated for 80 training trials for both the skewed condition and the uniform condition. The experimenter would click after each animation in order to move from the demonstration animation to the repetition animation. If the participants did not hear a verb during the demonstration phase; the experimenter would move on to the next animation. For the 1st person singular form, the first animation would show the "teacher" performing an action kicking a ball on the screen and the "teacher" would produce the form "*pateo*" (the "teacher" would still be seated next to the laptop). The second animation would show the child's head on another character performing the action. The participant would repeat this form. For the 2nd person singular form, the first animation would show the child kicking the ball on the screen and the action would be described as "*pateas*". The second animation would show the "teacher" performing the same action and the child would repeat the form "*pateas*". For the 1st person plural form, both the "teacher" and the child would each be kicking a ball. The action would be described by the "teacher" by producing the form "*pateamos*". The same animation would play, and the participant would imitate the taught verb. For the 3rd person singular form, a separate character would kick a ball and the "teacher" would produce the verb "*patea*". The animation would play again with the same character and the child would imitate the verb. Finally, for the 3rd person plural form, 2 new characters would appear and each kick a ball. Neither character was the child or the "teacher". The "teacher" would

produce the form "*patean*". The child would see the same animation again and repeat the verb.

The testing phase was the same as the previous studies, but the "teacher" would replace "Juan" for 2nd person forms.

6.2.1.4 Transcription and Coding

Spanish responses were audio-recorded and then transcribed by the researcher. Correct responses were coded as 1 and incorrect responses were coded as 0. When no answer was given, the cells were left blank and then removed from the analysis. Both the verb stem and the endings were coded. Responses were only included if a full verb was produced (i.e. both an eligible stem and an ending were produced). If either was missing, this was not counted as a complete trial. There were 1285 trials excluded, out of a total of 2815 trials, as follows – not producing a full verb form (i.e. not producing an eligible stem or ending) (157), not producing an eligible ending (44), not providing an answer (1125) and not producing an audible response (3). 20% of the children who took part in each condition were second coded by a research assistant who was blind to the condition. Inter-rater reliability was high (agreement = 90.2%; kappa = .844).

6.3 Results

For this study, 2 analyses were performed. We built a series of mixed effects logistic regression models to assess the effect condition and the identity of the target ending had on correct production. The second analysis assessed the effect these variables had on the rate of 3rd singular errors produced. Table 6.1 shows the proportion of

coded trials on which each response was given across the two conditions. Table 6.2 shows the proportion of ALL trials on which each response was given across the two conditions.

Table 6.1 - The proportion of coded trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.55(±.25)	0.03(±.05)	0.13(±.16)	0.18(±.18)	0.11(±.13)
Uniform	0.25(±.22)	0.10(±.18)	0.14(±.20)	0.20(±.18)	0.29(±.25)

Table 6.2 - The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced, separated by condition.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.35(±.14)	0.02(±.04)	0.11(±.13)	0.11(±.15)	0.07(±.07)
Uniform	0.12(±.13)	0.53(±.05)	0.08(±.08)	0.13(±.12)	0.15(±.10)

6.3.1 Correct Verb Ending on Day 3

The main aim of this analysis was to assess what effect target ending and condition would have on the production of verb endings. For this analysis, 1530 responses were analysed. Likelihood ratio tests were used to assess the fit of the model and whether removing these terms damaged the fit of the model to the data. Removing the different terms allowed us to measure how much explanatory value the different predictors provided. If removing a term provided a model with at least as good a fit, then that term was excluded from our final model. We first produced a mixed

effects logistic regression model that included an interaction between Condition and the Target Ending variable. Child identity was included as a random effect on the intercept. A random effect of participant on the slope for target ending was considered but excluded as it gave a singular fit. A random effect of target stem on the intercept was also considered but when fitting the model, it provided a singular fit, and thus was omitted in line with Barr et al. (2013).

As in our previous experiments, the experimental design included a test of whether the children could generalise to new verb-ending combinations. The results showed that a model with "unseen" trials included as a predictor variable did not provide a significantly better fit for the data than a null model ($X^2(1) = 3.03$, $p = .082$).
A full model which included target ending, condition and an interaction as fixed effects (Target Ending * Condition) was compared to a model with the interaction removed (Target Ending + Condition). The full model provided a significantly better fit to the data than a reduced model ($X^2(4) = 71.57$, $p <.001$). The full model was then compared to a model with only Target Ending included as a fixed effect. Again, the full model provided a significantly better fit to the data than a model with only Target Ending included as a fixed effect ($X^2(5) = 71.76$, $p <.001$). A full model was compared to a model with only Condition included as a fixed effect. The full model provided a significantly better fit to the data than a model with only Condition included as a fixed effect ($X^2(8) = 261.25$, $p <.001$). Finally, the full model was compared to a null model. The full model again provided a significantly better fit for the data than the null model ($X^2(9) = 261.58$, $p <.001$).

Figure 6.1 shows parameters derived from our final chosen model. For purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

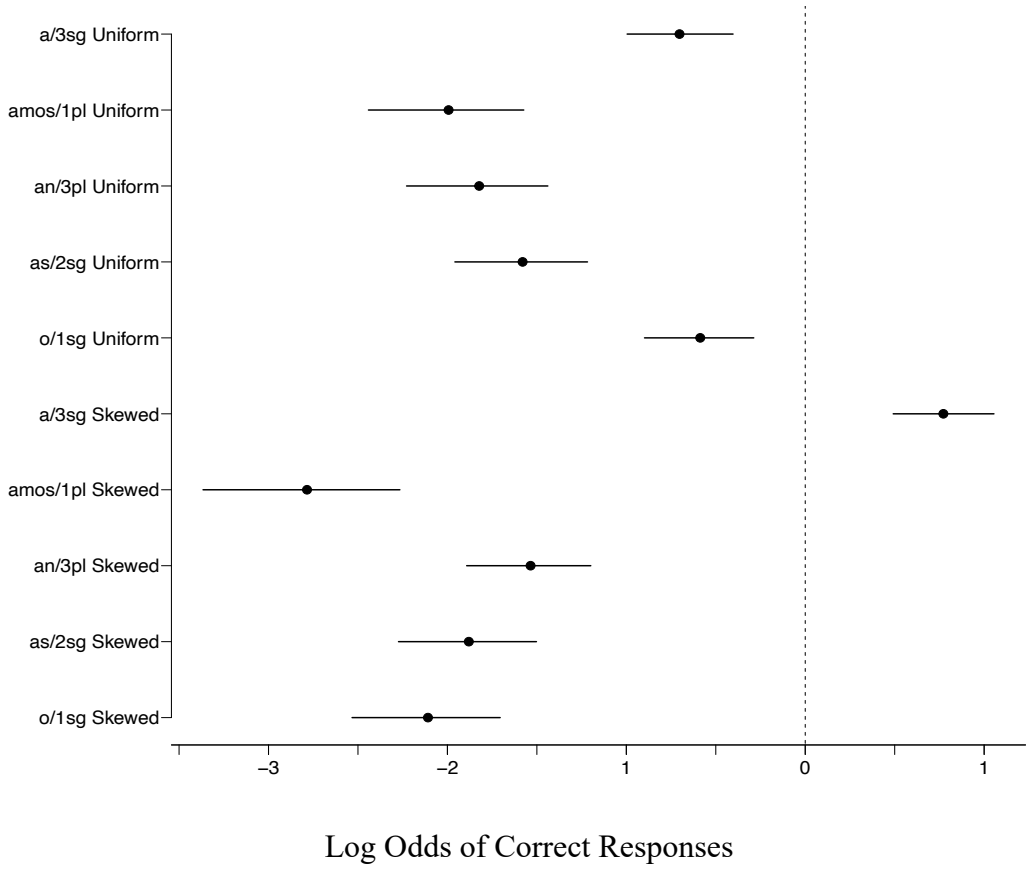

Figure 6.1 - The log odds and confidence intervals (95%) for correct verb endings produced on Day 3.

Figure 6.1 shows that in the skewed condition, the children in this study were significantly more likely to produce "-*a*" correctly than any other form. In the

uniform condition, children were more likely to produce "-*a*" correctly than they were"-*amos*", "-*an*" and "-*as*". However, children within the uniform condition were no more likely to produce "-*a*" correctly than "-*o*".

6.3.2 3rd Person Singular Error Production on Day 3

The main aim of this analysis was to assess what effect condition and target ending have on the production of 3rd person singular errors. This type of error is where "*a"* is produced instead of the target form, for example, producing "**baila*" instead of "*bailamos*" when "-*amos*" was the target ending. 704 responses on Day 3 (all responses for non-3SG trials) were analysed using mixed effects logistic regression models. Likelihood ratio tests were used to assess the fit of the models and whether each predictor term has explanatory value – whether removing these terms damaged the fit of the model to the data. We produced a full model that included an interaction between condition and the target ending variable. Child identity was included as a random effect on the intercept. A random effect of participant on the slope for target ending was considered but excluded as it gave a singular fit. A random effect of target stem on the intercept was also considered, but when fitting the model, it provided a singular fit, and so was omitted in line with Barr et al. (2013).

A full model where Condition, Target Ending and an interaction between the two were included as fixed effects (Target Ending * Condition) was compared to a model with the interaction removed (Target Ending + Condition only). The full model provided a significantly better fit to the data than a reduced model with the interaction removed ($X^2(3) = 11.26$, $p = .01$). The full model was compared to a

model with only Target Ending included as a fixed effect. The full model again provided a significantly better fit tor the data than a model with only Target Ending included as a fixed effect ($X^2(4) = 27.41$, $p < .001$). The full model was compared to a model with only condition included as a fixed effect. The full model provided a significantly better fit to the data than a model with only condition included ($X^2(6) = 63.80$, $p < .001$). A full model was finally compared to a null model where no fixed effects were present. The full model provided a significantly better fit to the data than a null model, ($X^2(7) = 80.58$, $p < .001$). It can be concluded that a full model (with target ending, condition and an interaction between them) gives the best fit to the data.

Figure 6.2 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which treat as its estimate. Where a mean parameter values lie outside of the 95% interval of another parameter, we can say that the former is different from the latter at $\alpha = 0.05$ (two-tailed).

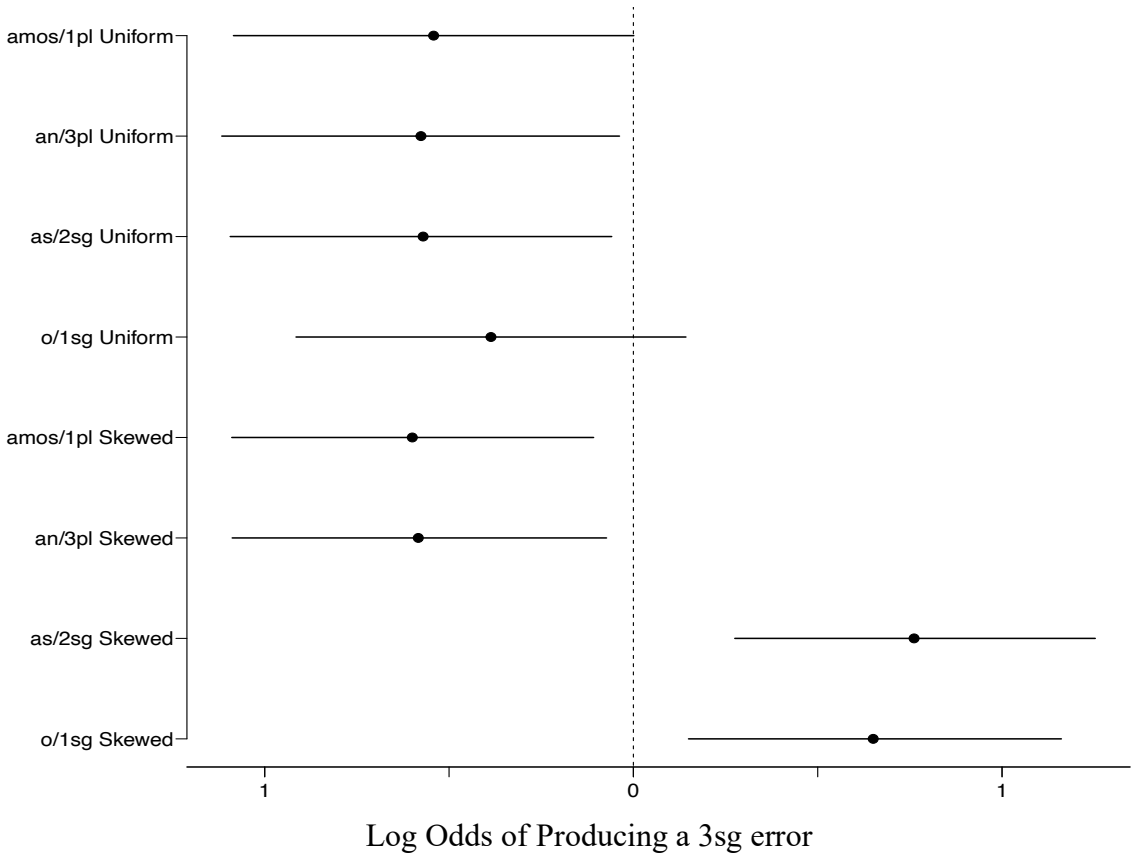

Figure 6.2 - The log odds and confidence intervals (95%) for 3rd person singular errors produced on Day 3.

Figure 6.2 shows that in the uniform condition participants were no more likely to produce a 3rd person singular "-*a*" error for any particular form than for any other. Within the skewed condition, children were more likely to produce a 3sg error when the target was a singular form (i.e. children were more likely to produce a 3sg error when the target was 2nd Person Singular "-*as*" or 1st Person Singular "-*o*").

6.4 Interim discussion

The current study aimed to measure the effect that frequency had on learning different morphological forms. Participants were split into two conditions, where one condition skewed the input frequency of the 3sg form, and the uniform condition

kept the input frequency equal across the different inflected forms. We predicted that there would be more correct responses for 3sg forms than the other targets but only in the skewed condition. We also predicted that there would be more 3sg errors in the skewed condition when compared to the uniform condition. This is expected as a result of the bias in the input towards the 3sg form. The results partially support the first hypothesis - unlike for the adults, the children appear to have a greater performance advantage (relative to the other forms) in the skewed than in the uniform condition. However, not consistent with our predictions, performance on 3sg is significantly better than on most other targets in the uniform as well as the skewed condition (the only difference is the magnitude of the performance advantage).

In the uniform condition, no difference was seen between the different target endings in the rate of 3sg errors produced. However, within the skewed condition children were more likely to produce a 3sg error when another singular form was the target form, suggesting that children can decipher the difference between plural and singular forms in this condition. The effect of condition supports the second hypothesis – participants were more likely to make a 3sg error when the input frequency was biased toward the 3sg form, but not when the input was equal across the different forms.

In order to address a potential confound, we again ran a second version of the uniform condition, where input frequency was equal across forms, but participants had 40 training trials rather than 80 training trials. This was compared to the same skewed data from Experiment 6. This allowed children in the uniform condition with

40 training trials to have the same relative frequency for each of the three non-privileged endings across the two conditions.

6.5 Experiment 7

6.5.1 Method

6.5.1.1 Participants

74 children who were monolingual speakers of English took part in the study. The same bootstrap power analysis for Experiment 6 was used here. Thirty-seven children were assigned to the skewed condition (taken from Experiment 6) and thirty-seven participants were assigned to uniform condition with 40 training trials, with a mean age of 9;4 (range 8;1-10;1). Participants were recruited from primary schools in the North West of England.

6.5.1.2 Design

The design and materials were the same as for Experiment 6, with the exception of the change in relative frequency of items in the uniform condition. Table 6.2 shows the proportional and relative frequencies across the different forms. Table 6.2 shows that the relative frequency is lower in this uniform condition with 40 training trials with 8 per inflection rather than 16 per inflection, as was seen in the uniform condition with 80 training trials.

Table 6.3 - The relative and proportional frequency of the different verb/ending combinations in the Uniform Condition with 40 Training Trials.

	3sg (a)	2sg (as)	1pl (amos)	3pl (an)	1sg (o)
Rasg-		1			1
Escal-			1	1	
Dibuj-		1	1		
Salt-				2	2
And-	1		1	1	1
Empuj-	1		1	1	1
Golpe-	1	1		1	1
Pate-	1	1		1	1
Dispar-	1	1	1		1
Bail-	1	1	1		1
Cant-	1	1	1	1	
Tir-	1	1	1	1	
FREQUENCY OF PERSON/NUMBER FORM	8	8	8	8	8
PROPORTION OF TOTAL TRIALS	0.2	0.2	0.2	0.2	0.2

6.5.1.3 Procedure

This was the same as for Experiment 6.

6.5.1.4 Transcription and Coding

This was the same process as for Experiment 6. 2914 trials were analysed and there were 1470 trials excluded for the following reasons – not producing a full verb form (i.e. not producing an eligible stem or ending) (155), not providing an answer (1300) and not producing an audible response (5). 20% of the children in each condition were second coded by a research assistant who was blind to the condition. Inter-rater reliability was high (agreement = 89.8%; kappa = .849).

6.6 Results

The analyses performed were the same as for Experiment 6. The skewed data reported here is that collected in Experiment 6. Table 6.4 shows the proportion of

coded trials on which each response was given across the two conditions. Table 6.5 shows the proportion of ALL trials on which each response was given across the two conditions.

Table 6.4 - The proportion of coded trials (±SD) on Day 3 in which each verb ending was produced, in the Skewed and Uniform Condition with 40 training trials.

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.54(±.25)	0.03(±.05)	0.13(±.16)	0.18(±.18)	0.11(±.13)
Uniform	0.43(±.29)	0.08(±.14)	0.16(±.23)	0.20(±.22)	0.12(±.15)

Table 6.5 - The proportion of ALL trials (±SD) on Day 3 in which each verb ending was produced, in the Skewed and Uniform Condition with 40 training trials..

	a (3sg)	amos (1pl)	an (3pl)	as (2sg)	o (1sg)
Skewed	0.35(±.14)	0.02(±.04)	0.11(±.13)	0.11(±.15)	0.08(±.07)
Uniform	0.16(±.13)	0.02(±.05)	0.04(±.06)	0.08(±.08)	0.05(±.06)

6.6.1 Correct Production on Day 3

In this analysis, 1400 trials were analysed using mixed effects logistic regression models. The random effects structure and the model comparison process was the same as for Experiment 6. We fitted a mixed effects logistic regression model that included Condition and the Target Ending as predictors, along with an interaction between them. Child identity was included as a random effect on the intercept. A random effect of participant on the slope for target ending was considered but excluded as it gave a singular fit. A random effect of target stem on the intercept was

also considered but when fitting the model, it provided a singular fit, and so was omitted in line with Barr et al. (2013).

The experimental design included a test of whether the children could generalise to novel verb-ending combinations. To examine the effect that exposure to a particular verb-ending combination had on the data, we compared a model with the exposure status (seen/unseen) of each trial as a predictor variable to a null model. The results showed that a model with the "Unseen" trials included as a predictor variable provided a significantly better fit to the data than a null model ($X^2(1) = 3.86$, $p = .0495$. This indicates that exposure status had a significant effect on correct production and this variable was therefore included as a fixed effect within our full model.

A full model, which included Target Ending, Condition, the "Unseen" trials and an interaction between Target Ending and Condition (Target Ending*Condition + Unseen), was compared to a model with the interaction removed (Target Ending + Condition + Unseen). The full model provided a significantly better fit to the data than a reduced model with the interaction removed, ($X^2(4) = 13.69$, $p = .008$). This suggests that the interaction explains variance over and above the other predictors. The full model was then compared to a model with Target Ending, Condition and an interaction between them (Target Ending*Condition). The "Unseen" fixed effect was removed in order to assess how much predictive value the unseen variable had for the data. The full model provided a significantly better fit to the data than a model with the "Unseen" predictor variable removed, ($X^2(1) = 10.09$, $p = .001$). This suggests that the "Unseen" predictor variable explains variance over and above the

other predictors. The full model was compared to a model with Condition and the Unseen variable included as fixed effects (Condition + Unseen). Again, the full model provided a significantly better fit to the data than a model with only Condition and the Unseen variable included as a fixed effect, ($X^2(8) = 211.98$, $p<.001$). The full model was then compared to a model which included Target Ending and the "Unseen" variable as a fixed effect. The full model again provided a significantly better fit to the data than a model with Target Ending and the "Unseen" variable included as the only fixed effect ($X^2(5) = 19.99$, $p = .001$). Finally, the full model was compared to a model that included only the "Unseen" variable as a fixed effect (equivalent to a null model for our purposes). The full model provided a significantly better fit to the data than a model that included only the "Unseen" variable as a fixed effect, ($X^2(9) = 228.01$, $p<.001$).

Figure 6.2 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean parameter value lies outside of the 95% interval of another parameter, we can say that the former is different for the latter at $\alpha = 0.05$ (two-tailed).

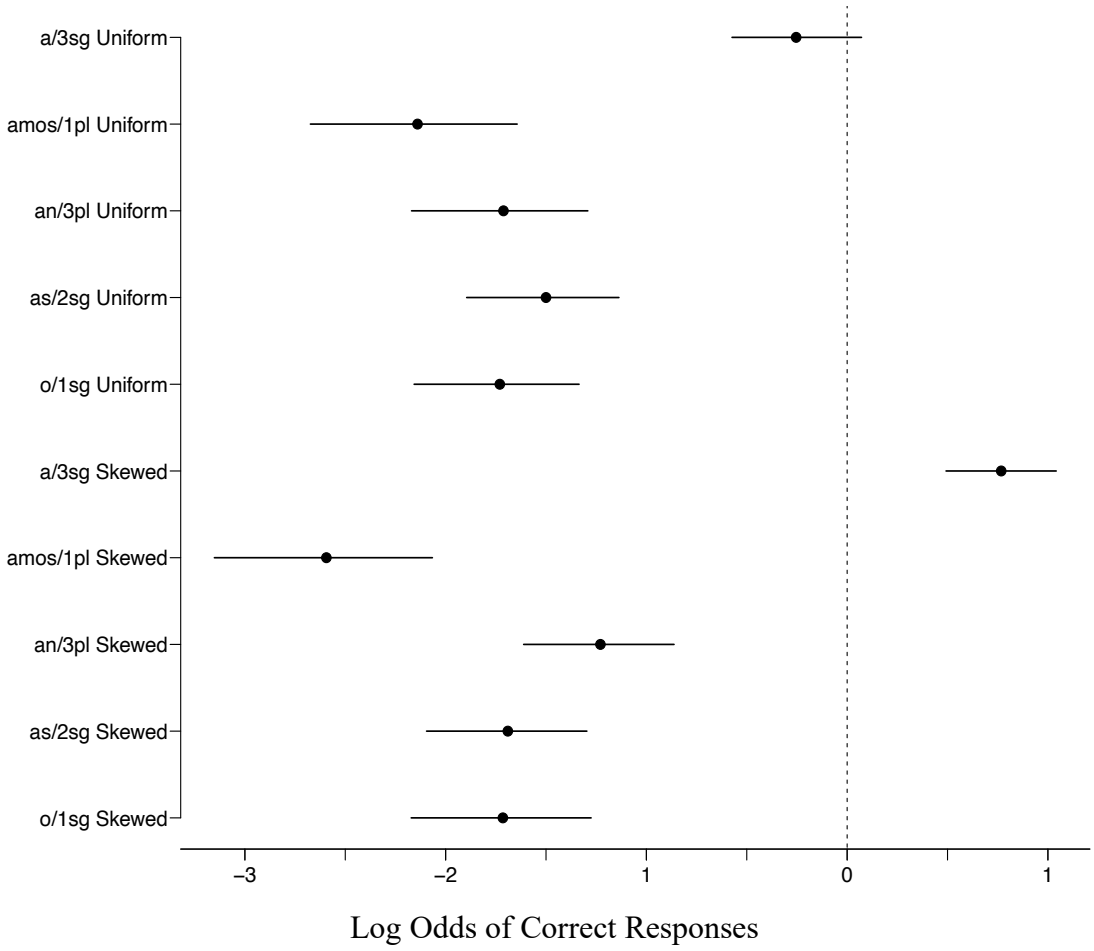

Figure 6.3 – The log odds and confidence intervals (95%) for correct verb endings produced on Day 3 (Unseen - β = -1.12, CI = -1.73 - -0.51).

Figure 6.3 shows that in the uniform condition, the children were significantly less likely to produce "-*amos*", "-*an*", "-*as*" and "-*o*" relative to "-*a*", and that this difference is even greater in the skewed condition.

6.6.2 3rd Person Singular Error Production on Day 3

The aim of this analysis was to assess what effect condition and target ending have on the production of the 3rd Person Singular Errors. 964 of all Day 3 responses on trials where the target was not a 3rd person singular form were analysed using mixed effects logistic regression models. The random effects structure and the model comparison process was the same as for the previous experiment. Child identity was

included as a random effect on the intercept. A random effect of participant on the slope for target ending was considered but excluded as it gave a singular fit. A random effect of target stem on the intercept was also considered but when fitting the model, it provided a singular fit, and thus was omitted in line with Barr et al. (2013).

A full model where Condition, Target Ending and an interaction between them were included as fixed effects (Target Ending*Condition) was compared to a model where the interaction was removed (Target Ending + Condition). The full model provided a significantly better fit to the data than a model with interaction removed, ($X^2(3) = 14.81$, $p=.002$). A full model was then compared to a model with only Target Ending included as a fixed effect. The full model provided a significantly better fit to the data than a model with only Target Ending included, ($X^2(4) = 17.23$, $p=.001$). A full model was compared to a model with only Condition included as a fixed effect. The full model provided a significantly better fit to the data than a model with only Condition included, ($X^2(6) = 57.43$, $p<.001$). Finally, the full model was compared to a null model with no fixed effects included. The full model provided a significantly better fit to the data than a null model, ($X^2(7) = 59.60$, $p<.001$).

Figure 6.4 shows parameters derived from our final chosen model. For the purposes of reporting we ran a Bayesian version of the model using STAN via the "rethinking" package (McElreath, 2015). The error bars show the 95% credible intervals for each parameter and the points show its mean value which, as is conventional we treat as its estimate. Where a mean value of a parameter lies outside

of the 95% interval of another parameter, we can say that the former is different from the latter at α = 0.05 (two-tailed).

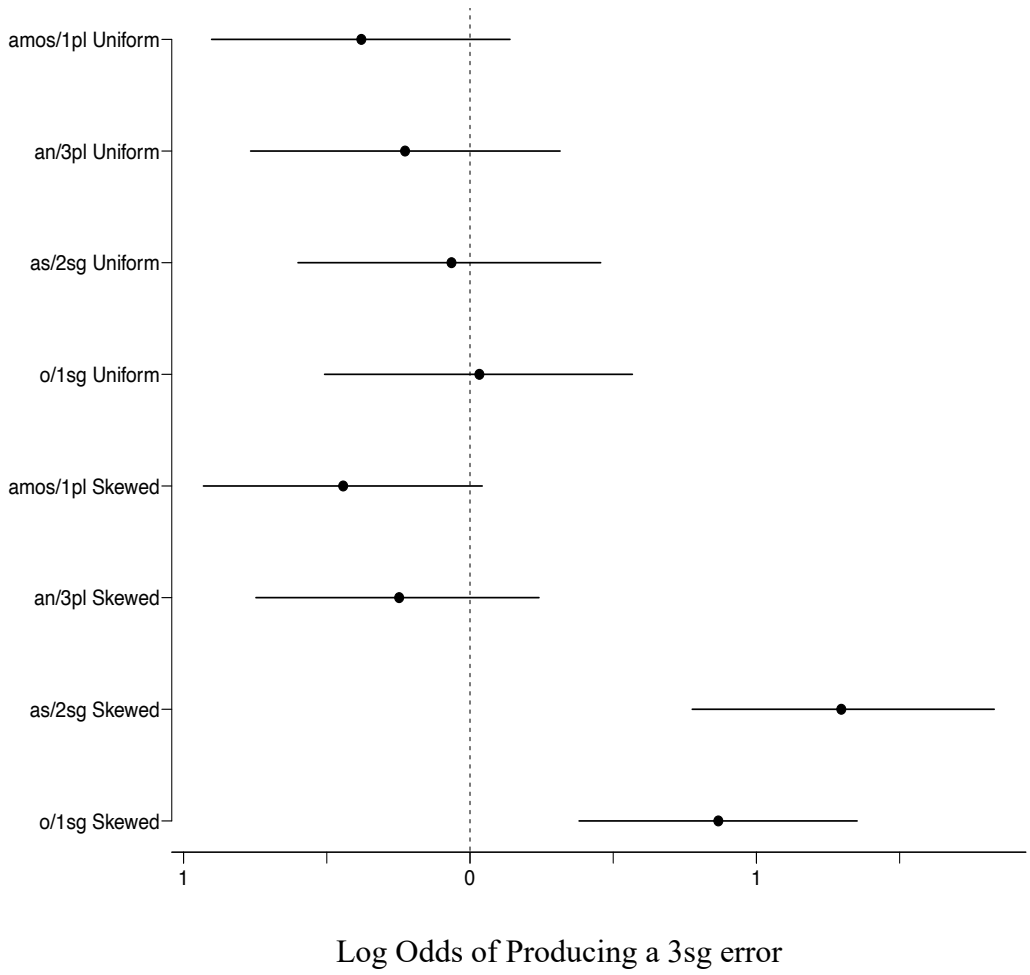

Figure 6.4 - The log odds and confidence intervals (95%) for the production of 3rd person singular errors on Day 3.

Figure 6.4 shows that in the uniform condition, children were no more likely to produce a 3sg error for any particular verb ending than any other. The chosen model also indicated that there was an interaction between Target Ending and Condition where Target Ending is affected by Condition. Figure 6.4 demonstrates that this effect is a result of children being more likely to produce a 3sg error when the target forms were "-*as*" and "-*o*" in the skewed condition but not the uniform condition.

Mirroring Experiment 6, children could decipher plural and singular forms, but there was confusability between singular forms.

6.7 How do children make use of the cues?

To assess how the participants are using the person and number cues, we fitted Bayesian Multinomial Regressions to the data from the three conditions – the uniform condition with 40 trials, the uniform condition with 80 trials and the skewed condition. The models are created using an R tool called *brms* (Burkner, 2017) which in turn makes use of STAN. Default weakly informative priors were used in all models. The 3sg form was set as the reference class.

6.8 Results

Using the data from Day 3, we fitted three Bayesian multinomial regression models that included Person and Number features as individual predictor variables. Participant identity was included as a random effect on the intercept. This was done separately for the three conditions. Leave-one-out cross validation (LOOIC; Gefland, Dey & Chang, 1992; Gefland, 1996) was used to assess the fit of the model and whether each predictor term had explanatory value. Removing the different terms allowed us to assess how much explanatory value the different terms provided. If removing a term produced a model with at least as good a fit, then that term was excluded from our final model. Three separate models were constructed in order to assess the effect of Person and Number features within each condition rather than comparing the conditions.

6.8.1 Uniform (80 Training Trials)

A model which included only Number as a fixed effect provided a better fit to the data (LOOIC = 1880.78, SE = 35.43) than a model that included both Person and Number features (LOOIC = 1895.38, SE = 35.96), a model that included only Person as a fixed effect (LOOIC = 1919.82, SE = 34.25) or a null model that included no fixed effects (LOOIC = 1911.19, SE = 33.40). These results demonstrate that participants within this condition use the Number cue to learn the different inflections, but not the person cue.

6.8.2 Uniform Condition (40 Training Trials)

A null model (which included no fixed effect) provided a better fit to the data (LOOIC = 1303.09, SE = 36.20) than a model with both Person and Number included (LOOIC = 1325.84, SE = 37.07), a model with only Person included (LOOIC = 1320.00, SE = 36.78) and a model with only Number included as a predictor variable (LOOIC = 1309.40, SE = 36.49). The selected model did not include either Person or Number as fixed effects, suggesting that children within this condition did not use the cues.

6.8.3 Skewed Condition

A model which included only Number as a fixed effect provided a better fit to the data (LOOIC = 1709.78, SE = 50.30) than a full model which included both Person and Number features as fixed effects (LOOIC = 1744.30, SE = 58.81), a model which included only Person as a fixed effect (LOOIC = 1819.20, SE = 71.20) and a null model which had no fixed effects (LOOIC = 1781.69, SE = 47.91). This

suggests that in the skewed condition, children could effectively use the Number cue to choose the different forms but did not use the Person feature.

The three multinomial regression analyses demonstrate that the condition to which a child was allocated had a significant effect on whether they could use the different Person and Number features to produce the different inflections. Within the uniform condition with 40 training trials, children struggled to use the Person and Number features to learn and produce the different inflections. However, in the skewed condition, children could effectively use the Number feature to learn the different inflections. This is perhaps reflected in the children producing 3sg errors when another singular form was the target, showing that children in the skewed condition could discriminate between the singular and plural forms but could not discriminate between the different person forms. The results indicate that children in the uniform condition with 80 training could also effectively use the number cues to learn the different inflections.

Figures 6.5-6.9 show the rate at which each of the five different inflections was erroneously produced in each of the four other contexts across all three conditions in experiments 6 and 7. The Figures show us that amongst the five different inflections, there is confusability amongst the different inflections. However, this mirrors the results found across the previous experiments, where both adult participants and children make errors by classifying on the basis of a single feature – the number feature. This is demonstrated clearly in Figure 6.5 and Figure 6.9 where children are more likely to produce a 3sg error and a 1sg error respectively, when another singular form was the target. The results are less clear with plural forms, but

production for the plural forms is extremely low for the children involved in this study (see Tables 6.1 and 6.3, where the production rate is extremely low for the 1pl form and relatively low for the 3pl form).

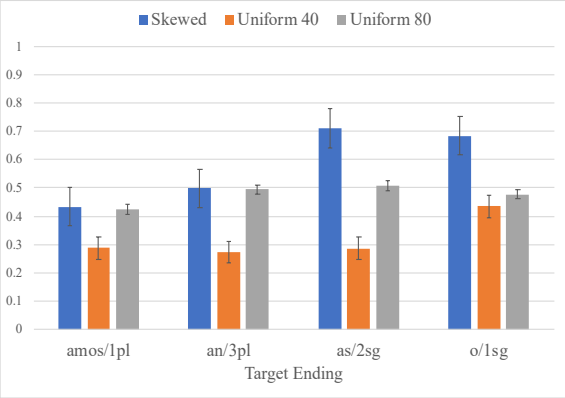

Figure 6.5 – Proportion of 3sg Errors produced on Day 3.

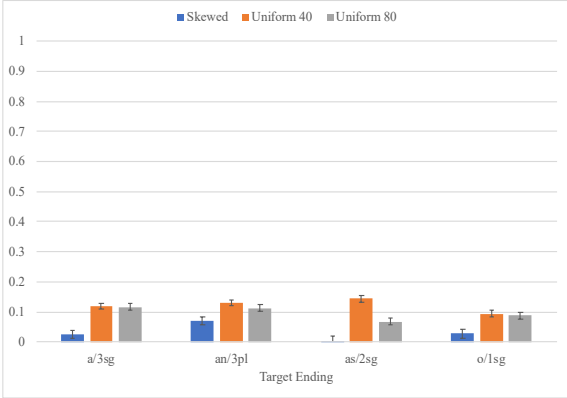

Figure 6.6 – Proportion of 1pl Errors produced on Day 3.

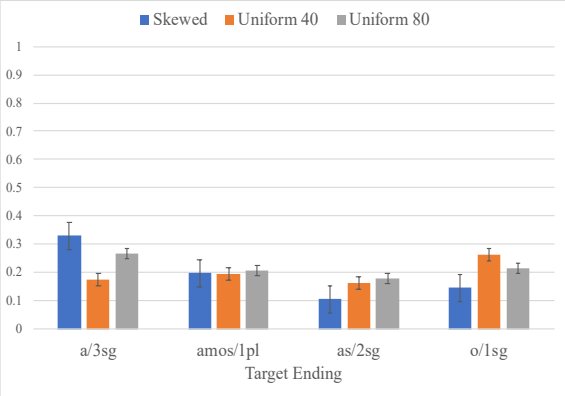

Figure 6.7 – Proportion of 3pl Errors produced on Day 3.

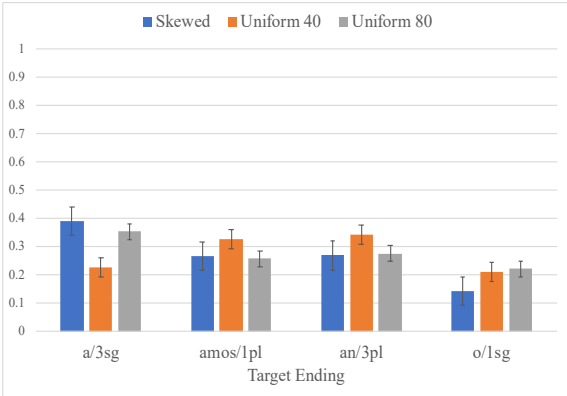

Figure 6.8 – Proportion of 2sg Errors produced on Day 3.

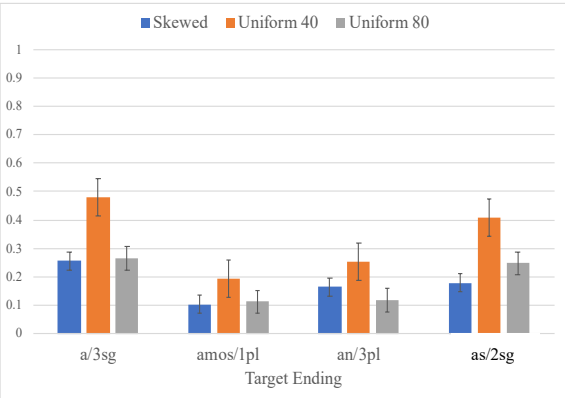

Figure 6.9 – Proportion of 1sg Errors produced on Day 3.

Figure 6.5 - 6.9 - Plots demonstrating where errors are made across each inflection learned on Day 3.

6.9 Discussion

The aim of the current studies was to increase the salience of the different linguistic features that are required for children to learn the different inflections. We hoped to improve learning in order to strength the claims we could make with regard to our main hypotheses.

Experiments 6 and 7 both indicated that there was a significant interaction between target ending and condition, where the 3sg form was produced significantly more successfully in the skewed condition than in the uniform condition (80 training trials). The 3sg form was also produced more successfully in the uniform condition (as seen in our previous experiments) than three of the other inflections (1pl, 3pl and 2sg – but not the 1sg form, which children in the uniform condition with 80 training trials produced more successfully). These results again highlight the important role that frequency plays in correct production, but suggest that the 3sg advantage cannot be fully explained by the frequency bias.

In terms of the production of 3sg errors, in the 80-training-trial uniform condition the production of 3sg errors did not differ across the different target endings. However, in the skewed condition participants were more likely to produce a 3sg error when the target was another singular form (i.e. 2sg or 1sg). This suggests that the bias in the input frequency increases the chances of producing a 3sg error, thus providing

support for the hypothesis that children would be more likely to produce a 3sg error in the skewed condition than in the uniform condition (with 80 training trials).
In our 3sg error analysis for Experiment 7, there was again an interaction between condition and target ending. In the skewed but not the uniform condition, participants were more likely to produce a 3sg error when a singular form was the target (as seen in Experiment 6).

The main reason for the methodological change in the current study was to make the Person cue more salient. To do this, we added a physical teacher (in the form of a stuffed animal) to the learning paradigm so the children had a physical reference point to establish when different characters were performing different actions. However, as for the studies reported in chapter 6, low overall accuracy was found across the three conditions, not being far above chance for any condition (Skewed, 24%; uniform (with 40 training trials), 21% and uniform (with 80 training trials), 23%).

In order to assess whether the children in this study used the different person and number cues to learn the different inflections, we fitted three Bayesian Multinomial Regression models for the three different conditions. The analysis showed that children in the uniform condition (with 40 training trials) could not effectively detect the cues to learn the different inflections. However, participants in the uniform condition with 80 training trials and the skewed condition used the Number feature, but not the Person feature to learn the different morphological forms.

With regards to the stressed/unstressed distinction, we find the same results as Chapter 5, where the lack of confusability between 3sg and 1pl is the product of a lack of production of 1pl rather than being able to distinguish between stressed and unstressed forms. There is confusability between 3sg and 1sg, which are highly confusable due to being unstressed/reduced in spoken Spanish and in our paradigm. However, there is also confusability between 3sg and the other forms which could be explained due to similarity in their phonological properties, thus reducing the effect that the stressed/unstressed distinction has.

Taken together, the results suggest that frequency is important for both successful and erroneous production, where participants in the skewed condition were more likely to produce the 3sg form successfully and more likely to produce the 3sg form when another target was required, thus generating a 3sg error. However, frequency cannot explain the higher performance for the 3sg form in both uniform conditions.

Part Three

7 General Discussion

The objective of this thesis is to explain a type of error that is commonly observed in the speech of children in a number of languages – the tendency to frequently deploy a particular person-number form of a verb in contexts where other forms are appropriate, a pattern that has been referred to as "defaulting". In particular, we focus on the tendency of Spanish-learning children to "default" to the third person singular (3sg) form of verbs. In order to study this, English speakers, both children and adults, were taught different Spanish verbs and inflections via a series of animated videos that displayed situations illustrating different person and number combinations (i.e. 3sg, 1pl, 3pl, 2sg and 1sg).

We devised a series of training studies that directly manipulated learners' exposure to the different forms. In a number of studies we manipulated the input frequency of the inflections the participants were taught. This allowed us to establish how much of an effect frequency has on morphological development, but also allowed us to explore different explanations for the "defaulting" phenomenon in Spanish. In another study we manipulated the phonological and semantic mapping of the Spanish inflectional paradigm. Each participant within this study was taught a unique inflectional paradigm, where the links between the phonological and semantic features of the inflections were different for each participant (e.g. *"-a"* would mean the 3sg form for one participant, but *"-a"* would mean 1pl for another participant. This allowed us to disentangle the effects of phonology and explore how much of an effect these factors have on the learning of the Spanish verb inflectional paradigm.

In the first part of this chapter, we will review each experiment in turn, detailing the aims of the experiment, the methodology used, the results and their implications. This will allow us to then explore the different reasons for children's 3sg errors in Spanish.

7.1 Experiments 1 and 2– Assessing the Effect of Skewed Frequency Distributions on the Learning of Spanish Person and Number Morphology.

Experiment 1 aimed to assess the role of frequency via a training study in which English-speaking adults were either assigned to a skewed condition in which the participants' input was biased towards the 3sg form or a uniform condition in which the participants' input was equal across the different inflections.

Both conditions involved 80 training trials on each of 3 days. We hypothesised in this study that participants would make more successful productions for the 3sg forms than for the other forms, but only in the skewed condition. We also hypothesised that participants would make more 3sg errors (e.g. where participants use the 3sg inflection when another inflection was the target).

In terms of correct productions of the 3sg form, the participants in this study were more successful at producing the 3sg form than the other endings. However, this was seen to apply in both conditions. Counter to our predictions, we found that participants were more likely to produce a 3sg error in the uniform condition than in the skewed condition. They were more likely to do so, in both conditions, when the target was another singular form. These result suggests that frequency is not driving

defaulting. A similar pattern of results was found for Experiment 2, where we sought to address a potential confound.

In order to establish what features participants use to learn the different person and number forms in Spanish, we conducted a Bayesian Multinomial Regression analysis. This allowed us to determine whether the participants could use the different person and number features that were embedded into our training study. Our analysis showed that participants could effectively use the different Person and Number cues to produce the different inflections. Further analysis, however, including reference to the 3sg error analysis described above, suggests that despite participants being able to use both Person and Number cues, many errors tended to occur as a result of classifying on the basis of a single cue: number.

7.2 Experiment 3 – Assessing the Contribution of Phonology and Semantics Factors in Spanish Learners Successful and Erroneous Production.

The results from experiments 1 and 2 suggest that, while frequency is a factor in both the successful and erroneous production of the 3sg form, it cannot explain the observed patterns in the use of the 3sg form that we are concerned with. Experiment 3 explored an alternative explanation – that the errors might be caused by phonology or semantics. In this experiment, English-speaking adults were again taught the different Spanish inflections, but for each participant there was a different randomly assigned relationship between form and meaning, so that, for example, for one participant the *"-a"* form would mean 3sg, where one person (not the participant or Juan, the teacher) would be performing the action, whereas for another participant the *"-a"* would mean a 3pl form, where two people (neither the participant nor Juan,

the teacher) would be performing the action. This allowed us to establish how much of an effect each factor has. Table 7.1 demonstrates example phonological/semantic mismatches.

Table 7.1 - The randomised phonological and semantic verb endings for four hypothetical participants.

Participant	3sg	1pl	3pl	2sg	1sg
Canonical Participant	a	amos	an	as	O
Participant A	amos	an	as	o	A
Participant B	an	as	o	a	Amos
Participant C	as	o	a	amos	An
Participant D	o	as	amos	an	A

In terms of correct productions, the results from Experiment 3 demonstrated that participants were more successful at learning and using the "-*a*" form than the other forms, regardless of the meaning that was assigned to them. With respect to semantics, participants were no more successful at producing the 3sg form than the other semantic forms (i.e. 1pl, 3pl, 2sg, 1sg) when the 3sg meaning was not associated with the 'a' form. These results suggest that it is the phonology of the 3sg form rather than its meaning (or its frequency), that is driving performance. However, participants were also less successful overall in the randomisation condition than in the canonical condition, suggesting that the real language structure of the paradigm, and the real language relationship between forms and meaning, also had a facilitative effect on learning.

7.3 Experiment 4 and 5– Assessing the role of a Skewed Frequency Distribution in the Learning of Spanish Person and Number forms in Children.

In experiments 4 and 5, we aimed to replicate experiments 1 and 2, but in school-aged (9 and 10 year old) children. The English-speaking children in this experiment were more successful at producing the 3sg form than the other forms. As in the adult experiments, there was evidence of this in both conditions, but unlike in the adult experiments there was a significant interaction, indicating that the 3SG advantage was significantly greater in the skewed condition. Thus, there was greater evidence in the children than in the adults that frequency affects production, but there was again a 3sg advantage even when there was no bias in the input. With regard to 3sg error, children were, in line with our predictions, and unlike the adults, more likely to produce 3sg errors in the skewed condition than in the uniform condition. In both the skewed and uniform conditions, children were more likely to produce 3sg errors when another singular form was the target, suggesting that children learn about and use number information, but may not be sensitive to person information.

Experiment 5 addressed the same potential methodological confound as Experiment 2, and the results were broadly the same as Experiment 4.

In order to assess what cues children use to learn the different inflections, we conducted a Bayesian Multinomial Regression analysis. This allowed us to determine whether the participants could use the different person and number features that were embedded into our training study. Our analysis showed children in the skewed condition and children in the uniform condition (with 40 training trials)

could use the Number cue to learn the different inflections. Children within the Uniform condition with 80 training trials could not effectively use either the Person or Number cues. Our analysis demonstrated that there was confusability amongst all the different inflections, but suggested that children were more likely to produce an error in relation to the Person feature, with children effectively discriminating between number features (i.e. between singular and plural forms), but not between person features (i.e. 1st, 2nd and 3rd person forms).

7.4 Experiment 6 and 7– Attempting to Improve Learning with the Use of a New Methodology.

The results from the previous studies broadly replicate those of the adults. However, the children in this study were less successful overall than the adults, and it appears from our analysis of how the children used the cues that they made limited use of grammatical person.

In order to try and improve performance for the children, we introduced a physical "teacher" in the form of stuffed animal to increase the salience of the person features. The expectation was that using a physically present "speaker" would enhance sensitivity to the social cognitive context that underlies grammatical person. Experiment 6 utilised this experiment to assess correct production and erroneous 3sg productions again, where we compared the skewed condition and the uniform condition with 80 training trials. Experiment 7 was the same as Experiment 6 but, as with previous chapters was included to address a possible confound. The results were broadly the same as for experiments 4 and 5, except for the additional clear

pattern that children were more likely to produce a 3sg error when another singular form was the target, suggesting overreliance on the number cue.

We also needed to establish what cues the children used to produce the different inflections. Our analyses revealed that children in the skewed condition and the uniform condition with 80 training trials, could effectively use the number features to learn the different inflections, but children in the uniform condition with 40 training trials could not effectively use either the person or number cues. Again, as also indicated in the 3sg analyses, participants seem adept at learning the different number features, but neglect the person features.

7.5 The Use of Person and Number Features in our Studies

Table 7.2 provides a graphic of the use of different cues within our studies. As part of the analysis, we assessed what cues the participants and children used in order to produce the different inflections that were included in the training study. Table 7.2 shows what cues the participants used.

In our experiments, we aimed to assess to what extent the participants and children within our study could use the different person and number cues to learn the different inflections that were taught. Our results showed that the adults within our study could effectively use the Person and Number cues to produce the different inflections within our study. However, the children within our study demonstrate that they are more inclined to use the number cue than the person cue to produce the different inflections.

Table 7.2 - The breakdown of the use of cues across the seven experiments.

Experiment	No. of Participants		Used Person	Used Number
Exp. 1- Adult Study	Skewed	21	✓	✓
Skewed vs Uniform 80	Uniform 80	25	✓	✓
Exp. 2 – Adult Study	Skewed	21	✓	✓
Skewed vs Uniform 40	Uniform 40	24	✓	✓
Exp. 3 – Adult Study	Phonology	20	-	-
Randomisation	Semantics	20	✓	✓
Exp. 4 – Child Study	Skewed	17	X	✓
Skewed vs Uniform 80	Uniform 80	20	X	X
Exp. 5 – Child Study	Skewed	17	X	✓
Skewed vs Uniform 40	Uniform 40	17	X	✓
Exp. 6 – New Design	Skewed	37	X	✓
Skewed vs Uniform 80	Uniform 80	36	X	✓
Exp. 7 – New Design	Skewed	37	X	✓
Skewed vs Uniform 40	Uniform 40	37	X	X

The results in our child studies show that children could only use the number feature effectively to produce the different inflections. This suggests that when producing inflections, children appear to classify the forms they produce by using one feature – the number feature. Children can effectively see when a singular form or a plural form is required. Thus, they can reduce their options down to the singular forms within their inflectional paradigm. However, they cannot discriminate between person features (i.e. they cannot differentiate between 1st, 2nd and 3rd person forms).

7.6 Overall Performance in the Skewed and Uniform Conditions

It is clear from our experiments that both children and adults are more successful at producing the 3sg form when compared to the other forms that they were learning. This result is increased in the skewed condition in the children, where the 3sg form

is produced more successfully than in both uniform conditions (with 40 training trials and 80 training trials).

It seems that, for adults, the frequency with which the learners heard the 3sg form affected overall accuracy. Participants were naturally more successful at producing the 3sg form in the skewed condition, but were also more successful at producing the other inflections too (1pl, 3pl, 2sg and 1sg) when compared to both uniform conditions. Participants in the skewed condition were exposed to the 3sg form in 50% of the trials, and thus were more likely to learn to use this form correctly and identify when the 3sg form was needed. This appears to result in performance being higher in the skewed condition for the other forms as well, due to participants being able to determine when a 3sg form should not be used and hence being less likely to produce a 3sg error.

However, the opposite is true for the children who participated in the experiments reported in chapters 5 and 6. Children in these experiments were more likely to produce a 3sg error in the skewed condition suggesting that the increased exposure to the 3sg form in that condition led them to "default" to this form and produce more errors. For adults, whilst input frequency appears to reduce the production of 3sg errors, for children, input frequency appears to increase the chances of producing a 3sg error. This difference could be explained through the learning potential of adults and children. Adults are able to learn when a 3sg form is required and when it is not, so can use that form correctly and are less inclined to use the 3sg form when that form is not the target. In contrast, children within the skewed condition can also

learn to use the 3sg form successfully, but struggle to learn to use the other forms, thus resort to using the simplest available form when they are unsure.

It is also important to highlight the results relating to the seen/unseen distinction within the experiments. The experimental design included a test of whether the participants could generalise to new verb-ending combinations. Whether these verb-ending combinations were seen or unseen during training allowed us to establish whether participants could generalise at testing. For the adults, including the distinction between seen and unseen trials in the model provided a better fit to that data, suggesting that adults struggled to generalise to forms they had not been taught before. For the children, removing the distinction between seen and unseen trials for three out of the four experiments provided a better fit to the data. On first view, this might seem to suggest that children are better than adults at generalising to forms they have not seen before. However, a more likely explanation is that children perform so poorly on seen forms that there is no difference between their performance on seen and unseen forms. In short, this result probably reflects the fact that the children's performance on both seen and unseen forms is only slightly above chance.

7.7 Differences between Adults and Children

It is important to recognise the difference in performance between the adults and children who were involved in the study. Firstly, the results indicate that adults were more likely to produce a 3sg error in the uniform condition than in the skewed condition compared to the children, where the opposite result was found (i.e. more

errors producted in the skewed than the uniform condition). These differences are important as they reflect the learning ability of the groups we tested.

As expected, the adults' performance was superior to that of the children. A potential reason for the difference in performance is the difficulty of the task for the children. The children in Experiments 4 and 5 used the same methodology as the adults in Experiment 1 and 2, where they learned a series of verbs and person and number combinations. They then had to remember the different combinations and when to use them (e.g. understanding when a 3sg animation was played and how to respond correctly). This was evidently difficult for the children, demonstrated by the low production rates seen across the child experiments.

Another factor to be acknowledged is the length of the task for the children, again potentially explaining the difference in performance between the adults and children. Each participant took part in three training sessions over three days with each session taking between 20 and 30 minutes to complete. The children may have developed a level of boredom towards the end of the study due to them having to complete the same task for each of the three days, thus reducing correct production during the testing phase.

7.8 Theoretical Implications

The results from the experiments have wider implications for the theories discussed in the introduction. The results indicate that the participants' pattern of performance was broadly similar to that shown by Spanish-speaking two-year-olds. That is to say, participants tended to be more accurate in 3sg contexts and to produce errors that

involved the incorrect use of 3sg forms in non-3sg contexts. These findings suggest that the pattern of performance shown by Spanish-speaking two-year-olds can be understood in terms of the interaction between domain-general learning mechanisms and the semantic-distributional properties of Spanish. They thus provide support for a constructivist analysis of this pattern and count against generativist models that see this pattern as a reflection of a maturationally controlled difference between the child and adult grammar.

The results also show that these effects reflect a combination of factors, including the relatively high frequency of 3sg forms in the training set, the phonological centrality of the 3sg -a ending within the present tense paradigm and an advantage for 3sg -a when it coded for 3sg as opposed to some other person number combination. These results provide further support for the idea that the pattern of performance in early child Spanish can be explained in terms of the semantic-distributional properties of child-directed Spanish, but suggest that 3sg errors in Spanish (and by implication subject-verb agreement errors in other languages) do not simply reflect a process of defaulting to the most frequent form of the verb.

The presence in a morphological paradigm of a single form that is not only the most frequent but the phonologically and semantically simplest form is related to the idea of "markedness". This is the observation that across languages, there are different linguistic elements that are neutral, highly frequent and basic, and thus are classified as unmarked as opposed to other forms which are "marked" (Zhang & Tian, 2015). Carminati (2005) proposes that 1st and 2nd person forms and 3rd person forms should be treated differently. 1st and 2nd person cues are stronger than 3rd person cues, and

therefore carry a greater cognitive weight. This could potentially explain the preference for the 3rd person singular form due to this form reducing the cognitive load on children.

Harley and Ritter (2002) propose that only 1st and 2nd person forms carry the status of a grammatical person. In contrast, 3rd person forms do not carry a person specification due to this form being outside of the speaker/addressee dyad. Thus, the 3rd person form is treated as being "unmarked". In the case of Spanish, the 3sg form is treated as an unmarked form due to it being semantically neutral, highly frequent and phonologically reduced (Battistella, 1990; Corbett, 2000). In short, of all the forms produced in the present-tense verb paradigm, 3sg could be considered the least marked form. Why languages contain marked and unmarked forms like this is not clear. However it has been proposed that it is the result of cognitive processes (see e.g Givon, 1991), a claim that our results would appear to support.

7.9 Overall Implications

In considering the overall implications of our results, we will first address the frequency effects found and what they mean for production. We will then highlight how different factors (i.e. Phonology and Semantics) come together to create the "defaulting" effects found in Spanish.

There are clear frequency effects found in our experiments. Children appear to "default" to the most frequent form, where, in Spanish, the most frequent form is the 3sg "a" form. This pattern of results could be expected because of the base rates of production, where children are more likely to produce a 3sg form (regardless of

condition) than they are to produce the other available endings. In the skewed condition, children are more likely to produce the 3sg form successfully when compared to both uniform conditions. However, children are also more inclined to produce a 3sg error when compared to the uniform conditions. However, the adults were less likely to produce a 3sg error within the skewed condition, despite there being a bias toward that form during the training phase.

The frequency effects we found support, to some extent, the notion that frequency both protects children from error and causes error (Ambridge et al., 2015), with children producing the most frequent form more successfully than the other forms in the paradigm, but also more likely to produce the 3sg form than the other forms when another form is the target, thus producing a "3sg error" (e.g. Aguado-Orea & Pine, 2015; Bybee & Hopper, 2001; Diessel & Tomasello, 2000; Rubino and Pine, 1998; Tasumi & Pine, 2017).

It is clear then that frequency has an effect on the learning of morphology. However, the results from the uniform conditions (with 40 training trials and 80 training trials) in our studies demonstrate that frequency alone cannot provide a full explanation of 3sg errors in Spanish. Our results show that there is still a bias toward the 3sg form despite there being no bias in the input. The base rates of production for the children in the uniform conditons provide insight into these results. The base rates of production in the uniform condition suggest that the children involved in this study were more likely to produce the 3sg form when compared to the other endings regardless of whether the form produced was correct or not. Therefore, an alternative

view of “defaulting” is that rather than “defaulting” to the highest frequency form children will “default” to the most phonologically simple form or prototypical form.

Experiment 3 illustrates the role of both phonology and semantics in the production of 3sg/”-*a*” errors. The results demonstrate that, in terms of phonology, “-*a*” errors were produced when the most phonologically similar form was the target – the “-*an*” suffix. Alternatively, in terms of semantics, participants produced errors ascribed to a confusability between both the person and number features, where participants would make more 3sg errors when the 1st person plural and 1st person singular forms were the target forms. These results suggest that “defaulting” behaviour can be attributed to the “-*a*” form being a phonologically simple form, as well as to participants classifying inflections on the basis of a single feature. For example, a child might know to produce a singular form as they can discriminate between singular and plural cues. However, they cannot discriminate between person cues, so they default to the most phonologically simple forms (i.e. the 3sg “-*a*” form and the 1sg “-*o*” form).

Another factor that might lead to “defaulting” is the shape of the phonological space of the morphological paradigm. Chapter 5 highlights the role of phonological distance between the “-*a*” form and the other forms that are available using the Levenshtein measure (Kessler, 1995; Sanders & Chin, 2009). This measure is based on the number of operations (i.e. deletions, insertions and substitutions) needed to modify the original form. The more operations required the greater the phonological distance from the original form. In the case of the “-*a*” form in Spanish, we propose that this is the central item as it requires the fewest operations to change it ino the the

other available forms (e.g. the phonological distance between "-*a*" and "-*an*" is one due to this change only requiring one insertion, whereas the phonological distance between "-*a*" and "-*amos*" is three due to this change requiring three insertions). Where participants make errors, they are most likely to switch to a similar phonological form. Since "-*a*" is the form that is, on average, nearest to the other forms, it is the most likely form to be erroneously produced. Thus, the observed defaulting errors may result from speakers confusing other forms with their nearest neighbour.

An important pattern we observe is that the erroneous production of 3sg forms is not indiscriminate but rather occurs most frequently when another singular form is the target. This suggests that errors are caused by children and adults selecting their inflections on the basis of a single feature – namely the number feature. This supports the claim of Bedore and Leonard (2001; 2005) that children make "near miss" errors, where the produced and the target form differ by only one feature (Engelmann et al. 2019; Savicute, Ambridge & Pine, 2018; Granlund et al. 2019). For example, speakers might produce a present tense third person singular form when a present tense first person singular form was the target. This type of single feature or "near miss" error is typical in our experiments, where the errors produced often differed by a single feature, usually the person feature.

Putting all of this together we propose that "defaulting" effects are a result of the phonological and semantic features of the morphological paradigm. Critically, we propose a centrality effect where the phonological "*a*" form and the 3rd person singular forms are central items within their inflectional paradigms. Phonologically,

"-*a*" is the most central item in relation to the other available inflections as it has close neighbours in "-*an*" and "-*as*" whilst being relatively close to "-*amos*". When we make errors, we move to the most similar item within our phonological space, where this item is the one which requires the fewest operations to derive it from the target form. A similar claim to centrality can be made with respect to semantics. As a consequence of the absence of the 2nd person plural form from our paradigm, the 1sg and 3sg have the largest number of semantic neighbours – other items that differ by only one feature. We can therefore think of them as the semantically central items. We propose that frequency along with these two factors work together in order to cause the high rate of 3sg errors observed.

7.10 Outstanding Issues and Future Research

One concern that might be raised concerning our experimental paradigm is that it is inherently third person in nature. It could be argued that our findings with respect to 3sg errors are a by-product of the task set-up, which encourages participants to treat the verb forms as third person descriptions of the animations even when those animations include pictures of themselves or their interlocutors. However, it is worth pointing out that Experiment 3 directly addresses this issue. If the tasks introduced a bias towards semantically 3rd person forms, such forms would be expected to be more readily produced by the participants in this experiment despite the mismatch between semantics and phonology. However, this trend was not found in the experiment. Instead, we found a bias toward the "-*a*" form, suggesting that phonological simplicity is the key factor responsible for the 3sg advantage in the other experiments.

The stressed/unstressed distinction was discussed in the previous chapters. We maintained the stressed/unstressed distinction in order to maintain the naturalistic properties of spoken Spanish. Within our paradigm, 1pl/-*amos* was stressed (constistent with spoken Spanish). In contrast, 3sg/-*a* and 1sg/-*o* are both unstressed, and thus are highly confusable. Our results demonstrate that there is little confusability between 1pl/-*amos* and 3sg/-*a* and there is confusability between 3sg/-*a* and 1sg/-*o,* supporting the idea that the participants in our study could distinguish forms based on whether they are stressed or unstressed. However, there are other reasons that are more plausible to account for these differences. Firstly, as demonstrated in Chapter 4, -*amos* is the most phonologically distinct form when compared to -*a.* The results in this chapter also show that when you randomise the phonological/semantic mapping, there is more confusability between -*a* and -*an.* We therefore suggest that rather than participants distinguishing between forms based on whether they are stressed or unstressed, they distinguish between forms based on other phonological and semantic properties.

Another consideration that must be made comes from both the adults and children who took part in the study. A potential problem with the adults who took part was that they would have had some experience of a foreign language prior to taking part in the study. Students in English schools are required to learn at least one modern foreign language before they take their GCSEs (e.g. French). We therefore excluded participants who had taken a foreign language at GCSE or even further in their studies (e.g. A Level). It must be acknowledged that the adult participants would have some experience of a language with Person and Number marking, which might have led them to treat the task as an explicit-paradigm filling task, rather than a test

of their inflectional understanding. That is to say their prior knowledge of morphological structure in a foreign language might have influenced their production in this study.

However, whilst this is a legitimate concern, we are assuming a certain amount of knowledge and application of their knowledge to our study. Firstly, we would be assuming that the adults who took part either implicitly or explicitly remembered the Person and Number markings from their previous forgeign language learning at school and, secondly, that they could apply their prior knowledge to our study in order for them to produce the different forms. If this was the case, it would be reasonable to predict that performance in both adult studies would be high. However, this is not the case, especially in the Randomisation studies, where production was considerably lower than in the first study.

In a similar vein, the children that took part in our studies were monolingual English speaking children. The children are therefore adept in a language that does mark Peron and Number features using auxiliaries. It is conceivable that the children's proficiency could have influenced their production due to their prior knowledge. However, similarly to the adults, we would be assuming a substantial amount of knowledge on the part of the children. First, we would be assuming that the children could link their knowledge of Person and Number to the Spanish forms in the task, and, second, we would be assuming that they could then transfer their knowledge to produce the desired forms. The poor performance in the child studies would suggest that this is not the case.

7.11 Future Research

The children in our experiments show low overall accuracy rates. This could be explained through the lack of feedback the child received from the teacher during the experiments. In real language learning, when a child makes an error, the adult involved in the dialogue would sometimes correct the error in the form of a "recast". A recast is where the error made is repeated immediately to the child, but in a corrected form (Cleave et al., 2015). For example, a child may produce the utterance "Him needs lunch" and the adult will correct this form immediately, "He needs lunch".

Recasts appear to facilitate language acquisition by highlighting the error the child has made and the particular feature that the child has yet to learn (Camarata & Nelson, 2006; Nelson, 1989). Saxton (1997) suggests that corrective recasts significantly influence a child's language development by providing negative feedback to the child, and also modelling the correct target form.

In our experiments, children were not provided with any feedback and so had no means of knowing whether the forms they were producing were correct or incorrect. There are two ways in which this could be modified. The participants in our expeirments received no feedback during the training phase, but were given examples at the beginning of the experiment. In future experiments, it may be effective to receive feedback on the forms they produce which could be done in two ways; by using recasts to provide the correct form when an error is produced during training or by providing feedback during an intermediate phase between the training and testing phases.

An interesting finding from this thesis is that 3sg errors are not produced at a high rate for all the different inflections but are specific to when the target inflection is another singular form. Bedore and Leonard (2001;2005) suggest that children make "near miss" errors where the erroneous forms children produce differ by one feature. Within our paradigm, children were more likely to produce a 3sg error when the target was another singular form, suggesting there is confusability between the person features, but not the number features. It would be valuable to explore this pattern further.

One issue with the randomisation experiment we conducted was that using real phonological forms potentially inhibited the learning of semantic cues as keeping "-*a*" allowed participants to use the simplest form rather than considering the semantics of the taught forms. In order to test this experimentally, we could remove the phonological structure from the learning paradigm by making the phonological forms equally similar or dissimilar to each other. This would enable us to explore how children use the semantic structure of their language to produce both correct and incorrect forms when there is no phonological bias; and thus what errors they make as a function of Person and Number features.

A necessary generalisation to this study is to extend the methodology used to examine the different conjugations within the Spanish language (i.e. the -*er* conjugation and the -*ir* conjugation). These conjugations pattern differently to the -*ar* conjugation but the predictions regarding frequency, phonology and semantics

remain the same. Thus, exploring these conjugations experimentally would highlight how these factors operate as a whole within the Spanish inflectional system.

The research in this thesis has focused on how children learn to produce different inflections in Spanish. However, within this study our focus has solely been on the inflections the participants produce as opposed to the verb stem and ending together. In naturalistic speech, the frequency of the verb ending used differs according to the verb stem that is produced alongside it (Lõo, Järvikivi & Baayen, 2018; Mosco del Prado Martin, Kostić & Baayen, 2004; de Jong, Schruder & Baayen, 2000; Karlson, 1986). An important consideration is that different endings do not occur equally with different stems in natural and this should be factored into any full model of learning. Therefore, varying the relationship between stems and endings at training to examine the types of forms the children produce would be a useful addition to the literature.

7.12 Conclusion

This thesis aimed to disentangle the different factors that contribute to children's higher-than-expected rate of production of the 3rd person singular form in Spanish. Previous research has highlighted the role of frequency in children's early production, where children appear to default to the highest frequency form, which in Spanish is the 3sg form "-*a*".

Our research found some effects of frequency experimentally. The children who took part in our studies clearly demonstrated the effect frequency had on both successful performance and on the production of 3rd person singular errors. Firstly, the children were more likely to produce a 3sg inflection in the skewed condition

(which biased the input frequency to the 3sg form) when compared to the uniform condition (where the input frequency of the inflections was equal). The children were more likely to produce a 3sg error when in the skewed condition too. This supports the previous findings where the bias toward the 3sg form in the input increases the chances of producing a 3sg form when another form (e.g. 1pl) should be produced. It could therefore be suggested that frequency is an important factor in the production of 3sg errors. However, the lck of such an effect for the children involved in our study leads us to be cautious when suggesting that it is purely input frequency that leads to "defaulting".

In fact, in some cases, increasing the frequency of 3sg forms in the input decreased the rate of 3sg errors (i.e. producing a 3sg form when another inflection was the target). And, critically, even when there was no such bias, participants also produced the 3sg form more successfully and produced more 3sg errors. Importantly, this result was found in both the adults and the children who took part in the experiments, leading us to explore other factors that could explain the production of 3sg errors in Spanish.

This led us to examine the role of phonology and semantics in the production of 3sg errors. Experiment 3 ,which randomised the phonological and semantic mapping for each adult participant, highlighted the role that both these factors play in terms of successful and erroneous production. We propose that the "defaulting" effects that have been found previously are at least partially the product of phonology and semantics. We posit that this is driven by a centrality effect where the fact that the "-*a*" form is the most central item within the phonological space occupied by the

paradigm and the 3rd person singular is the most central semantically, results in an increased rate of production of the 3sg form. Thus, we argue that it is phonological and semantic factors, that cause defaulting errors in Spanish-learning children's speech, with the bias in the input frequency increasing these effects further in children.

References

Aguado-Orea, J. & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *Plos 1, 10*(3), 1-21.

Aguado-Orea. J. (2004). *The acquisition of morpho-syntax in Spanish: Implications for current theories of development.* Unpublished doctoral dissertation, University of Nottingham.

Aguilar-Mediavilla, E., Sanz-Torrent, M. & Serra-Raventos, M. (2007). Influence of phonology on morpho-syntax in Romance languages in children with specific language impairment. *International Journal of Language and Communication Disorders, 42*(3), 325-347.

Aguirre, C. (2003). Early verb development in one Spanish-speaking child. In D. Bittner, W.U. Dressler & M. Kilani-Schoch (Eds.). *Development of Verb Inflection in First Language Acquisition: A cross linguistic perspective.* Mouton de Gruyter: New York.

Ambridge, B., Pine, J.M. & Lieven, E.V.M. (2014). Child language acquisition: Why universal grammar doesn't help you. *Language, 90*(3), 53-90.

Ambridge, B., Kidd, E., Rowland, C.F. & Theakston, A.L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language, 42*(2), 239-273.

Astington, J.W. (1994). *The Child's Discovery of the Mind (The Developing Child).* Cambridge, MA: Harvard University Press.

Austin, J. (2012). Markedness, input frequency and the acquisition of inflection: Evidence from Basque/Spanish bilingual children. *International Journal of Bilingualism, 17*(3), 259-283.

Bañón, J.A. & Rothman, J. (2019). Being a participant matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in Psychology, 10*(746), 1-17.

Barlow, J.A. & Pruitt-Lord, S. (2014). The role of phonological context in children's overt marking of '-s' in two dialects of American English. In A. Farris-Trimble & J.A. Barlow (Eds.). *Perspectives on Phonological Theory and Acquisition: In Honor or Daniel A. Dinnsen.* Philadelphia, PA: John Benjamins.

Barr, D.J., Levy, R., Scheepers, C. & Tily, H.J. (2013). Random effects structures for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Battistella, E.L. (1990). *Markedness: The Evaluative Superstructure of Language.* Albany: State University of New York Press.

Bedore, L.M. & Leonard, L.B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language and Hearing Research, 44*(4), 905-924.

Bedore, L.M. & Leonard, L.B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Linguistics, 26*(2), 195-225.

Blom, W.B.T. & Wijnen, F.N.K. (2013). Optionality of finiteness: Evidence for a no overlap stage in Dutch child language. *First Language, 33,* 225-245.

Brown, R. (1973). *A First Language: The Early Stages.* Cambridge, MA: Harvard University Press

Bybee, J. & Hopper, P. (2001). *Introduction to Frequency and the Emergence of Linguistic Structure.* In, J. Hopper and P. Hopper (Eds.), Frequency and the Emergence of Linguistic Structure, 1-27. Philadelphia, PA: John Benjamins.

Camarata, S.M. & Nelson, K.E. (2006). Conversational recast intervention with preschool and older children. In R.J. McCauley and M.E. Fey (Eds.), Treatment of language disorders in children, 237-264. Baltimore: Brookes.

Carminati, M.N. (2005). Processing reflexes of feature hierarchy (Person>Number>Gender) and implications for linguistic theory. *Lingua, 115*(3)*,* 259-285.

Chan, K.Y. & Vitevitch, M.S. (2010). The influence of the phonological neighbourhood clustering-coefficient on spoken word recognition. *Journal of Experimental Psychology, Human Perception and Performance, 35*(6), 1934-1949.

Chiat, S. (1996). *Personal Pronouns.* In P. Fletcher & M. Garman (Eds.), Language Acquisition: Studies in First Language Acquisition, 2nd Edition, 339-335. Cambridge, UK: Cambridge University Press.

Chomsky, N. (1957). *Syntactic Structures.* New York: Mouton de Gruyter.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory.* New York: Springer.

Chomsky, N. (1980). *Rules and Representations.* New York: Columbia University Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use.* New York: Praeger.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry, 36*(1), 1-22.

Clahsen, H., Aveledo, F. & Roca, I. (2002). The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language, 29*(3), 591-622.

Corbett, G.G. (2000). *Number.* Cambridge: Cambridge University Press.

Cowie, F. (1997). The logical problem of language acquisition. *Synthese, 111*(1), 17-51.

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling Optional Infinitive Phenomena. *Proceedings of the 3rd International Conference on Cognitive Modelling, 78-85.*

Dabrowska, E. (2015). What is exactly is Universal Grammar and has anyone seen it? *Frontiers in Psychology, 6,* 1-17.

Dabrowska, E. & Szczerbinksi, M. (2006). Polish children's productivity with case marking: The role regularity, type frequency and phonological diversity. *Journal of Child Langauge, 33,* 559-597.

Davis, B.L., Chenu, F. & Yi, H. (2018). Phonological Selection Patterns in Early Words: A preliminary cross-linguistic investigation. *Canadian Journal of Linguistics, 63*(4), 556-579.

de Jong, N.H., Schreuder, R. & Baayen, R.H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes, 15*(4/5), 329-365.

del Predo Martin, F.M., Kostic, A. & Baayen, R.H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition, 94*(1), 1-18.

Demuth, K. & Tremblay, A. (2008). Prosodically-condition variability in children's early production of French determiners. *Journal of Child Language, 35*(1), 99-127.

Demuth, K. & McCullough, E. (2009). The longitudinal development of clusters in French. *Journal of Child Language, 36*(2), 425-448.

De Villiers, J.G. (1985). Learning how to use verbs: Lexical coding and the influence of input. *Journal of Child Language, 12*(3), 587-595.

Diessel, H. (2007). Frequency effects in language acquisition, language use and diachronic change. *New Ideas in Psychology, 25*(2), 108-127.

Diessel, H. & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics, 11,* 131-151.

Efron, B. & Tibshirani, R.J. (1994). *An Introduction to Bootstrapping (Monographs on Statistics and Applied Probability).* London, UK: Chapman & Hall.

Engelmann, F., Granlund, S., Kolak, J., Zreder, M., Ambridge, B., Pine, J.M., Theakston, A.L. & Lieven, E.V.M. (2019). How input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology, 110,* 30-69.

Ettlinger, M. & Zapf, J. (2011). The role of phonology in children's acquisition of the plural. *Language Acquisition, 18*(4), 294-313.

Evans, K.E. & Demuth, K. (2012). Individual differences in pronoun reversal: Evidence from two longitudinal case studies. *Journal of Child Language, 39*(1), 162-191.

Ezeizabarrena, M.J. (1996). Adquisición de la morfologia verbal en euskra y castellano por niños bilingües. (Acquisition of verbal morphology in Basque and Spanish bilingual children). Bilbao, Spain: Servico Editorial de la Universidad del País Vasco. Cited in J. Austin, (2012). Markedness, input frequency and the acquisition of inflection: Evidence from Basque/Spanish bilingual children. *International Journal of Bilingualism, 17*(3), 259-283.

Ferguson, C.A. & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language, 51*(2), 419-439.

Freudenthal, D., Pine, J.M. & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6,* 17-15.

Freudenthal, D., Pine, J., Aguado-Orea, J. & Gobet, F. (2007). Modelling the developmental patterning of finiteness markings in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science, 31,* 311-341.

Freudenthal, D., Pine. J. & Gobet, F. (2006). Modelling the development of children's use of optional infitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277-310.

Freudenthal, D., Pine, J. & Gobet, F. (2007). Modelling tbe development of children's use of optional infinitives in Dutch and English using MOSAIC. *Cognitive Science, 30*(2), 277-310.

Freudenthal, D., Pine. J. & Gobet, F. (2009). Simulating the referential properties of Dutch, German and English root infinitves. *Language Learning and Development, 5,* 1-29.

Freudenthal, D., Pine. J. & Gobet, F. (2010). Explaining quantitative variation in the ration of optional infinitive errors across languages: A comparison of MOSAIC and the variation learning model. *Journal of Child Language, 36,* 643-669.

Freudenthal, D., Pine, J., Jones, G. & Gobet, F. (2015). Defaulting effect contribution to the simulation of cross-linguistic differences in optional infinitive errors. *In Proceedings of the 37th Annual Meeting of the Cognitive Science Society,* 746-751. *Pasadena.*

Garlock, V.M., Walley, A.C. & Metsala, J.L. (2001). Age-of-acquisition, word frequency and neighbourhood density effects on spoken work recognition by children and adults. *Journal of Memory and Language, 45*(3), 468-492.

Gathercoe, V.C.M., Sebastian, E. & Soto. P. (1999). The early acquisition of Spanish verbal morphology: Across-the-board or piecemeal knowledge? *International Journal of Bilingualism, 3*(2-3), 133-182.

Givón, T. (1991). Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *15*(2), 335-370.

Gobet, F. (1998). Memory for the meaningless: How chunks help. *Proceedings of the 20th Meeting of the Cognitive Science Society, Mahwah, NJ: Erlbaum ,* 398-403.

Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Lieven, E.V.M., Pine, J.M., Theakston, A.L & Ambridge, B. (2019). Language-generaland language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited production study of Polish, Finnish and Estonian. *Journal of Memory and Language, 107*, 169-194.

Grinstead, J.A.R. (1998). *Subjects, Sentenital Negation and Imperatives in Child Spanish and Catalan.* Ph.D. Thesis, UCLA.

Grinstead, J.A.R, De la Mora, J., Vega-Mendoza, M. & Florez-Avalos, B. (2009b). *An elicited production test of the optional infinitive stage in child Spanish*. In J. Crawford., K, Otaki and M Takahashi (Eds.), Proceedings of the 3rd Conference of Generative Approaches to Language Acquisition – North America. Casscadilla Press, University of Connecticut, Storrs.

Guasti, M.T. (1992). Incorporation, excorporation and lexical properties of causative heads. *The linguistic Review, 8.* 209-232.

Guasti, M.T. (1994). Verb syntax in Italian child grammar: Finite and non-finite verbs. *Language Acquisition, 3,* 1-40.

Guasti, M.T. (2002). *Language Acquisiton: The growth of grammar.* Cambridge: MIT Press.

Harley, H. & Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language, 78*(3), 482-526.

Harris, T. & Wexler, K. (1996). The optional infinitive stage in child English. In H. Clahsen (Ed.) *Generative Perspective in Language Acquisition,* Philadelphia: John Benjamins.

Hoekstra, T. & Hyams, N. (1998). Aspects of root infinitives. *Lingua, 106,* 81-112.

Hogan, T.P., Bowles, R.P., Catts, H.W. & Storkel, H.L. (2011). The influence of neighbourhood density and word frequency of phoneme awareness in 2nd and 4th grades. *Journal of Communication Disorders, 44*(1), 49-58.

Karlson, F. (1986). Frequency considerations in Morphology. *Zeitschrift für Phonetic, Sprachwissenschaft und Kommunikationsforschung, Berlin, 31*(1), 19-28.

Kessler, B. (1995). Computational dialectology in Irish Gaelic. *Proceedings of European ACL; Dublin, 60-67.*

Kliesch, C. (2012). Making sense of syntax – Innate or acquired? Contrasting universal grammar with other approaches to language acquisition. *Journal of European Psychology Students, 3*(1), 88-94.

Kuesser, J.B., Leonard, L.B. & Deevy, P. (2018). Third person singular -s in typical development and specific language impairment: Input and neighbourhood density. *Clinical Linguistics and Phonetics, 32*(3), 232-248.

Kunnari, S., Savinainen0Makkonen, T., Leonard, L.B, Maikinen, L., Tolonen, A-k., Luotonen, M. & Leinonen, E. (2011). Children with specific language impairment in Finnish: The use of tense and agreement inflections. *Journal of Child Language, 38*(5), 999-1027.

Lee, A., Hobson, R.P. & Chiat, S. (1994). I, you, me and autism: An experimental study. *Journal of Autism and Developmental Disorders, 24, 155-176.*

Legate, J.A. & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition, 14,* 315-344.

Leonard, L., Caselli, M.C. & Devescovi, A. (2002). Italian children's use of verb and noun morphology during the pre-school years. *First Language, 22*(3), 287-304.

Lewis, M. & Ramsay, D. (2004). Development of self-recognition, person pronoun use and pretend play during the 2nd year. *Child Development, 75*(6), 1821-1831.

Lõo, K., Järvikivi, J. & Baayen, H.R. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition, 175,* 20-25.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The Neighbourhood Activation Model. *Ear and Hearing, 19*(1), 1-36.

Mancini, S., Postiglione, F., Laudanna, A. & Rizzi, L. (2014). On the person-number distinction: Subject-verb agreement in processing Italian. *Lingua, 14* 28-38.

Mann, D. (2012). Chilean clitic reduplication: Implications for morphology and syntax. *Working Papers in TESOL & Applied Linguistics, 12*(2), 17-37.

Markova, G. & Smolík, F. (2014). What do you think? The relationship between person reference and communication about the mind in toddlers. *Social Development, 23*(1), 61-79.

Marshall, C.R. & van der Lely, H.K.J. (2009). The impact of phonological complexity on past tense inflection in children with grammatical-SLI. *Advances in Speech Language Pathology, 9*(3). 191-203.

Mazzaggio, G. (2016). The theory of mind's role in pronoun acquisition: The phenomenon of pronoun reversal in typically developing children. *Studies in the Linguistic Sciences: Illinois Working Papers. Conference Paper,* 55-69.

McElreath, R. (2015). *A Statistical Rethinking: A Bayesian Course with examples in R and Stan.* Florida: CRC Press.

Meisel, J.M. & Ezeizabarrena, M.J. (1996). Subject-verb and object-verb agreement in early Basque. In Clahsen, H. (Ed.), *Generative Perspectives on Language Acquisition: Empirical Findings, Theoretical Considerations and Crosslinguistic comparisons.* Amsterdam: John Benjamins, 201-239.

Metsala, J.L. (1997). An examination of word frequency and neighbourhood density in the development of spoken-word recognition. *Memory & Cognition, 25*(1), 47-56.

Munson, B. & Solomon, N.P. (2004). The effect of phonological neighbourhood density on vowel articulation. *Journal of Speech, Language and Hearing Research, 47*(5), 1048-1058.

Moreton, E. (2008). Analytical bias and phonological typology. *Phonology, 25,* 83-127.

Nelson, K.E. (1989). Strategies for first language teaching. In M.L. Rice & R.L. Schiefelbusch (Eds.), *The teachability of language, Baltimore: Brookes Publishing, 263-310.*

Newport, E., Gleitman, H. & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C.E. Snow & C.A. Ferguson (Eds.), *Talking to children.* Cambridge, UK: Cambridge University Press, 109-149.

Pereles, S., Liceras, J. & Bel, A. (2005). Some remarks on the realization of the realis/irrealis opposition in Child Language: Toward a universal characterisation of the root infinitive stage. *Ottowa Working Papers in Linguistics, 34,* 87-98.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1-2), 73-193.

Pizzuto, E. & Caselli, M.C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language, 19*(3), 491-557.

Poeppel, D. & Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language, 69,* 1-33.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioural Brain and Sciences, 1*(4), 515-526.

Räsänen, S.H.M., Ambridge, B. & Pine, J.M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest form? *Journal of Child Language, 41*(4), 756-779.

Räsänan, S.H.M., Ambridge, B. & Pine, J.M. (2016). An elicited-production study of inflectional verb morphology in child Finnish. *Cognitive Science, 40,* 1704-1738.

Ramscar, M. (2001). The influence of semantics on past tense inflection. In J.D. Moore and Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society.* London: Lawrence Erlbaum Associates.

Rizzi, L. (1994). Early null subjects and root null subjects. In T. Hoekstra and B.D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar.* Amsterdam: John Benjamins.

Roark, B. & Demuth, K. (2000) Prosodic constraints and the learner's environment: A corpus study. *Proceedings of the Annual Boston University Conference on Language Development, 24,* 597-608.

Rubino, R.B. & Pine, J.M. (1998). Subject-verb agreement in Brazilian-Portuguese: what low error rates hide. *Journal of Child Language, 25,* 35-59.

Salustri, M. & Hyams, N. (2003). Is there an analogue to the RI stage in the null subject language. In B. Beachley, A. Brown & F. Conlin (Eds.) *Proceedings of the BUCLD 27,* Sommerville, MA: Cascadilla Press.

Sanders, N.C. & Chin, S.B. (2009). Phonological distance measures. *Journal of Quantitative Linguistics, 16*(1), 96-114.

Savicute, E., Ambridge, B. & Pine, J.M. (2018). The role of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language, 45*(3), 641-672.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24,* 139-161.

Schlinger, H.D. (2009). Theory of Mind: An overview and behavioural perspective. *The Psychological Record, 59*(3), 435-448.

Schwartz, R.G. & Leonard, L.B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language, 9*(2), 319-336.

Shirai, Y. (1993). Inherent aspect and the acquisition of tense/aspect morphology in Japanese. In H. Nakajima and Y. Otsu (Eds.). *Argument Structure: Its syntax and acquisition.* Tokyo: Kaitakusha.

Shirai, Y. (1998). The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language, 26,* 185-221.

Shirai, Y. & Miyata, S. (2006). Does past tense marking indicate the acquistionof the concept of temporal displacement in children's cognitive development. *First Language, 26*(1), 45-66.

Slobin, D.I. (1973). Cognitive prerequisites for the development of grammar. In C.A. Ferguson and D.I. Slobin (eds), *Studies of Child Langugae Development.* New York: Holt, Rinehart & Winston.

Slobin, D.I. (1985). Crosslinguistic evidence fir the language-making capacity. In D.I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisiton Vol. 2.* Hillsdale NJ: Erlbaum.

Song, J.Y., Sundara, M. & Demuth, K. (2009). Phonological constraints on children's production of English third person singular -s. *Journal of Speech, Language and Hearing Research, 52,* 623-642.

Storkel, H.L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language and Hearing Research, 44*(6), 1321-1337.

Storkel, H.L. (2004). Do children acquire dense neighbourhoods? An investigation of similarity neighbourhoods in lexical acquisition. *Applied Linguistics, 25*(2), 201-221.

Storkel, H.L. & Rogers, M.A. (2000). Planning one syllable at a time: The reduced buffer capacity hypothesis in apraxia of speech. *Aphasiology, 13*(9-11), 793-805.

Tatsumi, T. & Pine, J.M. (2016). Comparing generativist and constructivist account of the use of the past tense form in early child Japanese. *Journal of Child Language, 43*(6), 1365-1384.

Tatsumi, T., Ambridge, B. & Pine, J.M. (2018). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production of Japanese. *Journal of Child Language, 45*(5), 1144-1173.

Tomasello, M. (1992a). *First verbs: A case study of early grammatical development.* Cambridge, UK: Cambridge University Press.

Tomasello, M. (1996). *The cultural roots of language.* In B.M. Velichkovsky & D.M. Rumbaugh (Eds.). Communicating meaning. The evolution and development of language (pp. 275-307). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tomasello, (2000). Do children have syntactic competence? *Cognition, 74*(3), 209-253.

Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.

Torrens, V. (1995). The acquisition of the functional category of inflection. *MIT Working Papers in Linguistics, 26,* 451-472.

Tsimpli, I.M. (1992). *Functional categories and maturation: The prefunctional stage of language acquisition.* Ph.D. Thesis, UCL.

van der Hulst, H. (2012). A minimal framework for vowel harmony. In B. Bert and R. Noske (eds.) *Phonological Explorations: Empirical, Theoretical and Diachronic Issues.* Berlin/Boston: De Gruyter.

Venexiano, E. & Parisse, C. (2010). The acquisition of early verbs in French: Assessing the role of conversation and of child-directed input. *First Language, 30*(3-4), 287-311.

Vihmam, M. & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics, 45*(4), 683-725.

Vitevitch, M.S. (1997). The neighbourhood characteristics of malapropisms. *Language and Speech, 40*(3), 211-228.

Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325-329.

Vitevitch, M.S. & Luce, P.A. (1999). Phonotactics, neighbourhood activation and lexical access for spoken words. *Brain and Language, 68*(1-2), 306-311.

Vitevitch, M.S. & Stamer, M.K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes, 21*(6), 760-770.

Vitevitch, M.S. & Rodriguez, E. (2005). Neighbourhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3*(1), 64-73.

Wellman, H.M. & Lui, D. (2004). Scaling of theory of mind tasks. *Child Development, 75*(2), 523-541.

Weschler, S. (2010). What 'you' and 'I' mean to each other: Person indexicals, self-ascription and theory of mind. *Language, 86*(2), 332-365.

Wexler, K. (1992). Optional Infinitives, head movement and the economy of derivations in child grammar. *Occasional Paper 45, Centre for Cognitive Science, MIT.*

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot and N. Hornstein (eds), *Verb Movement*, Cambridge: Cambridge University Press*,* 305-350.

Wexler, K. (1996). The development of inflection in a biologically based theory of language acquisition. In M.L. Rice (Ed), *Toward a Genetic Language*. Mahwah, NJ: Erlbaum. 113-114.

Wexler, K. (1998). Very early parameter session and the unique checking constraint: A new explanation of the optional infinitive stage, *Lingua, 106, 23-79.*

Wijnen, F.N.K., Kempen, M. & Gilles, S. (2001). Bare infinitives in Dutch early child language: An effect of input? *Journal of Child Language, 28*(3), 629-660.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Yang, C. (2002). *Knowledge and Learning in Natural Language.* New York: Oxford University Press.

Yang, C. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Science, 8,* 451-456.

Zhang, Y-Q. & Tian, F-J. (2015). Study on Markedness in linguistics. *Sino-US English Teaching, 12*(9), 667-671.

Appendices

Appendix 1: Full Bayesian Models cited in Chapter 3,4,5 and 6.

Experiment 1

Table 1 – Adult Participants – Skewed vs Uniform 80 - Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Unseen	-0.5287703	0.2702111	-0.9751898	-0.0824056
Uniform 80 *"-a"*	1.3972525	0.4167731	0.7537272	2.11057155
Uniform 80*"-amos"*	1.7696937	0.8721783	0.3746707	3.24760733
Uniform 80*"-an"*	0.2693893	0.6300831	-0.7791892	1.2853177
Uniform 80*"-as"*	-1.1326004	0.9310785	-2.6940363	0.37507996
Uniform 80*"-o"*	-0.5971507	0.7398935	-1.8427998	0.57254518
Skewed *"-a"*	2.6070544	0.4847898	1.8453613	3.44405891
Skewed *"-amos"*	1.63858	0.9291018	0.191241	3.16768398
Skewed *"-an"*	0.9002511	0.7342248	-0.2565631	2.12235943
Skewed *"-as"*	-0.6569587	1.019091	-2.335268	0.99263269
Skewed *"-o"*	-0.9446285	0.8359004	-2.3507014	0.35910111

Table 2 – Adult Participants – Skewed vs Uniform 80 – 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 80*"-amos"*	-0.6910329	0.5129042	-1.5198956	0.1650292
Uniform 80*"-an"*	-0.322345	0.4641081	-1.0857922	0.4338302
Uniform 80*"-as"*	0.5944782	0.4509849	-0.1373181	1.3433275
Uniform 80*"-o"*	0.4605682	0.459998	-0.2758552	1.2140976
Skewed *"-amos"*	-1.6084932	0.5271274	-2.4740478	-0.758041
Skewed *"-an"*	-1.6668256	0.5158927	-2.5201231	-0.8334441
Skewed *"-as"*	0.8878828	0.4849797	0.09650194	1.6723233
Skewed *"-o"*	0.5175615	0.4777087	-0.2698053	1.2883546

Experiment 2

Table 3 – Adult Participants – Skewed vs Uniform 40 – Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 40 *"-a"*	0.686033	0.3364915	0.1385478	1.24792961
Uniform 40 *"-amos"*	1.0611085	0.7927046	-0.1952561	2.37579819
Uniform 40 *"-an"*	-0.7972046	0.7117639	-1.9425216	0.3876663
Uniform 40 *"-as"*	-1.4667936	0.9778809	-3.1453747	0.07185627
Uniform *"-o"*	-0.9672808	0.7034048	-2.1380701	0.16870761
Skewed *"-a"*	2.5257401	0.4376584	1.8657809	3.29186921
Skewed *"-amos"*	1.5834265	0.8935853	0.1924907	3.13234019
Skewed *"-an"*	1.1219095	0.8036044	-0.1275139	2.49544519
Skewed *"-as"*	-0.6268646	1.0129145	-2.3242076	1.0375967
Skewed *"-o"*	-0.8181654	0.7680833	-2.1245155	0.40955192

Table 4 – Adult Participants – Skewed vs Uniform 40 – 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 40 *"-amos"*	-0.9325175	0.3916157	-1.582495	-0.2795451
Uniform 40 *"-an"*	-0.5675629	0.3270281	-1.1118877	-0.041338
Uniform 40 *"-as"*	0.8953268	0.3266359	0.3564493	1.43105576
Uniform 40 *"-o"*	1.3313997	0.3412513	0.7677699	1.89355721
Skewed *"-amos"*	-1.5255179	0.4130523	-2.2301799	-0.8562881
Skewed *"-an"*	-1.5645873	0.3997553	-2.2390265	-0.9350667
Skewed *"-as"*	0.8426685	0.3649106	0.242895	1.43336077
Skewed *"-o"*	0.4588879	0.3546796	-0.1195129	1.04249445

Experiment 3

Table 5 - Phonology Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Randomisation *"-a"*	-0.6289975	0.3895647	-1.28097	-0.0070158
Randomisation *"-amos"*	-3.0537	0.9923764	-4.7921934	-1.5778972
Randomisation *"-an"*	-3.7647196	0.7346124	-5.0625195	-2.686867
Randomisation *"-as"*	-3.1212555	0.8448852	-4.55057	-1.8303589
Randomisation *"-o"*	-1.8629582	0.5325629	-2.7516611	-0.9993939
Uniform *"-a"*	1.4123916	0.391113	0.777297	2.06728809
Uniform *"-amos"*	1.874434	0.8485179	0.618312	3.35541728
Uniform *"-an"*	0.2334602	0.4483808	-0.5156408	0.96569824
Uniform *"-as"*	-0.9737616	0.6464903	-2.0756631	0.05942921
Uniform *"-o"*	-0.4108097	0.4668461	-1.1856092	0.32161813

Table 6 - *"-a"* Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Randomisation *"-amos"*	-0.9550649	0.361001	-1.5585897	-0.385578
Randomisation *"-an"*	0.45123741	0.5711619	-0.4759777	1.3913356
Randomisation *"-as"*	-0.7900941	0.5006366	-1.6058312	0.00637198
Randomisation *"-o"*	-0.42478	0.5595615	-1.3569328	0.473605
Uniform *"-amos"*	-0.9567214	0.4277064	-1.6680359	-0.2760396
Uniform *"-an"*	-0.0813676	0.6372278	-1.1090414	0.97235051
Uniform *"-as"*	0.36202766	0.5140809	-0.4926749	1.20414103
Uniform *"-o"*	0.36110753	0.55599	-0.5370623	1.27021787

Table 7 - Semantic Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Unseen	-2.3230335	0.608858	-3.3231639	-1.3615345
Randomisation 3sg	-1.9886812	0.5125776	-2.8552599	-1.1992444
Randomisation 1pl	-1.736266	0.8162024	-3.0798329	-0.4253116
Randomisation 3pl	-2.5983455	0.6359406	-3.6753839	-1.6154054
Randomisation 2sg	-1.6321745	0.686651	-2.7662305	-0.5489415
Randomisation 1sg	-2.4592244	0.6011691	-3.4714866	-1.4907477
Uniform 3sg	1.52913	0.4344488	0.8136724	2.24591863
Uniform 1pl	1.7749169	0.7502635	0.6291512	3.05971609
Uniform 3pl	0.2159142	0.4675554	-0.5383475	0.9790695
Uniform 2sg	-0.9108879	0.6154897	-1.9268415	0.05854851
Uniform 1sg	-0.3848326	0.4720634	-1.168398	0.37618793

Table 8 - 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Randomisation 1pl	-1.2769537	0.4140875	-1.9673915	-0.6305367
Randomisation 3pl	-1.8479893	0.4120682	-2.5274704	-1.1754927
Randomisation 2sg	-1.9630604	0.4261296	-2.6703737	-1.2852931
Randomisation 1sg	-1.9097877	0.4172194	-2.5903728	-1.2280579
Uniform 1pl	-0.7566114	0.4459068	-1.4991184	-0.02429
Uniform 3pl	-0.3605937	0.4001871	-1.0205052	0.2969965
Uniform 2sg	0.514588	0.3699718	-0.072134	1.14431048
Uniform 1sg	0.384542	0.3812993	-0.234932	1.02000722

Experiment 4

Table 9 – Child Participants – Skewed vs Uniform 80 – Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 80 *"-a"*	-0.6286082	0.297217	-1.1391572	-0.150412
Uniform 80*"-amos"*	-3.6758922	0.6699146	-4.8461832	-2.6735976
Uniform 80*"-an"*	-1.929389	0.360788	-2.5448473	-1.3606769
Uniform 80*"-as"*	-1.1543019	0.3159411	-1.669926	-0.6371275
Uniform 80*"-o"*	-1.4169238	0.3403567	-1.9844644	-0.8738224
Skewed *"-a"*	0.9290656	0.2994344	0.4362069	1.4336504
Skewed *"-amos"*	-2.1293621	0.4037847	-2.8117462	-1.4847643
Skewed *"-an"*	-2.0888374	0.3868066	-2.7375918	-1.4821832
Skewed *"-as"*	-1.3879735	0.32501	-1.9330398	-0.8679983
Skewed *"-o"*	-2.0388219	0.3836708	-2.6810844	-1.4294412

Table 10 – Child Participants – Skewed vs Uniform 80 – 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 80 *"-amos"*	-0.7205263	0.2923339	-1.1988454	-0.2442483
Uniform 80 *"-an"*	-0.758444	0.3040539	-1.2553997	-0.2574528
Uniform 80*"-as"*	-0.1946653	0.32047	-0.7189665	0.3357555
Uniform 80 *"-o"*	0.3611428	0.3106793	-0.1482345	0.8805092
Skewed *"-amos"*	-0.3553882	0.3053939	-0.8583142	0.1454202
Skewed *"-an"*	-0.1845898	0.3020343	-0.6763103	0.310501
Skewed *"-as"*	0.9785571	0.3168698	0.4635268	1.5022593
Skewed *"-o"*	0.1132549	0.2929147	-0.3573244	0.5888873

Experiment 5

Table 11 – Child Participants – Skewed vs Uniform 40 – Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Unseen	-0.1413543	0.2648353	-0.5788691	0.2922013
Uniform 40 *"-a"*	-3.1618679	0.5532818	-4.1147691	-2.3067321
Uniform 40 *"-amos"*	-1.8341712	0.38625	-2.4845976	-1.226036
Uniform 40 *"-an"*	-1.3903052	0.3074739	-1.9052966	-0.8941302
Uniform 40 *"-as"*	-0.5656724	0.2716266	-1.0144517	-0.1315613
Uniform 40 *"-o"*	0.6898375	0.2421721	0.2966018	1.0924284
Skewed *"-a"*	-2.2594661	0.3630217	-2.8890335	-1.6799113
Skewed *"-amos"*	-2.5855382	0.4288706	-3.3291337	-1.9329458
Skewed *"-an"*	-1.6277298	0.2958327	-2.126561	-1.1521824
Skewed *"-as"*	-2.2296977	0.3487748	-2.8306691	-1.6791881
Skewed *"-o"*	-0.1413543	0.2648353	-0.5788691	0.2922013

Table 12 – Child Participants – Skewed vs Uniform 40 – 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 40 *"-amos"*	-1.6237489	0.3898229	-2.2860437	-1.0105148
Uniform 40 *"-an"*	-0.8687862	0.386811	-1.5161961	-0.245618
Uniform 40 *"-as"*	-0.6587849	0.3713397	-1.2847513	-0.071629
Uniform 40 *"-o"*	-0.347194	0.3999569	-1.0136804	0.30985815
Skewed *"-amos"*	-0.3714071	0.3385839	-0.9321688	0.18531556
Skewed *"-an"*	-0.1965467	0.3408732	-0.7481876	0.37124209
Skewed *"-as"*	1.0025297	0.3543245	0.4315973	1.60585336
Skewed *"-o"*	0.1264529	0.3278217	-0.4136089	0.65502408

Experiment 6

Table 13 – Child Participants (New Design) – Skewed vs Uniform 80 – Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 80 *"-a"*	-0.7007466	0.1826494	-1.00477	-0.3994417
Uniform 80 *"-amos"*	-1.9927969	0.2720281	-2.4536864	-1.5628708
Uniform 80 *"-an"*	-1.8264841	0.2459455	-2.239591	-1.4316161
Uniform 80 *"-as"*	-1.5805624	0.2306712	-1.968397	-1.2134615
Uniform 80 *"-o"*	-0.5904956	0.1851017	-0.9003803	-0.2933344
Skewed *"-a"*	0.7731067	0.1648375	0.5033999	1.0448695
Skewed *"-amos"*	-2.7985928	0.3424597	-3.3831589	-2.2725468
Skewed *"-an"*	-1.5360734	0.2098871	-1.8956599	-1.201725
Skewed *"-as"*	-1.8739273	0.237337	-2.2799444	-1.4969283
Skewed *"-o"*	-2.1034761	0.2473606	-2.5226098	-1.7125978

Table 14 – Child Participants (New Design) – Skewed vs Uniform 80 - 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 80 *"-amos"*	-0.5490374	0.3292886	-1.0960966	-0.0130939
Uniform 80 *"-an"*	-0.5729003	0.3351648	-1.1266655	-0.0379772
Uniform 80 *"-as"*	-0.5703823	0.3255744	-1.1115199	-0.0392261
Uniform 80 *"-o"*	-0.3928933	0.3288247	-0.9319616	0.14531052
Skewed *"-amos"*	-0.5914791	0.2984579	-1.1035644	-0.1066585
Skewed *"-an"*	-0.5829951	0.3028099	-1.0829707	-0.0958889
Skewed *"-as"*	0.7668714	0.2975378	0.2752883	1.24941721
Skewed *"-o"*	0.6497295	0.300894	0.1632786	1.14888889

Experiment 7

Table 15 – Child Participants (New Design) – Skewed vs Uniform 40- Correct Production – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 40 *"-a"*	-0.2590221	0.2014205	-0.5912785	0.07426683
Uniform 40 *"-amos"*	-2.1419426	0.3144558	-2.6768697	-1.6456505
Uniform 40 *"-an"*	-1.7032061	0.2720402	-2.1529183	-1.2675029
Uniform 40 *"-as"*	-1.4916496	0.2282332	-1.8716597	-1.1202188
Uniform 40 *"-o"*	-1.7323328	0.2544339	-2.1554818	-1.3148579
Skewed *"-a"*	0.7718294	0.16716	0.5003194	1.04896975
Skewed *"-amos"*	-2.594282	0.3426207	-3.1562882	-2.0556081
Skewed *"-an"*	-1.2394943	0.2286589	-1.6202686	-0.8728742
Skewed *"-as"*	-1.6819992	0.2389284	-2.0861541	-1.301184
Skewed *"-o"*	-1.7177548	0.275755	-2.1771073	-1.2654056

Table 16 – Child Participants (New Design) – Skewed vs Uniform 40 - 3sg Error – Full Bayesian Model (Day 3)

Condition and Ending	Mean	SD(±)	CI 5%	CI 95%
Uniform 40 *"-amos"*	-0.3827882	0.3171402	-0.8998237	0.13899148
Uniform 40 *"-an"*	-0.2299649	0.3403578	-0.7872362	0.3279065
Uniform 40 *"-as"*	-0.063268	0.3160458	-0.5718525	0.46049916
Uniform 40 *"-o"*	0.03050325	0.3246164	-0.5084962	0.56216017
Skewed *"-amos"*	-0.4455056	0.3015393	-0.9501022	0.03929706
Skewed *"-an"*	-0.254668	0.3121659	-0.7559805	0.25828133
Skewed *"-as"*	1.29221675	0.32468	0.7604383	1.83429958
Skewed *"-o"*	0.86391434	0.3102099	0.3617292	1.38091241

Appendix 2: Example Adult Participant Information Sheet and Consent Form (Chapter 3)

UNIVERSITY OF LIVERPOOL

Committee on Research Ethics

Participant Information Sheet

Connecting statistical learning and language use Version 1.0. 11th September 2015

You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends, relatives and GP, if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

1. **What is the purpose of the study?**

   The purpose of this study is to investigate how people learn to put the right endings on verbs in languages like French. For example, how do speakers learn that when they want to say 'I speak' in French, they need to say 'Je parle', but when they want to say 'We speak', they need to say 'Nous parlons'. Altogether, the study is expected to take around 2 hours; though this will be broken down into three shorter 40-minute sessions on different days.

2. **Why have I been chosen to take part?**

   Because you are a native speaker of English.

3. **Do I have to take part?**

   No - Participation is voluntary and participants are free to withdraw at any time without explanation and without incurring any disadvantage. Anyone who does not want to take part, or who, having started, does not want to continue, will not be coerced into doing so.

4. **What will happen if I take part?**

   You will be taught verbs from a language that they do not know by hearing these verbs paired with animations showing the relevant actions. The verbs will be presented in several different forms (I, you, s/he, we, they).

   You will take part in three sessions on different days. Each session will consist of a teaching phase and a testing phase. In the teaching phase, you will hear each verb form paired with an appropriate animation,and will then be asked to repeat that verb form. In the testing phase, you will be shown the relevant animation and asked if they can remember the word that went with that animation.

5. **Expenses and / or payments**

You will receive (…) EPR Points

6. **Are there any risks in taking part?**

   No risks are envisaged. If you experience any discomfort or anxiety you will not be asked to continue with the study. The researchers are research assistants, undergraduate Psychology students and Postgraduate Students from the University of Liverpool. All have received training from Professor Julian Pine, Professor of Developmental Psychology at the University of Liverpool.

7. **Are there any benefits in taking part?**

   This study has no specific educational benefits, but participants do generally enjoy taking part.

8. **What if I am unhappy or if there is a problem?**

   If you are unhappy, or if there is a problem, please feel free to let us know by contacting Dr Colin Bannard (0151 794 1198 / colin.bannard@liv.ac.uk) and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at ethics@liv.ac.uk. When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher(s) involved, and the details of the complaint you wish to make.

9. **Will my participation be kept confidential?**

   Your responses will be kept anonymous. Each individual data set will be given a participant number, which will be listed with his/her name on a "subject key" – this is simply to allow us to destroy your rating sheet if you withdraw consent for the data to be used after the study has ended. Only the researchers involved will have access to this key. After the study has been completed and written-up, all individual recordings, data files and record sheets will be destroyed. In the write-up of the research, the data will be presented completely anonymously, without referring to individual participants. Participation is entirely voluntary, and you may withdraw at any time without having to give a reason, and without detriment (if you withdraw after the study has begun we will destroy any data already collected).

10. **What will happen to the results of the study?**

    The results of the study will be published anonymously, and without reference to individual responses (e.g., "47% of participants produced errors such as 'Noun parle' instead of 'Noun parlons'").

11. **What will happen if I want to stop taking part?**

    You may withdraw at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and that no further use is made of them.

12. **Who can I contact if I have further questions?**

    The Principal Investigator: Colin Bannard (0151 794 1198 / colin.bannard@liv.ac.uk)

Committee on Research Ethics

PARTICIPANT CONSENT FORM

Title of Research Project: **Connecting statistical learning and language use (Version 1.0. 11th May 2015)**

Researcher(s): Julian Pine, Colin Bannard, Joseph Martin

Please initial box

1. I confirm that I have read and have understood the information sheet dated 11th May 2015 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected. ☐

3. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish. ☐

4. I agree to take part in the above study. ☐

_____	_____	_____
Name of Participant	Date	Signature

Do you speak, or have you ever had lessons in, any language(s) other than English (please tick)? ☐ ☐

Yes No

If yes, then please indicate the language(s) here:

_____	_____	_____
Name of Person taking consent	Date	Signature

_____	_____	_____
Researcher	Date	Signature

Version 1.0. 16/09/2015.

Appendix 3: Example Adult Participant Information Sheet and Consent Form (Chapter 4)

Committee on Research Ethics

Participant Information Sheet

Connecting statistical learning and language use Version 2.0. 1st February 2017

You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends, relatives and GP, if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

13. **What is the purpose of the study?**

    The purpose of this study is to investigate how people learn to put the right endings on verbs in languages like French. For example, how do speakers learn that when they want to say 'I speak' in French, they need to say 'Je parle', but when they want to say 'We speak', they need to say 'Nous parlons'. Altogether, the study is expected to take around 2 hours; though this will be broken down into three shorter 40-minute sessions on different days.

14. **Why have I been chosen to take part?**

    Because you are a native speaker of English.

15. **Do I have to take part?**

    No - Participation is voluntary and participants are free to withdraw at any time without explanation and without incurring any disadvantage. Anyone who does not want to take part, or who, having started, does not want to continue, will not be coerced into doing so.

16. **What will happen if I take part?**

    You will be taught verbs from a made up language that is based on but different from Spanish by hearing these verbs paired with animations showing the relevant actions. The verbs will be presented in several different forms (I, you, s/he, we, they). The different forms of the verb may have the same meaning that they have in Spanish, but will in many cases have different meanings.

    You will take part in three sessions on different days. Each session will consist of a teaching phase and a testing phase. In the teaching phase, you will hear each verb form paired with an animation, and will then be asked to repeat that verb form. In the testing phase, you will be shown the animations again and asked if you can remember the word that went with that animation.

17. **Expenses and / or payments**

    You will receive EPR Points for participation in all three sessions as indicated when you signed up for your timeslots.

18. Are there any risks in taking part?

No risks are envisaged. If you experience any discomfort or anxiety you will not be asked to continue with the study. The researchers are research assistants, undergraduate Psychology students and Postgraduate Students from the University of Liverpool. All have received training from Professor Julian Pine, Professor of Developmental Psychology at the University of Liverpool.

19. Are there any benefits in taking part?

This study has no specific educational benefits, but participants do generally enjoy taking part.

20. What if I am unhappy or if there is a problem?

If you are unhappy, or if there is a problem, please feel free to let us know by contacting Dr Colin Bannard (0151 794 1198 / colin.bannard@liv.ac.uk) and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at ethics@liv.ac.uk. When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher(s) involved, and the details of the complaint you wish to make.

21. Will my participation be kept confidential?

Your responses will be kept anonymous. Each individual data set will be given a participant number, which will be listed with his/her name on a "subject key" – this is simply to allow us to destroy your rating sheet if you withdraw consent for the data to be used after the study has ended. Only the researchers involved will have access to this key. After the study has been completed and written-up, all individual recordings, data files and record sheets will be destroyed. In the write-up of the research, the data will be presented completely anonymously, without referring to individual participants. Participation is entirely voluntary, and you may withdraw at any time without having to give a reason, and without detriment (if you withdraw after the study has begun we will destroy any data already collected).

22. What will happen to the results of the study?

The results of the study will be published anonymously, and without reference to individual responses (e.g., "47% of participants produced errors such as 'Noun parle' instead of 'Noun parlons'").

23. What will happen if I want to stop taking part?

You may withdraw at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and that no further use is made of them.

24. Who can I contact if I have further questions?

The Principal Investigator: Colin Bannard (0151 794 1198 / colin.bannard@liv.ac.uk)

Committee on Research Ethics

PARTICIPANT CONSENT FORM

Title of Research Project: **Connecting statistical learning and language use (Version 2.0. 1st February 2017)**

Researcher(s): Julian Pine, Colin Bannard, Joseph Martin

Please initial box

5. I confirm that I have read and have understood the information sheet dated 1st February 2017 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

6. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected. ☐

7. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish. ☐

8. I agree to take part in the above study. ☐

_____	_____	_____
Name of Participant	Date	Signature

Do you speak, or have you ever had lessons in, any language(s) other than English (please tick)? ☐ ☐

Yes No

If yes, then please indicate the language(s) here:

_____	_____	_____
Name of Person taking consent	Date	Signature

_____	_____	_____
Researcher	Date	Signature

Version 2.0. 1/02/2017.

Appendix 4 Example Child Participant Information Sheet and Consent Form (Chapter 5 & 6)

Dear Parent,

We are members of a research group at the University of Liverpool that studies how children learn to speak different languages. Your Headteacher has been kind enough to allow us to run one of our projects at your school.

In this project, we are investigating how children learn by copying others. Each child will take part in one session or in three sessions on different days. Each session will consist of a teaching phase and a testing phase. In the teaching phase, the child will hear an adult use words (e.g. to describe or request a picture or an object) or perform simple actions (e.g. making a necklace from beads) and may be asked to repeat it. In the testing phase, the child will be presented with the same or a very similar picture or object and asked to demonstrate what they have learned.

Children generally enjoy this kind of study and are very keen to take part. Further details about the study are given on the attached Parent Information Sheet.

If you WOULD like your child to take part in this study, please sign and return the consent form ASAP.

Participation is entirely voluntary and you may withdraw your child at any time without having to give a reason and without any disadvantage to you or your child. If you withdraw your child after the study has begun, we will destroy any data already collected. If your child decides that they do not want to take part on the day, they will be allowed to withdraw, even if you have given your consent.

We do hope that you will be happy for your child to take part in this enjoyable and interesting study.

Thank you in advance for your cooperation.
Yours sincerely,

Professor Julian Pine
Professor of Developmental Psychology
University of Liverpool

Committee on Research Ethics

Participant Information Sheet

How do children learn verb endings? **Version 1.0. 11th May 2015**

You are being invited to allow your child to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends, relatives and GP, if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

25. What is the purpose of the study?

The purpose of this study is to investigate how children learn to put the right endings on verbs in languages like French. For example, how do children learn that when they want to say 'I speak' in French, they need to say 'Je parle', but when they want to say 'We speak', they need to say 'Nous parlons'. Further information is given on the attached letter to parents. Altogether, the study is expected to take around 90 minutes per child; though this will be broken down into three shorter 30-minute sessions on different days.

26. Why have I been chosen to take part?

Because your child is a native speaker of English.

27. Do I have to take part?

No - Participation is voluntary and participants are free to withdraw at any time without explanation and without incurring any disadvantage. In addition to obtaining parental consent, we will ask each child if s/he wants to take part in the study. Any child who does not want to take part, or who, having started, does not want to continue, will not be coerced into doing so.

28. What will happen if I take part?

The child will be taught verbs from a language that they do not know by hearing these verbs paired with animations showing the relevant actions. The verbs will be presented in several different forms (I, you, s/he, we, they).

Each child will take part in three sessions on different days. Each session will consist of a teaching phase and a testing phase. In the teaching phase, the child will hear each verb form paired with an appropriate animation, and will then be asked to repeat that verb form. In the testing phase, the child will be shown the relevant animation and asked if they can remember the word that went with that animation.

29. Expenses and / or payments

None

30. Are there any risks in taking part?

No risks are envisaged. Any child who experiences any discomfort or anxiety will not be asked to continue with the study. The researchers are research assistants and/or undergraduate Psychology students from the University of Liverpool, and will obtain enhanced disclosure certificates before working with children. All have received

training from the researcher in overall charge of the study: Professor Julian Pine, Professor of Developmental Psychology at the University of Liverpool. Children will be seen individually, but always in an area that is in view of teachers, classroom assistants and/or other adults (e.g., a corridor, library or quiet corner).

31. Are there any benefits in taking part?

This study has no specific educational benefits, but children do generally enjoy taking part.

32. What if I am unhappy or if there is a problem?

If you are unhappy, or if there is a problem, please feel free to let us know by contacting Professor Julian Pine (0151 795 9402 / 07714587797 / julian.pine@liv.ac.uk) and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at ethics@liv.ac.uk. When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher(s) involved, and the details of the complaint you wish to make.

33. Will my participation be kept confidential?

Children's responses will be kept anonymous. Each individual child's data set will be given a participant number, which will be listed with his/her name on a "subject key" – this is simply to allow us to destroy your rating sheet if you withdraw consent for the data to be used after the study has ended. Only the researchers involved will have access to this key. After the study has been completed and written-up, all individual recordings, data files and record sheets will be destroyed. In the write-up of the research, the data will be presented completely anonymously, without referring to individual participants. Participation is entirely voluntary, and you or your child may withdraw at any time without having to give a reason, and without detriment (if you withdraw after the study has begun we will destroy any data already collected).

34. What will happen to the results of the study?

The results of the study will be published anonymously, and without reference to individual responses (e.g., "47% of children produced errors such as 'Noun parle' instead of 'Noun parlons'").

35. What will happen if I want to stop taking part?

You or your child may withdraw at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and that no further use is made of them.

36. Who can I contact if I have further questions?

The Principal Investigator: Julian Pine (0151 795 9402 / 07714587797 / julian.pine@liv.ac.uk)

Committee on Research Ethics

PARTICIPANT CONSENT FORM

Title of Research Project: ***How do children learn verb endings?*** **(Version 1.0. 11th May 2015)**

Researcher(s): Julian Pine, Colin Bannard, Joseph Martin

Please initial box

9. I confirm that I have read and have understood the information sheet dated 11th May 2015 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

10. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected. In addition, should my child not want to participate in all or part of the study, s/he is free to decline. ☐

11. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish. ☐

12. I agree to my child taking part in the above study. ☐

_____ _____

Child's Name Date of Birth

_____ _____ _____

Name of Parent/Guardian Date Signature

Does your child speak any languages other than English at home (please tick)? Yes ☐ No ☐

_____ _____ _____

Name of Person taking consent Date Signature

_____ _____ _____

Researcher Date Signature

Version 1.0. 11th June 2015.

Appendix 5: Imitation Headteacher Letter and Consent Form

To be sent as a letter or via email

Dear **Headteacher**

I hope you will not mind me contacting you out of the blue, particularly at this extremely busy time of year.

I am a researcher at the University of Liverpool, and am writing to ask whether you would be willing to consider helping me by allowing me to run a language acquisition study with some of the children in your school. Brief details of the study are given on the attached sample parent information sheet/ consent form and research proposal.

If you feel that you would be able to help, then please contact me on **07952228590 or jmartin9@liverpool.ac.uk. This study will be running from September until January 2019.** If you would like to take part, then the next step would be for me to visit the school to discuss the study with you and with the class teachers concerned, to distribute consent forms (to be sent home to the parents) and to meet the children who would be taking part (to reduce any later shyness, particularly with the younger children). Before beginning any testing, we would obtain written consent from both you and each child's parents.

I hope you will not mind if I telephone you some time over the coming weeks to ask if you would be able to consider helping with this study.

We look forward to hearing from you

Yours sincerely

Joseph Martin,
University of Liverpool

Prof Julian Pine,
Professor, University of Liverpool

UNIVERSITY OF LIVERPOOL

Committee on Research Ethics
HEADTEACHER CONSENT FORM

Title of Research Project: Children's Learning through imitation
Researcher(s): Colin Bannard, Julian Pine, Joseph Martin

I have read and understood the parent information sheet for the above study (Version 1.0; dated 14th October 2016) and consent to allow the researchers access to the children in my school.

_____	_____	_____
Headteacher Name	Date	Signature

_____	_____	_____
Researcher taking consent	Date	Signature

Supervisor:
Julian Pine, Psychological Sciences
Eleanor Rathbone Building, University of Liverpool
Bedford St. South, Liverpool, L69 7ZA
Tel: 0151 795 9402
Mobile:TBC
Email: jpine@Liverpool.ac.uk

Researcher
Joseph Martin
Address: As for Julian Pine

Email: jmartin9@liverpool.ac.uk

Version 1, 14th October 2016