Using Production and Online Sentence-Processing Paradigms to Investigate Young Children's Restriction of Linguistic Generalizations

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2016

Ryan Blything

School of Health Sciences

Contents

Word Count: 45,421

List of Tables

List of Figures

Abstract

Using Production and Online Sentence-Processing Paradigms to Investigate Young Children's Restriction of Linguistic Generalizations
Ryan Blything, The University of Manchester
For the degree of Doctor of Philosophy (PhD) 3rd July 2016

A crucial component of child language acquisition is successful generalization. First, a speaker must acquire abstract knowledge of how a particular linguistic-structure conveys meaning, and use this knowledge to generalize the structure to new lexical-items. For example, a speaker can use abstract knowledge of a SUBJECT-VERB-OBJECT structure to produce a sentence such as *The man rolled the ball*, even if the verb *roll* has never been encountered in this structure before. Second, a learner must appropriately restrict 'overgeneralizations' whereby a structure is used with an unsuitable verb (e.g. **The man fell the boy*). The most prominent theories regarding restriction of overgeneralizations are based on frequency of use and (semantic, phonological or pragmatic) compatibility between the item and construction. Since developmental evidence for these accounts is mostly limited to the judgment paradigm, which is unsuitable for testing children aged 5 and under, the aim of this thesis was to examine whether these restriction mechanisms are used by children as young as 3 or 4 – whose generalization mechanisms are at an earlier stage of development - and to develop new paradigms for doing so.

Study 1 used a production priming paradigm to examine children's (aged 3-4; 5-6) restrictions of verbal *un-* prefixation (e.g., **unbend*). Children's production probability of verbs in *un-* form (e.g., **unbend*) was negatively predicted by the frequency of the target verb in bare form (i.e., *bend/s/ed/ing*) and by the frequency of synonyms to a verb's *un-* form (e.g., *straighten* for **unbend*). Additionally, grammaticality judgments from older children (aged 5-6) revealed that preferences for *un-* forms were positively related to the extent to which the verb's semantics overlapped with a covert, probabilistic semantic "cryptotype" of meanings thought to be shared by verbs that are grammatical in *un-* form (e.g., *tie, pack, twist, screw, cover*).

Study 2 investigated whether overgeneralization errors in the domain of English past-tense (i.e., when 'regular' inflections are applied to verbs that require 'irregular' inflection; e.g., **sleeped*, **throwed*) are best attributed to analogy across exemplars, or to a default, "add *–ed*" rule applied regardless of a verb's memorized associations. Past-tense forms of novel verbs were elicited by showing children (aged 3-4; 5-6; 6-7; 9-10) animations of an animal performing a novel action described with a novel verb (e.g., *gezz; chake*) and asking what the animal 'did yesterday'. A verb's likelihood of receiving regular inflection (e.g., *gezzed*, *chaked*) was positively associated with its phonological similarity to existing regular verbs, consistent with the analogy-based approach.

Study 3 investigated the suitability of an online measure of sentence processing, namely Event Related Potentials (ERP), to investigate the role of verb-frequency in restricting transitive overgeneralizations. In line with previous studies, 'P600' and 'LAN' components were evoked in response to overgeneralization errors. However, the magnitudes of these components were not sensitive to a manipulation of verb-frequency (*e.g., *The clown laughed the boy* vs. **The clown giggled the boy*), raising doubt toward the suitability of ERP for examining the relative acceptability of overgeneralization errors.

Overall, the research indicates that even young children's generalizations are sensitive to the linguistic input (i.e., statistical regularities and generalized semantic or phonological patterns of use) and are not well explained by a system of abstract rules that act on discrete categories, whether this applies to syntactic categories (e.g., add *-ed* to any instance of the category "VERB") or discrete verb classes (e.g., a narrow-range rule that acts invariably on any verb that is part of an 'alternating' verb-class).

Declaration

This thesis is my own work and no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning;

SIGNATURE HERE DATE HERE

Copyright Statement

Acknowledgements

I have had a fantastic experience at the University of Manchester and have thoroughly enjoyed doing a PhD here. I am privileged to have worked under the supervision of Prof. Elena Lieven and Dr. Ben Ambridge and am hugely grateful for their insights on Language Acquisition Research that have helped shape my understanding of the subject – and for their encouragement that has kept me motivated throughout the PhD.

Additional thanks go to everyone at the Child Language Laboratory for providing a productive research environment that has been a pleasure to be a part of. I look forward to future collaborations with many of the researchers I have met during my PhD. I am also thankful to the schools and nurseries of Manchester for allowing me to collect data on site – this includes the Moss Hey Primary School, Dryden Street Nursery, and Tiny Toes Nursery. Great thanks must also go to the Economic and Social Research Council (ESRC) for funding my PhD work.

Finally, I thank my parents Fiona and Don, and Gran Bette for their support, good-humour and inspiration.

1. Review of Literature

1.1. Introduction

One of the most challenging tasks for any theory of syntax acquisition is to explain how a child uses knowledge of semantics-to-syntax mappings to produce novel utterances. Indeed, this ability is one of the most crucial aspects of communication – without it a speaker can produce only a set of rote-learned utterances. For example, an event in which a patient performs an action can be communicated using an intransitive-inchoative argument-structure (*e.g., The ball rolled*), and one can communicate the agent and patient of the action by using a transitive-causative argument-structure (*e.g., The man rolled the ball*). A child may acquire an abstract generalization of these semantics-to-syntax patterns to allow any verb to be used in a particular argument-structure. For example, an abstract representation of the transitive-causative structure (roughly SUBJECT-VERB-OBJECT) may be used to produce *The man snapped the twig* even if the child has never experienced *snapped* in that structure before (i.e., it is possible that *snapped* had only been heard in the intransitive-inchoative, *The twig snapped*, or even without an argument, *Snap!*).

However, these generalizations are only partially productive. That is, a child must learn that these generalizations can be applied to some verbs (*e.g., The boy snapped the stick*) and not others (*e.g., *The boy laughed the girl*). How does a child restrict her linguistic generalizations? A number of recent findings (see *Section 1.4*) have suggested that any theory that accounts for children's restriction of linguistic generalizations must include a role for the statistical properties of the verb itself (i.e., *entrenchment*; Braine & Brooks, 1995), the frequency of other formulations that convey the intended message (i.e., *pre-emption*; Clark & Clark, 1979; Goldberg, 1995), and the compatibility between the verb's semantic or phonological properties and those associated with the structure in which it appears (i.e., *semantic/phonological verb properties;* e.g., Ambridge & Lieven, 2011; Pinker, 1989). However, the majority of studies supporting this view have used grammaticality-judgment paradigms, which are thought to be unsuitable for detecting the restriction mechanisms employed by children younger than 5-6 (Ambridge, 2011), who are at the most crucial stage of forming and restricting linguistic generalizations. Examination of the restriction mechanisms employed by children younger than 5 is thus crucial to our understanding of syntax acquisition, and the aim of the thesis is to develop more suitable paradigms for investigating this age-group.

The thesis begins by reviewing two opposing theoretical frameworks of language acquisition: the generativist account (e.g., Chomsky, 1965; *Section 1.2*) and the constructivist account (e.g., Braine, 1976; Goldberg, 1995; *Section 1.3*). A review of these accounts is important with regard to the current investigation of overgeneralization errors because any successful theory of how these errors are restricted must be framed within a more general account of language acquisition that explains how speakers acquire and represent the abstract knowledge (e.g., SUBJECT-VERB-OBJECT) used to produce novel utterances (e.g., **The boy laughed the girl*). Thus, the purpose of *Section 1.2* and *Section 1.3* is to determine whether generativist or constructivist accounts are better placed to explain children's syntactic representations and in turn, which of these accounts should be used as a framework to investigate children's formation and restriction of linguistic generalizations. *Section 1.4* outlines the phenomenon of overgeneralization errors in more detail, reviewing several theories of how these errors are restricted. *Section 1.5* outlines the aims of the thesis in relation to testing the predictions made by accounts outlined in *Section 1.4*.

1.2. The Generativist Account of Syntax Acquisition

1.2.1. What linguistic system is acquired?

The Generativist account of language acquisition dates back to at least Chomsky's (1959) review of Skinner's (1957) *Verbal Behaviour*. Skinner's argument was that language can be acquired by simply repeating the utterances of other speakers (both single-words and whole-utterances) and having these utterances selectively reinforced. Chomsky argued that Skinner's proposal could not be correct because if it were, speakers would possess only a repertoire of rote-learned single-words and phrases. Instead, Chomsky stated that speakers must have a *generative* grammar that enables them to comprehend and produce novel utterances, and illustrated this point with sentences (1) and (2):

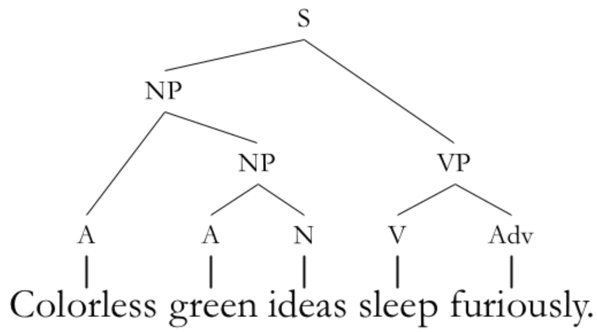
(1) *Colourless green ideas sleep furiously*

(2) **Sleep green colourless furiously ideas*

An English speaker can recognise that sentence (1) is grammatical and sentence (2) is ungrammatical. Since both of these sentences have a novel arrangement of lexical-items (assuming the speaker has never encountered these sentences before) and carry little if any meaning (thus preventing any semantically-motivated analysis of grammaticality), the speaker must have used an abstract understanding of grammar that enables comprehension of novel utterances.

Thus, Chomsky (1957, 1965, 1995) proposes a 'generativist' account of language which holds that sentences are best represented in terms of hierarchical phrase-structure nodes. For example, *Figure 1.1* displays a (simplified) phrase-structure tree for the sentence *Colourless green ideas sleep furiously*.

Figure 1.1 A phrase-structure tree for the sentence *Colourless green ideas sleep furiously*



Note that each lexical-item of a sentence is instantiated by a *lexical-category* (e.g., *colourless*=ADJECTIVE, *ideas*=NOUN), and each instance of a *lexical-category* forms part of a *phrasal-category* such as NOUN PHRASE (NP) or VERB PHRASE (VP)[1]. Most importantly, the phrase-structure tree is governed by a set of abstract rules that specify how *lexical-categories* can combine to form *phrasal-categories* (e.g., VP=V+Adv) and how *phrasal-categories* can combine to form sentences (e.g., S=NP+VP). Crucially then, generativist accounts of syntax acquisition propose that novel utterances are generated by combinatorial rules that act over abstract syntactic categories (i.e., *lexical-* or *phrasal-categories*) as opposed to individual words.[2]

1.2.2. How are phrase-structure rules acquired?

1.2.2.1. The Learnability Problem

A challenge for generativist accounts is to explain how speakers acquire a system of phrase-structure rules. If the learner approaches this task purely on the basis of input (i.e.,

[1] Most generativist theories embody the claim that each *phrasal-category* is built around a 'head'. The 'head' is the single lexical-item that determines the syntactic type of the *phrasal-category*. For example, the head of the NP *colourless green ideas* is the noun *ideas* whereas the head of the VP *sleep furiously* is the verb *sleep*.

[2] Rules that act over individual words as opposed to syntactic categories would be problematic for several reasons. Most notably, it would prevent a learner from knowing where to position *newly learned* words into sentences (for example, even if a learner knows that English places a transitive verb, *kick*, before its direct-object (*e.g., the ball*), they would have no rule that tells them this should also be the case for a newly learned transitive verb such as *push*). Conversely, under a generativist approach, an abstract rule that is formulated in terms of syntactic categories (e.g., VP = V + NP) provides a solution to the problem.

with no prior knowledge of phrase-structure), a 'learnability' problem arises. That is, when the learner is exposed to sentences in the input, an infinite number of hypotheses could be drawn about the phrase-structure rules that generate these sentences. Although the learnability problem applies to the acquisition of morphology and simple-syntax (both the focus of the current thesis), it is most commonly demonstrated at the level of complex syntax, specifically the formation of the English *yes/no* question. The generativist account assumes that English *yes/no* questions, given their non-canonical word-order, must be generated by an abstract *movement* rule that acts on a corresponding declarative sentence (e.g., *He is busy→Is he busy*?). One cannot plausibly learn the movement rule from the input alone because an infinite number of hypotheses can be made, only one of which is correct. For example, a speaker might first hypothesise a rule such as 'swap the subject and auxiliary.' Although such a rule works for the formation of *yes/no* questions with one auxiliary (*The girl is smoking→Is the girl smoking*?), it cannot be extended to *yes/no* questions with more than one auxiliary because the rule does not specify which auxiliary must be moved (e.g., Sentence 1 and Sentence 2).

(1) *The boy who is smoking is crazy→Is the boy who is smoking crazy*?
(2) *The girl is kissing the boy who is smoking →Is the girl kissing the boy who is smoking*?

The speaker may also hypothesise a rule such as 'move the last auxiliary', which works for sentence (1) but not (2) (**Is the girl is kissing the boy who smoking*?), or conversely hypothesise a rule such as 'move the first auxiliary' which works for (2) but not (1) (**Is the boy who smoking is crazy?*). The only way to arrive at the correct hypothesis would be to trial every possible rule, dismissing them one-by-one until the correct one is found. The problem is that the input provides an *infinite* supply of (incorrect) hypotheses (e.g., move the auxiliary which is closest to a determiner/ loudest/ quietest/ first/ last, etc.). How does a speaker acquire the correct rule, which is to move the auxiliary in the *main-clause* to the front of the sentence (see sentence (3) and (4))?

(3) *The boy [who is smoking]* ***is*** *crazy→****Is*** *the boy [who is smoking] crazy*?
(4) *The girl* ***is*** *kissing the boy [who is smoking] →****Is*** *the girl kissing the boy [who is smoking]*?

As a solution, Chomsky posited that all languages must be constrained to conform to a 'universal' grammar, and that children must have innate knowledge of these universals (for example, the relevant principle here – *structure dependence* - stipulates that rules operate on syntactic *structures* such as a main-clause, rather than individual words). Crucially, he posited that universal grammar tightly constrains hypotheses about any given language's phrase-structure rules, thus solving the learnability problem. In the case of yes/no questions, the universal grammar principle of *structure dependence* limits the number of hypotheses to 'move the auxiliary in the main-clause' or 'move the auxiliary in the subordinate clause', which can be tested against a small number of examples in the input. The challenge for generativist accounts is to specify the exact components of universal grammar that a child possesses, and how this may be used to acquire phrase-structure rules.

1.2.2.2. The Bootstrapping Problem

The bootstrapping problem arises from the generativist assumption that abstract combinatorial rules operate over abstract syntactic categories (see *Section 1.2.1*). Thus, to use phrase-structure rules, children must understand the concept of (*lexical* and *phrasal*) syntactic categories (e.g., NOUN, VERB, AUXILLARY, NOUN PHRASE, VERB PHRASE) and syntactic roles (e.g., SUBJECT, OBJECT). The problem here is circular because syntactic categories and syntactic roles are too abstract to be learned *without* phrase-structure rules. For example, there is no obvious semantic property shared by all nouns (take for example, *excitement*, *time,* and *bicycle*) and thus the concept 'noun' cannot be learned this way. Instead the concept can be learned only by recognising that all nouns conform to the same phrase-structure rules (e.g., nouns are words that can appear after a determiner and before an auxiliary). A similar case can be made for learning syntactic roles such as 'subject'. Again, this concept is too abstract to be learned by its semantic properties (i.e., a subject can be an agent as in *He pushed Bob*, an experiencer as in *He liked Bob,* a stimulus as in *He frightened Bob,* etc.) and thus can be learned only by recognising that all subjects conform to the same phrase-structure rules (e.g., all subjects can be marked with nominative case and appear first in canonical declarative sentences).

How can syntactic roles and syntactic categories be acquired without knowledge of phrase-structure rules? As a solution, generativists propose that *syntactic roles* and *syntactic categories* are part of universal grammar, and thus innate. The bootstrapping problem refers to the learner's challenge to link these concepts to real examples in the input language. In other words, innate syntactic categories are of no use to a learner unless he or she has a means of recognizing instances of them in the input. For example, universal knowledge of *lexical-categories* must be linked to the input language's unique lexical-items (e.g., *ball*=NOUN). Similarly, universal knowledge of *syntactic roles* must be linked to language-specific phrase-structure rules. For example, English uses a SUBJECT-VERB-OBJECT word order (e.g., *Peter* ***pushed*** *Bob*) whereas Korean uses a SUBJECT-OBJECT-VERB word order (e.g., *Peter Bob* ***pushed***: in both cases, *Bob* is the one being pushed).

In summary, the challenge for generativist accounts is to specify how universal grammar is used to: (i) assign *lexical-categories* to lexical-items, and (ii) acquire the phrase-structure rules of the input language.

According to Pinker (1984), children break into the phrase-structure of the input language by exploiting universal correspondences between semantics and prototypical instances of syntactic categories and roles. Less prototypical instances can then be learned independently by reading off the newly acquired phrase-structure rules. Although a number of generativist accounts have been proposed to solve the bootstrapping problem, the discussion below focuses only on Pinker's (1984) semantic bootstrapping hypothesis, which is the most specified generativist account of syntax acquisition (and thus the most testable), and is most directly linked with theories of retreat from overgeneralization (the central focus of the thesis). Conversely, other generativist accounts such as parameter-setting accounts (e.g., Chomsky, 1982; Baker, 2001) operate at a higher level of abstraction, and so do not attempt to account for the retreat from overgeneralization.

1.2.2.3. Pinker's (1984) Semantic Bootstrapping Hypothesis

Pinker's (1984) *semantic bootstrapping* account assumes that the following concepts are part of children's universal grammar:

(i) ***Syntactic categories,*** namely *lexical-categories* (e.g., NOUNS, VERBS), and *phrasal-categories* (e.g., NOUN PHRASE, VERB PHRASE);

(ii) ***Syntactic roles*** (e.g., SUBJECT, OBJECT), which represent the functional relationships between phrasal-categories;

(iii) ***Semantic roles*** (e.g., AGENT, PATIENT, ACTION), which represent the relationships between semantic participants observable from the event.

(iv) ***'Semantic-Syntactic' Linking Rules*** (e.g., AGENT→SUBJECT), which link the semantic information of an event to syntactic form.

The underlying assumption of *semantic bootstrapping* is that universal knowledge of correspondences between semantic and syntactic information can be used to 'bootstrap' into the phrase-structure of the input language. For example, suppose that a child has heard the English transitive sentence *The man kicked the ball,* and must acquire the following phrase-structure:

[[NP$_{subject}$ *The man*][VP [V *kicked*] [NP$_{object}$ *the ball*]]]

At this point, the child is assumed to have already learned the meaning of some lexical-items that describe physical objects (e.g., *man, ball*) and is able to match these items to entities in the real world (e.g., *ball* = spherical object). From here, Pinker posits that lexical-categories can be assigned to lexical-items by mapping PHYSICAL OBJECTS onto NOUNS (thus, *man* and *ball* are NOUNS), and mapping ACTIONS (or STATE-CHANGES) onto VERBS (thus, *kicked* is a VERB).

The child is also required to map semantic roles to syntactic roles, and in turn to *phrasal-categories*. First, the child must infer from a corresponding event that the lexical-item *man* refers to the AGENT (i.e., kicker), *ball* refers to the PATIENT (i.e., thing being kicked), and *kick* refers to the ACTION (kicking). Crucially, innate linking rules are posited to map the AGENT, ACTION and PATIENT to SUBJECT, VERB, and OBJECT roles respectively. This allows the SUBJECT-VERB-OBJECT word-order of – for this example, English – to be acquired. The child will now be able to infer on the basis of the surface distribution that *the dog* is a (SUBJECT) NOUN PHRASE (consisting of DETERMINER *the*, followed by NOUN *man*) and that *kicked the ball* is a VERB PHRASE consisting of a VERB (*kicked*) followed by (OBJECT) NOUN PHRASE (*the ball*). Thus, the child has 'bootstrapped' into the phrase-structure of their language.

However, it is important to note that semantic-syntactic linking rules such as those outlined above (e.g., AGENT→SUBJECT) are not sufficient to understand *every* sentence of the input. For example, passive sentences reverse the order of AGENTS and PATIENTS (*e.g., The wagon was pushed by Johnny*), and many sentences do not refer to an AGENT at all (*e.g., The girl received the book*).[3] Such sentences are broadly described as

[3] Pinker (1984) actually posits a hierarchy of semantic roles to address the fact that some sentences do not use AGENTS. The hierarchy is structured such that, when AGENTS are not present in an event, a parser moves down the hierarchy to map PATIENT or THEME to SUBJECT position (and GOAL/SOURCE/LOCATION to OBJECT position). However, Bowerman's (1990) study demonstrates that such a hierarchy is inconsistent with children's longitudinal spontaneous speech data (such that verbs which should benefit from 'canonical' mapping do not emerge before verbs that do not conform to mapping rules). Thus, Pinker's (1984) account is more viable if the initial

'noncanonical' because they do not conform to the 'canonical' semantic-to-syntactic linking patterns used to bootstrap into syntax (e.g., AGENT-ACTION-PATIENT→SUBJECT-VERB-OBJECT). Indeed, this motivates an important component of Pinker's account detailed below.

Specifically, Pinker proposes that once children have bootstrapped into phrase-structure using sentences that *do* conform to universal semantic-syntactic correspondences (e.g., AGENT→SUBJECT; PHYSICAL OBJECT→NOUN), children can use their newly-acquired knowledge of phrase-structure to understand sentences with noncanonical semantic-syntactic mappings. This is essentially a form of distributional learning. For example, a child may detect morphosyntactic similarities between items of the same syntactic category (e.g., NOUNS are likely to follow *the* or *a,* whereas VERBS are likely to end with *–ed* or *–ing*). Thus, a parser can assign syntactic categories to non-actional verbs (e.g., *imagine*) and abstract nouns (e.g., *happiness*) even if the child does not understand the meaning of these words.

With regard to mapping semantic roles to noncanoncial sentences, the child must learn the correct mapping by observing adult speakers' use of the relevant verb (or construction) and observing the corresponding event. For example, upon hearing *The girl received the book*, the child will 'read-off' phrase-structure rules to infer word order (SUBJECT=*the girl*, VERB=*received*, OBJECT=*the book*), and observe the (noncanonical) semantic roles from the event (i.e., *girl*=RECIPIENT; *book*=THEME). Thus, RECIPIENT is mapped to SUBJECT (*the girl*) and THEME to OBJECT (*the book*). Pinker hypothesises that hearing adults' use of noncanonical mapping for a particular verb (or construction) blocks default canonical mapping for that verb in the future.

1.2.2.4. Evaluation of Pinker's (1984) Semantic Bootstrapping

One challenge for Pinker's (1984) account is to explain how a child avoids using noncanonical sentences that would bootstrap into irregular phrase-structure (e.g., *You will get a spanking from me* = OVS). One plausible solution however, is that children use probabilistic learning such that more frequent mappings out-compete less frequent mappings (e.g., AGENT→SUBJECT mappings *usually* result in SVO structure) (e.g., Pinker, 1987).

'bootstrapping' phase relies only on prototypical AGENT-SUBJECT mappings rather than also positing THEME-SUBJECT mappings when AGENTS are not present.

A more serious problem for the account is that semantic-syntactic correspondences vary across languages (this is problematic because any linking-rule that bootstraps into syntax must be universal). For example, nominative-accusative languages such as English use the same nominative case-marker to mark the SUBJECT (i.e., AGENT) of transitive sentences (e.g., *She kicked the boy*) and the SUBJECT (i.e., ACTOR) of the intransitive sentences (*She danced*), but use an accusative case-marker to mark the OBJECT (i.e., PATIENT) of transitive sentences (*The girl kicked him*). In contrast, ergative-absolutive languages such as Dyirbal use ergative case-marking to mark the SUBJECT (i.e., AGENT) of a transitive sentence but use absolutive case-marking to mark the OBJECT (i.e., ACTOR) of the intransitive and the SUBJECT (i.e., PATIENT) of the transitive sentence.

In other words, ergative-absolutive languages map AGENTS and ACTORS to different syntactic roles (SUBJECT and OBJECT respectively), unlike nominative-accusative languages. If nominative-accusative language learners use innate linking rules to map AGENTS *and* ACTORS to SUBJECT, one would expect ergative-accusative language-learners to use the same rule. However, Pye (1990) analysed the speech of children learning K'ich'e (an ergative-absolutive language) and found that children never erroneously mapped ACTOR to SUBJECT. One possible explanation lies in Pinker's (1987) proposed probabilistic learning process. Under this notion, ACTOR→SUBJECT mappings may be treated like noncanonical mappings such that, at the bootstrapping level, they are outcompeted by AGENT→SUBJECT mappings, and must instead be learned *after* bootstrapping has taken place, via the input.

However, more problems arise when one considers split-ergative languages (e.g., Georgian, Hindi), which use ergative-absolutive markers in some contexts (e.g., present-tense) and nominative-absolutive markers in other contexts (e.g., past-tense). The problem here is that split-ergative languages do not seem to have any consistent relationship between semantic roles and syntactic roles (e.g., Siegel, 2000; Tomasello, 2005). Again, it is possible that probabilistic selection of innate linking mechanisms may account for such cases, with exceptions being learned from the input in the same way as other noncanonical mappings. However, critics (e.g., Bowerman, 1990; Ambridge, Pine & Lieven, 2014) argue that this proposal is only appealing if the overall number of noncanonical mappings is limited. Indeed, Bowerman (1990: 1260) remarks that "as the proportion of exceptions mounts, the benefits to be reaped from innate linking rules diminish and the idea becomes less plausible".

In summary, although strong regularities between semantic and syntactic roles clearly exist, it is unlikely they exist without exception across languages. Thus, it is unlikely semantic-to-syntactic linking rules are used as part of a universal grammar to bootstrap children into phrase-structure.

A further problem for semantic bootstrapping – and the generativist approach as a whole - is lexical specificity. Lexical specificity refers to the fact that children's early utterances appear to be limited to lexically-specific representations of argument-structure (e.g., KICKER *kick* KICKEE) (e.g., Dodson & Tomasello, 1998; Tomasello, 1992). Such lexically-specific use of language cannot be explained by generativist accounts because these accounts posit that children's utterances are generated by highly abstract representations of argument-structure ([[NPsubject][VP [V] [NPobject]]]), such that any NOUN and any VERB can be used in that structure. A more detailed summary of evidence for lexical-specificity is provided in *Section 1.3.3* as part of a review of 'constructivist' accounts of syntax acquisition, which are better-suited to explain lexically-specific effects. Constructivist accounts are also well-suited to explain cross-linguistic differences in semantic-syntactic linking regularities, which are posited to be learned rather than innately understood, allowing for substantial variation in the mappings across languages.

1.3.Constructivist Models of Syntax Acquisition

1.3.1. What linguistic-system is acquired?

The constructivist account (e.g., Ambridge & Lieven, 2015; Braine, 1992; Goldberg, 1995, 2006; Tomasello, 2003; Schlesinger, 1988) posits that syntax acquisition involves gradually learning a system of argument-structure constructions by induction. Argument-structure constructions can be described as abstract slot-and-frame schemas associated directly with a particular semantic function. The process of learning argument-structure constructions involves gradual abstraction of lexically-specific schemas (derived from an inventory of rote-learned utterances) into more abstract adult-like schemas.

1.3.2. How are argument-structure constructions learned?

Constructivist accounts posit that children's early knowledge of language is limited to an inventory of rote-learned utterances, each paired with their semantic function. Once the learner recognises that some rote-learned utterances perform a similar semantic function, these are analogised across to form a lexically-specific 'slot-and-frame' schema. For example, the child may analogise across utterances like *He's eating it, He's throwing it,* and *He's hitting it* to form a *He's [X]ing it* schema, where the slot [X] exhibits any property (usually phonological, semantic or pragmatic) shared by items that have appeared in that position as part of an utterance that formed the schema (e.g., *eating*, *throwing* and *hitting* contribute to 'action-like' properties of [X]). Importantly, the properties of schema-slots are 'fuzzy and probabilistic' such that only items whose properties overlap sufficiently with those of the slot can be used grammatically in that slot (e.g., Langacker, 2000).

As the child's inventory of lexically-specific schema grows, she analogises across schemas that share a semantic function to form more-abstract schemas (the term 'more-abstract' refers to the fact that slots become compatible with a broader range of properties, and thus lexical-items). For example, the child might analogise across a *He's [X]ing it* schema (where [X] exhibits 'action' properties) and a KICKER-KICKING-KICKEE schema to form a more-abstract AGENT-ACTION-PATIENT schema. Note that this AGENT-ACTION-PATIENT schema is not a fully adult-like abstract representation of the transitive construction: the AGENT slot exhibits properties that overlap with the properties of agent-subjects, but is not abstract enough to overlap with properties of noncanonical subjects such as EXPERIENCER (as in *She feared John*). Thus, at this stage, the learner has a 'weak' abstract schema that allows comprehension and production of novel transitive sentences that conform to the AGENT-ACTION-PATIENT schema (e.g., *She slapped John*) but not of novel transitive sentences with noncanonical items (e.g., *She feared John*).

Under most constructivist accounts, the last major challenge for the learner is to 'expand' the properties of the AGENT-ACTION-PATIENT schema to form a [X] [Y] [Z] schema (i.e., where [X], [Y], and [Z] exhibit probabilistic properties somewhat similar to the concepts of SUBJECT, VERB and OBJECT respectively).[4] Once this is achieved, children will understand that noncanonical items such as experiencers can be included in the slot previously reserved only for agents. Accounting for this transition is considered one of the biggest challenges for constructivist theorists. For example, *how* might an AGENT-ACTION-PATIENT schema 'expand' to form a SUBJECT-VERB-OBJECT schema? One theory (e.g., Tomasello, 2003) proposes that children must analogize across stored-utterances that 'align' at the structural level. For example, *She slapped John* and *She feared John* overlap at the functional level (both verbs relate to *She*) and the distributional level (e.g., ***She** slapp**ed**; **She** fear**ed***). However, the process behind this transition is still unclear and it is likely to be resolved only by computational modelling (although see Ambridge & Lieven, 2015, for some possibilities).

[4] Slots are labelled with letters as opposed to approximate syntactic or semantic categories to emphasise the assumption that slots exhibit fuzzy and probabilistic properties of items that have previously appeared in the slot.

Ultimately, the theory holds that a child will acquire a number of fully-productive, adult-like argument-structure constructions, such as the *DO-* dative ([P] [Q] [R] [S], where these slots probabilistically exhibit properties similar to the concepts of AGENT, VERB, RECIPIENT and THEME respectively) and *PO-* dative ([A] [B] [C] to [D], where these slots exhibit properties similar to the concepts of AGENT, VERB, THEME and LOCATION/RECIPIENT respectively)[5]. Thus, each argument-structure construction provides its own unique communicative function.

Perhaps most importantly, a constructivist theory posits that children learn argument-structure constructions so that they can be *used* to communicate, and this motive means that the learning mechanism is constrained to generalise across only linguistic units that perform a similar semantic function. The learner is posited to store each schema in memory along with the properties of each of its slots. Since the use of an item in a slot is restricted to items whose properties sufficiently overlap with items that have previously appeared in that slot, the appropriate word order is provided by the schema. Thus, under a constructivist account, there is no need to posit an *innate* understanding of categorical components of grammar such as NOUN and NOUN PHRASE. Instead, children learn constructions that are composed of slots which correspond roughly to traditional syntactic-categories (e.g., the [B] slot of the *PO-* dative corresponds roughly to [VERB]).

In summary, the 'learnability' problem (see *Section 1.2.2.1*) is solved because the learning mechanism is appropriately constrained from the beginning. Thus, there is no need to posit an innate universal grammar to bootstrap into syntax. Rather, the constructivist theory relies on basic social abilities and general cognitive mechanisms, most notably the ability to infer communicative intent (and thus recognise the semantic function of an utterance) and analogical reasoning (i.e., the ability to identify utterances that perform a similar semantic function, and to analogise across distributional similarities between these utterances).

1.3.3. Evaluation of the Constructivist theory

Constructivist theories hold that children's early linguistic representations are shaped by their (uneven) linguistic input, and thus based around lexically-specific schema (e.g., KICKER-KICKING-KICKEE) rather than adult-like abstract generalizations (e.g.,

[5] Letters are used to label each slot, this time to emphasise the point that, even though slot [R] of the *DO-* dative and [D] of the *PO-* dative both exhibit 'recipient-like' properties, the fine-grained differences between items that have appeared in [R] but not [D] (and vice-versa) will contribute to a slightly different constellation of properties.

SUBJECT-VERB-OBJECT). Thus, the theory makes two key predictions. First, if children are at an appropriately early stage of syntax acquisition, they should show better performance (i.e., in any task that requires understanding of correct word order) when they can make use of a lexically-specific schema. Second, even when the child has acquired adult-like abstract constructions, performance at tasks should be better with utterances that are prototypical instances of a construction (e.g., causative transitives vs. non-causative transitives) (because of the probabilistic properties associated with each slot of the schema). Importantly, this prediction contrasts directly with the generativist accounts, which posit that children's representations of language make use of lexically-independent phrase-structure rules (e.g., [[NPsubject][VP [V] [NPobject]]]).

Evidence for early lexically-specific schema has been demonstrated in naturalistic corpus studies. For instance, Pine and Lieven (1993) showed that 77% of children's naturalistic utterances could be generated by one of just ten lexically-specific schema (e.g., Wanna [X], me got [X]) (see also Macwhinney, 1975, Lieven, Pine, & Baldwin, 1997). More evidence is offered by computational modelling, most notably research by Bannard, Lieven and Tomasello (2009) which found that lexically-specific schemas inferred from just 26-to-28 hours of the speech of 2- and 3- year old children could be used by a model to account for 60-80% of 2 year olds' utterances in separately recorded events. Adding the abstract 'verb' category to the model did not improve coverage of 2 year olds' speech, but did improve coverage of 3 year olds' speech, indicating that children's productive schemas become more abstract with time.

Experimental studies testing for abstract (as opposed to lexically-specific) syntactic representations have typically used novel-verb paradigms, the logic being that if the novel verb can be used in a particular word-order without ever having been modelled in that order, this would demonstrate abstract knowledge of that word-order. Several examples are outlined below.

Akhtar (1999) taught children (age 2;1-3;1, 3;2-3;11, 4;0-4;9) novel verbs in either 'weird word order' (e.g., *dacked Elmo the car,* or *Elmo the car dacked*) or 'canonical word order' (e.g., *Elmo tammed the car*). Children were then asked to describe an action that corresponded to one of these novel verbs, with different event-participants. Four-year-old children corrected weird-word-order uses of novel verbs by using canonical word order in around 90% of trials (e.g., *Tigger dacked the fork*), whereas the younger age-groups did so on around 50% of trials (e.g., *Tigger the fork dacked*). It is tempting to conclude from this that older children's representations of transitive word-order are more abstract than younger children's (consistent with the notion that adult-like abstract representations develop gradually, emerging at around 4;0). However, Fisher (2002) argued that such findings can be attributed to younger children's performance limitations. In other words, null-effects cannot provide much insight into a young child's linguistic representations. A closer look at Akhtar's data provides more robust evidence for lexically-specific representations. When younger children *did* correct to canonical word-order, they used pronouns in approximately 50% of arguments (but never used pronouns when imitating weird-word-orders), thus indicating that lexically-specific slot and frame patterns were likely used for producing canonical word-orders (e.g., *He's [ACTION]ing it*). Similarly, Dodson and Tomasello (1998) showed that when novel verbs were modelled in a no-argument construction (e.g., *This is keefing*) children (aged 2;5-3;1) who generalised the novel verb to transitive word-order did so with pronoun arguments 90% of the time. More direct evidence for early lexically-specific schemas was offered by Childers and Tomasello (2001). In training, children (2;4-2;10) heard familiar verbs in transitive word order either exclusively with full NPs (*The cow's pulling the car*) or exclusively with pronouns (*He's pulling it*). Children were also taught novel verbs in non-transitive constructions. At test, 17/20 children in the pronoun group produced novel verbs in transitive word-order (using pronoun arguments 71% of the time) compared to just 9/20 children in the full-NP group (for this group, no pronoun arguments were produced in any trial). Again, this pattern of findings demonstrates that children aged 2-3 show a better understanding of the transitive-construction when they can make use of a lexically-specific schema that is prototypical of the construction (e.g., *He's [ACTION]ing it*).

More recently, however, researchers have argued that adult-like abstract representations may be evident at an earlier age if less-demanding tasks are used. To explore this possibility, Gertner, Fisher and Eisengart (2006) used a preferential looking paradigm, which requires a child to comprehend the word-order of a sentence and look significantly longer at a corresponding video than a non-corresponding video. Children (*M*=1;9) saw two videos side-by-side, one showing a girl as agent and a boy as patient (e.g., a girl pushing a boy's shoulder back and forth) and the other reversing these semantic roles (e.g., a boy rotating a girl in a chair). When hearing a transitive-causative sentence with a novel verb (e.g., *The girl gorped the boy*), children looked longer at the video in which the agent was the subject of the test sentence. Thus 21 month old children appeared to use abstract knowledge of transitive word order to comprehend a transitive sentence containing a novel verb – even though the arguments were full NPs.

Noble, Rowland and Pine (2011) replicated Gertner et al.'s finding with children aged 2;3, 3;4 and 4;3, using a pointing paradigm (which requires a child to point, rather than just look, at the corresponding video of a sentence). In a second experiment, Noble et al. tested the same participants' abstract knowledge of conjoined-agent intransitives (e.g., *The bunny and the duck are dacking*). Only children aged 3;4 and above could identify a corresponding video (i.e., two characters performing a non-causal action) significantly more than a non-corresponding video (i.e., two characters performing a causal action). The authors argued that the abstract representations demonstrated by 2 year olds to comprehend transitive and intransitive sentences cannot be fully adult-like because if this were the case, they would have been able to identify the correct referents of conjoined-agent intransitives as well. This developmental asymmetry indicates that children's abstract knowledge of transitive and intransitive argument-structure may be still developing between the ages of 2 and 3.

What conclusions can be drawn from the research outlined above? First, it is clear that children do have *some* abstract knowledge of verb-argument structure before the age of three (e.g., Gertner et al., 2006; Noble et al., 2011). However, there is also evidence that the abstract representations used by 2-year-olds have not reached an *adult-like* level of abstractness, and are instead either lexically-specific (e.g., Akhtar, 1999; Dodson & Tomasello, 1998; Pine & Lieven, 1993) or somewhat restricted to prototypical instances of argument-structure (e.g., Noble et al., 2011).

Whilst generativist accounts have trouble explaining children's lexically-specific knowledge of argument-structure, the challenge for constructivist accounts is to explain why 2 year-olds appear to have some abstract knowledge of transitive argument-structure. One possible solution lies in Tomasello and Abbot-Smith's (2002: 4) argument that "young 2-year-olds have…a weak transitive schema – one that enables certain kinds of linguistic operations but not others - whereas older children have a stronger and more robust schema based on a wider range of stored linguistic experience." The challenge for constructivist then, is to explain exactly how abstract young children's representation of argument-structure is, and how this abstractness develops. Nevertheless, the research outlined above is consistent with the notion that children learn argument-structure constructions that become more abstract with development.

The next section profiles a phenomenon of children's speech known as verb-argument structure overgeneralization errors, which can be described as when a verb is used in an infelicitous argument-structure (e.g., *The boy laughed the girl*). Such errors constitute evidence that children are productive with their syntactic representations (whether one posits a generativist or constructivist approach), such that they use a mechanism that generalizes uses of verbs into new argument-structures – an ability that is crucial for producing and comprehending novel utterances. Overgeneralization errors are relevant to the generativist versus constructivist debate because any successful theory of syntax acquisition must not only explain how abstract representations are acquired (see above), but also how generalizations of these representations to new lexical-items are (partially) restricted. The evidence discussed above indicates that constructivist (as opposed to generativist) accounts are better placed to explain children's syntactic representations and are thus better suited as a framework in which to explain children's formation and restriction of linguistic generalizations.

1.4 The Retreat from Overgeneralization Errors

The current section outlines the phenomenon of overgeneralization errors (*Section 1.4.1*) and follows this with a review of two major accounts of retreat from overgeneralization. Pinker's (1989) verb-class account (*Section 1.4.2*) holds that verbs must meet particular semantic, phonological, and/ or morphological criteria to be used grammatically in a particular argument-structure, and that children must learn these criteria by recognising which 'narrow-range' semantic classes of verbs are used in that construction. The distinction between these classes is fine-grained, and membership is discrete as opposed to probabilistic. For example, Pinker posits that verbs like *whisper* are less than fully-grammatical in the *DO*- dative (e.g., **Homer whispered Marge the story; cf. Homer whispered the story to Marge*) because these verbs are part of a *manner of speaking* class that cannot appear in the *DO*- dative. Conversely, verbs like *tell* are considered fully-grammatical in the *DO*- dative (e.g., *Homer told Marge the story*) because this verb is a member of a *verbs of communicated message* class. On the other hand, statistical-learning accounts such as pre-emption and entrenchment (*Section 1.4.3*) posit that a verb's argument structure preferences can be learned using an 'inference-from-absence' strategy such that repeated experience of a verb in another context (entrenchment) or repeated use of a synonymous formulation (pre-emption) constitutes as ever-strengthening probabilistic evidence that the verb is ungrammatical in a context in which it has not been heard. For example, the repeated presentation of *whisper* in the *PO*- dative (e.g., *Homer whispered the story to Marge*) constitutes as probabilistic evidence that it is less than fully-grammatical in the *DO*- dative (e.g., **Homer whispered Marge the story*).

As we shall see in the following sub-sections, the fact that previous research has supported both Pinker's verb-class account and statistical-learning accounts is problematic, since Pinker's account cannot explain the probabilistic nature of statistical effects, and statistical-learning accounts cannot explain a role for semantics/phonology. In *Section*

1.4.4., a hybrid account, known as *FIT* (e.g., Ambridge & Lieven, 2011) that explains both verb-class and statistical effects using a constructivist framework (taking into account the discussions in *Section 1.2* and *1.3*) is outlined. As will become clear, the main aim of the thesis is to examine whether *FIT* can be used to explain (i) the restriction mechanisms used by 3-4 and 5-6-year-old children, and (ii) whether these findings can be observed in the domain of morphological overgeneralizations, specifically verbal *un-* prefixation (e.g., **unsqueeze; *unopen*) (to investigate the role of a verb's statistic and semantic properties) and English past-tense (e.g., **drived; *sleeped*) (to investigate the role of a verb's phonological properties).

Section 1.4.2 and *1.4.3* outline existing evidence for the viability of Pinker's verb-class account and statistical-learning accounts respectively. However, evidence for these accounts has typically overlooked the domain of morphological overgeneralizations (hence the need for the thesis to investigate this domain), and have instead focussed on verb-argument structure overgeneralizations. Nevertheless, any theory of retreat from overgeneralization must account for morphological and verb-argument structure generalizations, and thus the evidence outlined in *Section 1.4.2* and *1.4.3* is directly relevant to the aims of the thesis.

1.4.1. Outline of Overgeneralization Errors in Child Language

To produce novel utterances, a child must use a generalization mechanism that allows lexical-items to be used in argument-structures in which they have never been attested. For example, a child may hear the verb *roll* used in the intransitive-inchoative (*The ball rolled*) and the transitive-causative (*The man rolled the ball*), and form a generalization that represents this process in abstract terms. For illustrative purpose, the generalization can be portrayed as follows:

[NP1] [VERB] ↔ [NP2] [VERB] [NP1]
the ball rolled   *the man rolled the ball*

Note that the precise mechanism that underlies this generalization differs under generativist and constructivist approaches. Under generativist approaches (e.g., Pinker, 1989), a lexical rule generates one argument-structure from another (bi-directionally) once it becomes evident that verbs can alternate between the two argument structures. Under constructivist approaches, the two constructions are learned independently as 'surface-constructions' that each denote their own functional purpose (e.g., Goldberg, 2005), and compete to convey the intended message (e.g., MacWhinney, 2004).

Crucially, both accounts posit that speakers make use of this generalization by applying it to lexical-items that have never been attested in the target argument-structure (e.g., *The pencil broke → The man broke the pencil*). Elicited-production studies with novel verbs have demonstrated that children are in use of generalization mechanisms from as young as 2;6 (e.g., *Big Bird is tamming→He's tamming it*; Dodson & Tomasello, 1998; Tomasello & Brooks, 1998) and further demonstrations of children's generalizations can be observed in spontaneous speech. Specifically, children are known to make 'overgeneralization' errors whereby generalizations are applied too broadly such that some verbs are used in infelicitous argument-structures. For example, *laugh* cannot be generalized to the transitive argument-structure because *laugh* cannot appear with a direct-object (*e.g., The man laughed→*The joke laughed the man*). Bowerman's (1988) corpus study reports multiple examples of transitive overgeneralization errors in spontaneous speech, such as **You just cried me* (cf. *You made me cry*), **Don't giggle her* (cf. *Don't make her giggle*), and **The cold stayed them awake* (cf. *The cold made them stay awake*) (grammatically-correct versions in parentheses).

The examples of overgeneralization outlined above have all been in the domain known as the 'causative-alternation' (where 'alternation' refers to a pattern in which a number of verbs systematically appear in two different argument-structures). Other argument-structure alternations have also given rise to overgeneralization errors, namely the dative alternation and the locative alternation. The dative alternation involves a generalization that allows verbs to be used in Prepositional Object (*PO*) datives (e.g., *John gave the flowers to Mary*) and Double Object (*DO*) datives (e.g., *John gave Mary the flowers*), for example:

[NP1] [VERB] [NP2] to [NP3] ↔ [NP1] [VERB] [NP3] [NP2]
John gave the flowers to Mary ↔ *John gave Mary the flowers*

These argument-structures can be differentiated by the fact that *PO*- datives follow a verb with a direct-object theme (e.g., *John gave <u>the flowers</u> to Mary*), whereas *DO*- datives follow the verb with an indirect-object recipient/goal (e.g., *John gave <u>Mary</u> the flowers*). An example of a dative overgeneralization error is when a verb that is grammatical only in the *PO*- dative (*I said no to her*) is erroneously generalized into the *DO*- dative (**I said her no*; e.g., Bowerman, 1988).

The locative alternation involves a generalization that allows verbs to be used in container-locatives (e.g., *Lisa sprayed the table with water*) and content-locatives (e.g., *Lisa sprayed water onto the table*), for example:

[NP1] [VERB] [NP2] [PREP] [NP3] ↔ [NP1] [VERB] [NP3] with [NP2]
Lisa sprayed water onto the table ↔ *Lisa sprayed the table with water*

Container-locatives and content-locatives can be differentiated by the fact that container-locatives mark the container as direct-object (e.g., *Lisa sprayed <u>the table</u> with water*) whereas content-locatives mark the content as direct-object (e.g., *Lisa sprayed <u>water</u> onto the table*). An example of a locative overgeneralization error is when a verb that is grammatical only in container-locatives (e.g., *Lisa filled the wagon with hay*) is erroneously generalized into the content-locative (e.g., **Lisa filled hay into the wagon*) or vice-versa, when a verb that is grammatical only in content-locatives (e.g., *Lisa poured water into the cup*) is erroneously generalized to the container-locative (**Lisa poured the cup with water*).

The fact that children produce overgeneralization errors (and show evidence of generalizing novel verbs; e.g., Dodson & Tomasello, 1998; Tomasello & Brooks, 1998) demonstrates that they must use a generalization mechanism to produce novel utterances. However, children must eventually reach an adult-like state, such that their generalization mechanism is restricted and converges on only grammatical utterances. The problem is compounded by research findings that it is not possible for all of children's errors to be corrected by caregiver feedback (e.g., McNeil, 1966; Demetras, Post & Snow, 1986). Although Chouinard and Clark's (2003) more recent corpus study demonstrated many instances whereby children's ungrammatical utterances are reformulated by their caregiver (e.g. *Child:*I want butter mine*; *Caregiver: ok give it here and I'll put butter on it*), it is clear that an additional mechanism is also required to account for how children understand the need to restrict erroneous generalizations of low-frequency verbs which almost certainly do not receive correction (*e.g., *Can I glow him?; *Salt clings it together;* Bowerman, 1988).

The fact that children cannot restrict overgeneralization errors based on caregiver feedback alone is known as the 'no-negative evidence problem' whereby the child faces a learning 'paradox' (e.g., Baker, 1979). The paradox is that a particular verb-structure combination cannot be deemed to be ungrammatical based on the fact that it is unattested because if this were true, the child would *never* generalize verbs to new structures. The child must restrict ungrammatical generalizations whilst retaining her ability to generalize; since this cannot be achieved using caregiver feedback (e.g., McNeill, 1966) or by registering unattested verb-structure combinations, overgeneralizations must be restricted based on evidence from *attested* verb-structure combinations (i.e., positive evidence) and/or some innate knowledge (e.g., semantic-primitives; see Section 1.4.2).

The leading solutions to the no-negative evidence problem are outlined below, beginning with Pinker's (1989) highly influential 'semantic verb-class hypothesis.' Note that although Pinker's hypothesis uses generativist assumptions (e.g., innate lexical rules and innate semantic-syntactic linking rules), a strength of the account is that its key assumption – that a verb's argument-structure is determined by whether it meets particular semantic, phonological, and/ or morphological criteria – is compatible with (and indeed, adopted by some) constructivist approaches (which assume these criteria are learned in a probabilistic manner, as opposed to a discrete manner).

1.4.2. Pinker's (1989) semantic verb-class hypothesis

Pinker's (1989) verb-class hypothesis assumes that linguistic generalizations can be explained by lexico-semantic rules that transform a verb's semantic-structure, and innate linking rules that map particular semantic-structures to particular argument-structures. To use the dative alternation as an example, a rule may transform the semantic-structure of the verb *give* from roughly "X causes Y to have Z" to "X causes Z to go to Y", with the former mapped to the *DO-* dative (e.g., *Jess gave Scott a present*), and the latter mapped to the *PO-* dative (e.g., *Jess gave a present to Scott*). Under this theory, children are credited with knowledge of semantic primitives (e.g., CAUSE, GO, ACT, BE, HAVE). A lexico-semantic rule can be applied to a verb only if it denotes the discrete semantic primitives associated with the relevant structure.

Pinker posits two types of lexico-semantic rules. The first is the *broad-range* rule, which is formed by a top-down process whereby the learner recognises that any verb which has undergone the alternation can take on two different semantic-structures. For example, any verb that appears in both the *DO-* dative and the *PO-* dative can take on the semantic-structure "X causes Y to have Z" and "X causes Z to go to Y". The broad-range rule is innately-specified to operate on a verb only if that verb is compatible with the discrete set of semantic structures defined by the alternation (e.g., if the verb can denote cause-to-have *and* cause-to-go). The logic here is that the broad-range rule ensures that children's generalizations are constrained by a discrete set of semantic criteria from the beginning and thus some verbs are restricted from ever undergoing the alternation (providing some resolution to the no-negative evidence problem).

However, broad-range rules do not specify *sufficient* criteria for verbs to undergo an alternation. For example, broad-range rules can be applied to *PO*-only verbs such as *shout* (e.g., *Jess shouted the information to Scott→*Jess shouted Scott the information*) because *shout* is compatible with both semantic-structures that are relevant to the dative alternation (one can *shout* someone information, causing the information to metaphorically *go* to the person, or causing the information to be possessed). Since judgement studies have demonstrated *shout* to be considered less-than-grammatical in *DO*- datives, unlike semantically-similar verbs such as *tell* and *read* (e.g., Ambridge, Pine, Rowland and Chang, 2012), an additional mechanism is needed to restrict these types of overgeneralization errors.

The second type of rule is a *narrow-range* rule, which specifies *sufficient* criteria for verbs to undergo an alternation. Specifically, the child must assign verbs that *do* alternate to a narrow-range semantic class. For example, verbs such as *tell* and *read* belong to the *verbs of communicated message* class and can all undergo the dative-alternation. Membership of narrow-range classes is generalized only to new verbs that are consistent with an existing class (e.g., *verbs of communicated message*). Narrow-range rules operate only on verbs that belong to an alternating class, not a non-alternating class (e.g., s*hout* and *whisper* cannot be used in *DO*- datives because they are members of the non-alternating *manner of speaking* class). Thus, *narrow-range* rules constrain overly-general *broad-range* rules.

Given these assumptions, children's linguistic generalizations are tightly constrained by semantic criteria from the beginning. Errors are posited to occur only when the child has (i) misunderstood the meaning of a verb, (ii) has yet to form the appropriate narrow-range semantic verb-class, or (iii) uses 'one-shot innovations', whereby the broad-range rule is used innovatively with inappropriate verbs either for emphatic or other special effect or because the suitable verb is not available to them. For example, Pinker (1989) observed "errors"/innovations in adults' speech such as **We'll credit you back the full purchase price* (cf. *We'll credit the full purchase price back to you*) and **He stripped him the ball* (cf. *He stripped the ball from him*) (where *stripped* refers to stealing the ball from a basketball player).

1.4.2.1. Evidence from different argument-structure alternations

The Dative Alternation:

Gropen, Pinker, Hollander, Goldberg and Wilson (1989) demonstrated evidence of children and adults' knowledge of the *broad-range* rule for the dative alternation, which can be summarised as follows:

(1) X causes Z to have Y by means of causing Y to go to Z (*DO- dative*) ↔
(2) X causes Y to go to Z (*PO- dative*)

The most stringent test of Pinker's theory is to examine children's use of novel verbs because any apparent effect of verb-semantics with familiar English verbs may simply arise from attested usage. Gropen et al.'s (1989) elicited-production experiment demonstrated that children (M= 6;11 or 8;3) produced novel verbs (each corresponding to a novel *motion* action) in *DO*- datives more often when the recipient of the action was highly animate (i.e., human) as opposed to moderately animate (i.e., a toy animal) or inanimate. The finding is evidence for the *broad-range* rule of the dative alternation because it demonstrates that children associated *DO*- datives with a 'cause to have' meaning and thus know that verbs used in the *DO*- dative must have an indirect-object argument that refers to a potential possessor and goal as opposed to an inanimate goal (e.g., *Jack sent <u>Tom</u> /*<u>Boston</u> a package*).

Evidence for the psychological reality of narrow-range semantic constraints on the dative alternation has been provided by Ambridge, Pine, Rowland and Chang (2012). Adults and children (5-6; 9-10) rated the acceptability of *PO*- and *DO*- dative sentences containing novel verbs or semantically-matched English verbs from *PO*-only (*accompanied motion; manner of speaking*) or alternating (*illocutionary communication; giving*) narrow-range classes. All age-groups rated *DO*- dative sentences as more acceptable when they contained a familiar verb from an alternating class (e.g., *Lisa told Bart the story*) as opposed to a *PO*- only class (e.g., **Lisa shouted Bart the instructions*). However, only adults showed the same sensitivity for novel verbs, which provide a more rigorous test of the psychological reality of semantic verb-classes.

The Locative Alternation:

Pinker's *broad-range* rule for the locative-alternation can be summarised as follows:

(3) X causes Y to move into/onto Z (*contents-locative*) ↔

(4) X causes Z to change state by means of moving Y into/onto it (container-locative)

Evidence for the psychological reality of locative *broad-range* rules was offered by Gropen, Pinker, Hollander and Goldberg (1991). Participants (aged 3-4; 5-6; 7-9; adults) were taught novel verbs that denoted *motion in a particular manner* (thus consistent with the semantic-structure of verbs that appear in the contents-locative (3)), or novel verbs that denoted *change of state* (thus consistent with the semantic-structure of verbs that appear in the container-locative (4)). All age-groups were more likely to produce *change of state* verbs in container-locatives, and more likely to produce *manner of motion* verbs in content-locatives, thus demonstrating sensitivity to the broad-range rule.

More recently, Bidgood et al. (2014) tested the psychological reality of specific narrow-range semantic classes posited for the locative alternation. Participants (age 5-6; 9-10; adults) were taught two novel verbs each from a *container-only* class (‘*fill*’ classes) a content-only class (‘*pour*’ classes) or an alternating class (‘spray’ classes). In a grammaticality judgment task, participants judged container-locative uses of novel *fill*-type verbs to be significantly more acceptable than content-locative uses of these verbs, with the opposite pattern observed for novel *pour*-type verbs. Note that the subtle differences between narrow-range classes of locative verbs (i.e., *pour*–type vs. *fill*-type) make the locative alternation a particularly strong test of the semantic verb class hypothesis.

The Causative Alternation:

The broad-range rule for the causative alternation has received less support, and can be summarised as follows:

(5) Y act/go (intransitive-inchoative) ↔

(6) X act on Y, thereby causing Y act/go (transitive-causative)

Recall that Pinker's strategy for solving the no-negative-evidence problem was to ensure that the *broad-range* rule is constrained to act only on verbs that are compatible with both semantic-structures relevant to an alternation (in this case, (5) and (6)). Thus, any verb that undergoes the *broad-range* rule for the causative alternation must be compatible with a dynamic event (i.e., the verb must have ACT or GO in its semantic-structure), and denote a direct, unmediated act. However, Bowerman’s (1988) corpus study reported that a relatively large proportion of children’s overgeneralizations are beyond the scope of this rule. Most notably, non-dynamic verbs (i.e. verbs with BE or HAVE in their semantic representation) such as *stay, be,* and *have* constituted around 30% of errors in the study (*e.g., *The cold stayed them awake*). Bowerman and Croft (2008) also argued that Pinker's theory finds it difficult to 'explain-away' these causative overgeneralizations as misinterpretations or one-shot innovations. For example, it is not clear how a child might interpret verbs like *stay* and *have* as denoting dynamic acts, and at the same time, these verbs are causativized over long periods of time (one child causativized *stay* 43 times between age 2 and 10).

However, evidence for narrow-range semantic classes involved in the transitive alternation is more convincing. In Brooks and Tomasello's (1999) novel-verb elicited production study children (aged 2-3, 4-5, 6-7) were taught two novel verbs, one of which belonged to the alternating *manner of motion* verb class (e.g. *tam* = novel swinging action) and the other to an intransitive-only *directed motion* verb class (e.g. *meek* = novel ascending action). Novel verbs were modelled exclusively in the intransitive construction. At test, the two older groups of children (but not the youngest group, who performed at chance) produced more transitive sentences with the *manner of motion* verb than with the *directed motion* verb. Ambridge, Pine and Rowland (2011) replicated these findings with children (aged 5-6; 9-10) and adults, using a grammaticality judgment task (and different novel actions).

1.4.2.2. General criticisms of Pinker's Semantic Verb-Class Hypothesis

Despite its influential contribution to research into the retreat from overgeneralization, Pinker's (1989) account suffers from a number of problems. First, as already noted, children sometimes go beyond the scope of the broad-range rule for the causative alternation, causativizing verbs such as *stay, ache, be,* and *have,* all of which violate the 'dynamic' constraint of the alternation (e.g., Bowerman, 1988). Another problem for the theory is its reliance on innate semantic-syntactic linking rules. The problem here is that there is considerable cross-linguistic variation in the way that some semantic roles are mapped onto syntactic-structure (see *Section 1.3.3* for an overview of these findings).

Goldberg (1995) rejects the assumption that lexico-semantic rules generate a new semantic representation of a verb because this means having to posit implausible verb meanings. For example, for the sentence *She baked him a cake*, one would have to generate a rather implausible semantic-representation of *bake* that means roughly "x intends to cause y to have z". Conversely, Constructivist accounts (see *Section 1.3.2*) naturally avoid the problem of positing implausible verb senses by positing that constructions have their own meaning that contributes to the meaning of a verb. The process is summarised by Goldberg (1995: 11-12): "the verb…is associated with one or a few basic senses which must be integrated into the meaning of the construction…a constructional approach requires that the issue of the interaction between verb meaning and constructional meaning be addressed." For example, the final interpretation of a sentence such as *She baked him a cake* can be explained in terms the *DO*- dative construction lending its 'cause to have' meaning to the verb *bake*.

Another weakness of Pinker's account is that it cannot account for verb-frequency effects. A large number of studies have demonstrated that children and adults rate overgeneralization errors as more acceptable for low-frequency verbs than higher-frequency verbs from the same semantic class (e.g. **The magician disappeared/vanished him*) (e.g., Ambridge, Pine, Rowland & Young, 2008; Bidgood, Ambridge, Pine & Rowland, 2014; Brooks, Tomasello, Dodson & Lewis, 1999; Brooks & Zizak, 2002; Robenalt & Goldberg, 2015; Theakston, 2004; Wonnacott, Newport & Tanenhaus, 2008). On Pinker's account, it is difficult to explain such frequency effects because membership of a verb-class is assumed to be discrete as opposed to probabilistic (a verb either satisfies the semantic primitives associated with a class, or it does not). For example, *vanish* and *disappear* are both members of the intransitive-only *'going out of existence'* verb-class and thus should be equally unacceptable when causativized. Evidence for verb-frequency effects is outlined in Section 1.4.3 in relation to statistical-learning accounts of retreat from overgeneralization, which are well-suited to explain such frequency effects.

The underlying weakness of Pinker's account is that it attempts to solve the no-negative-evidence problem by positing a set of *discrete* (semantic) criteria that verbs must satisfy in order to alternate between particular argument-structures. The problem here is lexical idiosyncrasy: there are a number of exceptions to the semantic criteria (e.g., Bowerman, 1988; Braine & Brooks, 1995; Goldberg, 1995) and more concerning, there are various degrees of grammaticality within verb-classes (see review of frequency effects in Section 1.4.3 and review of graded semantic effects in Section 1.4.4.3). Thus, an account that posits discrete constraints (where a verb either meets semantic criteria or it does not) cannot be a viable solution to the no-negative evidence problem. More viable solutions to the no negative-evidence problem must account for lexical-idiosyncrasies by positing a learning mechanism where grammaticality is determined not discretely, but probabilistically based on positive evidence (i.e., previous input) of how the relevant verb and similar verbs have been used.

1.4.3. Statistical-learning Accounts

Statistical-learning accounts (e.g., Braine & Brooks, 1995; Goldberg, 1995) posit that overgeneralizations are probabilistically blocked by a learning mechanism that is sensitive to the frequency of the relevant verb in the input. The two most prominent statistical-learning accounts hold different assumptions regarding the contexts to which this learning mechanism is sensitive. On the one hand, the entrenchment hypothesis (e.g., Braine & Brooks, 1995) holds that the learning mechanism is sensitive to the frequency of the relevant verb regardless of its context. For example, a transitive-causative overgeneralization (e.g., **The magician disappeared the rabbit*) is posited to be blocked by use of the verb in intransitive-inchoatives (*The rabbit disappeared*), periphrastic-causatives (*The magician made the rabbit disappear*), and even single-word constructions (*Disappear!*). The repeated use of the verb in a different construction constitutes ever-strengthening probabilistic evidence that the verb is ungrammatical in the construction in which it has not been heard. Thus, the prediction is that high-frequency verbs are more resistant to overgeneralization than low-frequency verbs. For example, since *disappear* is more frequent than *vanish,* and is thus more 'entrenched' in constructions such as periphrastic-causatives and intransitive-inchoatives, a speaker is less likely to overgeneralize *disappear* into the transitive-causative construction (despite these verbs belonging the same semantic verb-class).

On the other hand, pre-emption (e.g., Clark & Clark, 1979; Goldberg, 1995) holds that the learning mechanism is sensitive to the frequency of the verb in only near-synonymous constructions. For example, a transitive-causative overgeneralization (e.g., **The magician disappeared the rabbit*) is posited to be blocked by use of the relevant verb in a periphrastic-causative construction (e.g., *The magician made the rabbit disappear*), but not by use of the verb in less-synonymous constructions such as the intransitive-inchoative (e.g., *The rabbit disappeared*). Thus, under a pre-emption account probabilistic-blocking of a particular formulation occurs only in contexts where that formulation might be expected given its suitability to conveying a particular message.

However, pre-emption and entrenchment measures are difficult to separate empirically because potentially pre-empting constructions are always a subset of potentially entrenching constructions (for example, Ambridge, Pine, Rowland and Chang (2014) reported that the correlation between pre-emption and entrenchment measures for the dative alternation was r=0.9, p<0.001). Nevertheless, the accounts share the underlying prediction that high-frequency verbs are more resistant to overgeneralization than low-frequency verbs, and evidence for this prediction is outlined below.

The majority of evidence for statistical-learning comes from the transitive alternation. Most notably, Ambridge et al. (2008) demonstrated that children as young as 5-6 rated transitive-causative overgeneralizations as least acceptable with high-frequency verbs, more acceptable with low-frequency verbs, and most acceptable with novel verbs (e.g. **The man fell/tumbled/nunged Bart*), and that this effect held across two other verb classes, namely *going out of existence* (**The magician disappeared/vanished/tammed Bart*) and *semi-volitional emotional expression* (e.g. **The clown laughed/giggled/meeked Bart*). The findings corroborate findings of other judgment studies (e.g., Ambridge, Pine, Rowland, Jones & Clark, 2009; Theakston 2004), as well as those of an elicited production study (Brooks & Tomasello, 1999) in which children as young as 4;7 were less likely to produce transitive sentences with novel verbs if those verbs had been modelled in the periphrastic-causative construction.

Further evidence for statistical-learning accounts comes from the dative alternation. Ambridge, Pine, Rowland and Chang (2012) obtained grammaticality judgment ratings of *DO*-dative and *PO*- dative uses of 24 *PO*- only familiar verbs, varying in frequency. Children (5-6; 9-10) and adults showed a positive correlation between verb frequency and preference ratings for grammatical over ungrammatical uses of each verb. Pre-emption explains these results by positing that sentences such as **Homer screeched Lisa the instructions* are blocked only by use of the relevant verb in constructions with similar functional purpose such as *Homer screeched the instructions to Lisa*. Entrenchment posits such overgeneralization errors would also be blocked by use of the verb in sentences that do not serve a similar functional purpose, such as the simple transitive *Homer screeched the instructions*. However, the correlation between pre-emption and entrenchment measures in Ambridge et al.'s study was too high to distinguish between the two (r=0.9, p<0.001). Similar results have also been found in studies of the locative alternation, where children (age 5-6, 9-10) and adults rated locative overgeneralization errors as least acceptable with high-frequency verbs, more acceptable with low-frequency verbs, and most acceptable with novel verbs (e.g., **Lisa poured/dripped/naced the cup with water*) (Bidgood, Ambridge, Pine & Rowland, 2014). Note that the locative alternation provides a particularly stringent test of statistical learning because locative verbs are much lower in frequency compared to transitive verbs (and to some extent, dative verbs), and thus provide less opportunity for statistical-learning to occur.

1.4.4. A Role for Verb Semantics and Verb Frequency

1.4.4.1. Motivation for an integrated learning mechanism

Section 1.4.2 and *Section 1.4.3* outlined that children and adults use semantic and statistical restriction mechanisms in the retreat from overgeneralization, and this holds across three argument-structure alternations. However, it is clear that neither of these mechanisms can explain the retreat from error independently. Statistical-learning accounts cannot explain semantic effects because numerous studies have shown that semantic effects hold for novel (zero-frequency) verbs (e.g., Ambridge et al., 2008, Brooks & Tomasello, 1998; Theakston, 2004). At the same time, Pinker's semantic verb-class account cannot explain why overgeneralizations containing high-frequency verbs are rated as less acceptable than overgeneralizations containing low-frequency verbs from the same discrete semantic-class.

Such frequency effects cannot be explained away as 'semantic-effects in disguise,' whereby high-frequency verbs have better-learned semantics than low-frequency verbs, and are thus less acceptable in infelicitous constructions. The reason for this is that Wonnacott and colleagues (2008; 2011; 2012) have demonstrated that children (aged 4-7) and adults generalize to new constructions using statistical properties of the input even when semantics is controlled for using an artificial grammar learning (AGL) paradigm. AGL experiments require participants to learn a miniature language by exposing them to novel structures paired with a particular meaning. For example, Wonnacott, Newport and Tanenhaus (2008) exposed adult participants to two novel structures: VN_1N_2 and VN_2N_1, each mapped to identical semantics where N_1 was the agent of an action which affected the entity denoted by N_2 (note this design ensured that each structure was matched for semantics and thus generalization to one structure over the other could not be attributed to semantic properties associated with a structure). In Experiment 1, where the majority of verbs in the training phase did not alternate between the two novel structures (i.e., VN_1N_2 and VN_2N_1), participants were least likely to generalize verbs to the alternative construction when the verb was high frequency as opposed to low frequency (in training, high frequency verbs were presented 3 times as often as low frequency verbs). The results are taken as evidence that speakers use token frequency of a verb in a particular structure (i.e., entrenchment or pre-emption) to restrict generalization to a new structure. The findings have since been extended to six-year-old children who were exposed to a similar yet more simple AGL (Wonnacott, 2011) and thus, the restriction of linguistic generalizations cannot be explained by an account with no role for statistical-learning.

What is needed is a hybrid-account that includes a role for pre-emption/entrenchment and verb-semantics, and which specifies how these factors interact.

1.4.4.2. The *FIT* Account

One promising proposal is the *FIT* account (Ambridge & Lieven, 2011), which integrates a number of previous proposals in order to capture effects of pre-emption, entrenchment and verb-semantics. *FIT* is a constructivist-based account such that it assumes children learn constructions gradually through induction, and that each slot of a construction probabilistically exhibits any property (usually phonological, semantic or pragmatic) shared by items that have appeared in that position (see *Section 1.3* for a more detailed outline of construction-learning). For example, the intransitive-inchoative construction can be represented by the schema in (1) and the transitive-causative construction in (2):

(1) [V] (PATIENT/ACTOR) [W] (ACTION)

(2) [X] (AGENT) [Y] (ACTION) [Z] (PATIENT)

The slots are labelled with different letters rather than syntactic or semantic categories in order to capture the assumption that each slot exhibits its own probabilistic constellation of semantic properties (though, for ease of reference, prototypical semantic labels are given in parentheses). For example, [W] and [Y] can be thought of as “VERB” slots for the intransitive and transitive constructions respectively, but differ because [W] exhibits roughly ‘internally-caused action’ semantic properties whereas [Y] exhibits ‘direct, externally-caused action’ semantic properties.

The central assumption of *FIT* is that overgeneralization errors are a consequence of an item being used in a construction-slot in which it is less than optimally compatible (for example, when *laugh* is used in the transitive [X] [Y] [Z] schema, e.g., *[*The funny man*] [*laughed*] [*Bart*]). Crucially, the mechanism that gives rise to these errors is also the mechanism that eventually restricts these errors. The mechanism is termed ‘competition’ and is basically the notion that all constructions in a speaker’s lexicon compete to express the speaker’s intended message (e.g., MacWhinney, 2004). Early in development, an unsuitable (for adults) construction may outcompete others because the schema is more strongly represented (e.g., it may have higher token-frequency) and/ or the child has not yet developed knowledge of the properties associated with relevant schema slots or the verbs themselves. With linguistic experience, children fine-tune the properties exhibited by schema-slots to reflect semantic (or phonological/ pragmatic) restrictions on particular verbs. In turn, more appropriate constructions outcompete less appropriate constructions and errors cease.

The account yields entrenchment effects because the child develops probabilistic links between a verb and *all* the various constructions in which it appears (thus, errors are lower with highly-frequent verbs). For example, *laugh* is posited to have stronger links with the intransitive-inchoative and periphrastic-causative than do lower frequency verbs like *giggle* and *chuckle*, and thus *laugh* is more likely to be used in those constructions (and consequently less likely to be overgeneralized to the transitive-causative). Pre-emption occurs because constructions that a verb has been probabilistically-linked with receive more activation when they are *relevant* to the message (e.g., when one wishes to refer to the causer of a laughing action, the periphrastic-causative receives an activation-boost at the expense of the intransitive-inchoative). The account yields semantic effects because constructions receive more activation when the semantic properties exhibited by its 'verb' slot overlap with the semantic properties of the verb. Apparent 'verb-class' effects arise because verbs that have similar semantic properties are semantically compatible with a particular slot to a similar degree.

It is important to recognise why the constructivist framework assumed by *FIT* is potentially advantageous for explaining how children form and appropriately restrict linguistic generalizations. First, it assumes that semantics-to-syntax correspondences are learned rather than innate and thus does not suffer limitations posed by cross-linguistically different semantics-to-syntax mappings (see *Section 1.2.2.4* for summary). On a related note, the account dispenses with the need for an innately constrained broad-range rule, instead positing that generalization is determined by the probabilistic semantic 'fit' between construction slots and lexical-items (in most of the cases discussed above, verbs). This is important when one considers that causative overgeneralizations of non-dynamic verbs constituted around 30% of the errors in Bowerman's (1988) study (e.g., **The cold stayed them awake*) and thus, that any successful theory must explain how these errors are generated. The *FIT* account's assumption of probabilistic construction-slots, and the notion that fine-grained semantic properties associated with these slots develop gradually with time can potentially explain young children's tendency to causativize non-dynamic verbs like *stay* and *wait* (i.e., these children will not have an adult-like understanding of the construction-slot and are thus more likely to overgeneralize verbs into that construction). In contrast, Pinker's (1989) account has difficulty explaining the causativization of non-dynamic verbs because the semantics of these verbs are incompatible with the broad-range rule posited to generalize verbs into the transitive-causative (see *Section 1.4*.2.1).

As discussed in *Section 1.4.2,* constructivist frameworks (as adopted by *FIT*) posit that a verb is lent part of its meaning by the construction (e.g., Goldberg, 1995) as opposed to having its own meaning transformed by a lexico-semantic rule, and thus there is no need to posit implausible verb senses under the *FIT* account. Finally, the account allows for probabilistic statistical and verb-semantic effects to occur within one learning mechanism, and is thus consistent with previous research that has demonstrated children as young as 5-6 to use both of these factors to restrict overgeneralization (e.g., Ambridge et al., 2008).

1.4.4.3. Testing Predictions of the *FIT* Account

An important prediction of the account is that semantic effects should be graded and probabilistic such that the greater the extent to which a verb's semantic properties are incompatible with the semantic properties of the construction, the less likely it is to be produced, or rated as acceptable in that construction. At the same time, these semantic effects should be modulated by pre-emption and entrenchment effects.

Ambridge, Pine and Rowland (2012) investigated the predictions of the *FIT* account by collecting grammaticality judgements for the use of 142 locative verbs (60 for children) in content-locative and container-locative sentences. The verb-set varied in frequency, (corpus-based) preference for content-locative or container-locative constructions, and semantic 'fit' with content-locative or container-locative constructions. Semantic-fit was measured by obtaining ratings from a separate group of adult participants for the extent to which each test verb denoted semantic properties thought to be relevant to the locative constructions. Children's (age 5-6, 9-10) and adults' preferences for container-locative over content-locative uses of a verb were significantly correlated with the semantic-fit measure (for example, preference for container- over content- locative uses of a verb were positively related to the extent to which it denoted the 'end-state' of an action). Furthermore, acceptability ratings were significantly predicted by both the entrenchment measure (overall frequency regardless of the construction) and the pre-emption measure (frequency of a verb in content- or container- locative constructions only).

Similar evidence for the *FIT* account – in the form of probabilistic semantic-fit effects, entrenchment and pre-emption effects – has been observed in children's (age 5-6; 9-10) and adults' acceptability judgments of *DO*- dative and *PO*- dative sentences (for which [VERB] slots are probabilistically associated with meanings of possession-transfer and caused-motion respectively) (Ambridge, Pine, Rowland, Freudenthal & Chang, 2014).

The same study also tested the prediction that acceptability judgments would be related to the extent to which verbs denote morphophonological properties associated with [VERB] slots of *DO-* datives. The hypothesis comes from the observation that verbs of Latinate origin such as *donate* and *contribute*, despite being semantically suitable for *PO-* and *DO-* datives, seem to be restricted to appearing in *PO-* datives (e.g., *John donated a book to the library/ *John donated the library a book*). Since Latinate verbs can be identified by morphological properties (e.g., *-ute*, *-ate*) and phonological properties (e.g., disyllabic with second-syllable stress/ trisyllabic), it was posited that the *PO-* dative is associated with these morphophonological properties. The *DO-* dative is associated only with verbs that are monosyllabic or disyllabic with first-syllable-stress. Adults' preferences for *PO-* over *DO-* dative sentences were positively related to the extent to which verbs exhibited morphophonological properties hypothesised to be associated with [VERB] slots of *PO-* datives. The finding is consistent with the assumptions of the *FIT* account, specifically that [VERB] slots of argument-structure constructions reflect a learner's previous experience with the slot such that it is probabilistically associated with a complex set of (semantic and morphophonological) properties shared by verbs that have previously occupied that slot. For example, if the *PO-* dative construction is conceptualised as roughly [A] [B] [C] to [D], the verbal [B] slot probabilistically exhibits 'caused-motion' semantic properties *and* 'disyllabic with second-syllable stress/ trisyllabic' morphophonolgical properties. The fact that morphophonological constraints on generalizations of dative verbs has been demonstrated to be psychologically-real demonstrates that the *FIT* account naturally extends to constructions associated with properties other than semantic properties. The finding is particularly relevant to the present thesis because *Study 2* will investigate the English regular past-tense construction, which is posited to exhibit probabilistic phonological properties (see *Sections 1.4.4.4.* and *2.5*)

The *FIT* account has recently been instantiated as a connectionist model that simulates the formation and restriction of dative overgeneralization errors (Ambridge & Blything, 2015). The model used seven input units, each representing a semantic feature relevant to *DO-* or *PO-* constructions (e.g., causing to go, causing to have, mailing, etc.). Dative verbs (*PO-* only or alternating) were represented as a vector across the seven units, and each unit's activation was determined by using adult ratings of the extent to which each verb denoted these semantic properties (taken from Ambridge et al., 2014). Each presentation of a verb activated one of three output units that corresponded to either the *PO-* dative, *DO-* dative or 'any-other' construction, with activation determined by the verb's log frequency in the relevant construction according to the British National Corpus.

Pre-emption was implemented by an eighth input node (the 'message' node) which was activated only for presentations of a verb that activated the *PO-* or *DO-* output node (not the 'any-other' node), thus simulating a context in which the message (i.e., transfer) could only be formulated using a dative construction. The model was able to (i) simulate the pattern of overgeneralization then retreat, (ii) predict the argument-structure preferences of novel verbs based on *PO-* only or 'alternating' semantic properties, and (iii) predict the pattern of adults' by- verb grammaticality judgments (observed in Ambridge et al., 2014).

1.4.4.4. Morphological Overgeneralization Errors

Children must also be able to form and appropriately restrict generalizations at the morphological level such as with verbal *un-* prefixation (e.g., *twist→untwist; squeeze→*unsqueeze*) and English past-tense (e.g., *kick→kicked; drive→*drived*). So far, evidence for the role of pre-emption, entrenchment and verb-semantics in the retreat from overgeneralization has been discussed only with regard to verb-argument structure (i.e., transitive, dative and locative alternations) and for good reason, since research into the interactive role of these factors at the morphological-level of overgeneralization is relatively under-studied. However, it is important for any theory of retreat from overgeneralization to account for all types of these errors, at the syntactic and morphological level. Indeed, a major aim of the present thesis is to test the assumptions of *FIT* across different morphological and syntactic domains (see *Section 1.5*), and it is thus important to outline the main morphological domains of interest: English past-tense overgeneralizations, and verbal *un-* prefixation overgeneralizations.

English Past-tense Overgeneralization Errors

The English past-tense system is characterised by its distinction between regular (e.g., *kick/kicked; play/played*) and irregular (*sleep/slept; throw/threw*) patterns of inflection. Restriction of past-tense overgeneralization errors, where regular inflections are applied to verbs that require irregular inflection (e.g., *sleep/*sleeped; throw/*throwed*) is most naturally explained by pre-emption, since there usually exists a perfect synonym for the ungrammatical formulation (e.g., **sleeped* is pre-empted by *slept*, and the more one hears *slept* in contexts where they might expect **sleeped*, the more this constitutes probabilistic evidence that **sleeped* is ungrammatical). Thus, most of the research into children's past-tense overgeneralization errors has focused on the mechanism responsible for these overgeneralizations in the first place. On one hand, the *dual-route* model (e.g., Prasada & Pinker, 1993) posits that overgeneralization errors are best attributed to an innate context-free rule (i.e., add "*-ed*" to any instance of the category "verb"). On the other hand, the *single-route* model (e.g., Bybee & Moder, 1983) posits that overgeneralization reflects analogy across stored exemplars (e.g., *steal/*stealed* from *peel/peeled, heal/healed*). Consistent with the predictions of the single-route model only, children's (aged 9-10) acceptability judgments of novel regular past-tense forms (e.g., *queeded, nolded, chaked*) have been demonstrated to be positively related to a verb's similarity to existing regular verbs (Ambridge, 2010). Thus, it appears that children's morphological generalizations need not be attributed to a rule-based mechanism and are best explained by analogy across memorized exemplars. Consistent with the *FIT* account, this finding can be explained by positing that children have an inventory of morphological constructions, including 'regular' and 'irregular' past-tense schemas, where a [VERB] slot exhibits phonological properties of verbs that have previously been used in that slot. To be more precise, one can posit three 'regular' schemas, [W]-*d*, [X]-*t*, and [Y]-*əd,* to account for the fact that the regular past-tense morpheme may be realised as *–d*, (e.g., *played*), *-t* (e.g., *wish/wished*) or *-əd* (e.g., *hunt/hunted*). 'Irregular' schemas such as [Z]+*vowel-change* are also posited to account for particular irregular inflections such as *throw/threw* and *blow/blew*. Overgeneralization errors are assumed to occur when a child has not yet developed adult-like knowledge of the 'VERB' slot (and has not yet learned an appropriate pre-empting form), and consequently uses a phonologically inappropriate verb in that slot. 'Regular' overgeneralizations (e.g., *steal→*stealed*) are much more common than 'irregular' overgeneralizations (e.g., *heal→*hole*) because regular schemas have higher overall frequency (possibly meaning they have a higher resting activation level) and/ or the verb slots of regular schemas are more heterogeneous (i.e., 'open' to a broader range of phonological properties) than those of irregular schemas,

The domain of English past-tense is a particularly good test-case for the predictions of *FIT* because regular verbs do not share one definitive set of phonological properties, but rather cluster into phonological neighbourhoods. For example, the [Y] slot in the [Y]-*əd* schema exhibits phonological properties shared by a neighbourhood of verbs like *hunt/hunted, punt/punted* and *grunt/grunted*, as well as phonological properties shared by a neighbourhood of verbs like *bust/busted, trust/trusted* and *rust/rusted,* and a neighbourhood of verbs like *need/needed* and *plead/pleaded*. Thus, the *FIT* account assumes that there is no one phonological property that is necessary or sufficient for a verb to be used in regular form. Rather, the 'VERB' slot of each schema exhibits a fuzzy and probabilistic set of phonological properties denoted by verbs that have been used in the slot before. The extent to which a verb's own phonological properties overlap with the properties associated with the slot of a regular schema predicts the likelihood of it being used or rated as grammatical in regular past-tense form.

Verbal Un- Prefixation Overgeneralization Errors

Verbal *un-* prefixation overgeneralization errors involve the application of *un-* prefixation (e.g., *button* → *unbutton*) to incompatible verbs (e.g., **unbend; *uncome*; Bowerman, 1982). As with English past-tense, it is likely that pre-emption plays a large role in the retreat from *un-* prefixation errors (e.g., *deflate* is a direct synonym for **unblow*). However, not all *un-* forms have a perfect synonym (e.g., **unsqueeze* is not perfectly synonymous with formulations like *loosen*, *release* or *let go*) and it is thus likely that another mechanism is at least partly responsible for restriction of these errors. One compelling argument was made by Whorf (1956) who noted that verbs which are grammatical with *un-* (e.g., *unbuckle, untwist, uncover, unwrap, unleash*) do not seem to share any obvious semantic property but rather, denote a semantic "cryptotype" of meanings - where "cryptotype" refers to a covert category of overlapping, probabilistic meanings that are difficult to delineate (e.g., *covering, attaching, circular motion, change-of-state* and *enclosing*). Whorf's hypothesis is consistent with the *FIT* account's assumption that each slot of a construction (in this case, a morphological *un*[VERB] construction) exhibits a fuzzy and probabilistic cluster of properties commonly exhibited by lexical-items that have been experienced in that slot (conversely, Whorf's hypothesis would be difficult to explain using Pinker-style discrete verb-classes). Thus, the domain of verbal *un-* prefixation provides a particularly suitable test-case for the role of statistical-learning and verb-semantics (particularly *probabilistic* verb-semantics) in the retreat from overgeneralization, and both of these accounts have received empirical support (e.g., Ambridge, 2013; Clark, Carpenter & Deutsch; 1995; Ibbotson, 2013). Indeed, the domain of verbal *un-* prefixation is the focus of the first study in this thesis, and is discussed in more detail in *Section 1.5.2*.

1.5. The Thesis

1.5.1. Objectives of Thesis

The review of literature has outlined why any theory of the formation and restriction of linguistic generalizations must include a role for the statistical properties of the verb itself (i.e., *entrenchment*; Braine & Brooks, 1995), the statistical properties of competing formulations that convey the intended message (i.e., *pre-emption*; Clark &

Clark, 1979), and the compatibility between the verb's semantic or phonological properties and those associated with the construction in which it appears (e.g., '*fit'*; Ambridge & Lieven, 2011).

However, the majority of studies supporting this view have used a grammaticality-judgment paradigm which is limited to testing children aged 5-6 and above (Ambridge, 2011). The problem here is that children aged 5-6 are well-past the early stages of overgeneralization. An examination of corpus studies (e.g., Bowerman, 1988; Marcus, Pinker, Ullman, Hollander, Rosen, & Xu, 1992), demonstrates that dative overgeneralizations (**I said her no*; 3;1), locative overgeneralizations (**I'm gonna cover a screen over me*; 4;5), causative overgeneralizations (**Did she bleed it?*; 3;6), past-tense overgeneralizations (**singed*; 2;0) and verbal *un-* prefixation overgeneralizations (e.g., **uncapture me!*; 3;10) are evident in children's speech well before 5;0 years (age at time of utterance is shown next to each error). Thus, it is critical to examine the restriction mechanisms used by children aged 3 to 5, who are in the process of acquiring their generalization mechanisms, as opposed to older children, who are in use of a relatively mature system. With this in mind, the chief aims of the current thesis are to develop experimental paradigms that are more suitable for testing young children, and to examine what these paradigms tell us about the retreat from overgeneralization error. A parallel aim is to extend this investigation to the domain of derivational morphology (verbal *un-* prefixation) and inflectional morphology (English past-tense). The reason for this is that most previous evidence for the roles of statistical-learning and a verb's semantic/phonological properties in the restriction of overgeneralization errors has come from the domain of verb-argument structure (e.g., Ambridge et al., 2008, 2012, 2014; Bidgood, Ambridge, Pine & Rowland, 2014; Brooks, Tomasello, Dodson & Lewis, 1999; Brooks & Tomasello, 1998; Gropen et al., 1989; Robenalt & Goldberg, 2015; Theakston, 2004; Wonnacott, Newport & Tanenhaus, 2008), despite the fact that errors involving derivational and inflectional morphology provide compelling test-cases for these predictions. Indeed, given (a) the probabilistic cryptotype nature of the phonological/semantic generalizations, and (b) the frequent availability of perfectly pre-empting alternatives (e.g., *slept* for **sleeped*; *stand* for **unsit*) it could be argued that these domains constitute particularly useful test cases.

The first two studies address these aims by using a production paradigm to examine the restriction mechanisms used by children as young as 3-4 years to restrict productivity with derivational morphology (verbal *un-* prefixation; *Study 1*) and inflectional morphology (English past-tense; *Study 2*). *Study 3* investigates the suitability of an online measure of sentence processing, namely Event Related Potentials (ERP), to investigate the mechanisms used to restrict transitive-causative overgeneralizations. An outline of the rationale for *Study 1, Study 2,* and *Study 3* is provided in *Section 1.5.2, Section 2.6,* and *Section 3.6* respectively.

1.5.2. Prelude to *Study 1*

Verbal *un-* prefixation can be applied to many verbs to indicate a reversal action (e.g., *twist/ untwist; cover/uncover; leash/unleash*), but some verbs cannot be prefixed with *un-* even though they denote actions that can be reversed (e.g., **unbend, *unopen, *unsqueeze*). *Study 1* examines the roles of statistical-learning (both entrenchment and pre-emption) and verb-semantics in young children's restriction of such verbal *un-* prefixation overgeneralization errors. The domain of *un-* prefixation provides a compelling test of the role of verb-semantics, but has attracted less interest than verb argument-structure. Verbs that may and may not be used with *un-* prefixation do not appear to form discrete Pinker-style semantic classes. Rather, verbs which license *un-* cluster into a fuzzy "semantic cryptotype" of shared meanings (e.g., *covering, enclosing, attaching, circular motion, change of state, binding/locking*; e.g., Whorf, 1956). "Cryptotype" is a term used by Whorf to refer to a covert category of overlapping, probabilistic meanings that are difficult to delineate. No individual feature is necessary or sufficient to license *un-* prefixation; rather, the summed expression of these features reflects each verb's compatibility with the prefix.

The most relevant study that has investigated the role of verb-semantics in the restriction of *un-* prefixation is the grammaticality judgment study by Ambridge (2013) (see also Clark, Carpenter & Deutsch, 1995; Ibbotson, 2013). Children (age 5-6; 9-10) and adults rated the grammaticality of 48 *un-* prefixed verb forms on a 5-point scale; half grammatical (e.g., *unbutton; unlock*), half ungrammatical (e.g., **unfill; *ungive*). Each of 48 test verbs were rated for the extent to which they denoted 20 semantic features hypothesised by Li and MacWhinney (1996) to represent the semantic cryptotype (e.g., whether one object affects another, distorts it, contains it, etc.). For all age-groups (aged 5-6, 9-10; adults), a positive correlation was observed between the extent to which a verb was compatible with the semantic cryptotype and its rated grammaticality in *un-* form. An effect of pre-emption (measured by the frequency of alternative forms; e.g., *open* for **unclose*) and entrenchment (measured by the frequency of the bare form; e.g., *close*) were observed only for children aged 9-10.

Ambridge's (2013) study has important implications for theories of retreat from overgeneralization. The finding that all age-groups were sensitive to Whorf's hypothesised 'semantic-cryptotype' demonstrates that it is unrealistic to posit Pinker-like discrete verb classes and instead provides evidence for a fuzzy and probabilistic semantic constraint consistent with the notion of semantic 'fit' (e.g., Ambridge & Lieven, 2011; *Section 1.4.4.2*). The finding that effects of pre-emption and entrenchment were observed only in older children may demonstrate that a role for statistical-learning in restriction of overgeneralization errors emerges at a later stage than a role for verb-semantics, but may also suggest that the judgment paradigm used was too noisy or too difficult for detecting the full-range of restriction mechanisms used by the younger age-group.

A significant limitation of Ambridge's (2013) study, and the overwhelming majority of previous studies that have investigated the roles of statistical-learning and verb-semantics, is that the use of a judgment paradigm is limited to testing children aged 5-6 and above, and thus neglects younger children who are at a more crucial stage of learning semantic or statistical constraints. Indeed, children as young as 3;2 have been found to overgeneralize the application of *un-* prefixation to incompatible verbs (e.g., **unbend; *uncome*; Bowerman, 1988). Thus, it is important to examine (i) whether younger children's productivity is restricted by pre-emption, entrenchment and semantic-fit and (ii) whether these mechanisms can be extended to the domain of morphological verbal *un-* prefixation. To address these aims, *Study 1* uses a production-priming and grammaticality-judgment paradigm to examine the mechanisms employed by children aged 3-4 and 5-6 in their restriction of overgeneralization errors involving verbal *un-* prefixation.

2. Children use statistics and semantics in the retreat from overgeneralization (*Study 1*)

The manuscript has been published: Blything, R.P., Ambridge, B., Lieven, E.V.M. Children Use Statistics and Semantics in the Retreat from Overgeneralization. PLoS ONE. 2014: 9(10): e110009. doi: 10.1371/journal.pone.0110009

The presentation and reference style used for the manuscript has been altered in the interest of consistency with other chapters of this thesis.

2.1. Abstract

How do children learn to restrict their productivity and avoid ungrammatical utterances? The present study addresses this question by examining why some verbs are used with *un-* prefixation (e.g., *unwrap*) and others are not (e.g., **unsqueeze*). Experiment 1 used a priming methodology to examine children's (3-4; 5-6) grammatical restrictions on verbal *un-* prefixation. To elicit production of *un-*prefixed verbs, test trials were preceded by a prime sentence, which described reversal actions with grammatical *un-* prefixed verbs (e.g., *Marge folded her arms and then she unfolded them*). Children then completed target sentences by describing cartoon reversal actions corresponding to (potentially) *un-* prefixed verbs. The younger age-group's production probability of verbs in *un-* form was negatively related to the frequency of the target verb in bare form (e.g., *squeez/e/ed/es/ing*), while the production probability of verbs in *un-* form for both age groups was negatively predicted by the frequency of synonyms to a verb's *un-* form (e.g., *release/ *unsqueeze*). In Experiment 2, the same children rated the grammaticality of all verbs in *un-* form. The older age-group's grammaticality judgments were (a) positively predicted by the extent to which each verb was semantically consistent with a semantic "cryptotype" of meanings - where "cryptotype" refers to a covert category of overlapping, probabilistic meanings that are difficult to access - hypothesised to be shared by verbs which take *un-*, and (b) negatively predicted by the frequency of synonyms to a verb's *un-* form. Taken together, these experiments demonstrate that children as young as 4;0 employ pre-emption and entrenchment to restrict generalizations, and that use of a semantic cryptotype to guide judgments of overgeneralizations is also evident by age 6;0. Thus, even early developmental accounts of children's restriction of productivity must encompass a mechanism in which a verb's semantic and statistical properties interact.

2.2. Introduction

An essential component of language acquisition is a speaker's ability to move beyond the linguistic input and use words in novel ways. For example, when verbs are observed in both the intransitive and transitive construction (e.g., *The ball* bounced → *The man* bounced *the ball*), a speaker may form an abstract linguistic generalization (e.g., [NOUN PHRASE1] [VERB] → [NOUN PHRASE 2] [VERB] [NOUN PHRASE1]) that allows other verbs to be used this way even if they are unattested in that form (e.g., *The stick* broke → *The man* broke *the stick*). A fully adult-like command of language is achieved only when such generalizations are restricted to verbs that are grammatical in the target construction; failure to do so will yield 'over-generalization' errors (e.g., *The woman* laughed → **The man* laughed *the woman*). The current paper aims to elucidate the mechanisms employed by children to restrict their linguistic generalizations. Specifically, we examine young children's (age 3-4; 5-6) restrictions of verbal *un*- prefixation (e.g., *squeeze→*unsqueeze*); a domain that has been observed to yield overgeneralization errors in both corpus (e.g., **unbend, *uncome, *unhate, *unpress, *uncapture*; Bowerman, 1982) and production studies (e.g., **unstick*, **uncrush*, **unbury*, **unbend*, **unsqueeze*; Clark, Carpenter & Deutsch, 1995), with children as young as three years old.

The retreat from overgeneralization cannot be explained in its entirety by negative-evidence (McNeill, 1966) which holds that these errors cease as a consequence of a caregiver's corrective feedback (e.g., if a child says *The man laughed the woman* then the caregiver may offer a correction such as *The man made the woman laugh*). Specifically, it is not feasible for every possible overgeneralization to be corrected and this position is supported by findings that overgeneralizations containing novel verbs are recognised as ungrammatical by children and adults (e.g., Ambridge, Pine, Rowland & Young, 2008). Rather, a number of recent findings (for review, see Ambridge, Pine, Rowland, Chang & Bidgood, 2013) have suggested that any theory that accounts for children's retreat from overgeneralization errors must include a role for the statistical properties of the verb itself (i.e., entrenchment; Braine & Brooks, 1995), the potential competing formulations that convey the desired message (i.e., pre-emption; Clark & Clark, 1979), and the relationship between the verb's semantic properties and those associated with the construction in which it appears (e.g., Pinker, 1989). However, the majority of studies supporting this view have used a grammaticality-judgment paradigm which is thought to be unsuitable for children younger than 5-6, and even children at this age showing somewhat inconsistent results (Ambridge, 2011; Wonnacott, 2011). Examination of whether mechanisms of pre-emption, entrenchment and verb-construction semantics are also employed by younger children is crucial to our understanding of children's retreat from overgeneralization and thus of language acquisition as a whole. Before discussing this issue, it is necessary to outline the specific factors that each of these mechanisms is assumed to involve.

In pre-emption (Clark & Clark, 1979), the repeated presentation of a verb in a particular construction constitutes ever-strengthening probabilistic evidence that non-attested alternative formulations which express the same intended meaning are ungrammatical. For example, transitive uses of the verb *laugh* (e.g., **The man laughed the woman*) are posited to be blocked by periphrastic causative uses of that verb (e.g., *The man made the woman laugh*) because both formulations convey a similar meaning (i.e., external causation). However, the theory holds that transitive uses of *laugh* are not blocked by intransitive uses (*The woman laughed*) because the intransitive structure conveys a different meaning (internal causation). For example, children as young as 4;7 have been shown to be less likely to produce transitive sentences with novel verbs if those verbs have been modelled in the periphrastic causative construction (Brooks & Tomasello, 1999). Furthermore, evidence for pre-emption has been observed in children's (aged 5-6 and 9-10) and adults' judgments of overgeneralizations involving the dative construction (e.g., **Bart whispered Lisa the secret*; Ambridge, Pine, Rowland, Freudenthal & Chang, 2014).

Conversely, in entrenchment (Braine & Brooks, 1995), the repeated presentation of a verb in *any* context constitutes ever-strengthening probabilistic evidence that non-attested uses of that verb are ungrammatical. For example, transitive uses of the verb *laugh* are posited to be blocked by both periphrastic and intransitive uses of the verb (i.e., *The man made the woman laugh; The woman laughed*), and indeed any other uses (*He laughed it off*; *You're laughing at it; Laughing!* etc.). Evidence for this theory was demonstrated by a study in which children aged 3;4 were less likely to produce transitive causative overgeneralization errors with high frequency verbs (e.g. *come*) than with low frequency verbs (e.g. *arrive*; Brooks, Tomasello, Dodson & Lewis, 1999). Evidence for entrenchment has also been observed in children's (aged 5-6 and 9-10) and adults' judgments of overgeneralizations involving transitive (Ambridge, Pine, Rowland & Young, 2008), dative (Ambridge, Pine, Rowland, Freudenthal & Chang, 2014) and locative constructions (e.g., **Marge splashed the carpet with juice*; Ambridge, Pine, & Rowland, 2012)).

A semantically-focused approach arises from the claim that each construction is associated with particular semantic features. For example, the transitive-causative is associated with *direct external causation* (e.g., *X broke Y*), whereas the intransitive is associated with *internal causation* (e.g., *Y broke*). Pinker's (1989) semantic verb class hypothesis theorised that each verb in a speaker's lexicon is assigned to a 'narrow-range' semantic class, with particular classes semantically consistent with – and hence grammatical with – particular sets of constructions. For example, verbs like *ascend* and *rise* belong to a *motion in a particular direction* class that is semantically consistent with the semantics of the intransitive construction but not the transitive construction (*ascending* and *rising* can be internally caused but not directly externally caused). Conversely, verbs like *swing* and *bounce* belong to a *manner of motion* class that is semantically consistent with the semantics of both the intransitive and transitive constructions (these verbs having elements of both internal and external causation), and can thus freely alternate between the two constructions. Evidence for this proposal was demonstrated in a study which found that children as young as 4;7 were more likely to produce transitive causative sentences with novel verbs consistent with a *manner of motion class* as opposed to a *motion in a particular direction* class (Brooks & Tomasello, 1999).

In its original form, Pinker's (1989) discrete class-based proposal (either a verb is a member of a compatible semantic class, or it is not) does not naturally explain the finding that grammatical acceptability appears to be a probabilistic, graded phenomenon, whereby grammaticality depends on the *extent* to which a verb's semantics are consistent with those of the target construction. For example, the greater the extent to which a verb has semantic properties associated with the transitive, locative, and dative constructions, the greater the extent to which it is felicitous in those constructions, as rated by children (aged 5-6 and 9-10) and adults (e.g., Ambridge, Pine, Rowland, Freudenthal & Chang, 2014; Ambridge, Pine, & Rowland, 2012; Ambridge, Pine, & Rowland, 2011)). Thus, previous literature regarding verb-argument structure overgeneralization errors points to a role for pre-emption, entrenchment and probabilistic verb-and-construction semantics.

However, the problem of retreat from overgeneralization applies not just to syntax (i.e. verb-argument structure), but to morphology as well. Additionally, a truly developmental understanding of the retreat from error can only be achieved by investigating children of all ages – including those younger than 5;0 who have been neglected by the type of judgment studies outlined above. To illustrate these points, children as young as 3;2 have been found to overgeneralize the application of *un-* prefixation to incompatible verbs (e.g., **unbend; *uncome*; Bowerman, 1982) and it is therefore important to examine (i) whether younger children's productivity is restricted by pre-emption, entrenchment and verb-and-construction semantics, and (ii) whether these mechanisms can be extended to the domain of morphological verbal *un-* prefixation (note that the only studies to our knowledge that have investigated the role of pre-emption, entrenchment, or verb-and-construction semantics in children less than 5 years old (Brooks

& Tomasello, 1999; Brooks, Tomasello, Dodson & Lewis, 1999), have focused on the transitive alternation).

Ambridge (2013) investigated whether children's (aged 5-6; 9-10) and adults' restrictions on *un-* prefixation could be explained by the mechanisms outlined above. For pre-emption to apply to the domain of *un-* prefixation, it is necessary for ungrammatical *un-* forms (e.g., **unsqueeze*) to be pre-empted by near synonyms (e.g., *release, loosen*). Thus the hypothesis predicts that errors will be less common for verbs with frequently occurring (near) synonyms to their *un-* form. In contrast, the entrenchment hypothesis holds that such errors will be less common for verbs that occur frequently without the *un-* prefix. Ambridge offered evidence that both mechanisms can be extended to the domain of verbal *un-* prefixation. Participants rated the grammaticality of 48 *un-* prefixed verb forms on a 5-point scale; half grammatical (e.g., *unbutton; unlock*), half ungrammatical (e.g., **unfill; *ungive*). Frequency counts of (a) verbs in bare form (e.g., *squeez-e-es-ed-ing*) and (b) synonyms of their *un-* form (e.g., *release* and *loosen* for **unsqueeze*) were obtained to examine the entrenchment and pre-emption accounts respectively. The findings for 9-10 year olds supported these hypotheses, with both frequency counts negatively predicting the rated acceptability of ungrammatical *un-* forms. However, neither entrenchment nor pre-emption were supported for the youngest children (aged 5-6). Thus, Ambridge demonstrated a successful extension of entrenchment and pre-emption to verbal *un-* prefixation, but only for later stages of development. One possibility is that sufficient entrenchment/ pre-emption had not yet occurred; another is that these younger children simply struggled with the judgment task. The present study picks apart these possibilities by running a judgment task and a production task designed to be less-demanding for this age group.

How can the semantic approach be applied to verbal *un-* prefixation? Verbs that do and do not take the prefix do not appear to form discrete Pinker-style semantic classes. Rather, verbs which license *un-* cluster into a fuzzy "semantic cryptotype" of shared meanings (e.g., *covering, enclosing, attaching, circular motion, change of state, binding/locking*; (Li & Macwhinney, 1996; Whorf, 1956). "Cryptotype" is a term used by Whorf to refer to a covert category of overlapping, probabilistic meanings that are difficult to access relative to overt prototypical grammatical categories (e.g., for the transitive construction). No individual feature is necessary or sufficient to license *un-* prefixation; rather, the summed expression of these features reflects each verb's compatibility with the prefix. To underline this point, Whorf noted that "we have no single word in the language that can give us a clue to its meaning;...hence the meaning is subtle, intangible, as is typical of cryptotypic meanings."

Ambridge's (2013) grammaticality judgment study of verbal *un-* prefixation examined the psychological reality of Whorf's (1956) semantic crytpotype. Each of 48 test verbs were rated for the extent to which they denoted 20 semantic features hypothesised by Li and MacWhinney (1996) to represent the semantic cryptotype. For all age-groups (aged 5-6, 9-10; adults), a positive correlation was observed between the extent to which a verb was compatible with the semantic cryptotype and its rated grammaticality in *un-* form, constituting clear evidence for the graded probabilistic use of verb semantics by children as young as 5-6.

To summarise, recent findings suggest a role for pre-emption, entrenchment and probabilistic verb-and-construction semantics in the retreat from overgeneralization, at least for children aged 5-6 and older. However, this research has mainly been limited to judgment studies, which themselves may be inappropriate for children younger than 5 years. Furthermore, judgment studies have yielded mixed findings for 5-6 year olds, with this age-group demonstrating effects of statistical learning (i.e., pre-emption and/ or entrenchment) in judgments of transitive (Ambridge et al., 2008), dative (Ambridge, Pine, Rowland, Freudenthal & Chang, 2014) and locative constructions (Ambridge, Pine, & Rowland, 2012) but not verbal *un-* prefixation (Ambridge, 2013). Although it is possible that children were too young for the relevant lexical items to have undergone sufficient entrenchment/pre-emption, an alternative possibility is that, for these younger children, the judgment paradigm was too demanding, insensitive or noisy to detect statistical learning effects. In the present study, we investigate the possibility that a potentially-easier experimental task - elicited production - may be more likely to detect the full range of restriction mechanisms employed by younger children. This was achieved by having the same children (aged 3-4 and 5-6) complete both a Production (Experiment 1) and Judgment study (Experiment 2).

2.3. Experiment 1: Production Study

2.3.1. Participants

Participants were 20 children aged 3;6-4;7 (*M*=4;0) and 20 children aged 5;6 to 6;6 (*M*=6;0). An additional four children from the youngest age group were recruited but excluded because they did not comply with the procedure. All participants were monolingual and did not possess any known language impairment. The children were recruited from nurseries and schools in Manchester and were tested at those locations in a separate room.

2.3.2. Design

Participants were divided into one of four counterbalanced groups which differed according to which verb-set was used in target sentences (*verb-set* "*A*" or "*B*"; see *Procedure and Materials*) and whether the production task (Experiment 1) preceded or followed the judgment task (Experiment 2). The dependent variable was whether or not the child produced the target verb in *un-* form on each trial. We used the same independent variables as (Ambridge, 2013) so that a fair comparison could be made with that study. The first three independent variables were employed as control measures to ensure that any effect of pre-emption, entrenchment or verb-and-construction semantics (we will henceforth use the term "semantic-cryptotype" when referring this concept in the domain of *un-* prefixation) could not be attributed to one of these measures.

- ***Corpus presence of un-form* (*Verb-type*).** Each test verb's existence/ non-existence in *un-* form within the British National Corpus [spoken and written]; BNC) was recorded to control for the possibility that verbs which are attested in *un-* form are more likely to be produced in *un-* form. The BNC was used to obtain all frequency counts in the current study because corpora of children's speech (such as CHILDES (Macwhinney, 2000) registered many acceptable *un-* forms as having zero-frequency despite being perfectly acceptable in *un-* form.
- ***Corpus frequency of un- form (log transformed).*** Each verb's frequency in *un-* form within the BNC was recorded in order to control for the likelihood that verbs that have been frequently encountered in *un-* form are more likely to be produced in this form.
- ***Reversibility Measure* (*log transformed*).** In order to control for the possibility that acceptability in *un-* form is simply a proxy for the reversibility of the action denoted, Ambridge (2013) had 15 adult participants rate the extent to which each test verb (presented in bare form only) was reversible using a 7-point scale (for details see Ambridge, 2013).
- ***Pre-emption measure (log transformed).*** This was the summed frequencies of the two most commonly-suggested synonyms for each verb's *un-* form (e.g., *empty* and *drain* for **unfill*) in the BNC. Ambridge (2013) asked 15 adults to suggest potential synonyms (other than *un-* forms) for the reversal of a verb's bare form.

- ***Entrenchment measure (log transformed).*** This was simply the frequency of each verb's bare (i.e., NOT *un-* prefixed) form (all inflected forms; e.g., *fill/fills/filled/filling*) in the BNC (all texts).
- ***Semantic-cryptotype measure.*** This was a composite measure (created using Principal Components Analysis; PCA) of the extent to which each verb was rated (by a separate group of adults) as instantiating each of 20 semantic features proposed by Li and MacWhinney (1996) to collectively characterise the semantics of verbs that may be grammatically prefixed with *un-*, based mostly on Whorf's (1956) cryptotype (see Ambridge, 2013)). The 20 semantic features were as follows (note that as a consequence of PCA, only 9 features comprised the final semantic cryptotype measure – all identified in bold font): (1) **Mental Activity**, (2) Manipulative Action, (3) **Circular Movement**, (4) **Change of location**, (5) Change of state, (6) Resultative, (7) A affects B, (8) **A touches B**, (9) A distorts B, (10) A contains B, (11) **A hinders B**, (12) A obscures B, (13) A surrounds B, (14) A tightly fits into B, (15) **A is a salient part of B**, (16) **A and B are separable**, (17) A and B are connectable, (18) **A and B are interrelated**, (19) A and B are in orderly structure, (20) **A and B form a collection**. The loadings of the nine original features on the composite semantic-cryptotype measure were as follows: Change of Location (0.92), A and B are separable (0.91), A touches B (0.78), Mental Activity (0.72), A and B are interrelated (0.71), A hinders B (0.69), Circular Movement (0.68), A is a salient part of B (0.63), A and B form a collection (-0.51).

2.3.3. Procedure and Materials

The experiment used a production priming paradigm. Children were asked to take turns with the experimenter to describe cartoon picture sequences on a laptop (this arrangement allowed for the experimenter's description to serve as a 'prime' sentence and the child's description to serve as a 'target' sentence). All prime and target sentences corresponded to a cartoon sequence of an action followed by a reversal of that action. Each prime sentence was read-aloud in full by the experimenter and consisted of a verb that was grammatical in *un-* form (e.g., *Homer buckled his belt and then he unbuckled it*). The target sentence was begun by the experimenter (e.g., *Lisa squeezed the sponge and then she…*) but was completed by the child, such that she was responsible for describing the reversal action of the cartoon (e.g., *…*unsqueezed/ loosened/ released it*). Half of the target sentences contained verbs that are grammatical in *un-* form, half ungrammatical; the rationale was that children's restrictions on verbal *un-* prefixation would dictate whether the reversal action was – or was not – described with the target verb's *un-* form.

To ensure that the paradigm was age-appropriate, the experiment took the guise of a bingo game similar to that used by a recent developmental structural priming study (Rowland, Chang, Ambridge, Pine, & Lieven, 2012) whereby a confederate would pseudo-randomly hand 'bingo cards' to players following a prime sentence or target sentence. The bingo cards (i.e., tokens) matched the sentence that had been spoken and served as rewards for completing a trial and thus helped keep the children engaged in the game throughout the study. The first player to fill up his or her bingo grid won the game (every session was fixed such that the participant would win the bingo game on the final target trial of the session).

Target Verbs. Forty-eight target sentences were created, each with a different target verb (note that to allow for the most meaningful comparison, the target verbs were the same as test verbs used in Ambridge's (2013) judgment study). A check of the CHILDES database (Macwhinney, 2000) – whereby we extracted frequencies at which verbs are produced by, and heard by children aged one to seven years old - revealed that the majority of the verbs used occurred frequently in child-directed speech, and – indeed – were often used by the children themselves (see *Table A2.1* in Appendix 1). We thank Dave Ogden for making available to us a spreadsheet containing the frequencies of each individual lexical item in the entire CHILDES database. It is also worth noting that in even our Judgment study (which is a relatively difficult task for young children), examination of "zero" verbs (i.e., those that cannot take *un-*) revealed that each age-group misidentified no more than three of these verbs as being more acceptable in *un-* form than their bare form – see *Figure A2.1* [Appendix 4] and *Figure A2.2* [Appendix 5] respectively. Additionally, all verbs were accompanied by picture sequences to demonstrate each verb's meaning (in both Experiment 1 and 2), and on no occasions did children indicate to the researcher that they were unsure of a verb's meaning. Thus, we can be confident that most children were familiar with and understood the majority of these verbs (allowing us to use the same set as Ambridge (2013) - so as to ensure comparability across studies).

Half of the target verbs were grammatical in *un-* form (*"un-* verbs") and half ungrammatical in *un-* form ("*zero-* verbs"), as classified by Li and MacWhinney (1996):

- ***un- verbs (N = 24)****: Bandage, Buckle, Button, Chain, Cork, Crumple, Delete, Do, Fasten, Hook, Lace, Latch, Leash, Lock, Mask, Pack, Reel, Roll, Screw, Snap, Tie, Veil, Wrap, Zip.*

- ***zero- verbs (N = 24)****: Allow, Ask, Believe, Bend, Close, Come, Embarrass, Fill, Freeze, Give, Go, Lift, Loosen, Open, Press, Pull, Put, Release, Remove, Sit, Squeeze, Stand, Straighten, Tighten.*

It is important to note that nothing hinges on the accuracy of this classification of verbs as *un-/zero* (the classification was not used as a predictor in any statistical analysis). The point is simply that roughly half of the target *un-* forms were broadly-speaking grammatical, meaning that children could not usefully adopt a task-dependent strategy of treating all as grammatical (or ungrammatical). In order to reduce the number of trials completed by children, each child was assigned only one of two sets of 24 target verbs (Verb-set A/Verb-set B; see *Table A2.2* in Appendix 2), each containing 12 randomly selected *un-* verbs and 12 randomly selected *zero* verbs.

Prime Verbs. There were also 24 prime sentences for each participant with the caveat that no verb served as both a prime and target verb for the same participant. Thus, the 12 grammatical *un-* verbs used as target verbs in Verb-set A were used as prime verbs for Verb-set B, and vice versa. Twelve additional verbs (mostly taken from Li and MacWhinney (1996) and all grammatical in *un-* form) were used as prime verbs for all participants, in order to make up the total of 24 primes per participant.

Sentences. For each verb (both prime and target) we created a sentence of the form *[CHARACTER] [VERB-ed] and then (s)he un-[VERB-ed]* (for a full list, see *Table A2.2* in Appendix 2), and a corresponding sequence of still cartoon pictures. Four different characters (*Homer, Bart, Lisa* and *Marge*) were used. An additional three prime and target sentences plus corresponding sequences were created for the practice session; all used verbs that were grammatical in *un-* form (this served to encourage production of *un-* forms before testing began) and did not form part of the test sets. The prime and target sentences were randomly selected for each trial; we did not use pre-specified prime and target pairs. To avoid the task becoming too arduous for children, the test session was divided into two sessions of 12 prime-target trials, with a rest period between each session.

2.4.4. Coding

Coding was based on the child's first response only. Responses were coded as "*un-* form", "not *un-*" or "other" (i.e., excluded) according to the following criteria:

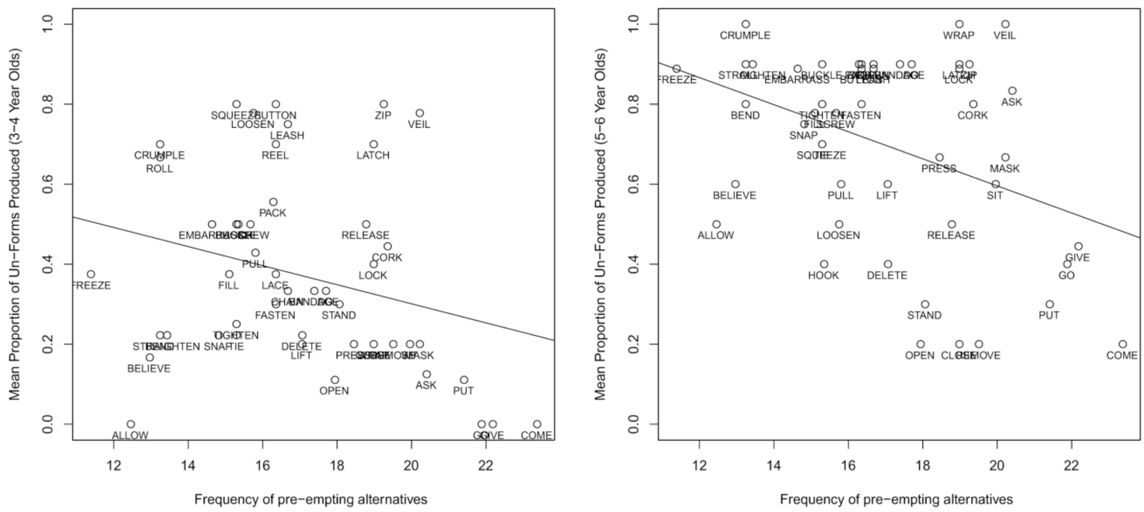
- ***"Un-form":*** if the target verb was produced with *un-* prefixation (e.g., EXP: *Homer wrapped the present and then he…* CHI: *unwrapped it*).

- ***"Not un-":*** if the participant described the reversal action accurately without using the target verb in *un-* form (e.g., *took the wrapper off*).
- ***"Other":*** Responses were excluded from analyses if: (i) there was experimenter error, or (ii) the response did not accurately describe a reversal of the action denoted by the target verb; this criterion includes responses in which a general reversal term (e.g., *didn't*) was used without any relevance to the specific reversal action (e.g., *Marge allowed Bart some chocolate and then she…didn't*).

2.5.5. Results and Discussion

The current study used an elicited production paradigm to investigate children's (aged 3-4; 5-6) grammatical restrictions on verbal *un-* prefixation. Collapsing across all verbs, responses were coded as "Other" for 9.79% of 3-4 year olds' trials and 4.38% of 5-6 year olds' trials (out of 47 trials excluded for 3-4 year olds, 35 were due to the child's response being an inaccurate description of the reversal action, or use of a general reversal term [e.g., "*didn't*"], 10 were due to no response being given, and 2 were due to experimenter error; out of 21 trials excluded for 5-6 year olds, 10 were due to an inaccurate description of the reversal action or use of a general reversal term, 5 were due to no response being given, and 6 were due to experimenter error). Once these trials were excluded from the denominator, 3-4 year olds and 5-6 year olds produced *un-* forms of the target verb on 37.64% (SD= 48.5) and 69.06% (SD=46.27) of trials respectively. Given (a) the low rate of excluded "Other" responses, and (b) the fact that only around 50% of target verbs are grammatical in *un-* form, these totals indicate that the production priming paradigm was highly successful at eliciting both *un-* forms and alternative reversal verbs. Furthermore, examination of *zero*-verbs only (i.e. verbs that do not take *un-*) revealed that the younger age group produced *un-* forms on 23.38% [SD=42.46] of these trials (older group = 50.31% [SD=50.16]). Thus, we can also be sure that both age-groups were over-generalizing *un-* prefixation to verbs that do not take un (i.e., *zero*-verbs).

Figure 2.1. Mean proportion of *un-* forms produced for each verb by age group as a function of the pre-emption predictor (age 3-4 on the left; age 5-6 on the right)



Results were analysed using binomial linear mixed effects models (*glmer* from package *lme4*; Bates, Maechler & Bolker, 2011) in the *R* environment (R Development Core Team, 2013). Mixed-effects models predict individual trials rather than averaging over trials, and offer the added benefit of treating both participant and item as random effects (i.e., the model creates an intercept for each participant and each item, thus removing variation within each of these factors). They are also robust against missing data (Baayen, Davidson & Bates, 2008). All models outlined below included random intercepts for participants and verbs, as well as by-participant and by-verb random slopes where applicable. Following recommendations outlined by Barr, Levy, Scheepers and Tily (2013), a maximal random effects structure (i.e., all random intercepts and slopes) was attempted for each model. If maximal models did not converge, the model was simplified using the *best-path* algorithm (Barr et al., 2013). Under *best-path*, the inclusion of each random slope was determined by whether its addition to a random-intercept only model improved the model's fit of the data (using likelihood ratio tests and anti-conservative alpha values of $p<0.25$) (slopes that satisfied this criteria were added one-by-one until the model failed to converge). In line with the recommendations of a recent paper (Wurm & Fiscaro, 2014), we used simultaneous regression models with neither residualization nor centering.

Collapsing Across Age. The first model collapsed across age-group and is shown in Table 2.1 (because all predictors were entered in a single step, the order in which they are listed is arbitrary). Fixed effects were measures of (a) *verb-type* (dummy-coded, with 'zero-verbs' as default) (b) *frequency of verb in un- form*, (c) *reversibility*, (d) *pre-emption*,

(e) *entrenchment*, (f) *semantic-cryptotype* and (g) *age* (dummy-coded, with 3-4 year olds as default) (see *Method* section for details of each measurement). The outcome variable was whether the child produced a "*un-*" or "not-*un*" response on each trial ("other" responses were excluded). A positive beta (β) value indicates a positive correlation between the predictor and the likelihood of a verb being produced in *un-* form – as expected for *semantic-cryptotype*. A negative β value indicates a negative correlation between the predictor and the likelihood of a verb being produced with *un-* prefixation – as expected for measures of *pre-emption* and *entrenchment*. The model also included interaction terms for *age*pre-emption, age*entrenchment,* and *age*semantic-cryptotype*. By-participant random slopes for *semantic-cryptotype* were included (the inclusion of any other slopes failed convergence). Considering first the control predictors, a main effect of *frequency of un- form* was observed, such that production of *un-* forms was positively related to the target verb's corpus frequency in *un-* form. The other control predictor – *reversibility* – exerted a *negative* effect, such that *un-* forms were more likely to be produced with verbs that were *less* reversible – this emphasises that the *semantic-cryptotype* measure could not have been a proxy for a verb's reversibility. Turning now to the predictors of interest, production probability of *un-* forms was negatively related to the frequency of both pre-empting forms (*pre-emption*) and the verb's bare form (*entrenchment*). Neither *pre-emption* or *entrenchment* effects interacted with *age*, indicating that these effects were evident in both age-groups.

Table 2.1. Mixed Effects Models for production data (Collapsed Across Age)

Fixed effects	*Beta (β)*	*SE*	*z value*	*HPD95 lower*	*HPD95 upper*	*p*
(Intercept)	3.86	1.75	2.21	0.432	7.293	0.027
Verb Type	-0.78	0.54	-1.45	-1.836	0.276	0.148
Freq of *un-* form	**0.34**	**0.10**	**3.32**	**0.137**	**0.533**	**0.001**
Reversibility	**-0.65**	**0.26**	**-2.54**	**-1.154**	**-0.148**	**0.011**
PreEmption	**-0.16**	**0.08**	**-1.99**	**-0.308**	**-0.002**	**0.047**
Entrenchment	**-0.27**	**0.11**	**-2.35**	**-0.491**	**-0.044**	**0.019**
Semantics	0.32	0.28	1.16	-0.222	0.868	0.246
Age	3.33	1.82	1.83	-0.234	6.903	0.067
PreEmption:Age	-0.05	0.09	-0.54	-0.216	0.123	0.593
Entrenchment:Age	0.02	0.13	0.16	-0.225	0.265	0.873
Semantics:Age	0.04	0.33	0.11	-0.609	0.684	0.910

Follow-up models were conducted for each age-group and are shown in *Table 2.2*. These models were set up in the same way as the model outlined above but excluded *Age* and its interactions as fixed effects. The model of 3-4 year olds' data included by-participant random slopes for *Frequency of un-form, Pre-emption, Entrenchment,* and *Semantic-cryptotype*. The model of 5-6 year olds' data included by-participant random slopes for *Verb-type, Frequency of un-form, Entrenchment,* and *Semantic-cryptotype.*

Table 2.2. Mixed Effects Models for production data (Age 3-4; Age 5-6)

	Age 3-4						**Age 5-6**					
Fixed effects	*Beta (β)*	*SE*	*z value*	*HPD95 lower*	*HPD95 upper*	*p*	*Beta (β)*	*SE*	*z value*	*HPD95 lower*	*HPD95 upper*	*p*
(Intercept)	4.73	2.00	2.36	0.80	8.66	0.02	7.80	2.32	3.36	3.25	12.35	0.00
Verb Type	-0.87	0.71	-1.24	-2.26	0.51	0.22	-1.12	0.89	-1.26	-2.86	0.62	0.21
Freq of *un-* form	0.24	0.14	1.70	-0.04	0.52	0.09	**0.49**	**0.17**	**2.84**	**0.15**	**0.83**	**0.00**
Reversibility	-0.33	0.33	-1.00	-0.97	0.31	0.32	**-1.06**	**0.41**	**-2.57**	**-1.87**	**-0.25**	**0.01**
PreEmption	-0.17	0.09	-1.87	-0.35	0.01	0.06	**-0.21**	**0.10**	**-2.15**	**-0.41**	**-0.02**	**0.03**
Entrenchment	**-0.36**	**0.15**	**-2.44**	**-0.64**	**-0.07**	**0.01**	-0.23	0.16	-1.40	-0.55	0.09	0.16
Semantics	0.37	0.30	1.23	-0.22	0.96	0.22	0.46	0.39	1.18	-0.30	1.22	0.24

Age 3-4. Again considering first the control predictors, a main effect of *frequency of un- form* was observed, such that production of *un-* forms was positively related to the target verb's corpus frequency in *un-* form. The other control predictor – *reversibility* – did not exert any significant effect, indicating that the study's *semantic-cryptotype* measure did not serve as a proxy for reversibility. Turning to the predictors of interest, production probability of *un-* forms was negatively related to the frequency of the verb's bare form (*entrenchment*). Production probability of *un-* forms was negatively related to the frequency of both pre-empting forms (*pre-emption*; see Figure 2.1), but this effect narrowly missed significance (p=0.06).

It was somewhat surprising that the pre-emption effect narrowly missed significance for this age-group despite yielding a significant effect in the full model with no significant *age* interaction (Table 2.1). One reason may be that the model for 3-4 year olds required more random-slopes than the model that collapsed across age, and so can be considered a more conservative estimate of the pre-emption effect. Given this, a separate analysis was conducted with only pre-emption as a fixed-effect (with random intercepts for verb and participant and by-participant random-slopes for pre-emption) and production of *un-* forms as the outcome variable. A significant effect of pre-emption was observed, $\beta=-0.24$ $(SE=0.09), z=-2.54, p=0.01$). Thus, whilst the effect of pre-emption could not be observed over and above the rest of the predictors in 3-4 year olds' data, it is clear that this factor is important in 3-4 year olds' restriction of *un-* prefixation.

Age 5-6. A main effect of *frequency of un- form* was observed, such that production of *un-* forms was positively related to the target verb's corpus frequency in *un-* form. As with the full model that collapsed across age, a *negative* effect of *reversibility* was observed, again emphasising that the *semantic-cryptotype* measure could not have been a proxy for a verb's reversibility. Turning to the predictors of interest, a significant negative correlation was observed between the proportion of *un-* forms produced and *frequency of pre-empting forms* (*pre-emption*; see Figure 2.1), but not entrenchment.

The results outlined above demonstrate that both 3-4 year olds and 5-6 year olds use *pre-emption,* such that production of *un-* forms was negatively predicted by corpus frequency of synonyms for the target verb's *un-* form. An effect of *entrenchment* – such that production of *un-* forms was less likely when the target verb was highly frequent in bare form - was observed for 3-4 year olds but not 5-6 year olds. The effect of *semantic-cryptotype* failed to reach significance for either age-group.

The finding of no *semantic-cryptotype* effect for the 5-6 year olds is at odds with that of Ambridge (2013) who did find such an effect. A possible explanation for this pattern is that – for older children - a grammaticality judgment task – as used in this previous study – is better suited to detecting fine-grained semantic effects than is a binary production task. On the other hand, an effect of *pre-emption* for 5-6 year olds was observed in the present study, but not the judgment study of Ambridge, possibly because the *semantic-cryptotype* effect observed in this previous study left less variance to be explained by *pre-emption*. Another possibility is that a production task encourages children to search their lexicon for pre-empting alternatives to ungrammatical *un-* forms to a greater extent than does a judgment task.

In order to examine these possibilities, and to investigate the relationship between production and grammaticality judgment data more generally, we investigated whether the children who participated in the current production study would show a similar pattern of data in a grammaticality judgment task.

2.4. Experiment 2: Grammaticality Judgement Study

2.4.1. Participants

Participants were the same as those who took part in the production study. The two studies were completed at least one week apart, in counterbalanced order.

2.4.2. Design

Participants remained in their counterbalanced groups (e.g., participants exposed to Verb-set A in the production study were asked to judge the grammaticality of target verbs from that set, in both *un-* and bare form). The dependent variable was the acceptability rating of each *un-* form on a scale of 1 to 5 (log transformed). The judgment study used the same predictor variables as the production study, plus one additional predictor:

- ***Acceptability of bare form (log transformed).*** Participants rated the acceptability of each verb's bare form (e.g., *squeeze*) to control for the possibility that individual participants would show general (dis)preferences for particular verbs, perhaps based on semantic or phonological properties, regardless of form (*un-*/bare).

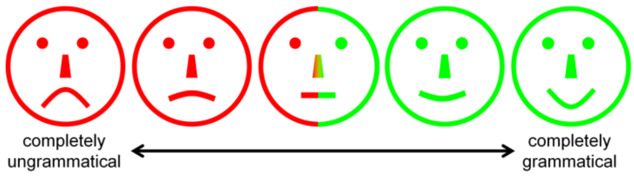
2.4.3. Procedure and Materials

All sentences were presented in audio form. To make the task more engaging, children were introduced to a toy dog that was 'learning to speak English.' The child was asked to help the dog to speak properly by telling him which words sounded "right" and which words sounded "wrong and a bit silly" (for full details see, Ambridge et al., 2008; Ambridge, 2013). Children were then introduced to a five-point smiley-face scale (Figure 2.2) which would be used to rate sentences in a graded manner. In short, the process involved a child taking a green or red counter to indicate grammatical or ungrammatical items respectively and placing the counter on the scale to indicate the degree of grammaticality (5 = perfectly grammatical; 1= very ungrammatical). To familiarise themselves with the rating scale, participants first completed a practice session comprised of seven sentences, each including either a correct past-tense form or an over-regularization error (e.g., **Homer breakded the cup*), accompanied by an appropriate picture sequence. Participants were asked to rate the *verb only*: After the participant had heard the full sentence, the experimenter repeated the verb in isolation and asked participants to indicate its grammaticality). The subsequent two test sessions took the same format as the practice session.

2.4.4. Sentence Stimuli

Each verb was presented in two separate trials: once in bare form to obtain a control rating (e.g., *Lisa squeezed the sponge*) and once in *un-* form (**Lisa unsqueezed the sponge*). There were thus 96 test trials (48 bare- and 48 *un-* forms) in the judgment study as opposed to just 48 in the production study. Children remained in their counterbalanced groups (Verb-set A or Verb-set B) and were thus only required to complete 48 judgment trials each (24 bare forms and 24 *un-* forms). The test session was split into two separate sessions of 24 trials to avoid the task becoming too arduous. Each test session contained a verb's bare and *un-* form but these forms were never presented in consecutive trials; with this caveat in mind, all trials were presented in a random order for each participant. For a full list of practice and test sentences, see Table A2.3 in the Appendix 3.

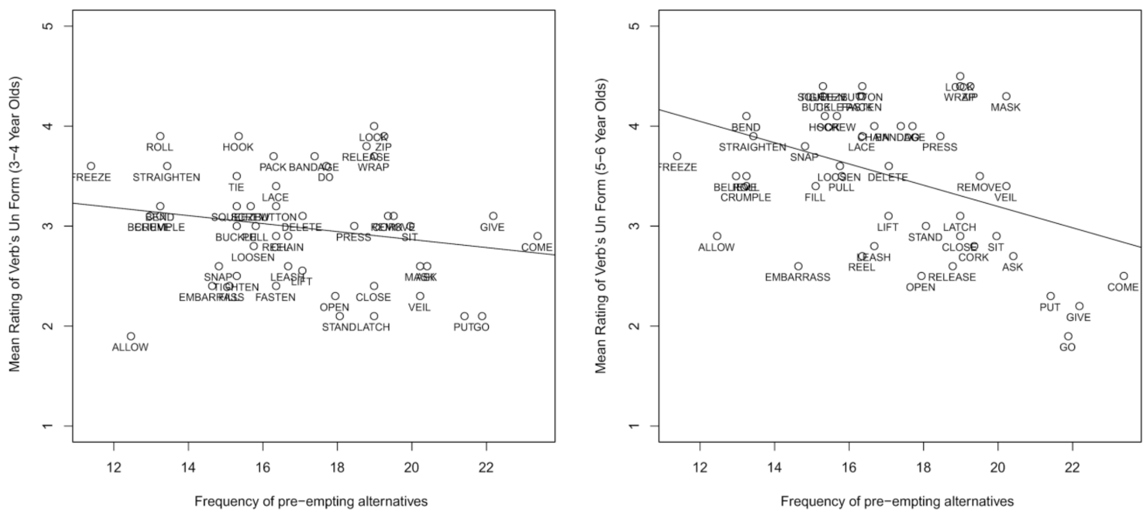
Figure 2.2. The 5-point smiley face scale used by participants to rate the relative acceptability of the *un-* prefixed and bare verb forms (reproduced from Ambridge et al., 2008: 105, by permission of Elsevier).



2.4.5. Results and Discussion

The purpose of the grammaticality judgment study was to examine the possibility that, compared with a production task, a judgment task is more likely to detect fine-grained semantic-cryptotype effects (due to its greater sensitivity). As well as an exploration of the relationship between production and judgment data more generally, it also served as an investigation of whether the graded grammaticality judgment paradigm could be extended to children aged 3-4.

Figure 2.3. Mean acceptability rating for each verb's *un-* form as a function of the pre-emption predictor (age 3-4 on the left; age 5-6 on the right).



Results were analysed using linear mixed effects models (*lmer* from package *lme4*; Bates, Maechler & Bolker, 2011). The dependent variable was the acceptability rating for each verb's *un-* form. All models included random intercepts for participants and verbs and random slopes where applicable (fitted using the *best-path* method outlined in Experiment 1). The models used the same fixed effects as the production study, plus one additional fixed effect which was employed as a control variable: *acceptability ratings for each verb's bare form*. Results of the judgment analysis collapsed across age-groups are shown in Table 2.3., and results of the analyses by-age-group are shown in Table 2.4. The significance of each fixed-effect was determined by p- values calculated based on Satterthwaite's approximation (in accordance with a recent stats paper; see Luke, 2016) using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2016).

Table 2.3. Mixed Effects Models for Judgment Data (Collapsed Across Age)

Collapsed Across Age

	Beta (β)	*SE*	*t*	*HPD95 lower*	*HPD95 upper*	*p*
(Intercept)	0.75	0.55	1.37	-0.32	1.82	0.17
Verb Type	-0.10	0.19	-0.51	-0.48	0.28	0.61
Freq of *un-* form	**0.09**	**0.04**	**2.67**	**0.03**	**0.16**	**0.01**
Rating For Bare Form	**0.16**	**0.03**	**4.84**	**0.10**	**0.23**	**0.00**
Reversibility	-0.09	0.09	-1.04	-0.27	0.08	0.30
PreEmption	-0.04	0.03	-1.40	-0.08	0.01	0.17
Entrenchment	0.01	0.04	0.36	-0.06	0.09	0.72
Semantics	0.11	0.09	1.16	-0.07	0.28	0.25
Age	**1.18**	**0.51**	**2.30**	**0.17**	**2.18**	**0.02**
PreEmption:Age	-0.05	0.03	-1.92	-0.11	0.00	0.06
Entrenchment:Age	0.01	0.04	0.20	-0.07	0.09	0.84
Semantics:Age	**0.23**	**0.10**	**2.24**	**0.03**	**0.42**	**0.03**

Collapsing Across Age. The model included by-participant random slopes for *Frequency of un- form* and *Semantic Cryptotype*. Judgment data revealed a significant main effect of *Rating of bare form* such that judgments of *un-* forms were more favourable when a verb was rated favourably in bare-form (justifying its use as a control variable). A main effect of *Age* was observed, and this effect significantly interacted with the effect of *Semantic-cryptotype,* warranting further investigation by age-group.

Table 2.4. Mixed Effects Models for Judgment Data (Age 3-4; Age 5-6)

	Age 3-4						Age 5-6					
	Beta (β)	*SE*	*t*	*HPD95 lower*	*HPD95 upper*	*p*	*Beta (β)*	*SE*	*t*	*HPD95 lower*	*HPD95 upper*	*p*
(Intercept)	0.80	0.63	-0.43	2.02	1.27	0.21	**1.89**	**0.55**	**0.82**	**2.96**	**3.45**	**0.00**
Verb Type	-0.12	0.26	-0.62	0.38	-0.47	0.64	-0.07	0.21	-0.49	0.35	-0.33	0.75
Freq of *un-* form	0.08	0.05	0.00	0.17	1.86	0.07	**0.10**	**0.04**	**0.03**	**0.18**	**2.65**	**0.01**
Rating For Bare Form	**0.20**	**0.05**	**0.11**	**0.29**	**4.35**	**0.00**	**0.15**	**0.05**	**0.05**	**0.24**	**3.03**	**0.00**
Reversibility	-0.10	0.12	-0.33	0.13	-0.83	0.41	-0.08	0.13	-0.34	0.18	-0.60	0.55
PreEmption	-0.03	0.03	-0.09	0.02	-1.21	0.23	**-0.09**	**0.02**	**-0.14**	**-0.04**	**-3.77**	**0.00**
Entrenchment	0.00	0.04	-0.08	0.09	0.09	0.93	0.03	0.04	-0.05	0.11	0.77	0.45
Semantics	**0.11**	0.10	-0.08	0.30	1.16	0.25	**0.33**	**0.09**	**0.15**	**0.52**	**3.58**	**0.00**

Age 3-4. The model included by-participant random slopes for *Verb Type*. Judgment data for the youngest group revealed no significant effects of *semantic-cryptotype*, *pre-emption* or *entrenchment*. Rather, the only significant predictor of grammaticality ratings for *un*-forms was *ratings for bare form*. Thus, it seems that the judgment paradigm underestimated 3-4 year olds' grammatical knowledge (relative to the production study), given that it failed to yield any significant effects of *pre-emption, entrenchment* and *semantic-cryptotype,* the former two of which were present for the same participants in the production study. The likelihood is that judgment data from the younger age group were too noisy for detection of any mechanisms of restriction.

Age 5-6. The model included by-participant random slopes for *Reversibility, Semantic-cryptotype and Entrenchment*. Considering first the control predictors, the older age group's judgments of *un-* forms were positively correlated with *frequency of un- form*, and *rating for bare form*. Turning to the predictors of interest, judgments of a verb's *un-*form were positively correlated with the extent to which verbs denoted semantics of *un-*prefixation (*semantic-cryptotype*) and negatively correlated with the frequency of pre-empting forms (*pre-emption;* see *Figure 2.3*). There was no effect of the *entrenchment* measure.

In summary, 3-4 year olds' judgment data appeared too noisy to yield any effects any of the proposed restriction mechanisms. Thus our knowledge of this age-group's restriction mechanisms must be taken from production data, which revealed effects of *entrenchment* and *pre-emption*, but not *semantic-cryptotype*. The older age-group (5-6 year olds) used both *pre-emption* and *semantic-cryptotype* to guide grammaticality judgments of *un-* prefixed verbs. The *pre-emption* effect persisted in this age-group's production data but the *semantic-cryptotype* effect did not, possibly because semantic effects are more fine-grained and thus harder to detect in production tasks. Taken together, Experiment 1 and 2 indicate that children as young as 3-4 are using *pre-emption* and *entrenchment* to guide productivity of verbal *un-* prefixation, and that use of a *semantic cryptotype* – a category that encompasses the semantics shared by verbs that have previously appeared in the same context – has emerged by 5-6 years old.

Comparison between Judgment and Production Data

We suggested above that judgment paradigms may be an unsuitable measure of 3-4 year olds' grammatical knowledge. To examine the validity of this claim, we compared judgment data and production data. We expected to find a correlation between production probability and judgments of *un-* forms for 5-6 year olds, but not 3-4 year olds, on the assumption that only for the older group is the judgment paradigm a suitable measure of the grammatical knowledge that drives production

For both age groups, we ran a mixed-effects model (*glmer*) with children's mean proportion of *un-* forms produced (i.e. production data) as the outcome measure and ratings of a verb's *bare form* (a control variable for judgment data) and *un-* form (the predictor of interest) as fixed effects. All models included participants and items as random effects.

Age 3-4. Younger children's production of *un-* forms was negatively predicted by their *ratings for bare forms,* β =-0.45 (*SE* = 0.15), *p*=0.003, but was not predicted by their *ratings for un- forms, B*=-0.00, *SE*=0.14, p=0.97. These data suggest that 3-4 year old children's ratings of *un-* prefixed verbs were determined by baseline (dis)preference for individual verbs (in their canonical bare form) rather than their knowledge of restrictions on *un-* prefixation, rendering the grammaticality judgment paradigm unsuitable for younger children (at least, for this particular study). Recall that the production data did indeed suggest knowledge of restrictions on *un-* prefixation for this age group.

Age 5-6. Older children's production of *un-* forms was not related to their *ratings for bare form,* β = -0.24 (*SE*=0.16), *p*= 0.15) but was positively predicted by *ratings for un-form,* β = 0.34 (*SE* = 0.15), *p*=0.023, such that the more likely a verb was rated as grammatical in *un-* form, the more likely they were to produce that verb in *un-* form.

We can conclude that that the judgment paradigm was unsuitable as a measure of 3-4 year old children's grammatical knowledge. The judgment paradigm can be considered a reasonably valid measurement of 5-6 year olds' grammatical knowledge given that judgments of verb *un-* forms predicted the likelihood a verb would be produced in *un-*form. Moreover, the paradigm yielded effects of *pre-emption* and *semantic-cryptotype* for this age-group, the latter of which was not detected by the production paradigm. Thus, when used with a suitable age-group, the judgment paradigm may be a more sensitive measure of children's use of a semantic-cryptotype in their restriction of *un-* prefixation.

2.5. General Discussion

Recent research has demonstrated that any complete account of the retreat from overgeneralization must incorporate roles for pre-emption, entrenchment and verb-and-construction semantics (e.g., Ambridge, Pine, Rowland, Freudenthal & Chang, 2014; Ambridge, Pine, & Rowland, 2012; Ambridge, Pine, & Rowland, 2011; Ambridge, 2013). However, the roles played by these mechanisms in the early stages of retreat from error are less clear. In the Introduction, we outlined a recent grammaticality judgment study of overgeneralization errors involving verbal *un-* prefixation (Ambridge, 2013), in which 5-6 year old children demonstrated use of a semantic "cryptotype" hypothesised to represent verbs that take *un-* (Whorf, 1956), but no use of pre-emption or entrenchment. The current study investigated the possibility that judgment paradigms may underestimate young children's grammatical knowledge, and hence obscure pre-emption/ entrenchment effects that may be present at this age and younger. To address this possibility, we employed what we hope was a less demanding production paradigm to examine young children's (3-4; 5-6) restrictions on verbal *un-* prefixation.

In Experiment 1, children were asked to describe reversal actions of verbs that were or were not grammatical in *un-* form (e.g., *unwrap; *unsqueeze*), the rationale being that the likelihood of that verb being produced in *un-* form would be dictated by the verb's semantic properties, its entrenchment in other contexts, and/ or the frequency of pre-empting formulations. In Experiment 2, the same children were asked to give grammaticality judgment ratings for each verb's *un-* form so that findings from production and judgment paradigms could be compared.

Looking first at 3-4 year old children, production of *un-* prefixed verbs was negatively predicted by (a) the target verb's frequency in bare form (i.e. not *un-* form; e.g. *squeez/e/s/d/ing*) and (b) the frequency of synonyms to the target verb's *un-* form (e.g., *release* + *loosen* for **unsqueeze*) – demonstrating use of entrenchment and pre-emption respectively. Thus, production data provides evidence that entrenchment and pre-emption are operational for children as young as 3-4 (M=4;0). However, 3-4 year olds' judgment data were deemed too noisy to detect any use of restriction mechanisms.

Examination of 5-6 year old children's data revealed that the pre-emption measure predicted judgments and production of *un-* prefixed verbs, confirming that use of this mechanism persists into this later developmental stage. A semantic-cryptotype effect was evident amongst 5-6 year olds, such that judgments of *un-* prefixed verbs were positively related to the extent to which each verb denoted a semantic cryptotype hypothesised to represent properties shared by verbs that licence *un-* (e.g. Whorf, 1956).

Taken together, the present experiments indicate a role for pre-emption, entrenchment and verb-and-construction semantics from an early age. Further, it appears that children may initially learn verbs' restrictions by monitoring the distributional patterns of the verb in other contexts [entrenchment], as well as those of the verb's competing formulations that convey similar meaning [pre-emption], with a role for verb-and-construction semantics (or more specifically, in this study's case, Whorf's (1956) hypothesised "semantic cryptotype") emerging by 5-6 years old. Although it may be tempting to conclude that these results support a "statistics-before-semantics" approach whereby use of a verb's statistical properties precedes use of its semantic properties (e.g.,Tomasello, 2003), caution must be taken in adopting this perspective. The reason is that both pre-emption and entrenchment have underlying semantic motivation. For pre-emption to operate, a speaker must recognise that a pre-empting alternative exhibits appropriate semantics that convey the same message as the target verb's *un-* form. Entrenchment can also be argued to have underlying semantic motivations, since any lexical item's entrenchment is a consequence of a verb exhibiting suitable semantics to convey the desired message (when placed in a suitable sentence construction). Thus, evidence for children's use of entrenchment or (especially) pre-emption demonstrates the ability to use a verb's statistical and – in some sense - semantic properties to restrict productivity, with the current study indicating that this ability is evident from 3-4 years old. Acknowledging previous literature that demonstrates pre-emption, entrenchment and verb-and-construction semantics to persist into later stages of development (e.g., Ambridge, 2013; Ambridge et al., 2011; 2012), it is clear that children's restriction mechanisms involve an interactive process in which 'statistical' and 'semantic' effects cannot be picked apart so easily.

One framework that may be useful for understanding these results is the *FIT* account outlined in Ambridge and Lieven (2011). A more detailed description of how this account can yield entrenchment, pre-emption and verb-and-construction semantic effects in the domain of *un-* prefixation is given in Ambridge (Ambridge, 2013). In brief, the central idea is that all constructions in a speaker's lexicon compete for activation (Macwhinney, 2014); i.e., for selection to express the speaker's intended message (e.g., the reversal of a squeezing action). The most relevant "constructions" in this context are whole words (e.g., *release, loosen*) and the morphological *un-* prefixation construction (*un-[VERB]*).

The account yields pre-emption effects because the greater the frequency of competing forms (e.g., *release, loosen*), the greater their activation, and hence the lower the activation of the competing potential *un-* form (e.g., **unsqueeze*). The account yields entrenchment effects due to the assumption that *every* construction in the speaker's inventory competes for selection, with the activation determined by – amongst other things – their relevance to the speaker's message. For example, if the message is the reversal of a squeezing action, the competitors will be not only *release, loosen* and *unsqueeze*, but *squeeze* itself. Entrenchment effects occur because the activation of each alternative is determined not only by its relevance, but also its input frequency (and hence the strength of its trace in the lexicon). Because pre-empting forms (e.g., *release, loosen*) are better (i.e., more relevant) competitors for a given *un-* error (e.g., **unsqueeze*) than are entrenching forms (e.g., *squeeze*), this account may be able to explain the present finding that pre-emption appears to be more important than entrenchment. Future modelling work should attempt to clarify whether or not such an account can in fact yield this pattern (for preliminary modelling work in this domain, see Li & Macwhinney, 1996; Ambridge, Freudenthal, Pine, Mills & Clark et al., 2009).

The account yields verb-and-construction semantic effects due to the assumption that the *un-[VERB]* construction, like all abstract constructions, is acquired by abstracting across memory-traces of stored exemplars of this construction in memory (e.g., Abbot-Smith & Tomasello, 2006), in this case, individual *un*-forms (*unscrew, unbutton* etc.). Thus the *[VERB]* slot of this construction probabilistically exhibits the averaged semantic properties of every item that has previously occupied that slot (e.g., Langacker, 2000)). The greater the overlap between the semantic properties of this slot and a putative filler (e.g., *squeeze*), the greater the activation of the relevant *un-* form. Again, preliminary computational models of the acquisition of *un-* prefixation (Ambridge, 2013; Abbot-Smith & Tomasello, 2006), show this type of semantic generalization. We are agnostic with regard to the question of whether the *un-[VERB]* construction is represented independently of the exemplars that instantiate it (i.e., between prototype and exemplar models). However, the assumption that a prerequisite for this generalization is a set of stored exemplars, may be able to explain the present finding that statistical effects appear to emerge before verb-and-construction semantic effects (though – as we have just seen – not before *all* types of semantic effect): Effects of pre-emption and entrenchment can arise on the basis of the stored exemplars themselves; verb-and-construction semantic effects only as the result of some kind of generalization or analogy across these exemplars. However, to address this question more definitively, more modelling work will be needed.

One issue that we should address is that the lack of filler trials in the production study (such that all prime sentences featured reversal actions described with a *un-* prefixed verb) may have led to an unrealistic 'over'-production of *un-* forms that was not representative of levels of *un-* prefixation in children's spontaneous speech. However, this paradigm was indeed designed to pull children towards using *un-* prefixation, the rationale being that a child's command of verbs' distributional and semantic properties should guide their productivity, thus providing a window into restriction mechanisms employed by these children. Since we obtained a number of results that differentiated between verbs, the use of this method appears to be justified. Indeed, using a method that led to lower rates of *un-* prefixation would most likely have significantly reduced the possibility of observing the by-verb differences that are required in order to test the pre-emption, entrenchment and verb-and-construction semantics hypotheses. We must also acknowledge that – on the one hand – only a judgment paradigm was sufficiently sensitive to detect semantic effects in 5-6 year olds, but – on the other – only a production paradigm was sufficiently simple to detect pre-emption and entrenchment effects in 3-4 year olds. Thus a profitable direction for future research is to employ paradigms such as eye-tracking or Event Related Potentials (ERP), that are sufficiently sensitive to detect fine-grained effects, but that can be combined with tasks that are very simple for young children.

In conclusion, the present findings indicate that children as young as 3-4 are guided by pre-emption and entrenchment in their production of verbal *un-* prefixation. By age 5-6, children also show use of a complex 'cryptotype' of semantic properties thought to be representative of verbs that licence *un-*. Together, these findings reflect a complex interaction between statistical and semantic properties of competing lexical items that we have posited to be operational within one interactive framework.

2.6. Prelude to Study 2

Study 1 demonstrated that verb-semantics and statistical-learning mechanisms are used to restrict generalizations of verbal *un-* prefixation from an early age. The finding that generalizations of *un-* prefixation were positively related to the extent to which a verb is compatible with Whorf's (1956) semantic cryptotype (a cluster of semantic features typically associated with *un-* prefixable verbs, but which cannot be easily defined) is difficult to explain in terms of discrete semantic classes (e.g., Pinker, 1989). Rather, the finding is better explained by use of a productive *un*[VERB] schema, where the [VERB] slot exhibits fuzzy and probabilistic properties shared by clusters of verbs that have previously occupied that slot, and the likelihood of a verb being used with this schema is determined probabilistically by the semantic 'fit' between the semantic properties of the verb and those of the [VERB] slot.

The explanation outlined above is consistent with the *FIT* account (Ambridge & Lieven, 2011), which can also explain the effects of statistical-learning in terms of a probabilistic learning mechanism where different formulations (i.e., lexical, morphological and syntactic constructions) compete to convey the intended message (see *Section 1.4.4* of the thesis).

Another important prediction of the *FIT* account is that it is possible for constructions to exhibit not just semantic properties but also phonological properties, and that these properties can govern the grammaticality of a formulation. In a description of the constructivist framework that has been adopted by the *FIT* account, Ambridge and Lieven describe the property of a construction slot as:

> a weighted average of all the items that have appeared in this position in the input utterances that gave rise to the schema[...]. But a weighted average of *which* properties: their meanings, their sounds, their stress patterns? In principle, over any of these things; indeed, over any properties that the child can perceive. If the items that appear in a particular position in the source utterances are similar with respect to a given property (e.g., meaning), then the slot in the resultant schema will exhibit this property. (2015:5)

In the case of the English past-tense, verbs that have appeared in a particular past-tense form tend to share phonological properties as opposed to semantic properties. For example, verbs like *click/clicked, trick/tricked, kick/kicked* are examples of a phonologically-similar set of verbs that are used in 'regular' past-tense form (i.e., with an '*-ed'* inflection). Thus, it is possible that speakers generate English past-tense forms (and overgeneralizations) based on a verb's phonological properties.

English past-tense is characterised by its distinction between regular (e.g., *kick/kicked; play/played*) and irregular (*sleep/slept; throw/threw*) patterns of inflection. Under the constructivist framework adopted by *FIT*, one can posit an inventory of 'regular' and 'irregular' morphological schemas that instantiate the distinction between regular (e.g., *kick/kicked; play/played*) and irregular (*sleep/slept; throw/threw*) past-tense forms. More precisely, one can posit three 'regular' schemas, [VERB]-*d*, [VERB]-*t*, and [VERB]-*əd,* to account for the fact that the regular past-tense morpheme may be realised as –*d*, (e.g., *played*), -*t* (e.g., *wish/wished*) or -*əd* (e.g., *hunt/hunted*). 'Irregular' schemas such as [VERB]+*vowel-change* are posited to account for particular irregular inflections such as *throw/threw* and *blow/blew*. Each of the different [VERB] slots exhibit a different constellation of phonological properties, determined by the properties shared by verbs that have previously appeared in the slot. For example, the [VERB]-*t* past-tense schema exhibits the phonological properties shared by *miss/missed, kiss/kissed,* and *hiss/hissed,* as well as the phonological properties shared by *wish/wished, fish/fished,* and *dish/dished*.

Rather like the semantic 'cryptotype' posited to represent verbs that take *un-*, the phonological properties posited to be exhibited by past-tense schemas are held to be a fuzzy and probabilistic cluster of properties, none of which are deemed necessary or sufficient for a verb to be used in the schema. The greater the extent to which a novel verb's phonological properties overlap with the constellation of phonological properties associated with a past-tense schema, the more likely it is to be generalised into that schema. The domain of English past-tense thus provides a compelling test of *FIT*: the greater extent to which a verb's phonological properties overlap with a 'regular' or 'irregular' schema, the more likely it is to be produced in 'regular' or 'irregular' form respectively.

More broadly, the psychological-reality of phonological 'fit' between a verb and a regular past-tense schema (e.g., [VERB]-*t*) is the subject of an intense debate amongst psycholinguists. Researchers are divided as to whether the formation of 'regular' past-tense forms (e.g., *clicked, walked, covered*) is best attributed to a context-free rule (i.e., add "-*ed*" to any instance of the discrete category "verb"; e.g. Prasada & Pinker, 1993; Marcus et al, 1992) or to analogy across phonologically-similar regular past-tense forms in associative memory (e.g., *click/clicked* from *trick/tricked, kick/kicked*; e.g., Bybee & Moder, 1983; Rumelhart & McClelland, 1986). The notion of a context-free rule applied to the discrete category "VERB" is akin to the generativist assumption that language is represented as a set of phrase-structure rules that act on abstract syntactic categories (e.g., add "*–ed*"). Conversely, the notion of analogy across phonologically-similar past-tense forms is akin to the constructivist/*FIT* assumption that learners possess 'regular' past-tense schemas such as [VERB]-*t*, which are formed by analogising across phonological properties shared by verbs that have previously been used in this manner). Understanding which of these cognitively distinct mechanisms is responsible for the formation of 'regular' past-tense forms can provide valuable insight into the way language is represented and generalised.

Support for the analogy-based (as opposed to rule-based) account of regular past-tense formation has been offered by Albright and Hayes (2003). Adults were more likely to produce and favourably judge regular past-tense forms of novel verbs that were phonologically-similar to existing 'regular' English verbs. A similar finding was demonstrated in Ambridge's (2010) grammaticality-judgement study for 9-10-year-old children. However, for the 6-7-year-olds in Ambridge's study, acceptability judgements of novel verbs' 'regular' past-tense forms were not influenced by their phonologically-similarity to existing 'regular' English verbs. It is difficult to know whether this reflects 6-7-year-olds' less-adultlike memorization of regular verbs (which would impact on analogy-based generalizations), or difficulties with the judgment task. The fact that developmental evidence for an analogy-based account is limited to children aged 9-10 is problematic because children of this age are well past the peak rate of overgeneralization (e.g., Maratsos, 2000; Maslen, Theakston, Lieven & Tomasello, 2004), meaning that these data reflect the output of a relatively mature system, rather than tapping directly into acquisition. Addressing this shortcoming, *Study 2* uses an elicited production paradigm to investigate the role of analogy in children's (age 3-4, 5-6, 6-7 and 9-10) formation of English past-tense forms.

3. Children's acquisition of English past-tense: novel verb production data from young children (*Study 2*)

Text as it appears in:

Blything R.P., Ambridge B, Lieven EVM. Children's Acquisition of the English Past-Tense: Evidence for a Single Route Account from Novel Verb Production Data (*submitted to Cognitive Science; status: revise and resubmit*).

The presentation used for the manuscript has been altered slightly in the interest of consistency with other chapters of this thesis.

3.1. Abstract

The current study adjudicates between two opposing accounts of morphological productivity, using English past-tense as its test case. The single-route model (e.g., Bybee & Moder, 1983) posits that both regular and irregular past-tense forms are generated by analogy across stored exemplars in associative memory. In contrast, the dual-route model (e.g., Prasada & Pinker, 1993) posits that regular inflection requires use of a formal 'add -*ed'* rule that does not require analogy across regular past-tense forms. Children (aged 3-4; 5-6; 6-7; 9-10) saw animations of an animal performing a novel action described with a novel verb (e.g., *gezz; chake*). Past-tense forms of novel verbs were elicited by prompting the child to describe what the animal 'did yesterday'. Collapsing across age-group (since no interaction was observed), the likelihood of a verb being produced in regular past-tense form (e.g., *gezzed; chaked*) was positively associated with the verb's similarity to existing regular verbs, consistent with the single-route model only. Results indicate that children's acquisition of the English past-tense is best explained by a single-route analogical mechanism that does not incorporate a role for formal rules.

3.2. Introduction

A major debate amongst cognitive scientists is whether a speaker's ability to produce novel sentences and morphologically inflected forms should be attributed to symbolic rules that act over abstract categories (e.g., Chomsky, 1965; Prasada & Pinker, 1993) or to a mechanism that analogises across stored exemplars (e.g., Langacker, 2000; Rumelhart & McClelland, 1986). English past-tense morphology, which is characterised by its clear distinction between regular (e.g., *kick/kicked; play/played*) and irregular (e.g., *keep/kept; steal/stole*) patterns of inflection, provides a particularly suitable framework to adjudicate between these two approaches. Specifically, the emergence of 'overregularization' errors in children's speech (when regular inflections are applied to verbs that require irregular inflection; e.g. **keeped,* **stealed,* **hitted*) has given rise to two opposing accounts of children's morphological productivity: the *dual-route* model (e.g., Prasada & Pinker, 1993) which posits that overregularization reflects the use of a formal morphological rule (i.e., add '*–ed'*) and the *single-route* model (e.g., Bybee & Moder, 1983) which posits that overregularization reflects analogy across stored exemplars (e.g., *peel/peeled, heal/healed, steal/*stealed*). The current study uses an elicited production paradigm to investigate which of these assumptions best accounts for young children's morphological productivity. First though, it is necessary to outline each model's predictions regarding irregular and regular past-tense forms.

For irregular past-tense forms (e.g., *kept, stole*), the single- and dual-route models assume, in effect, identical mechanisms: Both accounts assume that these forms are stored in associative memory and used in analogical generalization. If an irregular form cannot be retrieved directly from memory (i.e., if the representation is weak or absent), it is generated by analogy to phonologically-similar stored forms (e.g., if the past-tense of *know* has not yet been learnt, *knew* may be generated by analogy to *blow/blew* and *throw/threw*). Thus, with regard to novel-verb studies, the single- and dual-route models make an identical prediction: the likelihood of a novel verb being produced in irregular past-tense form is positively associated with its phonological similarity to existing irregular verbs. Although this prediction is supported in elicitation studies with adults (e.g., Prasada & Pinker, 1993; Albright & Hayes, 2003), it is almost impossible to test with children, by virtue of the fact that they almost never offer irregular past-tense forms of novel verbs in such studies, due to priming of regulars (Ramscar, 2002) (though children do show this effect in a judgment task; Ambridge, 2010). In any case, since the single- and dual-route models make identical predictions with regard to the effect of similarity-to-irregulars, even if children did produce a sufficient number of irregulars to show this effect, it would not help to adjudicate between the two accounts.

To do so, we need to investigate a phenomenon for which the two accounts make different predictions: production of regular inflection. The single-route model (e.g., Bybee & Moder, 1983) holds that all morphological productivity – both regular and irregular – can be attributed to a single associative process (essentially the same process that the dual-route model assumes for irregulars only). Initially, an attempt is made to retrieve a verb's (regular or irregular) past-tense form directly from memory (e.g., *knew*). If retrieval fails, a past-tense form is generated by analogy to phonologically-similar verbs *regardless of their regularity* (e.g., *knew* may be generated by analogy with irregular verbs like *throw/threw* and *blow/blew* but **knowed* may be generated by analogy with regular verbs like *show/showed* and *glow/glowed*). Thus, the single-route model predicts that the likelihood of a novel verb being produced in regular past-tense form is positively associated with its phonological similarity to existing regular verbs.

Under the dual-route model (Prasada & Pinker, 1993) regular forms are generated via the application of a default rule that adds the suffix '*–ed*' to a verb's root-form (e.g., *know/*knowed*). This default rule steps in whenever an irregular form fails to be either retrieved directly from memory or generated by analogy with stored irregulars. Crucially, the 'add *–ed'* rule can be applied to any verb "regardless of anything else that might happen to the stem as a result of other rules or memorized associations" (Berent, Pinker & Shimron, 2002, p.463). Thus the dual-route model does not share the prediction of the single-route model that the likelihood of a novel verb being produced in regular past-tense form is positively associated with its phonological similarity to existing regular verbs. Rather, Prasada and Pinker (1993, p.22) predict that "the goodness of the suffixed past-tense forms does not decline as a function of distance from known suffixed forms". Indeed, Prasada and Pinker (1993) take this very finding (from their adult judgment study) as evidence for the dual-route model.

There have been attempts to modify the dual route model to allow for storage of at least some high-frequency regulars (e.g., Alegre & Gordon, 1999; Pinker & Ullman, 2001). In our view such modification renders the model empirically indistinguishable from the single-route model. We come back to this in the Discussion but for now, focus on the clearly contrasting predictions of each account. In summary, setting aside irregulars (for which the models share a prediction that is almost impossible to test in elicited production studies with children) and modified versions of the dual-route model (whose predictions are less straightforward), the contrasting prediction of the accounts is clear: The single-route model predicts a positive association between the likelihood of a novel verb being produced in regular-past-tense form and its phonological similarity to existing regular forms. The dual-route model does not. Note that it is generally agreed that this prediction must be tested with novel verbs (e.g., Albright & Hayes, 2003; Ambridge, 2010; Prasada & Pinker, 1993) which require the use of a generalization mechanism (i.e., either rules or analogy) rather than allowing for the retrieval of rote-learned forms from memory (assumed under both accounts).

Two previous novel-verb studies have challenged the claim of the dual-route model that regular inflection requires a context-free 'add –*ed*' rule. Albright and Hayes (2003) implemented a computational model that produced novel verbs by "concatenating combinations of relatively common syllable onsets and syllable rhymes" (p.135) in 4,253 English verbs. This process ensured that the phonological properties of novel verbs varied in the extent to which they were similar/ dissimilar to existing phonological neighbourhoods without violating the rules of English phonology (this avoids confounds encountered by a previous study (Prasada & Pinker, 1993) in which a verb like '*ploamphed*' was judged less favourably than '*plipped'* - not because of the former's greater phonological distance from existing regulars, but because it was less well-formed [i.e., *ploamph/plip*]). The model output four types of novel verbs that were similar to (i) regular verbs only (ii) irregular verbs only, (iii) regular verbs and irregular verbs, or (iv) neither regular nor irregular verbs. Consistent with predictions of both single-route and dual-route models, adult English speakers were more likely to produce and favourably judge irregular past-tense forms of novel verbs that were similar to existing irregulars. Crucially, participants were also more likely to produce and favourably judge regular past-tense forms of novel verbs that were similar to existing regulars, consistent with the predictions of the single-route but not the dual-route model.

Second, Ambridge (2010) replicated the judgment component of Albright and Hayes (2003) with children aged 6-7 and 9-10. For both ages, the acceptability of novel irregular past-tense forms increased as a function of the verb's similarity to existing irregular verbs (consistent with both models). However, consistent with the predictions of the single-route model only, acceptability of novel regular past-tense forms increased as a function of the verb's similarity to existing regular verbs (though for the older group only).

The novel verb findings of Albright and Hayes (2003) and Ambridge (2010), for adults and children respectively, constitute preliminary support for the single-route over the dual-route model. However, at present, this support must be considered tentative for four reasons. First, the populations which have shown the crucial effect of similarity to regulars (adults and 9-10 year olds) may well be subject to demand characteristics in both production and judgment studies. For example, they might assume that the experimenter does not intend for them to produce – or give maximum acceptability ratings to – regular forms on every trial, and so base their regular productions and judgments on analogy to known regular forms, even though they may not necessarily do so in normal circumstances. Second, evidence from children is limited to judgment data. This is problematic because overregularization errors are a production phenomenon, and providing acceptability judgments can be argued to be a metalinguistic task. Third, the 6-7 year olds studied by Ambridge (2010) failed to show a significant effect of similarity to regulars (despite an effect of similarity-to-irregulars). It is difficult to know whether this pattern reflects this age-group's relatively incomplete formation of phonological neighbourhoods for regular verbs (which generally have lower token frequency than irregular verbs), or difficulties with the judgment task. Note that such difficulties need not necessarily affect regular and irregular forms equally (e.g., perhaps some children simply gave higher ratings to regulars across the board). Finally, developmental data are limited to children 6-7 and older (presumably because it is difficult to obtain meaningful graded judgments from younger children; e.g., Blything, Ambridge & Lieven, 2014). This is problematic because children of this age are well past the peak rate of overgeneralization (e.g., Maratsos, 2000; Maslen, Theakston, Lieven & Tomasello, 2004), meaning that these data reflect the output of a relatively mature system, rather than tapping directly into acquisition.

Addressing these shortcomings, the current study employed an elicited production paradigm using novel verbs (the core-set from Albright & Hayes, 2003)[6] to investigate the generalization mechanisms underlying 3-4, 5-6, 6-7, and 9-10 year olds' productivity with English past-tense. Novel verbs were elicited using a sentence-completion task (e.g., *The duck likes to bredge. Look, there he is bredging. Everyday he bredges. So yesterday he...*). Although, in principle, both the single- and dual-route model predict a positive association between the likelihood of a novel verb being produced in irregular-past-tense form and its phonological-similarity to existing irregular forms, it is likely that children will not produce a sufficient number of irregular forms to allow this prediction to be tested. Thus the focus of the present study is on testing the crucial contrasting predictions of the two accounts: The single-route model predicts a positive association between the likelihood of

[6] Albright and Hayes demonstrated that adults' phonological ratings of stem-forms were poorly correlated with ratings of past-tense forms (r=0.006), thus indicating that their design successfully minimised any possible of phonological confounds in the stimuli.

a novel verb being produced in regular-past-tense form and its phonological similarity to existing regular forms; the dual-route model predicts no such effect.

3.3. Method

3.3.1. Participants. Eighteen children were recruited from each age group (3-4 [M= 3;8], 5-6 [M=5;9], 6-7 [M=7;3], and 9-10 [M=10;4] year olds). An additional three children from the youngest age group were excluded because they did not comply with procedure. All participants were monolingual English speakers and had not been diagnosed with any language impairment. All participants were recruited from schools or nurseries in Manchester and testing occurred at those locations in a private room.

3.3.2. Materials. The study used 40 novel verbs, all sourced from Albright and Hayes' (2003) computational model. This model generated 2,344 'candidate forms' (e.g., *bize, rife, chool, spling*) by "concatenating combinations of relatively common syllable onsets and syllable rhymes" (p.135) in 4,253 English stem/ past-tense pairs (4035 regular; 218 irregular)[1]. The model produced judgment ratings for each candidate's regular (e.g., *bized, rifed, chooled, splinged*) and irregular past-tense form (e.g., *boze, rofe, chole, splung*) and a 'core-set' of 40 verbs were chosen that varied orthogonally and continuously according to their phonological-similarity to existing English (i) regular and (ii) irregular verbs. Phonological-similarity scores were based on judgment ratings of each verb provided by Albright and Hayes' computational model. Note that Albright and Hayes implemented two computational models (thus producing two sets of similarity scores): a *multiple-rules* model, and an *analogical* model. Similarity scores used in the current study were sourced from the analogical model which, according to Albright and Hayes (2003), most closely implements the notion of analogy relevant for the theoretical accounts under investigation here (i.e., the dual-route and single-route models).

3.3.3. Design. Each participant was presented with 40 verbs in a within-subjects design. To combat fatigue effects, verbs were equally divided into Set 'A' and 'B', to be completed on different days. Each verb-set contained an equal number of verbs that were similar to existing *Irregulars-Only, Regulars-Only, Regulars-and-Irregulars* or *Neither-Regulars-nor-Irregulars* (classification based on judgment ratings from Albright and Hayes' model – see Table A3.1 in *Appendix 6* for classification). Children were assigned to one of two counterbalanced groups which ensured that half were first exposed to Verb-Set A (followed by B on a different day) and the remaining half were first exposed to Verb-Set B (followed by A on a different day).

The (trial-level) outcome variable was whether the novel verb was produced in (a) regular form (e.g., *bized*) or (b) irregular form (e.g., *boze*). The predictor variables were (a) phonological-similarity to existing regular verbs, (b) phonological-similarity to existing irregular verbs – both continuous predictors – and (c) participant age group (3-4, 5-6, 6-7, 9-10 years).

3.3.4. Procedure. Children were told that the experimenter would describe cartoons on the laptop and that they should 'fill-in-the-blanks' when the experimenter stopped talking. This format was designed to elicit plural forms of nouns in the practice session and past-tense forms of verbs in the test session.

Practice Session. For each practice trial, the experimenter described the first image of a single item (e.g., "*Here is one mouse*") and then began to describe a second image that depicted more than one instance of that item ("*and now there are three...*"). In order to ensure that the child was comfortable with producing both regular and irregular responses, two of the four practice trials used nouns with regular plural forms (*shoe/shoes; car/cars*) and the remaining two trials displayed images of nouns with 'irregular' plural forms (*man/men; mouse/mice*). Overregularization errors (e.g., "*mouses*") were corrected by the experimenter.

Test Session. The child was told that she would see cartoon videos of animals 'doing some funny things.' Forty video animations were created using *Anime Studio Pro 6*, each featuring one of four animals (*duck, bunny, frog, or bear*) performing a novel intransitive action. Animations were played using *Apple QuickTime*. Each trial presented a verb in bare-stem (non-finite), present-progressive (*-ing*), and simple present-tense form (-s), using the following template: "*The duck/bear/frog/bunny likes to VERB. Look, there he is VERBing. Everyday he VERBs. So yesterday he...*". Animation-verb pairings were random, and different for each participant, in order to control for any possible semantic effects on the use of regular versus irregular past-tense forms (e.g., Ramscar, 2002). Verbs were presented in pseudo-random order, different for each participant, with the constraint that no more than two consecutive trials featured a verb from the same 'island of reliability' (using Albright and Hayes' classification, see *Table A3.1 in Appendix 6*.

Responses were coded according to whether the response was a (i) Regular Form (e.g., *bized*) (ii) Irregular Form[7] (e.g., *boze*), or (iii) Other (i.e., *No-change* [e.g., *bize*]; *Third-person-present* [e.g. *bizes*]; *Past-progressive* [e.g., *bizing*]). Responses that could not be classified into any of these categories were excluded (the percentage of excluded trials for 3-4, 5-6, 6-7 and 9-10 year olds was 16.9%, 8.93%, 4.03%, and 7.65% respectively).

3.4. Results

***Table 3.1*. Mean percentage [SE in brackets] of Regular, Irregular, No Change, 3rd Person Present, and Past Progressive Forms produced by each age-group**

	Regular (vs all)	Irregular (vs all)	No Change (vs all)	3rd Person Present (vs all)	Past Progressive (vs all)
3-4 Years	33.39 (1.95)	4.62 (0.87)	8.9 (1.18)	44.35 (2.06)	8.73 (1.17)
5-6 Years	41.69 (1.93)	2.15 (0.57)	2.62 (0.63)	51.69 (1.96)	1.85 (0.53)
6-7 Years	86.38 (1.31)	4.06 (0.75)	1.16 (0.41)	2.17 (0.56)	6.21 (0.92)
9-10 Years	88.79 (1.23)	6.21 (0.94)	1.36 (0.45)	0.61 (0.3)	3.03 (0.67)
All Ages	63.78 (0.95)	4.26 (0.4)	3.33 (0.35)	23.76 (0.84)	4.88 (0.42)

Table 3.1 shows the mean percentage of (i) *Regular*, (ii) *Irregular*, (iii) *No-change*, (iv) *Third-person-present*, and (v) *Past-progressive* forms produced by each age-group. Analysis was split into two parts: one that used proportion of irregular forms (vs. regular, no-change, third-person-present, and past-progressive forms) as its outcome measure (to test for the effect of *similarity to existing irregulars*) and another that used proportion of

[7] Irregular forms of novel verbs were classified based on the irregular form(s) that received the most favourable judgment score(s) from Albright and Hayes's (2003) computational model. As a consequence of this selection criterion, the irregular form of five verbs was homophonous with the stem or 'root' form (e.g., *glit/glit*) on analogy with no-change forms (e.g., *hit/hit*). Thus, for these five verbs, such forms, when produced were coded as 'irregular'. In the case of some verbs, the model gave a favourable judgment to more than one irregular form of the verb; for such cases, only the form that was most frequently produced (by participants in the current study) was coded as 'irregular' and the form that was produced less frequently was excluded (note this strategy led to the exclusion of just 3 trials). See *Appendix 6* for a full list of regular and irregular forms.

regular verbs (vs. irregular, no-change, third-person-present, and past-progressive forms) as its outcome measure (to test for the effect of *similarity to existing regulars*).

***Figure. 3.1*. Mean proportion of (i) verbs that received irregular inflection as a function of their similarity to existing irregular verbs [on the left], and (ii) verbs that received regular inflection as a function of their similarity to existing regulars [on the right]**

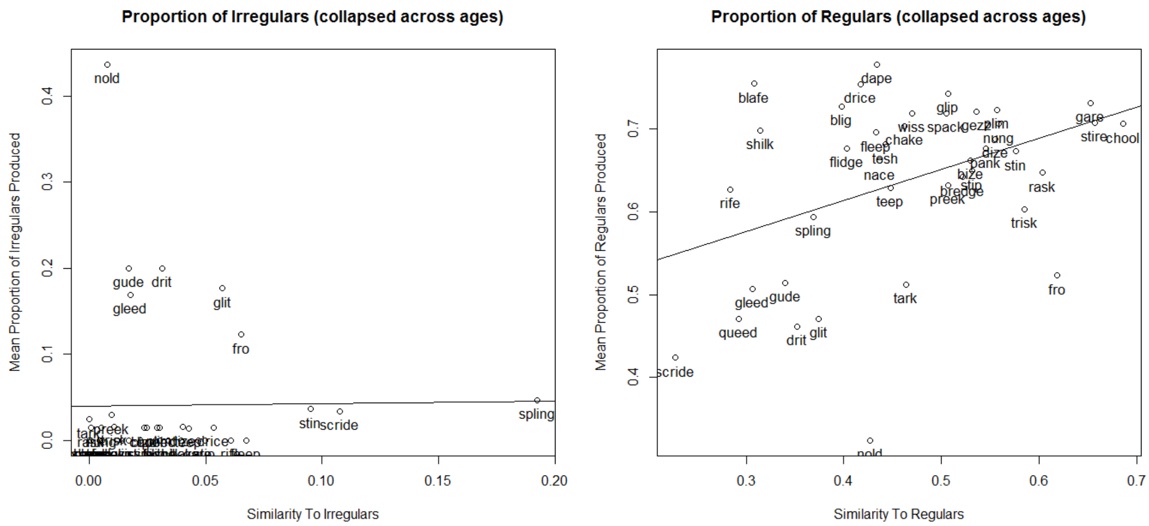


Figure 3.1 plots the rate of (left-hand panel) irregular and (right-hand panel) regular inflection as a function of similarity to existing irregulars and regulars respectively (collapsing across age). As expected, the rate of irregular inflection for all but handful of novel verbs was at or close to zero, with the consequence that we cannot meaningfully test the shared prediction of the two accounts with regard to irregular inflection. On the other hand, visual inspection of Figure 3.1 suggests that, as predicted by the single-route model only, the likelihood of a verb being produced in regular-past-tense form is positively associated with its phonological-similarity to existing regular verbs, as predicted by the single-route model only.

To investigate whether this trend was statistically-significant, results were analysed using binomial linear mixed-effects models (*glmer* from package *lme4;* Bates, Maechler, Bolker & Walker, 2015) in the R environment (R Development Core Team, 2011). The main advantages of Linear Mixed-Effects Models (LMEMs) are that they predict individual trials rather than average over trials[8], are robust against missing data and treat both participants and items as random effects (Baayen, Davidson, & Bates, 2008).

[8] This underlines the point that *Figure 3.1*, which plots *mean* production scores, should be taken as only a rough indication of the relationship between variables. Since LMEMs predict individual trials rather than averaging over items or participants, a more meaningful interpretation of the

All LMEMs were fitted with random intercepts for participants and verbs and, where applicable, by-participant and by-verb random slopes. Random-intercepts offer the benefit of removing idiosyncratic variation within each random factor (i.e., participants and verbs), whereas random-slopes control for the possibility that treatment effects may vary within each random factor.

In accordance with recommendations outlined by Barr, Levy, Scheepers and Tily (2013), the inclusion of each random slope was determined by whether its addition to a random-intercept only model improved the model's fit of the data (using likelihood ratio tests and anti-conservative alpha values of $p<0.25$) (slopes that satisfied this criterion were added one-by-one until the model failed convergence). Results were analysed using binomial linear mixed effects models (*glmer* from package *lme4*; Bates, Maechler & Bolker, 2011) in the *R* environment (R Development Core Team, 2013). Mixed-effects models predict individual trials rather than averaging over trials, and offer the added benefit of treating both participant and item as random effects (i.e., the model creates an intercept for each participant and each item, thus removing variation within each of these factors). They are also robust against missing data (Baayen, Davidson & Bates, 2008).

All models included as fixed-effects a measure of *similarity-to-regulars* and a measure of *similarity-to-irregulars*, thus ensuring that any effect of *similarity-to-regulars* on the likelihood of a verb's regular inflection was over and above any effect of *similarity-to-irregulars* (and that the *similarity-to-regulars* measure was not a proxy measure of a verb's dissimilarity to existing irregular verbs).

3.4.1. Analysis 1: Production of Irregular Forms

The first analysis used as its outcome variable *production of irregular forms (vs. all other forms)*. Fixed-effects were entered into the model simultaneously (thus, the order in which predictors are listed is arbitrary): (a) *age* (centred to stabilise beta terms of the model), (b) *similarity to existing regulars* (to control for trade-off effects), (c) *age: similarity to existing irregulars,* and (d) *similarity to existing irregulars*. By-participant random slopes were included for *similarity-to-regulars* and *similarity-to-irregulars*, and by-verb slopes included for *Age*. The LMME output is shown in *Table 3.2*. The effects of *similarity to existing regulars* and *similarity to existing irregulars* did not reach

relationship between variables is to look at the β values. A positive β value indicates a positive correlation between the predictor and the likelihood of a verb being produced in the relevant form; a negative β value indicates a negative correlation between the predictor and the likelihood of a verb being produced in the relevant form.

significance: the likelihood of a verb being produced in irregular past-tense form was not influenced by its similarity to existing regular or irregular verbs. Neither the main effect of *Age*, nor the interaction of *Age* by *similarity to existing irregulars* was significant; thus we did not conduct separate analyses for each age group. The lack of significant findings for this analysis was expected given that very few irregular forms were produced (see *Figure 3.1*) and there was thus insufficient variance between data points to observe any underlying effects.

***Table 3.2*. Mixed-effects models fitted to production probability of irregular forms**

Fixed effects	*Beta (B)*	*SE*	*z*	*2.50%*	*97.50%*	*p*
(Intercept)	-2.52	2.20	-1.15	-6.83	1.79	0.252
Age	0.14	0.16	0.83	-0.19	0.46	0.406
Similarity to Regulars	-8.60	4.48	-1.92	-17.37	0.18	0.055
Age*Similarity to Irregulars	0.70	2.68	0.26	-4.55	5.95	0.794
Similarity to Irregulars	-6.16	15.57	-0.40	-36.68	24.37	0.693

3.4.2. Analysis 2: Production of Regular Forms

The second analysis used as its outcome variable *production of regular forms (vs. all other forms)*. Fixed-effects were entered into the model simultaneously: (a) *age* (centred to stabilise beta terms of the model), (b) *similarity to existing irregulars* (to control for trade-off effects), (c) *age: similarity to existing regulars*, and (d) *similarity to existing regulars*. By-participant random slopes were included for *similarity-to-regulars* and *similarity-to-irregulars*, and by-verb slopes included for *Age*.

The LMME output is shown in *Table 3.3*. A trade-off effect of *similarity to existing irregulars* (in the predicted direction) did not reach significance (p=0.3). Crucially, and in support of the single-route model only, a main effect of *similarity to existing regulars* was observed: the likelihood of a verb being produced in regular past-tense form was positively associated with its phonological-similarity to existing regular verbs. A main effect of *Age* was observed such that the likelihood of a verb being produced in regular-past-tense form increased with participants' age; inspection of *Table 3.1* indicates this can be attributed to the fact that 3-4 and 5-6 year-olds produced a much greater proportion of *third-person-present* forms (mean = 44.35% and 51.69% respectively) than 6-7 and 9-10 year olds (mean = 2.17% and 0.61% respectively), thus diluting the percentage of regular-past-tense forms produced by younger children. The effect of *Age* did not interact with the effect of *similarity to existing regulars*; thus we did not conduct separate analyses for each age group.

***Table 3.3*. Mixed-effects models fitted to production probability of regular forms**

Fixed effects	*Beta (B)*	*SE*	*z*	*2.50%*	*97.50%*	*p*
(Intercept)	-0.50	0.84	-0.60	-2.14	1.14	0.549
Age	**0.61**	**0.22**	**2.80**	**0.18**	**1.03**	**0.005**
Similarity to Irregulars	-4.98	4.78	-1.04	-14.35	4.39	0.298
Age*Similarity to Regulars	0.70	0.46	1.53	-0.20	1.59	0.127
Similarity to Regulars	**4.11**	**1.66**	**2.47**	**0.85**	**7.37**	**0.013**

3.5. Discussion

Recent judgment and production studies with novel verbs (e.g., Albright & Hayes, 2003; Ambridge, 2010) have provided some evidence that morphological productivity, specifically for the English past-tense, can be explained by a single-route analogical process (e.g., Langacker, 2000; Bybee & Modor, 1983), without the need for a context-free default rule (e.g., Pinker, 1999; Prasada & Pinker, 1993). However, developmental data from studies that have used novel verbs have been limited to judgment data from children aged 6-7 and 9-10 (Ambridge, 2010). The current study used a production task (which is presumably less demanding, as well as enjoying greater ecological validity) to investigate whether evidence for the single-route model extends to younger children (age 3-4; 5-6; 6-7; 9-10); a population who are still in the early stages of morphological acquisition and whose data are perhaps most crucial for any theory of morphological productivity to explain. Past-tense forms were elicited by showing children an animation of an animal performing a novel action described with a novel verb (e.g., *gezz; chake*) and prompting the child to describe what the animal 'did yesterday'. Collapsing across age, the likelihood of a novel verb being produced in regular past-tense form was positively associated with its phonological-similarity to existing regular verbs. This effect did not interact with age, indicating that children as young as 3-4 are in use of a morphological system that generates regular past-tense forms by analogising across stored exemplars of existing regular verbs.

The results extend previous support (e.g., Ambridge, 2010) for the single-route model (e.g., Bybee & Moder, 1983) to a younger population, and are inconsistent with the dual-route model (e.g., Prasada & Pinker, 1993). Specifically, the single-route model assumes that regular past-tense forms are generated by analogy across stored exemplars of regular verbs (i.e., the same memory-based associative process that is posited by both models to be responsible for generation of irregular past-tense forms). In contrast, the dual-route model holds that regular past-tense forms are generated by a separate 'add *–ed*' rule, yielding the prediction that regular inflection is not influenced by a verb's phonological-similarity to existing regular verbs. Recall that Prasada and Pinker (p.22) took as evidence for the model their finding that "the goodness of the suffixed past-tense forms does not decline as a function of distance from known suffixed forms". Crucially then, only the single-route model can explain the current study's finding that children were more likely to produce a verb with regular inflection when the verb was phonologically-similar to existing regular verbs. This finding has important developmental implications for the past-tense debate because it is the first time a novel verb paradigm (the most stringent test of the models in question) has been used to demonstrate that overregularization errors made by children as young as 3-4 need not be attributed to a rule-based mechanism (as under the dual-route model), and, indeed, are better explained in terms of analogy across stored exemplars (as under the single route model).

A possible objection to this conclusion is that while the present finding of an effect of similarity to regulars on regular inflection is inconsistent with the original "strong" version of the dual-route model (Prasada & Pinker, 1993), it is not necessarily inconsistent with more recent versions that do allow for some regular storage. For example, Pinker and Ullman (2002, p.458) state that the model "does not posit that they [i.e., regular past-tense forms –RB] are *never* stored, only that they do not *have* to be", thus leaving open the possibility for a modified version of the model that stores and analogises across regular past-tense forms (e.g., Alegre & Gordon, 1999; Hartshorne & Ullman, 2006). In this case, regular past-tense forms of novel verbs could be generated by rules *or* analogy. Note that such a model is almost impossible to distinguish empirically from the single-route model, because – in effect – it *is* the single-route model (i.e., an analogical system that operates on regulars and irregulars alike) with a default rule bolted on (and no clear specification of when this rule should be used instead of analogy).

However, one possible way to distinguish between the single-route model and the regular-storage dual-route model is as follows. Since, under the latter, irregular past-tense forms are *always* generated by analogy, while regular forms are *at least sometimes* generated by the default rule instead, the regular-storage dual-route model would seem to predict that the effect of similarity-to-irregulars on irregular production will be larger than the effect of similarity-to-regulars on regular production. Testing this prediction is beyond the scope of the current study, given the floor effect observed for irregular production. Although the child judgment data reported by Ambridge (2010) are consistent with this possibility, the adult judgment data reported by Albright and Hayes (2003) are not, with partial *r* values of 0.49 and 0.58 for the effects of similarity to irregulars and regulars respectively. While further research should attempt to further elucidate the relative effect sizes of irregular and regular analogy, it is already clear that the only way to reconcile the dual-route model with the findings of the present study (and those of Albright & Hayes, 2003, and Ambridge, 2010) is to have it adopt the core mechanism of the single-route model, making the two almost indistinguishable.

Can a rule-based account of morphological productivity be salvaged? Of most promise, the multiple-rules model (Albright & Hayes, 2003) posits multiple 'micro-rules' that register the phonological properties of every encountered verb stem and its past-tense form, regardless of the verb's regularity. A process known as *minimal generalization* recognises shared phonological features amongst micro-rules and retains these shared segments within new, more general rules. Each of these rules is assigned a graded confidence value according to how accurately it yields a correct prediction, and rules with high values are more likely to be used again. Albright and Hayes compared their multiple rules model against their 'analogy' model (i.e., the analogy model that provided similarity-ratings for the present study), finding that the multiple-rules model predicted adult judgment data more accurately. However, the analogy model used by Albright and Hayes was, in the authors' words, a 'more naive model of analogy' (p.122) such that it worked by finding phonological-similarities indiscriminately across all parts of a word: it was not biased towards computing the analogy across the end of the word as would be the case with more sophisticated analogy models. For example, the past-tense form of *string* is *strung* based on the fact it shares the same *final* rhyme as verbs like *fling/flung* and *cling/clung* - the analogy-based model used by Albright and Hayes would be somewhat misled by paying equal attention to verbs that share the same beginning as *string*, e.g., *start/started*, thus reducing the accuracy of its similarity measure). In contrast, the multiple-rules model works on the basis of finding 'structured similarities' where a micro-rule operated only if verbs shared an identical phonological pattern, thereby sidestepping the problem of unrestricted analogy (e.g., *fling*, *cling* and *string* all end in *ing* and thus the micro-rule would be based on this similarity alone). Since more sophisticated instantiations of analogy-based models could be biased or restricted to attend to the same similarities that the micro-rules were sensitive to, it is best to reserve judgement with regard to whether data is more accurately predicted by analogy-based or rule-based models.

Whilst the multiple-rules model may redeem rule-based accounts, one can argue that its adoption of probabilistic outcomes and graded rules that strengthen with experience risk making it empirically indistinguishable from analogy models. McClelland and Patterson (2002, p.471) state that a multiple-rules model "characterizes an underlyingly connectionist processing system at a higher level of analysis, with rules providing descriptive summaries of the regularities captured in the network's connections". Moreover, the psychological principles underlying the multiple-rules model are unclear, since – unlike is sometimes assumed for the defaulting mechanism of the dual-route model - it is implausible for a system of tens or hundreds of micro-rules to be innate. But, as McClelland and Patterson (2002) note, the alternative – that they are learned and probabilistic – renders these multiple rules nothing more than descriptions of the analogical process hypothesized under a single-route model.

To sum up, the current study has allowed us to take another step closer to understanding the mechanism that underlies morphological productivity. Elicited production data from children aged 3-4, 5-6, 6-7 and 9-10 revealed that the likelihood of a novel verb being produced in regular-past-tense form was positively associated with its phonological-similarity to existing regular verbs. These data are inconsistent with the assumption of the dual-route model of a context-free rule that "can apply regardless of anything else that might happen to the stem as a result of other rules or memorized associations" (Berent, Pinker & Shimron, 2002, p.463). Rather, it would seem that any account of the developing morphological system must incorporate the assumption that language learners store, and compute phonological analyses across, regular past-tense forms.

3.6. Prelude to Study 3

Taken together, *Study 1* and *Study 2* demonstrate that children as young as 3-to-4-years probabilistically restrict generalizations of verbs based on the frequency of other formulations that convey the intended message (pre-emption), the verb's frequency in other contexts (entrenchment) and the extent to which the properties of the verb overlap with properties of the construction ('fit'). Thus, it appears that children are in use of an interactive, probabilistic learning mechanism that yields all three of these effects from an early age.

However, evidence for this position (from the current thesis and at least the majority of previous literature that has investigated older children and adults) has been limited to behavioural paradigms such as production and judgment methods. The problem here is that behavioural paradigms rely on overt responses, which (i) contribute only indirect information about the processing of a given linguistic stimulus, and (ii) place extra demands on attention and working memory, which (a) may encourage processing strategies that are not true reflections of everyday language use, and (b) render the task unsuitable for children younger than around 3;0, thus neglecting younger children whose data may provide further insight into the initial stages of linguistic generalization (e.g., the possibility that one mechanism emerges before another).

Thus, a profitable direction for future research is to develop online and implicit measures of the hypothesised processes (pre-emption, entrenchment and 'fit'), which have lower task demands relative to behavioural measures. *Study 3* investigates the suitability of the Event-Related-Potentials (ERP) paradigm for this goal. ERP measures patterned voltage changes on the scalp that are directly associated with the brain's response to a particular experimental manipulation (e.g., a grammatical violation), and can thus offer a relatively direct observation of cognitive activity underlying the processing of overgeneralization errors. An ERP waveform typically displays a number of positive and negative peaks at specific latencies and scalp-distributions; if a peak is known to reliably co-occur with a particular event, it is recognised as an electrophysiological marker of a cognitive, sensory or motor process associated with that event. For example, the *P600* component reliably co-occurs with syntactic violations, and gets its name because it is a *positive*-going component that often peaks around *600* milliseconds.

Study 3 investigates whether ERP components that are known to mark overgeneralization errors (e.g., P600) are sensitive to statistical restriction mechanisms (i.e., pre-emption/ entrenchment) that have previously been detected only by behavioural paradigms. The study uses transitive argument structure overgeneralization errors as a test-case, because these types of errors have most robustly yielded effects of verb-frequency in behavioural studies, such that errors with high-frequency verbs (e.g., **The clown laughed the boy*) are rated as more unacceptable than errors with low-frequency verbs (**The clown giggled the boy*) (e.g., Ambridge, Pine, Rowland & Young, 2008; Brooks & Tomasello, 1999; Brooks, Tomasello, Dodson & Lewis, 1999; Theakston, 2004). The study tests a sample of healthy adults (since previous ERP literature into the processing of grammatical violations has focussed on adults, and the study is thus more informed by that literature), with a view to applying the ERP paradigm to other test-groups such as children in the future. In doing so, our aim is to investigate the suitability of ERP for detecting fine-grained differences in the relative acceptability of overgeneralization errors.

4. Using Event-Related Potentials to Investigate the Retreat from Overgeneralization Error (*Study 3*)

Text as it appears in:

Blything R.P., Ambridge B, Lieven EVM. Using Event Related Potentials (ERP) to investigate the effect of verb frequency on processing verb-argument structure violations (*submitted to Language, Cognition and Neuroscience*).

The presentation used for the manuscript has been altered slightly in the interest of consistency with other chapters of this thesis.

4.1. Abstract

A crucial component of language acquisition is the ability to generalize verbs into new verb-argument structures. However, these generalizations must also be restricted to avoid some verbs being 'overgeneralized' into infelicitous verb-argument structures (e.g., **It always sweats me* ['*It always makes me sweat*']). The primary aim of the current study was to develop an Event-Related-Potentials (ERP) paradigm for testing the statistical restriction mechanisms posited to guide the retreat from verb-argument structure overgeneralization errors. The paradigm provides a valuable alternative to behavioural methods because it provides an online and implicit measure of sentence-processing with low task demands. The current study therefore measured ERPs in healthy adults as they were presented with transitive verb-argument structure overgeneralization errors containing high frequency verbs (e.g., *On Monday Bob <u>laughed</u> the girl in the garden*) or low frequency verbs (e.g., *On Monday Bob <u>giggled</u> the girl in the garden*). As expected, ERP components that typically mark syntactic violations (Left-Anterior-Negativity and P600) were observed in response to these overgeneralization errors. However, the magnitude of these components was not influenced by the manipulation of verb-frequency, raising doubts regarding the suitability of ERP for detecting fine-grained differences in the relative acceptability of errors of this type.

Keywords: Verb-argument structure overgeneralization; Event Related Potentials; Statistical-learning; P600; Left anterior negativity (LAN);

4.2. Introduction

The ability to produce novel utterances is one of the most important attributes of human communication. Underpinning this ability is a mechanism that allows verbs to be generalized from one verb-argument structure to another, but which also must be restricted to avoid ungrammatical utterances. For example, when verbs are experienced in the intransitive-inchoative (e.g., *The ball rolled*) and the transitive-causative (e.g., *The man rolled the ball*), an abstract generalization may be formed (i.e., [NOUN PHRASE 1] [VERB] → [NOUN PHRASE 2] [VERB] [NOUN PHRASE1]) that allows any verb to be generalized in the same way (e.g., *The ball rolled → The boy rolled the ball*).[9] However, intransitive-only verbs cannot be used with a direct-object, and appropriate restriction mechanisms must prevent such verbs from being 'over-generalized' to transitive-causative verb-argument structures (e.g., **The man fell the boy*). The present study develops the ERP paradigm to investigate whether it is suitable for investigating restriction mechanisms that have previously been studied using only behavioural paradigms.

The account under investigation, known broadly as *statistical-learning* (e.g., Braine & Brooks, 1995; Goldberg, 1995), posits that verb-argument structure overgeneralizations are probabilistically blocked by the repeated presentation of a relevant verb in another construction. To understand the mechanism posited to underlie *statistical-learning*, one must refer to two prominent accounts: *pre-emption* and *entrenchment*. *Pre-emption* (e.g., Clark & Clark, 1979; Goldberg, 1995) holds that learning is sensitive to the frequency of the verb in only near-synonymous constructions. For example, a transitive-causative overgeneralization (e.g., **The magician disappeared the rabbit*) is posited to be blocked by use of the relevant verb in a periphrastic-causative construction (e.g., *The magician made the rabbit disappear*), but not by use of the verb in less-synonymous constructions such as the intransitive-inchoative (e.g., *The rabbit disappeared*). *Entrenchment* (e.g., Braine & Brooks, 1995) holds that learning is sensitive to *any* use of the verb, regardless of the construction (e.g., *The rabbit disappeared; The magician made the rabbit disappear; Disappear!*). Although *pre-emption* and *entrenchment* differ with regard to the constructions to which the learning mechanism is sensitive, the present study does not differentiate between the two approaches. Rather, the aim is to examine the central

[9] In this example, a lexical rule generates transitive-causatives from intransitive-inchoatives (and vice-versa) once it becomes evident that verbs can alternate between the two argument structures. Another possibility is that each construction is learned independently as a 'surface generalization', and that verbs are used in constructions that serve a similar functional purpose (e.g.,Goldberg, 2002). The problem of overgeneralization arises no matter which of these theories is assumed.

prediction of a generalized account of statistical-learning (that collapses across *pre-emption* and *entrenchment*): namely that high frequency verbs will be more resistant to overgeneralization than low-frequency verbs.

Evidence for the psychological reality of statistical restriction mechanisms has been offered by behavioural studies which have demonstrated that overgeneralization errors containing high-frequency verbs (e.g., **The magician disappeared the rabbit*) receive less favourable grammaticality judgment ratings, and are less likely to be produced, than overgeneralizations containing low-frequency verbs (e.g., **The magician vanished the rabbit*) (e.g., Ambridge, Pine, Rowland & Young, 2008; Bidgood, Ambridge, Pine & Rowland, 2014; Blything, Ambridge & Lieven, 2014; Brooks, Tomasello, Dodson & Lewis, 1999; Brooks & Zizak, 2002; Robenalt & Goldberg, 2015; Theakston, 2004; Wonnacott, Newport & Tanenhaus, 2008). However, evidence does not extend beyond the behavioural paradigm, which is limited by its reliance on overt responses to linguistic stimuli, and thus contributes only indirect information about the processing of a linguistic stimulus. Moreover, behavioural paradigms can be problematic for important test groups such as young children, early second-language learners, and brain-damaged patients (e.g., aphasics) because the extra demands placed on their attention and working memory may encourage processing strategies that are not true reflections of everyday language use.

To address these limitations and to complement existing behavioural evidence, there is a need to develop a new dependent measure for investigating restriction mechanisms, which monitors the processing of overgeneralizations online without necessitating a behavioural response. For this reason, the current study uses Event-Related Potentials (ERPs), so-called because they are a measure of electric potentials associated with a particular event (e.g., a syntactic violation). In a typical experiment, scalp-recorded ERPs are time-locked to the onset of critical stimuli (e.g., the precise time that a sentence becomes ungrammatical) and monitored with high temporal-accuracy over a defined time-window. An averaged waveform typically displays a number of positive and negative peaks at specific latencies and scalp-distributions; if a peak is known to reliably co-occur with a particular event, it is recognised as an electrophysiological marker of a cognitive, sensory or motor process associated with that event. For example, the *P600* component reliably co-occurs with syntactic violations and gets its name because it is a *positive*-going component that often peaks around *600* milliseconds.

Before reviewing ERP studies that are most relevant to the current study, it is important to outline the three major language-related ERP components: N400, Left-Anterior-Negativity (LAN) and P600. The N400 component is a centro-parietally distributed negative-going wave that peaks around 400 ms after word-onset. It was first described at sentence-level when Kutas and Hillyard (1980) demonstrated that semantically inappropriate words elicited a more negative N400 compared to semantically appropriate words (e.g., *I take coffee with cream and dog vs. He returned the book to the library;* see also DeLong, Urbach, & Kutas, 2005; Kutas & Federmeier, 2011; Kutas & Hillyard, 1984). It is now well-established that the magnitude of the N400 effect marks the difficulty with which a given lexical-element can be integrated into a preceding semantic context (e.g. DeLong et al., 2005; Wlotko & Federmeier, 2012).

Syntactic manipulations are associated with two ERP components: the LAN and P600. The LAN is a negative-going wave that occurs in a similar time-window as N400 (i.e., 300 to 500 ms) but is observed primarily at left-anterior electrode sites. The LAN has been observed for outright (morpho-)syntactic violations such as violations of local phrase-structure (e.g., **The broker hoped to sell the stock was sent to jail;* Osterhout & Holcomb, 1992) and agreement (e.g., **The plane took we to paradise;* Coulson, King & Kutas, 1998). The LAN is typically followed by the P600: a late-positive wave that occurs between 500 and 1000 ms post-stimulus onset and is maximal at centro-parietal sites. Unlike LAN, P600 also occurs with dispreferred (i.e., 'garden-path') or complex syntactic continuations that are grammatical (e.g., Kaan, Harris, Gibson & Holcomb, 2000; Mecklinger, 1995; Osterhout & Holcomb, 1992). Thus, LAN is posited to reflect online processing of outright syntactic violations whereas P600 is posited to reflect integration difficulties arising from any surprising or salient sentence continuations (e.g., Kotz & Friederici, 2003).[10]

A number of adult ERP studies have also demonstrated language-related ERP components to occur in response to verb-argument structure overgeneralizations (compared to well-formed sentences). The consistent finding amongst these studies is a negativity-positivity ERP pattern whereby a negative component (LAN or N400) is followed by a P600. In Dutch, Hagoort and Brown (2000) reported a LAN-plus-P600 pattern when intransitive-only verbs were used with a post-verbal object NP (e.g., **De*

[10] Of further note, there is interesting debate regarding the precise cognitive mechanisms underlying these language-related components, with some researchers arguing that they index domain-general purposes such as working memory load (in the case of LAN) and subjective salience (in the case of P600) as opposed to serving language-specific purposes (e.g., Coulson et al., 1998; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Sassenhagen, Schlesewsky & Bornkessel-Schlesewsky, 2014). Nevertheless, these components are universally regarded as valuable tools for testing linguistic theory because they have been demonstrated to systematically co-vary with specific syntactic or semantic manipulations, and can thus be regarded as reliable *markers* of syntactic or semantic processing (even if the precise cognitive events underlying the components have not been identified).

zoon van de rijke industrieel pocht *de auto* *van zijn vader [*The son of the rich industrialist boasts* *the car* *of his father]*). Here, the LAN can be taken to reflect the fact that a NP could not be assigned a syntactic role (i.e., direct-object) by the verb, and the P600 to reflect integration difficulties that follow the detection of this syntactic anomaly. Similarly, Rosler, Putz, Friederici and Hahne (1993) reported a LAN when German passive sentences containing intransitive verbs that could not passivize (e.g., **Der Mann wurde gefallen [*The man was fallen down]*) were compared to well-formed passives containing transitive verbs (*Der Mann wurde begrüßt [The man was greeted]*). A small positivity was also visible in the P600 time-region, but did not reach significance and was not discussed further by the researchers.[11]

In partial discrepancy to studies outlined above, a N400-plus-P600 pattern with *no* associated LAN, has been observed in response to German passive sentences containing intransitive-only verbs (e.g., **Der Gartenwurde oft* *gearbeitet* *[*The garden was often* *worked]*; Frisch, Friederici & Hahne, 2004), and German active transitive sentences containing intransitive-only verbs (**Heute trodelte (V) der Cousin (NOM)* *dem Geiger (DAT)* *am Aufzug [*Today dawdled (V) the cousin (NOM)* *the violinist* *(DAT) at the lift]*; Friederici & Frisch, 2000). For English transitive verb-argument structure violations (e.g., **John sneezed* *the doctor* *and the nurse vs. John visited* *the doctor* *and the nurse*), Kielar, Meltzer-Aascher and Thompson (2012) reported no significant effects in the P600 time-window (500-700ms) but, in the earlier 300-500ms window, observed a significant negativity at right-frontal and left-parietal sites, as well as a positivity at posterior sites. These effects were interpreted as a N400-P600 pattern, as observed in previous studies (e.g., Frisch et al., 2004).[12] The N400 can be taken to reflect the fact that the NP could not be assigned a thematic role (e.g., patient) by the verb, and the P600 taken to reflect integration difficulties that follow the detection of this semantic anomaly.

It is unclear why these studies of overgeneralization have observed different distributions of negativity (i.e., left-anterior distribution [LAN] or bi-hemispheric, centro-parietal distribution [N400]). One possibility is that particular design elements of the studies encourage more emphasis on syntactic or semantic processing (yielding LAN or

---

[11] Despite not reaching significance in Rosler et al.'s (1993) analysis, this positivity has widely been reported as a P600-like component in various ERP papers (e.g., Friederici & Frisch, 2000; Frisch et al., 2004; Hahne, Wolf, Muller, Murbe, Friederici, 2012).

[12] The central purpose of Kielar et al.'s study was to examine the processing of verb-argument structure violations in individuals with agrammatic aphasia. The results reported here are from the control group of healthy young adults (N = 15; Mean Age = 22), described as "native speakers of English" with "no history of neurological, psychiatric, speech, language, or learning disorders" (p. 3322).

N400 respectively), but the number of studies in the literature is too limited to draw any firm conclusions.[13] Indeed, the issue of whether negativity is indicated by a LAN or N400 is orthogonal to the main purpose of the current study and is only returned to in the discussion.

The aim of the current study is to investigate whether the amplitude of ERP components known to mark overgeneralization errors (i.e., LAN/N400 and P600) can be used as a continuous dependent measure for investigating the mechanisms used to restrict overgeneralization errors. Given that (i) behavioural evidence demonstrates that high-frequency verbs are more resistant to overgeneralization than low-frequency verbs; (ii) a negativity (LAN/N400) – positivity (P600) ERP pattern systematically co-occurs with manipulations of verb-argument structure overgeneralization errors (vs. well-formed sentences) (e.g., Hagoort & Brown, 2000; Kielar et al., 2012; Friederici & Frisch, 2000; Frisch et al., 2004; Rosler et al., 1993); and (iii) the amplitude of these ERP components are sensitive to varying levels of violation (e.g., Gouvea et al., 2010; Gunter, Friederici, & Scriefers, 2000; Osterhout, Holcomb, and Swinney, 1994), it is expected that an experimental manipulation of verb-frequency will modulate the amplitude of the ERP components triggered by overgeneralization error.

In the current study we manipulated *grammaticality* by presenting intransitive-only verbs (e.g., *laugh*) in either intransitive (grammatical) or transitive (ungrammatical) argument structures, and manipulated *frequency* by using either high (HF) or low frequency (LF) verbs in those sentences. In line with previous behavioural studies that have controlled for semantics while examining the role of verb-frequency (e.g., Ambridge et al., 2008), HF and LF verbs were matched as closely as possible for semantics (e.g., *laugh* [HF] vs. *giggle* [LF]).

An effect of *grammaticality* is predicted, such that comparison of ungrammatical sentences against grammatical sentences will yield a negative component in the 300-500ms

[13] One line of argument draws a distinction between two classes of verb-argument structure violation. Specifically, mismatches between the case-marking of a given argument and that required by the verb (i.e., 'case-marking' argument structure violation) have exclusively elicited LAN rather than N400 (e.g., **the plane took <u>we</u> to paradise*; Coulson et al., 1998; Friederici & Frisch, 2000). In contrast, mismatches between the number of arguments in a sentence and the number licensed by the verb (i.e., 'number-of-arguments' argument structure violation) tend to elicit N400 more often than a LAN (e.g., Friederici & Frisch, 2000). The latter type of violation – number of arguments – is of main interest to the current study as this constitutes a verb-argument structure overgeneralization error; since these types of error have been found to elicit N400 (Frisch et al., 2004; Kielar et al, 2012) and LAN (e.g., Hagoort and Brown, 2000) depending on the study, the distribution of negativity cannot be considered clear-cut.

time-window (LAN/N400), followed by a positive component in the 500-800ms window (P600). Furthermore, based on behavioural evidence for a restriction mechanism that is guided by statistical learning (e.g., Braine & Brooks, 1995; Goldberg, 1995), a *grammaticality* x *frequency* interaction is predicted such that the *grammaticality* effect is greater for sentences containing *HF* verbs than sentences containing *LF* verbs.

4.3. Method

4.3.1. Participants

Nineteen adults (*M* = 20 years, Range = 18-32; 14 Females) participated in the experiment. All participants were undergraduate Psychology students recruited from the University of Manchester's course-credit scheme and were right-handed, monolingual English speakers. The study was approved by the University of Manchester ethics committee.

4.3.2. Design and Materials

In a repeated measures design, the independent variables were *Grammaticality* (*Grammatical* [intransitive] vs. *Ungrammatical* [transitive]) and *Verb Frequency* (*High Frequency [HF]* vs. *Low Frequency [LF]*). Thus, there were four conditions: *Ungrammatical/HF* (e.g., Sentence 1), *Grammatical/HF* (e.g., Sentence 2) *Ungrammatical/LF* (e.g., Sentence 3), and *Grammatical/LF* (e.g., Sentence 4). The full list of test-stimuli is provided in *Table A4.1 in Appendix 7*.

(1) **On Wednesday (PP1) Bob (NP1) laughed (V) the girl (NP2) in the kitchen (PP2)*
(2) *On Wednesday (PP1) Bob (NP1) laughed (V) in the kitchen (PP2)*
(3) **On Wednesday (PP1) Bob (NP1) giggled (V) the girl (NP2) in the kitchen (PP2)*
(4) *On Wednesday (PP1) Bob (NP1) giggled (V) in the kitchen (PP2)*

Grammaticality Manipulation. To manipulate *Grammaticality*, intransitive-only verbs were used in either intransitive-inchoative (Grammatical) or transitive-causative (Ungrammatical) argument-structures. Transitive-causatives (e.g., Sentences 1 and 3) consisted of an initial Prepositional Phrase (PP1) followed by an initial Noun Phrase (NP1), an intransitive-only Verb (V), a second Noun Phrase (NP2) and a final Prepositional Phrase (PP2). The structure of intransitive-inchoative sentences (e.g., Sentences 2 and 4) was identical to transitive-causatives, except that they excluded NP2.

Verb-frequency Manipulation. Verb-frequency was treated as a binary factor so that sentences containing *HF* verbs (e.g., Sentence 1 and 2) were compared to sentences containing *LF* verbs (e.g., Sentence 3 and 4). It was necessary to treat *verb-frequency* as a binary factor because it is generally recommended (e.g., Kaan, 2007) that each condition in an ERP study contain at least 40 items in order to obtain a good signal-to-noise ratio (thus, continuous treatment of *verb-frequency* would have required too many trials).

Verbs were selected using the criterion that any given *HF* verb was semantically-matched (as close as possible) to a *LF* verb (in line with previous behavioural studies that have controlled for semantics whilst examining the role of verb-frequency; e.g., Ambridge et al., 2008). The need to control for verb-semantics limited the selection of test-verbs to four *HF* verbs (*laugh* [4.77], *fall* [5.01], *disappear* [4.21], *smile* [4.71]) and four *LF* verbs (*giggle* [3.55], *tumble* [4.5], *vanish* [3.25], *grin* [3.55]). The frequency counts, shown next to each word in brackets, are based on Zipf SUBTLEX-UK frequency scores (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The mean syllable-count for *HF* verbs (1.75) was equal to that for *LF* verbs (1.75), thus controlling for possible effects of word-length. To reach the required 40 trials per condition, each *HF* verb was used 10 times in *Grammatical/HF* and *Ungrammatical/HF* conditions, and each *LF* verb was used 10 times in *Grammatical/LF* and *Ungrammatical/LF* conditions, providing a total of 160 test sentences (80 grammatical; 80 ungrammatical). An additional pair of semantically-matched verbs (*stay*/*wait*) was also used in 10 grammatical and 10 ungrammatical sentences each (creating a total of 40 additional sentences), but later examination revealed their frequency scores (*stay* [5.37], *wait* [5.39]) were too close to be categorised as *HF* or *LF* verbs; thus these sentences served only as fillers.

Sentence-content. Since each verb was used in each sentence structure (transitive/intransitive) 10 times, it was necessary to ensure variation in the content of these sentences. Thus, for each sentence-structure, each verb was used with 10 NP1s (always a common one-syllable male name; i.e., *Bob, Scott, Rob, Matt, Mike, Rick, Jack, Jake, Mark, Luke*), five PP1s (i.e. *On Monday/ On Tuesday/ On Wednesday/ On Thursday/ On Friday*) and two semantically-suitable PP2s (e.g., sentences containing *disappeared* always ended with *as if by magic* or *at the picnic*). A decision was made to keep NP2s (i.e., the point of violation) the same in all transitive sentences because more salient or varied NP2s (e.g. *the king, the aunt*) may have caused unwanted confounds; rather, '*the girl*' was used as NP2 for all transitive sentences due to its relatively less salient properties.

Controlling content across conditions. At the same time, we ensured that each individual sentence from each condition corresponded to a sentence in the other three conditions, such that all were identical in lexical content except for the two crucial manipulations. Taking Sentences 1-4 as an example, all have identical content until the manipulation of *verb-frequency* (i.e., *laugh/giggle*) or *grammaticality* (i.e., the onset of the post-verbal lexical item), thus ensuring that different responses between conditions can be attributed only to the experimental manipulation.

It is also possible for acoustic anomalies (e.g., prosodic variation) to emerge when ungrammatical sentences are read aloud, which may provide listeners with cues to grammaticality (e.g., Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013). For this reason, test-sentences were constructed using a cross-splicing procedure, which ensured that corresponding sentences from different conditions (e.g., Sentences 1-4) were also *acoustically* identical up to the manipulation. First, a 'dummy' sentence was created for each set of corresponding sentences that essentially provided an acoustic-frame that was identical for each corresponding sentence other than the manipulation (subsequently spliced in). The dummy sentence featured a suitable transitive-verb within a transitive-structure (e.g., *On Wednesday Bob amused the girl in the kitchen*). To create transitive test-sentences, a recording of the test-verb was spliced into the dummy sentence (e.g., *On Wednesday Bob ~~amused~~ the girl in the kitchen → On Wednesday Bob laughed/giggled the girl in the kitchen*). To create intransitive test-sentences, the NP2 was simply spliced out of the transitive test-sentence (e.g., *On Wednesday Bob laughed/giggled ~~the girl~~ in the kitchen*). Splicing points were determined by a careful visual inspection of the oscillogramms and were placed at zero-crossings only, using the audio-editing software, *Audacity 2.1.1* (Audacity Team, 2015). All sentences were spoken by a male native speaker of English at natural sentence prosody and digitally recorded at a sampling rate of 44.1 kHz in a sound-proof booth.

4.3.3. Procedure

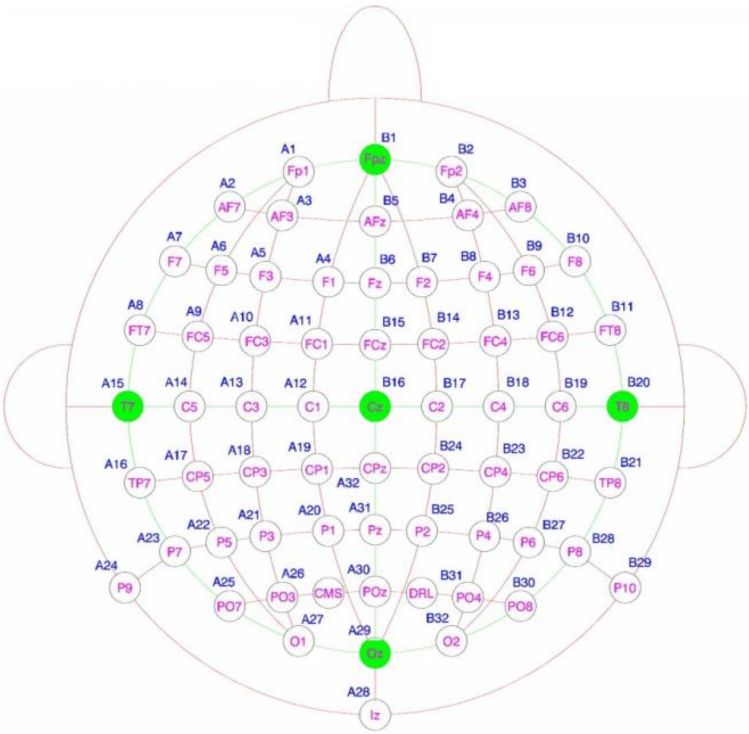
All sentences were presented aurally over loudspeakers, using *E-prime 1.0*. The order of presentation was pseudo-randomised for each participant, with the caveat that successive sentences never featured the same verb or its semantically-matched verb (e.g., a sentence containing *disappeared* was never followed by a sentence containing *disappeared* or *vanished*). Each sentence began playing at the same time that an asterisk appeared on the centre of the computer screen. The asterisk remained on the screen until 800ms after the offset of the sentence, at which time it was replaced by a question-mark that prompted participants to perform a binary judgement of the sentence's grammaticality (by pressing a green button to indicate grammaticality or a red button to indicate ungrammaticality). Assignment of left and right buttons to indicate grammaticality was counterbalanced across participants. Following the button press, the question mark was replaced by a feedback screen that indicated the percentage of correct answers that the participant had given (this feature served to maintain participants' task attentiveness, which has been linked to the reliability of ERP components). After a 1000ms wait, the next trial began. Trials were presented in blocks of 10, and participants were encouraged to take 20-30 second breaks at the end of each block. Test sessions usually lasted 90 minutes including electrode application.

4.3.4. ERP Recording and Pre-Processing

As stimuli were presented to participants, continuous EEG was recorded from 64 scalp locations using Ag-AgCl electrodes that were attached to an elastic cap and positioned according to the 10-20 system (see *Figure 4.1*). The *Active Two* system (Biosemi, Amsterdam), which does not require gain adjustment or measurement of impedance, was used for recording. Two additional scalp electrodes (CMS/DRL) were used as the online reference (for a complete description, see Schutter, Leitner, Kenemans, & van Honk, 2006; www.biosemi.com) and four additional electrodes used to record the horizontal and vertical electrooculogram (EOG) so that eye-movements and blinks could be monitored. Band-pass filters were set at 0.16-100Hz, with a sampling rate of 1000Hz (later downsampled to 200 Hz).

Pre-processing of data was performed using SPM8 (Litvak et al., 2011). ERPs were extracted from the continuous EEG by marking the onset of the critical word in each sentence and dividing trials into epochs from -200 to 1000ms relative to the critical word's onset. The critical word in *ungrammatical* conditions was the post-verbal determiner (e.g., *On Wednesday Bob laughed* <u>*the*</u> *girl in the kitchen*), and the critical word in *grammatical* conditions was the post-verbal preposition (*On Wednesday Bob laughed* <u>*in*</u> *the kitchen)*. To combat artefacts, and in accordance with Picton et al.'s (2000) guidelines, trials in which the peak-to-peak amplitude exceeded 200 μV in any of the recording channels (including EOG) were rejected (a total of 24.37% of trials). Single-trial ERPs that were retained were aligned to a 200-ms baseline prior to the critical word's onset.

***Figure 4.1*. Display of all (64 + 2) electrodes used in data collection.** Midline Analysis was conducted on 5 midline electrodes (*Fz, FCz, Cz, CPz, Pz*). Lateral Analysis was conducted on six Regions of Interest, each containing six electrodes: six electrodes: *left anterior* (F3, F5, F7, FC3, FC5, FT8), *right anterior* (F4, F6, F8, FC4, FC6, FT8), *right central* (C4, C6, T8, CP4, CP6, TP8) *left posterior* (P3, P5, P7, PO3, PO7, O1) and *right-posterior* (P4, P6, P8, PO4, PO8, O2).



4.4. Results

Statistical analyses were conducted with Linear Mixed Effects Models (LMEMs) in the *R* environment (*R* Development Core Team, 2011) using either *lmer* (for continuous ERP data) or *glmer* (for binomial behavioural data) from the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). LMEMs predict individual trials rather than averaging over trials and are robust against missing data and violations of sphericity (e.g., Baayen, Davidson, & Bates, 2008; Quene & Van den Bergh, 2008).

All LMEMs were fitted with random intercepts for participants and verbs and, where applicable, by-participant and by-verb random slopes. Random-intercepts offer the benefit of removing idiosyncratic variation within each random factor (i.e., participants and verbs), whereas random-slopes control for the possibility that treatment effects may vary within each random factor.

In accordance with recommendations outlined by Barr, Levy, Scheepers and Tily, (2013), a maximal random effects structure (i.e., all random intercepts and slopes) was attempted for each model. If maximal models did not converge (as was often the case given the complexity of the models in the present study), the model was simplified using the *best-path* algorithm (Barr et al., 2013). Under *best-path*, the inclusion of each random slope was determined by whether its addition to a random-intercept only model improved the model's fit of the data (using likelihood ratio tests and anti-conservative alpha values of p<0.25) (slopes that satisfied this criterion were added one-by-one until the model failed to converge).

All fixed-effects were entered into a model simultaneously. For *glmer* models (binomial behavioural data), the significance of each fixed-effect was determined by p-values produced from the *glmer* output. For *lmer* models (continuous ERP data), the significance of each fixed-effect was determined by p- values calculated based on Satterthwaite's approximation (in accordance with a recent stats paper; see Luke, 2016) using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2016).

4.4.1. Behavioural Data (manipulation check)

Before considering the data of primary interest (i.e., the ERP data) we first consider the behavioural data (recall that participants provided a binary grammaticality judgment for each sentence). This allows us to check that we successfully varied sentence structure (intransitive-inchoative/transitive-causative) and verb-frequency (high/low) in such a way as to manipulate the perceived grammaticality of the sentences as intended (i.e., as in previous behavioural studies such as that of Ambridge et al, 2008). *Table 4.1a* displays the mean percentage of sentences judged as "grammatical" for each of the four sentence-types (collapsed across verbs).

Table 4.1a. Mean (+/- SE) Percentage of Sentences Rated as Grammatical (first two columns) and Difference Scores (final column)

	Transitive (Ungrammatical)	Intransitive (Grammatical)	Difference Score (Intransitive-Transitive)
High Frequency Verbs	1.58 (0.00)	99.08 (0.00)	97.50
Low Frequency Verbs	6.58 (0.00)	98.68 (0.00)	92.10

Table 4.1b. Mean (+/- SE) Percentage of Sentences Rated as Grammatical (first two columns) and Difference Scores (final column): 1st Trials Only

	Transitive (Ungrammatical)	Intransitive (Grammatical)	Difference Score (Intransitive-Transitive)
High Frequency Verbs	0 (0)	100 (0)	100.00
Low Frequency Verbs	5.26 (0.03)	97.37 (0.02)	92.11

The LMEM included *Grammaticality* (*Grammatical* vs. *Ungrammatical*) and *Frequency* (*HF Verb vs. LF Verb*) as fully-crossed fixed-effects (dummy-coded, with ungrammatical high-frequency verbs as default-level), with by- subject and by- verb fully-maximal random-slopes. The outcome variable was each participant's binary grammaticality judgment (i.e., *Grammatical/Ungrammatical*) for each test-trial. As displayed in *Table 4.2*, a significant main-effect of *Grammaticality* was observed, with no significant effect of *Frequency* and no significant *Grammaticality*Frequency* interaction.

Table 4.2. Linear Mixed-Effects Model for Behavioural Data

Fixed effects	*Beta (b)*	*SE*	*z*	Pr(>|z|)
(Intercept)	7.35	1.88	3.91	0.00
GrammaticalityTransitive	-13.93	2.59	-5.37	0.00
FrequencyLow	-1.53	2.18	-0.70	0.48
GrammaticalityTransitive:FrequencyLow	3.30	2.96	1.11	0.27

One explanation for the null *Grammaticality*Frequency* interaction may be the fact that each verb was presented in grammatical and ungrammatical sentences on ten occasions in order to reach the recommended number of trials for an ERP study (as opposed to typical judgement studies in which each verb is presented just once grammatically and once ungrammatically). To investigate this possibility, an extra analysis was conducted on a subset of behavioural data that only included each participant's *first* response to each verb in a grammatical or ungrammatical sentence (thus ensuring a fairer manipulation check by levelling the playing-field with previous behavioural studies). *Table 4.1b* displays the mean percentage of sentences judged as "grammatical" for each of the four sentence-types (collapsed across verbs) within the subset. A repeated-measures ANOVA (as opposed to an LMEM, which was too complex to model singularity in the data) was conducted with *Grammaticality* and *Frequency* as fully-crossed independent variables. The outcome variable was the mean percentage of "grammatical" judgments for each sentence type (by-participant). A significant effect of *Grammaticality* ($F(18)=17.53$, $p<0.001$), and a significant *Grammaticality*Frequency* interaction ($F(18)=8.31$, $p=0.009$) were observed. To investigate if the effect of grammaticality was different for high- and low- frequency verbs, a follow-up LMEM was conducted on the difference scores for each verb (i.e., difference between judgments for grammatical sentences vs. ungrammatical sentences) with *Frequency* as the fixed-effect (slopes for this effect did not converge). A significant main-effect of *Frequency* was observed ($b=-0.08$, $SE= 0.03$, $z= -2.54$, $p=0.012$), indicating that the preference for grammatical (vs. ungrammatical) uses of high-frequency verbs was significantly greater than that for low-frequency verbs. Thus, the manipulation of *Grammaticality* and *Frequency* employed by the present study held-up to the manipulations employed by previous behavioural studies that have demonstrated verb-frequency to modulate the grammaticality of verb-argument structure overgeneralizations. The question now is whether the ERP paradigm can detect such fine-grained differences in the acceptability of overgeneralization errors.

4.4.2. ERP Data

Figure 4.2 illustrates averaged Event Related Potentials (ERPs) post-onset of the critical word in intransitive (*grammatical*) and transitive (*ungrammatical*) sentences (collapsed across *HF* and *LF* verbs) while *Figure 4.3* plots the difference between these conditions separately for sentences containing *HF* verbs and sentences containing *LF* verbs. For presentation only, ERPs were filtered offline with a 10 Hz low pass (statistical analyses were carried out on original data).

Figure 4.2. **Averaged ERPs post onset of the critical word in transitive (ungrammatical) sentences and intransitive (grammatical) sentences**

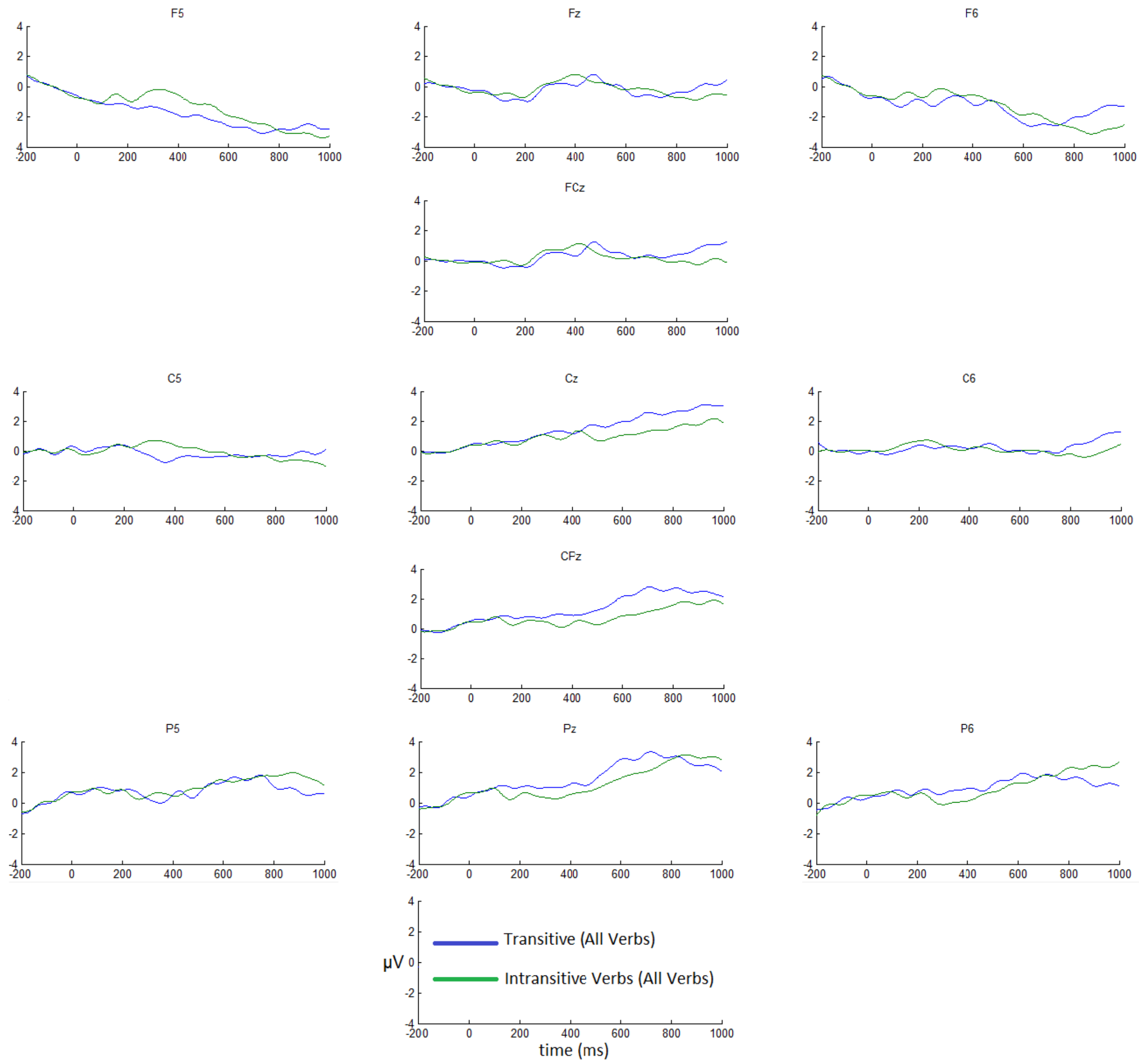
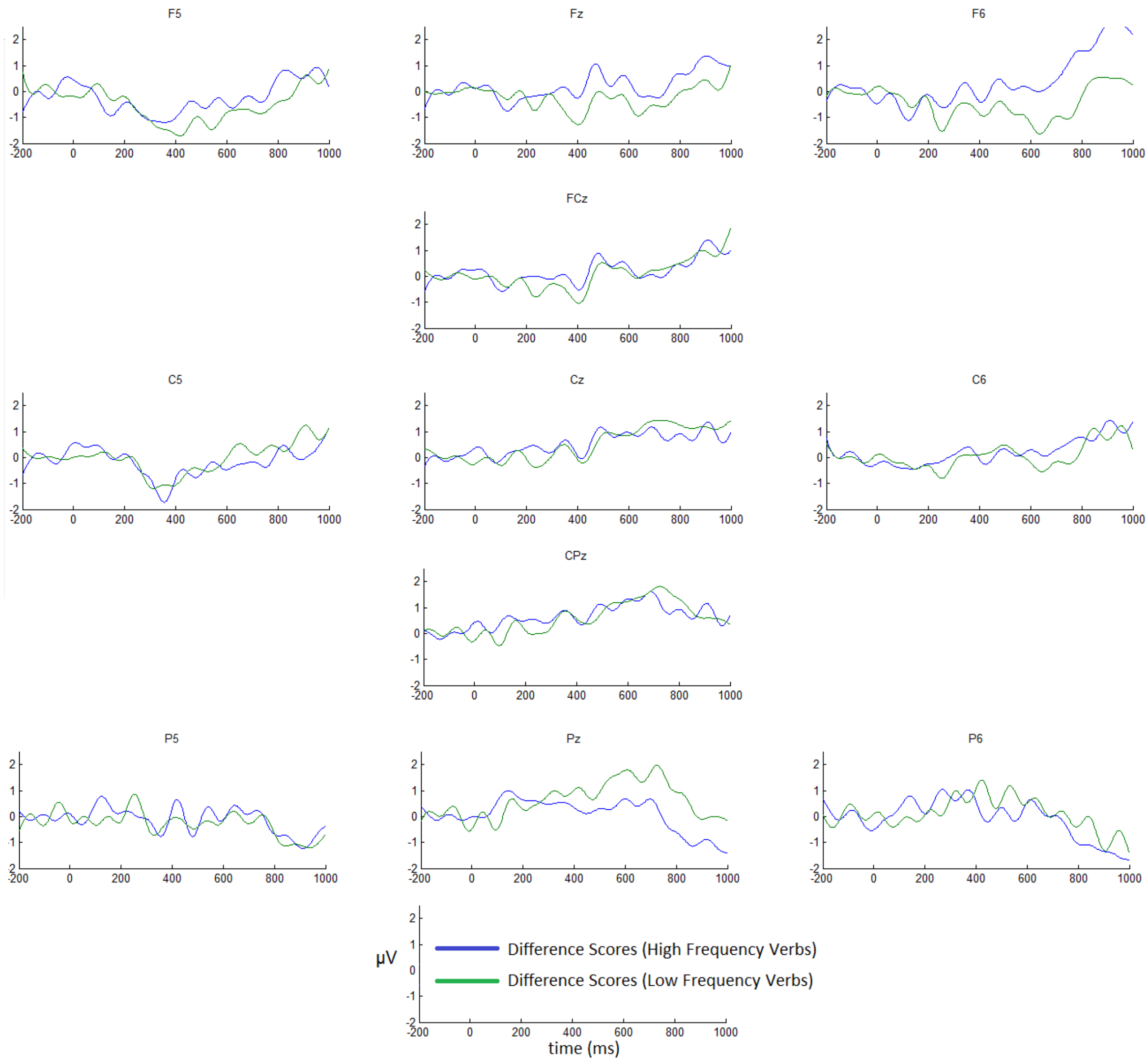


Figure 4.3. **Difference waves displaying the difference between averaged ERPs for transitive (ungrammatical) sentences and intransitive (grammatical) sentences, for sentences containing high-frequency verbs and sentences containing low-frequency verbs**



ERP data analysis was divided into two sections to accomodate the different latencies of different ERP components. *Analysis 1* examined the *300-500 ms* window to investigate negativity effects (LAN/ N400) whereas *Analysis 2* examined the *500-800 ms* window to investigate P600 effects. The time-windows were selected based on previous ERP studies that have found negativity (LAN/N400) and P600 effects.

As with previous studies (e.g., Frisch et al., 2004), both time-windows used a separate analysis for *midline* and *lateral* electrode sites. For the *Midline* analysis, *Electrode* (i.e., 5 midline electrodes: Fz, FCz, Cz, CPz, Pz) was used as an interaction-term (i.e., if a factor interacts with *Electrode*, this may reveal distributional differences in its effect). For the *Lateral* analysis, six *Regions of Interest (ROIs)* were created by crossing the factors

Region (*Anterior, Central, Posterior*) and *Hemisphere* (*Left, Right*). Each *ROI* included six electrodes: *left anterior* (F3, F5, F7, FC3, FC5, FT8), *right anterior* (F4, F6, F8, FC4, FC6, FT8), *right central* (C4, C6, T8, CP4, CP6, TP8) *left posterior* (P3, P5, P7, PO3, PO7, O1) and *right-posterior* (P4, P6, P8, PO4, PO8, O2) (again, if a factor interacted with *Region* or *Hemisphere*, this may reveal distributional differences in its effect).

All analyses proceded from a 'global' model that included all fixed-effects and interactions, to follow-up analyses at individual electrodes/ROIs if significant interactions with distributional factors were observed. *Tables 4.3* and *4.4* summarise the outcome of LMEMs for *midline* and *lateral* analyses respectively in the 300-500ms time window, and *Table 4.5* provides a summary of relevant mean ERP amplitudes for this time-window. *Tables 4.6* and *4.7* summarise the outcome of LMEMs for *midline* and *lateral* analyses respectively in the 500-800ms time-window, and *Table 4.8* provides a summary of relevant mean ERP amplitudes for this time-window. Note that for all LMEMs, fixed-effects were dummy-coded (i.e., such that each level was compared to a default-level). The default-level for each fixed-effect is provided in brackets as follows: *Grammaticality* (Grammatical Intranstives), *Frequency* (High), *Electrode* (CPz), *Hemisphere* (Left), and *Region* (Anterior). The random effects structures used are provided in a footnote for each analysis.

4.4.3. Analysis 1: LAN/ N400 time-window (300-500ms)

The outcome measure for all LMEMs in *Analysis 1* was the mean ERP amplitude in the *300-500 ms* window for each individual trial (measured for each participant at each electrode).

Midline Analysis[14] *(Table 4.3).* The global LMEM treated *Grammaticality* (*Grammatical* vs. *Ungrammatical*) and *Frequency* (HF verb vs. LF verb) as fully-crossed fixed effects. *Electrode* (i.e., 5 midline electrodes: Fz, FCz, Cz, CPz, Pz) was also entered as an interaction term but was not entered as a main effect because the outcome of this would not be of interest to the hypothesis at hand. Most relevant to the main aim of the study, the global analysis did not reveal a significant *Grammaticality*Frequency* interaction, nor a *Grammaticality*Frequency*Electrode* interaction. Thus, no follow-up analyses were conducted on this data.

[14] Random-effects structure for Midline Global Analysis: *(1|PPT) + (0+ Grammaticality+ Grammaticality:Frequency |PPT) + (1|Verb)*

Lateral Analysis[15] *(Table 4.4).* The global LMEM conducted for *Lateral* electrode sites used *Grammaticality* and *Frequency* as fully-crossed fixed effects, and also included *Region* and *Hemisphere* as interaction terms for each of these factors. A main effect of *Grammaticality* was observed, but this effect must be interpreted in light of a significant *Grammaticality*Hemisphere*Region* interaction. Thus, follow-up analyses at each ROI were conducted (treating *Grammaticality* as the sole fixed-effect since no interaction with Frequency was observed). A main effect of *Grammaticality* was observed at the *Left-Anterior* ROI (*Table 4.2*), such that ungrammatical (transitive) sentences yielded significantly more negative ERP than grammatical (intransitive) sentences. The restriction of this negativity to left-hemispheric and anterior regions of the scalp is consistent with the characteristics of a LAN component, and thus supports the first hypothesis that a negative ERP component would be observed in the 300-500ms time-window, in response to overgeneralization errors (compared to well-formed sentences). Crucially, the second hypothesis of the study was that the magnitude of this component would be sensitive to the manipulation of verb-frequency. However, the fact that *Freqeuncy* did not yield any significant interactions in the Global analysis indicates that the LAN was not modulated by verb frequency, and may not be a suitable measure of the fine-grained differences in relative acceptability of overgeneralization errors.

Additionally, an effect of *Grammaticality* was observed at the *Posterior Right* ROI, but given its distribution, the direction of the effect was unexpected, such that ungrammatical (transitive) sentences yielded significantly more *positive* ERP than grammatical (intransitive) sentences. This effect is not easy to explain given previous literature, but of most relevance to the present study, it did not interact with *Frequency*, (given the Global analysis) and thus did not inform our main line of inquiry.

[15] Random Effects Structure for Lateral Analyses: (i) Global: *(1| PPT)+ (0+Grammaticality+ Frequency+ Frequency:Hemisphere|PPT)+(1|Verb);* (ii) AnteriorLeft: *(1| PPT)+(1 | Verb); (iii) AnteriorRight: (1 | PPT)+(0 + Grammaticality | PPT)+(1| Verb);* (iv) CentralLeft: *(1| PPT)+(1 | Verb)+(0+Grammaticality|Verb)*; (v) CentralRight: *(1 | PPT)+(0 + Grammaticality | PPT);* (vi): PosteriorLeft: *(1 | PPT)+(0 + Grammaticality | PPT)+(1+Grammaticality | Verb)*; (vii) PosteriorRight: *(1| PPT)+ (1 | Verb)*

Table 4.3. Linear Mixed-Effects Models for Midline sites in the 300-500ms window

Model	Fixed effects	*Beta (b)*	*SE*	df	t	Pr(>|t|)
Global	(Intercept)	0.29	0.34	101	0.85	0.40
	GrammaticalityTransitive	0.56	0.45	200	1.26	0.21
	FrequencyLow	0.38	0.43	183	0.88	0.38
	GrammaticalityIntransitive:ElectrodeCz	0.58	0.41	11770	1.42	0.16
	GrammaticalityTransitive:ElectrodeCz	0.46	0.41	11770	1.12	0.26
	GrammaticalityIntransitive:ElectrodeFCz	0.16	0.41	11770	0.40	0.69
	GrammaticalityTransitive:ElectrodeFCz	-0.22	0.41	11770	-0.53	0.60
	GrammaticalityIntransitive:ElectrodeFz	-0.02	0.41	11770	-0.06	0.95
	GrammaticalityTransitive:ElectrodeFz	-0.29	0.41	11770	-0.71	0.48
	GrammaticalityIntransitive:ElectrodePz	0.29	0.41	11770	0.72	0.48
	GrammaticalityTransitive:ElectrodePz	-0.05	0.41	11770	-0.12	0.90
	FrequencyLow:ElectrodeCz	0.16	0.58	11770	0.28	0.78
	FrequencyLow:ElectrodeFCz	0.47	0.58	11770	0.81	0.42
	FrequencyLow:ElectrodeFz	0.38	0.58	11770	0.66	0.51
	FrequencyLow:ElectrodePz	-0.29	0.58	11770	-0.50	0.62
	GrammaticalityTransitive:FrequencyLow	-0.14	0.60	499	-0.23	0.82
	GrammaticalityTransitive:FrequencyLow:ElectrodeCz	-0.07	0.82	11770	-0.09	0.93
	GrammaticalityTransitive:FrequencyLow:ElectrodeFCz	-0.57	0.82	11770	-0.70	0.49
	GrammaticalityTransitive:FrequencyLow:ElectrodeFz	-0.89	0.82	11770	-1.08	0.28
	GrammaticalityTransitive:FrequencyLow:ElectrodePz	0.67	0.82	11770	0.82	0.41

Table 4.4. Linear Mixed-Effects Models for Lateral sites in the 300-500ms window

Model	Fixed effects	*Beta (b)*	*SE*	*df*	*t*	*Pr(>|t|)*
Global	(Intercept)	-0.16	0.18	47	-0.92	0.36
	GrammaticalityTransitive	**-1.05**	**0.18**	**574**	**-5.68**	**0.00**
	FrequencyLow	-0.14	0.23	65	-0.60	0.55
	GrammaticalityIntransitive:HemisphereRight	-0.40	0.25	49	-1.58	0.12
	GrammaticalityTransitive:HemisphereRight	**0.76**	**0.25**	**49**	**2.99**	**0.00**
	GrammaticalityIntransitive:RegionCentral	**0.77**	**0.17**	**85150**	**4.46**	**0.00**
	GrammaticalityTransitive:RegionCentral	**1.05**	**0.17**	**85150**	**6.05**	**0.00**
	GrammaticalityIntransitive:RegionPosterior	**0.92**	**0.17**	**85150**	**5.30**	**0.00**
	GrammaticalityTransitive:RegionPosterior	**1.81**	**0.17**	**85150**	**10.36**	**0.00**
	FrequencyLow:HemisphereRight	0.46	0.35	52	1.31	0.20
	FrequencyLow:RegionCentral	-0.33	0.25	85150	-1.33	0.18
	FrequencyLow:RegionPosterior	-0.31	0.25	85150	-1.24	0.21
	GrammaticalityTransitive:FrequencyLow	-0.12	0.25	85180	-0.49	0.63
	GrammaticalityIntransitive:HemisphereRight:RegionCentral	-0.29	0.25	85150	-1.19	0.24
	GrammaticalityTransitive:HemisphereRight:RegionCentral	-0.36	0.25	85150	-1.48	0.14
	GrammaticalityIntransitive:HemisphereRight:RegionPosterior	-0.27	0.25	85150	-1.12	0.26
	GrammaticalityTransitive:HemisphereRight:RegionPosterior	**-0.85**	**0.25**	**85150**	**-3.45**	**0.00**
	FrequencyLow:HemisphereRight:RegionCentral	0.14	0.35	85150	0.41	0.68
	FrequencyLow:HemisphereRight:RegionPosterior	-0.14	0.35	85150	-0.40	0.69
	GrammaticalityTransitive:FrequencyLow:HemisphereRight	-0.28	0.35	85170	-0.80	0.42
	GrammaticalityTransitive:FrequencyLow:RegionCentral	0.48	0.35	85150	1.36	0.17
	GrammaticalityTransitive:FrequencyLow:RegionPosterior	0.47	0.35	85150	1.33	0.18
	GrammaticalityTransitive:FrequencyLow:HemisphereRight: RegionCentral	0.00	0.49	85150	0.01	0.99
	GrammaticalityTransitive:FrequencyLow:HemisphereRight: RegionPosterior	0.31	0.49	85150	0.63	0.53
Anterior Left	(Intercept)	-0.20	0.34	14.00	-0.58	0.57
	GrammaticalityTransitive	**-1.13**	**0.13**	**14188**	**-8.50**	**0.00**
Anterior Right	(Intercept)	-0.36	0.30	21.09	-1.20	0.25
	GrammaticalityTransitive	-0.10	0.25	18.28	-0.41	0.69
Central Left	(Intercept)	0.39	0.24	22.17	1.64	0.12
	GrammaticalityTransitive	**-0.61**	**0.27**	**7.05**	**-2.26**	**0.06**
Central Right	(Intercept)	-0.02	0.20	9.22	-0.09	0.93
	GrammaticalityTransitive	0.37	0.28	7.81	1.30	0.23
Posterior Left	(Intercept)	0.50	0.31	11.99	1.62	0.13
	GrammaticalityTransitive	0.01	0.24	10.16	0.03	0.98
Posterior Right	(Intercept)	-0.01	0.25	10.00	-0.05	0.96
	GrammaticalityTransitive	**0.62**	**0.13**	**14200**	**4.89**	**0.00**

Table 4.5. Mean (+/- SE) ERP Amplitude in MicroVolts (300-500 ms Time-Window) For Grammatical Sentences and Ungrammatical Sentences Containing High or Low Frequency Verbs

Analysis	Grammatical (HF Verb)	Grammatical (LF Verb)	Ungrammatical (HF Verb)	Ungrammatical (LF Verb)	Grammatical (All Verbs)	Ungrammatical (All Verbs)
Midline (Global)	0.6 (0.05)	1.14 (0.05)	0.81 (0.05)	1.04 (0.05)	0.87 (0.09)	0.6 (0.05)
Fz	0.43 (0.43)	1.15 (0.33)	0.65 (0.36)	0.24 (0.34)	0.79 (0.23)	0.43 (0.43)
Fcz	0.62 (0.38)	1.47 (0.32)	0.67 (0.36)	0.84 (0.32)	1.05 (0.21)	0.62 (0.38)
Cz	0.96 (0.3)	1.68 (0.37)	1.36 (0.33)	1.7 (0.32)	1.32 (0.19)	0.96 (0.3)
CPz	0.44 (0.32)	0.83 (0.35)	0.79 (0.3)	1.13 (0.33)	0.64 (0.2)	0.44 (0.32)
Pz	0.53 (0.39)	0.56 (0.33)	0.57 (0.45)	1.32 (0.35)	0.55 (0.22)	0.53 (0.39)
Lateral (Global)	0.09 (0.13)	-0.03 (0.13)	-0.1 (0.13)	-0.09 (0.13)	0.03 (0.04)	0.09 (0.13)
Anterior Left	-0.15 (0.14)	-0.17 (0.13)	-1.18 (0.13)	-1.48 (0.13)	-0.16 (0.1)	-0.15 (0.14)
Anterior Right	-0.51 (0.13)	-0.16 (0.12)	-0.36 (0.13)	-0.47(0.13)	-0.34 (0.09)	-0.51 (0.13)
Central Left	0.59 (0.11)	0.09 (0.11)	-0.12 (0.11)	-0.28 (0.12)	0.34 (0.09)	0.59 (0.11)
Central Right	-0.1 (0.1)	0.04 (0.11)	0.23 (0.1)	0.49 (0.1)	-0.03 (0.07)	-0.1 (0.1)
Posterior Left	0.72 (0.14)	0.13 (0.12)	0.52 (0.14)	0.44 (0.14)	0.42 (0.09)	0.72 (0.14)
Posterior Right	0 (0.13)	-0.09 (0.12)	0.32 (0.12)	0.74 (0.13)	-0.04 (0.09)	0 (0.13)

4.4.4. Analysis 2: P600 time-window (500 to 800ms)

The outcome measure for all LMEMs in *Analysis 2* was the mean ERP amplitude in the *500-800 ms* window for each individual trial (measured for each participant at each electrode). Otherwise, *Analysis 2* followed an identical procedure to *Analysis 1*.

Midline Analysis[16] (*Table 4.6*). The Global LMEM for midline analysis revealed a marginally significant main effect of *Grammaticality* (p=0.06). *Frequency* did not exert a significant main effect, nor did it significantly interact with any other factors. To examine whether the *Grammaticality* effect takes on the typical-distribution of a P600 component, it must be interpreted in terms of a significant *GrammaticalityxElectrode* interaction. To investigate this interaction, separate mixed-effects models at each midline electrode site (*Fz, FCz, Cz, CPz, Pz*) employed *Grammaticality* as a fixed effect. *Grammaticality* exerted no effect at *Fz* or *FCz*, as can be expected given that P600 is maximal at centro-parietal

[16] Random Effects Structure for Models: (i) Global: *(1|PPT)+(0+Grammaticality+Grammaticality:Frequency| PPT)+ (1|Verb);* (ii) Fz: *(1| PPT)+ (0+Grammaticality|PPT)+ (1 | Verb)+(0+Grammaticality|Verb)* ; (iii) FCz: *(1 | PPT)+(0 + Grammaticality | PPT)+ (1+Grammaticality | Verb)* ; (iii) Cz: *(1 | PPT)+ (1 | Verb) + (0+ Grammaticality|Verb);* (iv) CPz: *(1 | PPT)+(0 + Grammaticality | PPT)+(1 | Verb)+(0 + Grammaticality | Verb);* (v) Pz: *(1| PPT)+(0+Grammaticality|PPT)+ (1 | Verb)+(0+Grammaticality|Verb) ;*

sites. A significant main effect of *Grammaticality* was observed at centro-parietal sites: *Cz, CPz,* and *Pz*. In all cases, the *Grammaticality* effect was such that ungrammatical (transitive) sentences yielded greater positivity than grammatical (intransitive) sentences. Thus, consistent with the first hypothesis, overgeneralization errors (compared to well-formed sentences) yielded a positive ERP in the 500-800ms time-window that can be taken to resemble the P600 response. However, the second hypothesis of the study, namely that the magnitude of the P600 evoked by overgeneralization errors would be sensitive to the manipulation of verb-frequency, was not supported: *Grammaticality* did not significantly interact with *Frequency* in the Global model. Thus, the P600 may not be sensitive to the relative acceptability of overgeneralization errors, and thus may not be a suitable measure of the mechanims employed to restrict these types of error.

Lateral Analysis[17] *(Table 4.7).* The global LMEM for lateral analysis revealed a significant *GrammaticlalityxRegionxHemisphere* interaction as well as a significant *GrammaticalityxFrequencyxRegion* interaction. To investigate these interactions, follow-up analyses at each ROI included both *Grammaticality* and *Frequency* as fully-crossed fixed-effects. However, at all ROIs, no significant effect of Gramamticality, Frequency, or most crucially, *Grammaticality*Frequency* interactions were observed.

---

[17] Random Effects Structure For Models: (i) Global: *(1| PPT)+(0+Frequency | PPT)+ (1|Verb);* (ii) AnteriorLeft: *(1| PPT)+(0+Grammaticality:Frequency | PPT)+(1| Verb)+(0+Grammaticality| Verb); (iii)* AnteriorRight: *(1| PPT)+ (0+Grammaticality:Frequency|PPT)+(1 | Verb);* (iv) CentralLeft: *(1| PPT)+ (0+Grammaticality|PPT)+(1 | Verb);* (v) CentralRight: *(1| PPT)+ (0+Grammaticality+ Grammaticality:Frequency|PPT)+ (1 | Verb)+ (0+Grammaticality|Verb);* (vi): PosteriorLeft: *(1| PPT)+ (0+Frequency+Grammaticality:Frequency|PPT)+(1 | Verb);* (vii) PosteriorRight: *(1| PPT)+ (0+Grammaticality|PPT)+ (1 | Verb)+ (0+Grammaticality|Verb)*

Table 4.6. Linear Mixed-Effects Models for Midlne sites in the 500-800ms window

Model	Fixed effects	*Beta (b)*	*SE*	df	t	Pr (>|t|)
Global	(Intercept)	1.21	0.50	47	2.42	0.02
	GrammaticalityTransitive	**1.09**	**0.57**	**84**	**1.92**	**0.06**
	FrequencyLow	-0.24	0.60	68	-0.40	0.69
	GrammaticalityIntransitive:ElectrodeCz	0.06	0.47	11740	0.13	0.89
	GrammaticalityTransitive:ElectrodeCz	-0.12	0.47	11740	-0.26	0.79
	GrammaticalityIntransitive:ElectrodeFCz	**-1.32**	**0.47**	**11740**	**-2.82**	**0.00**
	GrammaticalityTransitive:ElectrodeFCz	**-2.04**	**0.47**	**11740**	**-4.35**	**0.00**
	GrammaticalityIntransitive:ElectrodeFz	**-1.61**	**0.47**	**11740**	**-3.46**	**0.00**
	GrammaticalityTransitive:ElectrodeFz	**-2.53**	**0.47**	**11740**	**-5.41**	**0.00**
	GrammaticalityIntransitive:ElectrodePz	**1.04**	**0.47**	**11740**	**2.23**	**0.03**
	GrammaticalityTransitive:ElectrodePz	0.26	0.47	11740	0.56	0.57
	FrequencyLow:ElectrodeCz	0.24	0.66	11740	0.37	0.71
	FrequencyLow:ElectrodeFCz	0.85	0.66	11740	1.29	0.20
	FrequencyLow:ElectrodeFz	1.01	0.66	11740	1.53	0.13
	FrequencyLow:ElectrodePz	-0.41	0.66	11740	-0.62	0.54
	GrammaticalityTransitive:FrequencyLow	0.30	0.78	97	0.38	0.70
	GrammaticalityTransitive:FrequencyLow:ElectrodeCz	0.05	0.94	11740	0.05	0.96
	GrammaticalityTransitive:FrequencyLow:ElectrodeFCz	-0.47	0.94	11740	-0.50	0.62
	GrammaticalityTransitive:FrequencyLow:ElectrodeFz	-0.90	0.94	11740	-0.97	0.33
	GrammaticalityTransitive:FrequencyLow:ElectrodePz	0.81	0.94	11740	0.86	0.39
Fz	(Intercept)	-0.06	0.42	16	-0.14	0.89
	GrammaticalityTransitive	-0.15	0.36	274	-0.42	0.67
Fcz	(Intercept)	0.21	0.39	16	0.53	0.61
	GrammaticalityTransitive	0.26	0.35	18	0.75	0.46
Cz	(Intercept)	1.36	0.41	21	3.34	0.00
	GrammaticalityTransitive	**1.05**	**0.30**	**241**	**3.47**	**0.00**
CPz	(Intercept)	1.12	0.43	18	2.63	0.02
	GrammaticalityTransitive	**1.19**	**0.38**	**18**	**3.17**	**0.01**
Pz	(Intercept)	1.85	0.40	14	4.60	0.00
	GrammaticalityTransitive	**0.87**	**0.35**	**441**	**2.45**	**0.01**

Table 4.7. Linear Mixed-Effects Models for Lateral sites in the 500-800ms window

Model	Fixed effects	*Beta (b)*	*SE*	df	t	Pr (>|t|)
Global	(Intercept)	-1.32	0.17	121	-7.78	0.00
	GrammaticalityTransitive	**-0.72**	**0.20**	**85200**	**-3.64**	**0.00**
	FrequencyLow	0.00	0.21	190	-0.01	0.99
	GrammaticalityIntransitive:HemisphereRight	**-0.51**	**0.20**	**85180**	**-2.59**	**0.01**
	GrammaticalityTransitive:HemisphereRight	0.32	0.20	85180	1.63	0.10
	GrammaticalityIntransitive:RegionCentral	**1.70**	**0.20**	**85180**	**8.56**	**0.00**
	GrammaticalityTransitive:RegionCentral	**2.19**	**0.20**	**85180**	**10.99**	**0.00**
	GrammaticalityIntransitive:RegionPosterior	**3.31**	**0.20**	**85180**	**16.72**	**0.00**
	GrammaticalityTransitive:RegionPosterior	**3.97**	**0.20**	**85180**	**19.94**	**0.00**
	FrequencyLow:HemisphereRight	**0.88**	**0.28**	**85180**	**3.12**	**0.00**
	FrequencyLow:RegionCentral	-0.37	0.28	85180	-1.31	0.19
	FrequencyLow:RegionPosterior	**-0.63**	**0.28**	**85180**	**-2.23**	**0.03**
	GrammaticalityTransitive:FrequencyLow	-0.05	0.28	85210	-0.19	0.85
	GrammaticalityIntransitive:HemisphereRight:RegionCentral	0.06	0.28	85180	0.22	0.83
	GrammaticalityTransitive:HemisphereRight:RegionCentral	-0.24	0.28	85180	-0.87	0.39
	GrammaticalityIntransitive:HemisphereRight:RegionPosterior	0.17	0.28	85180	0.61	0.54
	GrammaticalityTransitive:HemisphereRight:RegionPosterior	**-0.62**	**0.28**	**85180**	**-2.19**	**0.03**
	FrequencyLow:HemisphereRight:RegionCentral	-0.19	0.40	85180	-0.48	0.63
	FrequencyLow:HemisphereRight:RegionPosterior	-0.66	0.40	85180	-1.67	0.10
	GrammaticalityTransitive:FrequencyLow:HemisphereRight	**-0.77**	**0.40**	**85180**	**-1.93**	**0.05**
	GrammaticalityTransitive:FrequencyLow:RegionCentral	0.37	0.40	85180	0.93	0.35
	GrammaticalityTransitive:FrequencyLow:RegionPosterior	0.40	0.40	85180	1.01	0.31
	GrammaticalityTransitive:FrequencyLow:HemisphereRight: RegionCentral	0.24	0.56	85180	0.43	0.67
	GrammaticalityTransitive:FrequencyLow:HemisphereRight: RegionPosterior	0.81	0.56	85180	1.43	0.15
Anterior Left	(Intercept)	-1.29	0.54	12	-2.38	0.04
	GrammaticalityTransitive	-0.74	0.57	12	-1.31	0.21
	FrequencyLow	0.06	0.82	14	0.07	0.95
	GrammaticalityTransitive:FrequencyLow	-0.10	0.88	15	-0.12	0.91
Anterior Right	(Intercept)	-1.86	0.40	21	-4.68	0.00
	GrammaticalityTransitive	0.19	0.47	18	0.41	0.69
	FrequencyLow	0.96	0.47	16	2.06	0.06
	GrammaticalityTransitive:FrequencyLow	-0.92	0.56	18	-1.63	0.12
Central Left	(Intercept)	0.42	0.29	19	1.47	0.16
	GrammaticalityTransitive	-0.23	0.23	36	-1.00	0.33
	FrequencyLow	-0.36	0.30	9	-1.19	0.26
	GrammaticalityTransitive:FrequencyLow	0.31	0.26	14185	1.21	0.23
Central Right	(Intercept)	-0.16	0.38	18	-0.42	0.68
	GrammaticalityTransitive	0.34	0.46	7	0.74	0.48
	FrequencyLow	0.38	0.43	12	0.88	0.40
	GrammaticalityTransitive:FrequencyLow	-0.24	0.71	9	-0.35	0.74
Posterior Left	(Intercept)	2.07	0.39	15	5.27	0.00
	GrammaticalityTransitive	-0.13	0.35	17	-0.38	0.71
	FrequencyLow	-0.67	0.51	12	-1.33	0.21
	GrammaticalityTransitive:FrequencyLow	0.32	0.43	18	0.74	0.47
Posterior Right	(Intercept)	1.63	0.37	13	4.40	0.00
	GrammaticalityTransitive	-0.06	0.44	12	-0.15	0.89
	FrequencyLow	-0.42	0.42	6	-1.01	0.35
	GrammaticalityTransitive:FrequencyLow	0.39	0.51	6	0.77	0.47

***Table 4.8.* Mean (+/- SE) ERP Amplitude in MicroVolts (500-800 ms Time-Window) For Grammatical Sentences and Ungrammatical Sentences Containing High or Low Frequency Verbs**

Analysis	Grammatical (HF Verb)	Grammatical (LF Verb)	Ungrammatical (HF Verb)	Ungrammatical (LF Verb)	Grammatical (All Verbs)	Ungrammatical (All Verbs)
Midline (Global)	0.88 (0.15)	1.02 (0.15)	1.46 (0.15)	1.7 (0.15)	0.95 (0.11)	1.58 (0.10)
Fz	-0.4 (0.38)	0.26 (0.36)	-0.11 (0.36)	-0.13 (0.34)	-0.07 (0.26)	-0.12 (0.25)
Fcz	-0.07 (0.37)	0.57 (0.33)	0.37 (0.34)	0.81 (0.33)	0.25 (0.24)	0.59 (0.24)
Cz	1.31 (0.31)	1.55 (0.30)	2.32 (0.31)	2.6 (0.3)	1.43 (0.22)	2.46 (0.22)
CPz	1.33 (0.31)	1.15 (0.33)	2.32 (0.29)	2.41 (0.31)	1.24 (0.22)	2.36 (0.21)
Pz	2.25 (0.34)	1.58 (0.39)	2.4 (0.34)	2.83 (0.34)	1.91 (0.26)	2.61 (0.25)
Lateral (Global)	0.15 (0.06)	0.14 (0.06)	0 (0.06)	-0.03 (0.06)	0.14 (0.04)	-0.01 (0.04)
Anterior Left	-1.26 (0.16)	-1.33 (0.16)	-2.04 (0.15)	-2.09 (0.16)	-1.29 (0.11)	-2.06 (0.11)
Anterior Right	-1.89 (0.16)	-0.94 (0.14)	-1.63 (0.14)	-1.6 (0.14)	-1.42 (0.11)	-1.62 (0.10)
Central Left	0.47 (0.13)	0.02 (0.13)	0.23 (0.12)	0.1 (0.13)	0.25 (0.09)	0.17 (0.09)
Central Right	-0.2 (0.12)	0.29 (0.13)	0.16 (0.12)	0.32 (0.12)	0.05 (0.09)	0.24 (0.08)
Posterior Left	2.12 (0.15)	1.45 (0.14)	1.91 (0.15)	1.57 (0.15)	1.79 (0.10)	1.74 (0.11)
Posterior Right	1.62 (0.14)	1.33 (0.14)	1.37 (0.13)	1.56 (0.13)	1.47 (0.10)	1.46 (0.09)

4.4.5. Summary of ERP data

Ungrammatical sentences (realised by verb-argument structure overgeneralizations) yielded a significant negativity in the 300-500 ms window and a significant positivity in the 500-800 ms window that resembled Left Anterior Negativity (LAN) and P600 components respectively, thus corroborating previous evidence for this outcome. However, the effects of *Grammaticality* did not interact with *Frequency*. Thus, the LAN and P600 components evoked by overgeneralization errors (vs. well-formed sentences) were not sensitive to the manipulation of verb-frequency. Possible explanations as to why this may be the case are considered in the *Discussion*.

4.5. Discussion

The present study investigated whether ERP components (LAN and P600) that are known to mark verb-argument structure overgeneralizations are sensitive to statistical restriction mechanisms that have previously been detected only by behavioural paradigms.

Specifically, a large number of behavioural studies have demonstrated that statistical restriction mechanisms restrict verbs from being generalized into infelicitous verb-argument structures (e.g., Ambridge, Pine, Rowland & Young, 2008; Bidgood, Ambridge, Pine & Rowland, 2014; Blything, Ambridge & Lieven, 2014; Brooks, Tomasello, Dodson & Lewis, 1999; Brooks & Zizak, 2002; Robenalt & Goldberg, 2015; Theakston, 2004; Wonnacott, Newport & Tanenhaus, 2008), such that verbs that are highly frequent in the input are more resistant to overgeneralization than low-frequency verbs (e.g., **The clown laughed the boy* receives less favourable grammaticality ratings and is less likely to be produced than **The clown giggled the boy*). Such findings provide support for a *statistical-learning* account of retreat from overgeneralization (e.g., Braine & Brooks, 1995; Goldberg, 1995), which posits that the probabilistic restrictions of verb-argument structure overgeneralization errors is influenced by the frequency of the verb in grammatical constructions.

The present study sought to extend this investigation to the ERP paradigm for two primary reasons. First, ERPs measure patterned voltage changes on the scalp that are associated with the experimental manipulation, and thus offer a direct observation of cognitive activity underlying the processing of overgeneralization errors (e.g., Frisch et al., 2004; Hagoort & Brown, 2000) (in contrast to behavioural studies, which rely on an overt response that comes *after* the stimulus has been processed). Second, as discussed in the Introduction, the ERP paradigm is an online measurement that does not necessarily require an overt response, and can thus potentially be applied to key testing populations who are more susceptible to the limitations of behavioural paradigms (e.g., young children and second-language learners).

Until now, the ERP paradigm had not been applied to measure the restriction mechanisms thought to operate on verb-argument structure overgeneralizations. The present study measured ERPs in healthy adults as they were presented with transitive verb-argument structure overgeneralization errors containing high frequency verbs (e.g., *On Monday Bob <u>laughed</u> the girl in the garden*) or low frequency verbs (e.g., *On Monday Bob <u>giggled</u> the girl in the garden*). Consistent with previous studies (e.g., Hagoort & Brown, 2000; Rosler et al., 1993), a LAN-P600 ERP pattern was observed in response to overgeneralization errors. However, most relevant to the goals of the present study is the finding that the magnitude of these components did not appear to be influenced by the manipulation of verb-frequency.

The fact that the LAN and P600 components were not modulated by verb-frequency is contrary to what was expected under the rationale of the current study. Previous studies suggest that (i) these components systematically mark the occurrence of overgeneralization errors (including in the current study) (e.g., Hagoort & Brown, 2000; Kielar et al., 2012; Friederici & Frisch, 2000; Frisch et al., 2004; Rosler et al., 1993); (ii) the amplitude of these components is sensitive to different extents of syntactic violation (e.g., Gouvea et al., 2010; Gunter, Friederici, & Scriefers, 2000; Osterhout, Holcomb, and Swinney, 1994); and (iii) grammaticality ratings and production rates of overgeneralization errors are modulated by verb-frequency (e.g., Ambridge et al, 2008). Consequently, we predicted that the size (i.e., mean amplitude) of the ERP components observed in the present study should be sensitive to the manipulation of verb frequency.

One possible explanation for this unexpected null result may be that the P600 and LAN components are not, in fact, suitable measures of fine-grained linguistic processes underlying the processing of these types of errors. First, it is clearly possible for the amplitude of the P600 to be sensitive to different extents of syntactic manipulations (although less is known about the sensitivity of LAN). For example, Gouvea et al. (2010) demonstrated that the P600 evoked by long distance *wh*- dependencies was smaller than the P600 evoked by syntactic ambiguities and syntactic violations, and thus showed that the P600 can be sensitive to different extents of syntactic integration difficulties. Furthermore, Osterhout, Holcomb and Swinney (1994) demonstrated a larger P600 to disambiguating words like *was* in sentences such as *the reporter saw the story was big* compared to *the reporter believed the story was big* (*saw* is more transitively-biased than *believed* and thus more likely to garden-path a reader towards the direct-object rather than the complement clause), thus demonstrating that the P600 can be sensitive to the lexical bias of the verb (under syntactically-ambiguous conditions). A number of studies have also demonstrated that the amplitude of the P600 in response to outright syntactic violations can be sensitive to the probability and salience of the violation (e.g. Coulson, King and Kutas, 1998; Gunter, Friederici, & Scriefers, 2000). Nevertheless, to our knowledge, no study has yet demonstrated the P600 to index the role of verb-bias in the modulation of outright errors (as opposed to syntactically-ambiguous conditions). Indeed, in their review of the component, Gouvea et al. (2010: 5) note "a scarcity of continuous measures of 'syntactic fit', in contrast to the readily available continuous measures of semantic fit that have been exploited extensively in studies of N400."

Another possible explanation for this null result is that certain design aspects of the present study may have reduced the sensitivity of ERP components to fine-grained grammaticality effects. First, the requirement for each verb to be presented 10 times in each level of the *Grammaticality* manipulation (transitive/intransitive) (necessitated by the requirement for each level of an ERP condition to consist of at least 40 items; e.g., Kaan, 2007) might have led to a repetitiveness that caused a more passive approach to online-processing (thus desensitizing online ERPs). Second, the stimuli were presented in the auditory domain (i.e., continuous speech). Although this method of presentation is very common in ERP studies (indeed, it has been used in previous studies of verb-argument structure overgeneralizations; e.g., Frisch et al., 2004; Kielar et al., 2012), ERP studies that have detected fine-grained effects of grammaticality (as opposed to simply comparing a violation against a well-formed sentence) have typically used visual word-by-word presentation of stimuli. Audio-presentation is arguably less suited to yielding fine-grained differences in ERP amplitudes (such as the modulating effect of frequency on grammaticality) because of (i) the rapid presentation of words, which may lead to the overlap of ERP components to successive words (thus reducing sensitivity of the measure); and (ii) difficulty in identifying the onset of the critical-word (as is the case with voiceless-fricatives such as *the*, which was the critical-word in the present study), which can cause temporal jitter in the temporal locking of ERP and thus reduce the sensitivity of the measure.

A second unexpected, though more incidental, finding was that overgeneralizations yielded a LAN-P600 effect as opposed to the N400-P600 effect observed in some previous studies of verb-argument structure overgeneralization (e.g., Friederici & Frisch, 2000; Frisch et al., 2004). Although some previous studies have demonstrated a LAN-P600 in response to these errors (e.g., Hagoort & Brown, 2000; Rosler et al., 1993), the N400-P600 pattern has been more common, and it is thus instructive to consider why the N400 was not elicited in the current study. One plausible explanation lies in the design of the study. Specifically, all ungrammatical sentences were transitive-causative sentences in which a noun-phrase (*the girl*) followed an intransitive-only verb. In contrast, all grammatical sentences were intransitive sentences in which a suitable prepositional phrase followed the intransitive verb. Thus, participants may have engaged in a processing strategy whereby ungrammatical sentences were detected by listening-out for the post-verbal determiner; a strategy that would be more likely to yield syntactic integration difficulties characterised by a LAN as opposed to the thematic integration difficulties that would have presumably yielded an N400.

To sum up, counter to our predictions, we found no evidence that the amplitudes of the ERP components (LAN and P600) that mark verb argument structure overgeneralizations are sensitive to verb frequency. As this is, to our knowledge, the first study to directly address this question, it would be wise to reserve judgment with regard to the question of whether this null finding is real or a consequence of particular design features of the present study (e.g., audio vs. visual presentation). One profitable direction for future research would be to conduct a meta-analysis of previous ERP verb-argument structure violation studies that examines whether the amplitudes of components observed in those studies are sensitive to the frequency of verbs used to realise the violation. Such a finding would be beneficial to the field, as it would demonstrate whether or not it is theoretically possible to use ERP components as a continuous dependent measure of the restriction mechanisms used to guide generalization of verbs. For now, one must be cautious with regard to the suitability of the ERP paradigm for examining the effect of statistical restriction mechanisms in the retreat from overgeneralization.

5. General Discussion

5.1. Recap of Thesis Objectives and Summary of Findings

The aim of the thesis was to examine whether pre-emption, entrenchment and semantic/ phonological verb properties are used by children as young as 3-4-years-old, and to develop new paradigms for doing so. The rationale for this aim is that previous research that has demonstrated a role for these mechanisms has been limited to using the grammaticality-judgment paradigm, which is unsuitable for testing children younger than 5-6, neglecting an age-group whose generalization mechanisms are at an earlier and more crucial stage of development. Another reason to develop alternative paradigms for examining these theories is that the judgment paradigm may not be sensitive to the full-range of restriction mechanisms used by children younger than 9-10, particularly in the case of verbal *un-* prefixation (where a previous judgment study found an effect of entrenchment and pre-emption for 9-10-year-olds but not 5-6-year-olds; Ambridge, 2012), and English regular past-tense (where a previous judgment study found an effect of the verb's phonological properties for 9-10 year olds but not 6-7 year olds; Ambridge, 2010). Indeed, the domains of verbal *un-* prefixation and English past-tense are the focus of *Study 1* and *Study 2* respectively because together, they provide compelling test-cases for the role of statistical-learning (entrenchment and pre-emption) and semantic/ phonological verb properties in the retreat from overgeneralization.

Study 1 used a production paradigm to examine the roles of statistical-learning and verb semantics in children's (age 3-4; 5-6) restriction of verbal *un-* prefixation overgeneralization errors (e.g., **unsqueeze; *unopen*). The youngest children's production of verbs in *un-* form (e.g., **unbend*) was negatively predicted by the frequency of the target verb in bare form (i.e., *bend/s/ed/ing*) (entrenchment) and by the frequency of synonyms to a verb's *un-* form (e.g., *straighten* for **unbend*) (pre-emption). Grammaticality judgments collected from older children (aged 5-6) revealed that the preference for a verb in *un-* form was positively related to the extent to which it denoted Whorf's (1956) 'cryptotype' of meanings (a probabilistic cluster of semantic features typically associated with *un-*prefixable verbs, but which cannot be easily defined). The findings demonstrate that children use probabilistic verb-semantics and statistical-learning mechanisms to restrict generalizations of *un-* prefixation from an early age.

Study 2 investigated children's generalizations of English past-tense, a domain which constitutes a particularly suitable test-case for the restrictive role of a verb's phonological properties because verbs which take 'regular' past-tense form appear to cluster into phonological neighbourhoods (e.g., *click/ clicked; trick/ tricked*), in a similar manner as verbs which take 'irregular' form (e.g., *sleep/ slept; keep/ kept*) (the theories under investigation agree that irregular past-tense forms are generalized based on a verb's phonological properties, but disagree as to whether this is the case for regular past-tense forms; e.g., Bybee & Moder, 1983; Prasada & Pinker, 1993). The likelihood of children's (aged 3-4; 5-6; 6-7; 9-10) producing novel verbs in regular past-tense form (e.g., *gezzed, chaked*) was positively associated with the novel verb's phonological similarity to existing regular verbs, consistent with the theory that regular past-tense forms are analogically generated on the basis of their phonological-similarity to other 'regular' verbs stored in associative memory.

Study 3 examined the suitability of the Event Related Potentials (ERP) paradigm for detecting restriction mechanisms used to avoid overgeneralization errors. The paradigm provides a valuable alternative to behavioural methods because it provides an online and implicit measure of sentence-processing with relatively low task demands. Previous ERP studies with adults (e.g., Hagoort & Brown, 2000; Rosler et al., 1993) have demonstrated that ERP components known as P600 and LAN are observed in response to verb-argument structure overgeneralizations (compared to well-formed sentences). *Study 3* sought to extend these studies by examining whether P600 and LAN are sensitive to statistical restriction mechanisms. Adult participants were presented with verb-argument structure overgeneralization errors containing high frequency verbs (e.g., *On Monday Bob laughed the girl in the garden*) or low frequency verbs (e.g., *On Monday Bob giggled the girl in the garden*). P600 and LAN were observed in response to overgeneralization errors, but the magnitude of these components was not influenced by the manipulation of verb-frequency, raising doubts regarding the suitability of ERP for detecting fine-grained differences in the relative acceptability of errors of this type.

5.2. Theoretical and Methodological Implications.

Taken together, the findings of *Study 1 and Study 2* indicate that children as young as 3-to-4- years-old restrict generalization of verbs to linguistic constructions based on the frequency of other formulations that convey the same intended message as the generalization (pre-emption), the verb's frequency in other constructions (entrenchment), and the extent to which the properties of the verb overlap with properties of the construction (fit). Implications of these findings are considered below with regard to accounts of retreat from error referred to in the Introduction. Since *Study 3* is more methodologically-motivated, its findings are considered separately under *Methodological Implications*.

5.2.1. Implications for statistical-learning accounts

Statistical-learning accounts posit that linguistic generalizations are probabilistically restricted by the statistical properties of the input. The precise input to which this learning mechanism is sensitive depends on whether one assumes learning by pre-emption or entrenchment, and thus it was important to investigate both of these factors. The pre-emption account (Clark & Clark, 1979; Goldberg, 1995) proposes that a linguistic-generalization is restricted based on the frequency with which the generalization is *not* attested when its communicative-function is required. Thus, the account holds that repeated use of an alternative formulation that serves an identical (or nearly-identical) communicative-function as the non-attested linguistic generalization constitutes ever-strengthening probabilistic evidence that the non-attested generalization is unacceptable. Conversely, the entrenchment account (e.g., Braine & Brooks, 1995) proposes that a linguistic generalization is restricted by the frequency with which the relevant verb is used in any other construction - regardless of the construction's communicative-function. The theory holds that a verb becomes increasingly entrenched in constructions in which it is repeatedly used and that this constitutes as ever-strengthening probabilistic evidence that the verb should not be generalised to a construction in which it has never been used.

Study 1 provided evidence that 3-4-year-old children use both pre-emption and entrenchment to restrict overgeneralizations of verbal *un-* prefixation. The likelihood of a verb's production in *un-* form (e.g, **unfill*) was negatively related to the frequency of the verb in other constructions (e.g., *fill[/s/ed/ing]*) and to the frequency of potentially pre-emptive forms (e.g., *empty/ drain*). Thus, it appears that even at the earliest stages of forming linguistic generalizations, a child's learning mechanism is sensitive to (i) the communicative-function of a formulation and the frequency with which it occurs (pre-emption), and (ii) the frequency of the verb in each construction (entrenchment).

One can argue that *Study 1* provides particularly convincing evidence for the co-existence of pre-emption and entrenchment because the domain of verbal *un-* prefixation allows a direct-comparison of each account. That is, the pre-emption measure (frequency of synonyms to a verb's *un-*form) is independent of the entrenchment measure (frequency of the relevant verb in other constructions) thus avoiding problems of covariance. By way of comparison, in the domain of verb-argument-structure (the chosen domain of most research into pre-emption and entrenchment; see *Section 1.4.3*), the pre-emption measure is the frequency of the relevant verb in the most-nearly-synonymous construction (e.g., transitive-causative overgeneralizations [*The clown laughed Mummy] are probabilistically-blocked by the frequency of the relevant verb in the periphrastic-causative [*The clown made Mummy laugh*]), and the entrenchment measure is the frequency of the relevant verb in *any* alternative construction (e.g., **The clown laughed Mummy* is probabilistically-blocked not only by use of periphrastic-causatives, but also intransitive-inchoatives [*Mummy laughed*]). Thus, in the domain of verb-argument-structure, measures of pre-emption are a subset of- and thus highly correlated with- measures of entrenchment. Indeed, a study of dative overgeneralizations reported a very high correlation between these measures (*r*=0.9; Ambridge et al., 2014), as did a study of locative overgeneralizations (*r*=0.7; Ambridge et al., 2012). In summary, *Study 1* indicates that 3-4-year-old children use pre-emption and entrenchment to restrict linguistic-generalizations, and the domain of verbal *un-* prefixation investigated by this study seems to be the most reliable direct-measure of pre-emption and entrenchment effects, at least in terms of familiar English constructions.

A broader question relates to how the effects of pre-emption and entrenchment interact. In other words, it is not enough for a statistical-learning account to simply posit that both effects exist (a position that is supported by the findings of *Study 1*). Statistical-learning accounts must specify the relative weighting of each effect and whether this weighting changes under different situations. For example, while the youngest children in *Study 1* showed both pre-emption and entrenchment effects, older children demonstrated effects of pre-emption only. Indeed, it is likely that the relative influence of pre-emption varies not only between age-groups, but also between different domains of overgeneralization. Whilst pre-emption seems to work very well for morphological overgeneralizations (since there is usually a direct-synonym for errors like **unopen* [*close*] and **sleeped* [*slept*]), it may play a less-important role in restricting some overgeneralizations at the syntactic-level. For example, a pre-emption account holds that transitive-causative overgeneralizations of *laugh* are blocked only by uses of the periphrastic-causative construction (see above). However, the periphrastic-causative construction does not offer a perfectly synonymous communicative-function [it denotes an agent indirectly - as opposed to directly - causing a patient to perform an action] and is quite rare in child-directed speech, thus providing little opportunity to learn by pre-emption.[18] On the other hand, if one posits learning by entrenchment, one can learn to

[18] A recent demonstration of this is in Ambridge et al.'s (2015) study of transitive overgeneralization errors, where the authors compiled frequency-counts of each test-verb (*laugh*,

restrict transitive overgeneralizations by witnessing the verb in highly-frequent constructions like the intransitive, thus providing plentiful opportunities to learn this restriction. Ambridge et al. (2015) demonstrated that, when controlling for covariance (by running separate analyses for each predictor) and testing across a range of argument-structure constructions, adults' and children's (age 9-10) acceptability judgments of overgeneralization errors were negatively predicted by the frequency of the verb in all other constructions (entrenchment) but not by the frequency of the verb in only the most-nearly-synonymous construction (pre-emption).

Thus, whilst *Study 1* has demonstrated that any statistical-learning account must incorporate pre-emption and entrenchment effects (even at the earliest stages of forming generalizations) another challenge is to explain why under some circumstances, entrenchment seems to play a more robust role than pre-emption (e.g., at the argument-structure level) and in other circumstances, pre-emption may play a more important role (e.g., at the morphological-level and perhaps at different stages of development). The answer is likely to lie in an account that posits competition between 'pre-emptive' constructions (including in the morphological domain, fully-lexically specified constructions; e.g., *open* pre-empts **unclose*) and 'entrenched' constructions. Activation of a pre-emptive or entrenched construction is determined not only by the verb's frequency in that construction, but by other factors that vary based on the domain of overgeneralization and age-of-participant. One such factor may be the *extent* to which a construction can be considered synonymous (in terms of communicative-function) to the overgeneralized construction. That is, pre-emption may have more explanatory power than entrenchment when the pre-emptive construction is considered highly-synonymous. Another factor may be each construction's overall frequency in the input: A pre-emptive construction must be frequently available in the input in order for pre-emptive learning to take place. One framework that is compatible with these possibilities is the *FIT* account (outlined in *Section 1.4.4*), and I explain why this account is well-suited to explaining the findings of the present thesis – that is, pre-emption, entrenchment and probabilistic semantic/ phonological verb effects - in section *5.2.3*.

---

giggle, fall, tumble) from SUBTLEX-UK corpus data; corpus uses of each test-verb were classed as 'transitive', 'intransitive', 'passive', 'periphrastic', or 'other construction', with the sum frequency-count (i.e., verb-use in all constructions) instantiating the entrenchment measure and 'periphrastic' frequency-count instantiating the pre-emption measure. Just 2.5% of verb-uses were 'periphrastic', compared to 93.3% 'intransitive'.

5.2.2. Implications for Pinker's verb-class account and rule-based account

The major claim of Pinker's (1989) verb-class account is that a verb's acceptability in a particular argument-structure is determined by its membership of a suitable narrow-range verb-class, and that a verb can be assigned to one of these classes only if it meets fine-grained semantic or morphophonological criteria. For example, verbs from a *motion in a particular manner* class are acceptable in the transitive-causative structure (e.g., *The man rolled the ball*), but verbs from a *semi-volitional expression* class are not (e.g., **The man laughed the woman*). Membership of a verb-class is discrete as opposed to probabilistic, and thus any two verbs from the same verb-class that are overgeneralized to a particular argument-structure (e.g., **The man laughed/giggled/chuckled the woman*) are predicted to be equally ungrammatical in that structure.

Although Pinker does not explicitly use the verb-class account to explain restriction of verbal *un-* prefixation errors, he does make explicit his aim to "leave no negative exceptions" to this account (Pinker, 1989:103). With this in mind, together with the fact that Pinker does not propose a separate account for *un-* prefixation errors, the verb-class account must be extended to explain restriction of *un-* prefixation errors.

At first glance, one might infer that the effect of verb-semantics observed in *Study 1*, where the acceptability of a verb's *un-* form was positively predicted by the extent to which that verb denotes semantic properties associated with verbal *un-* prefixation, is in line with the verb-class account. However, the verb-class account struggles to explain the precise relationship between a verb's semantic properties and its grammaticality in *un-* form. First, there is no obvious proposal for how *un-* verbs (i.e., verbs that can be prefixed with *un-*) may cluster into any set of discrete semantic-classes (unlike, for example, *verbs of contained motion in a particular manner,* which is a class of verbs that can be used in the transitive argument-structure [e.g., *roll, skid, bounce, float*]). Rather, it appears that *un-* verbs probabilistically cluster into a fuzzy "semantic cryptotype" of meanings (e.g., *covering, enclosing, attaching, circular motion, change of state, binding/locking*; e.g., Whorf, 1956) where no individual semantic property is either necessary or sufficient for a verb to be used with *un-* (emphasising why it is difficult to posit a discrete verb-class to which a verb is or is not a member). Second, even if one could propose a suitable set of verb-classes to define *un-* verbs, Pinker's account holds that membership of such a class is discrete as opposed to probabilistic and that this should be reflected in grammaticality judgements such that any two verbs from a (hypothetical) verb-class which licences *un-* prefixation should be rated as equally grammatical in *un-* form. *Study 1* indicates that this is not the case, given the significant positive correlation between a verb's acceptability in *un-* form and the extent to which it denotes Whorf's cryptotype.

The effect of verb-semantics observed in *Study 1* is better explained by children's use of a productive *un*[VERB] construction, which is formed by analogising across stored exemplars of that construction (e.g., *unroll, unbutton*, etc.). Thus, akin to Whorf's hypothesised semantic cryptotype, the [VERB] slot of this schema exhibits fuzzy and probabilistic semantic properties shared by items that have previously occupied that slot in input utterances. The likelihood of a verb being produced or rated as grammatical in *un-* form is determined probabilistically by the overlap between the semantic properties of the verb and those of the [VERB] slot in the *un*[VERB] schema (e.g., Langacker, 2000).

The position taken here is that generalization of a verb to a construction is determined by analogy to stored-exemplars of that construction (which may be instantiated as a probabilistic schema such as *un*[VERB]), as opposed to categorical membership of a discrete verb-class. The greater the extent to which a verb shares properties with other verbs that have been attested in a construction, the more likely the verb will be used or rated as grammatical in that construction.

The same analogy-based approach is well-suited to explain the findings of *Study 2*, in which the likelihood of a novel verb's regular past-tense form being produced (e.g., gezz/*gezzed*, *chake/chaked*) was positively associated with the novel verb's phonological similarity to familiar verbs that have been attested in regular past-tense form. Here, one can posit that children are in use of three 'regular' past-tense schemas, [VERB]-*d*, [VERB]-*t*, and [VERB]-*əd,* to account for the fact that the regular past-tense morpheme may be realised as *–d*, (e.g., *played*), *-t* (e.g., *wish/wished*) or *-əd* (e.g., *hunt/hunted*). The [VERB] slots of these schemas exhibit fuzzy and probabilistic phonological properties shared by verbs previously used in that schema, and the likelihood of a verb being produced in regular past-tense form is determined probabilistically by the overlap between the phonological properties of the verb and those of the [VERB] slot in a regular schema. The reason why generalization is determined by a verb's phonological properties as opposed to semantic properties relates to a key assumption of the constructivist approach (see *Section 1.2*). That is, constructions are acquired by abstracting across stored exemplars of the construction, with each slot of a construction exhibiting any property (e.g., semantic, phonological, pragmatic, etc.) that is shared by lexical-items previously used in that slot.[19]

[19] The [VERB] slot of a construction does not have to exhibit wholly semantic or wholly phonological properties. For example, the [VERB] slot of the *PO-* dative construction is posited to exhibit semantic properties (e.g., caused-motion) and morphophonological properties (disyllabic

Since verbs that are used in regular past-tense form cluster into different phonological neighbourhoods (e.g., the [VERB]-*t* construction is formed by *miss/missed; kiss/ kissed; hiss/ hissed* but also *wish/wished and dish/dished*), the [VERB] slot of a construction exhibits a fuzzy and probabilistic constellation of phonological properties shared amongst these verbs.[20]

The analogy-based approach outlined above to explain the findings of *Study 2* is consistent with the assumptions of the *single-route* model (e.g., Bybee & Moder, 1983), which attributes the formation of regular past-tense forms to analogy across phonologically-similar regular past-tense forms in associative memory (e.g., *wiss/wissed* from *miss/missed; kiss/ kissed*). Conversely, it is difficult to explain these findings in terms of the *dual-route* model (e.g., Prasada & Pinker, 1993), which posits that formation of regular past-tense forms are best attributed to an innate context-free rule (i.e., add "*-ed*" to any instance of the category "VERB"). More broadly, this relates to a theoretical divide between generativist and constructivist approaches with regard to how language is represented in the brain. On the one hand, the notion of a context-free rule applied to the discrete category "VERB" is akin to the generativist assumption that language is represented as a set of phrase-structure rules that act on abstract syntactic categories (e.g., add "*–ed*"). On the other hand, the notion of analogy across phonologically-similar past-tense forms (supported by the findings of *Study 2*) is akin to the constructivist assumption that there is no need to posit abstract rules that operate over discrete abstract categories and rather, language use is best explained by analogy across memorized exemplars.

Overall it appears that children as young as 3-4-years-old form and restrict linguistic generalizations by analogising to stored exemplars of a construction, with the likelihood of a verb's generalization to that construction dependent probabilistically on its similarity to any property that is shared by verbs previously used in the construction. Thus, the relevant judgment and production data are not well explained by a system of abstract rules that act on discrete categories – whether this applies to syntactic categories (e.g., add

with second-syllable stress/ trisyllabic) (see *Section 1.4.4.3*). Such cases demonstrate an advantage of positing a probabilistic learning mechanism, where the likelihood of generalization is determined by a number of different competing factors such as a verb's phonological and semantic properties. Predicting the relative contribution of phonological and semantic properties is likely to require computational modelling.

[20] The reason why a verb's semantic properties are not posited to influence the likelihood of its generalization to regular or irregular past-tense constructions is that verbs which have been experienced in the relevant [VERB] slot (e.g., [VERB]-*t*) exhibit heterogeneity with regard to semantic properties and thus, the [VERB] slot also exhibits heterogeneity with respect to that property (that is, past-tense constructions such as [VERB]-*t* are open to a wide range of semantic properties).

"-*ed*" to any instance of the category "VERB") or semantic verb classes (e.g., a narrow-range rule that acts on any verb that is part of an 'alternating' verb-class).

5.2.3. Implications for the FIT account/ hybrid accounts

The discussion outlined above demonstrates that young children's formation and restriction of linguistic generalizations must be explained by a hybrid-account that includes a role for pre-emption, entrenchment and graded semantic/ phonological effects, and which specifies how these factors interact. In line with this view, the *FIT* account (e.g., Ambridge & Lieven, 2011) explains the retreat from error by a probabilistic process of competition between constructions. A speaker is assumed to possess an inventory of constructions that compete to convey the intended message, with the winning construction determined by (i) the extent to which the communicative-function of the construction is relevant to the intended message ('*relevance'*), (ii) the frequency with which the relevant verb has been used in the construction ('*item-in-construction frequency*'), and (iii) perhaps most importantly, the extent to which properties of the verb overlap with those exhibited by the slot in which it is to be used ('fit').

The assumption that constructions receive activation based on *item-in-construction frequency* yields entrenchment effects because every use of a verb in a particular construction strengthens its link with that construction at the expense of its link with other constructions. The assumption that activation is determined additionally by the *relevance* of the construction yields pre-emption effects (independent from entrenchment effects) because constructions that are relevant to the message receive an activation boost (adding to any activation a construction might already have from *item-in-construction frequency*). Indeed, the combination of *item-in-construction frequency* and *relevance* can potentially explain why the effect of pre-emption is more robust than entrenchment (and vice-versa) in different domains of argument-structure – a point that is returned to in *Section 5.3: Future Directions*.

The graded effects of a verb's semantic properties (*Study 1*) and phonological properties (*Study 2*) can be explained under the notion of 'fit', which can be thought of in the same way as the analogy-based approach described in *Section 5.2.2.*. The account holds that each construction is formed by analogising across stored exemplars of that construction in associative memory and thus, that a [VERB] slot in that construction exhibits a fuzzy and probabilistic set of any perceivable properties shared by verbs that have previously occupied that slot.[21] The greater extent to which a verb's properties 'fit' with those of the [VERB] slot of the construction, the more likely the verb is to be generalized to, and/ or rated as grammatical in that construction.

Although the *FIT* account is, in the author's words, " 'new' only to the extent that it combines elements of previous proposals in a novel way" (Ambridge et al., 2012: 24), it is the only account to my knowledge that can potentially explain the findings of this thesis. Indeed, the thesis provides compelling support for this framework in three ways. The first relates to the evidence for children's use of 'fit' in domains that serve as ideal test-cases for this notion: verbal *un-* prefixation and English regular past-tense. No semantic or phonological property is deemed necessary or sufficient for a verb to be used in these constructions (unlike constructions such as the transitive, which requires relatively clear-cut semantic criteria) and rather, these constructions are associated with a probabilistic cryptotype set of semantic or phonological properties. Second, the finding that children use pre-emption and entrenchment to restrict use of verbal *un-* prefixation (recall that this domain provides arguably the most reliable independent-measure of pre-emption and entrenchment effects) indicates use of a probabilistic restriction mechanism that yields these effects in addition to semantic/ phonological fit (as posited under the *FIT* account). Third, the effects of pre-emption, entrenchment and fit have all been demonstrated by children as young as 3-4-years-old, an age-group that for methodological reasons has largely been neglected by previous research. It was important to extend previous findings of pre-emption, entrenchment and fit to this age-group because these children are in the process of acquiring their generalization mechanisms, as opposed to older children who are in use of a relatively mature system that may not reflect how children have learned to restrict errors in the first place.

The findings of semantic and phonological fit must also be discussed from a developmental perspective. Whilst 3-4-year-old children demonstrated effects of phonological fit (*Study 2*) but *not* semantic fit (*Study 1*), 5-6-year-old children demonstrated both of these effects. Thus, whilst it is clear that both age-groups can restrict linguistic generalizations by analogising across stored exemplars of the relevant construction, the younger age-group showed sensitivity only to phonological regularities (across exemplars of regular past-tense forms), and not to semantic regularities (across

[21] The account is agnostic with regard to whether speakers use constructions that are formed 'on-the-fly' by analogising across stored exemplars (i.e., an exemplar model) or whether these constructions 'stand-alone' as representations that are independent of the stored exemplars (i.e., a prototype model).

exemplars of *un*- forms). The delayed effect of semantic fit can potentially be explained by reasoning that the slot in the *un*[VERB] construction is associated with an especially complex set of semantic properties (hence Whorf's use of the term 'cryptotype') which take longer to learn than the relatively simple set of phonological properties associated with the [VERB] slots of regular past-tense constructions (i.e., [VERB]-*d*, [VERB]-*t*, [VERB]-*əd*). Indeed, this reasoning is consistent with the *FIT* account's assumption that the properties associated with construction slots are built gradually over time, and thus, whilst knowledge of slot-properties is non-adult-like, children will be unaware of a suboptimal fit between a verb and its slot. Future studies should investigate whether 3-4-year-old children demonstrate effects of semantic fit with other constructions such as the transitive-causative construction, which is associated with a perhaps-simpler set of semantic properties that may be learned relatively quickly.

Overall, the findings of the thesis demonstrate that the retreat from overgeneralization error involves an interactive, probabilistic learning mechanism that yields effects of pre-emption, entrenchment and semantic/ phonological fit. One caveat to this conclusion is that, as yet, there is little evidence that it can be extended to more implicit, online measurements of sentence-processing and thus this could be a valuable direction for future research to explore. For this reason, *Study 3* of the thesis investigated the suitability of the Event Related Potentials (ERP) paradigm for investigating the relative-acceptability of overgeneralization errors; its findings are considered below.

5.2.4. Methodological Implications

A key aim of this thesis was to develop existing experimental paradigms for the purpose of investigating the retreat from overgeneralization error amongst younger children who had largely been neglected by previous research. *Study 1* and *Study 2*, which used production-priming and elicited-production paradigms respectively, achieved this by adapting behavioural paradigms to be suitable for testing children as young as 3-4-years-old. The purpose of *Study 3* was to extend the investigation of retreat from overgeneralization beyond the behavioural paradigm, namely to the Event Related Potentials (ERP) paradigm, which provides an online and implicit measure of sentence-processing with relatively low task demands. *Study 3* investigated whether ERP components that are known to mark overgeneralization errors can be used to detect the mechanisms demonstrated by behavioural paradigms to restrict these types of errors.

Since the central purpose of *Study 3* was methodological, the implications of its findings have not been discussed until now. The finding that P600 and LAN components were evoked by verb-argument structure overgeneralization errors (vs. well-formed sentences) was expected given previous studies of the same type (e.g., Rosler et al., 1993; Hagoort & Brown, 2000). Based on the traditional interpretation of these components, one can posit that the LAN reflected the online detection of the violation, whereas the P600 reflected integration difficulties arising from this violation.

However, the finding that the magnitudes of P600 and LAN were not sensitive to a manipulation of verb-frequency (e.g., **The clown* <u>*laughed*</u> *the boy* vs. **The clown* <u>*giggled*</u> *the boy*) raises doubt regarding the suitability of these components for detecting fine-grained differences in the relative acceptability of overgeneralization errors. Since a large number of previous behavioural studies have demonstrated that participants use verb-frequency to restrict overgeneralizations (e.g., Ambridge, Pine, Rowland & Young, 2008; Bidgood, Ambridge, Pine & Rowland, 2014; Brooks, Tomasello, Dodson & Lewis, 1999; Brooks & Zizak, 2002; Robenalt & Goldberg, 2015; Theakston, 2004; Wonnacott, Newport & Tanenhaus, 2008), it is worth reconsidering the plausibility of the claim that P600 or LAN can be used to detect these fine-grained differences. A good starting point is to review the circumstances in which the components are *known* to occur.

Both P600 and LAN occur in response to outright (morpho-)syntactic violations such as violations of local phrase-structure (e.g., **The broker hoped to sell the stock was sent to jail;* Osterhout & Holcomb, 1992), pronoun case-marking (e.g., **The plane took we (=us) to paradise*), subject-verb number agreement (e.g., * *Every Monday he *mow (=mows) the lawn;* Coulson, King & Kutas, 1998) and verb-argument-structure (e.g., **The man was fallen down*; e.g., Rosler et al., 1993). Unlike LAN however, P600 has also been demonstrated to occur in response to (grammatically acceptable) syntactically-ambiguous sentences such as garden-path sentences, and syntactically-complex sentences such as *wh*-dependencies (e.g., Gouvea, Phillips, Kazanina, & Poeppel, 2010; Osterhout & Holcomb, 1992). In line with these findings, the consensus is that LAN reflects the online detection of a (morpho)syntactic violation, whereas P600 reflects integration-difficulties that arise not only from these violations, but also any surprising or salient sentence continuations (e.g., Kotz & Friederici, 2003). With these definitions in mind, it is perhaps not surprising that the only evidence for these components' sensitivity to the relative acceptability of different syntactic manipulations (i.e., syntactically-ambiguous vs. outright-syntactic violations; violation vs. violation [e.g., agreement violations vs. case-marking violations]) comes from studies of the P600. Even so, there have been only a handful of such studies, leading Gouvea et al. (2010: 5) to remark that there is a "scaricity of continuous measures of 'syntactic fit', in contrast to the readily available continuous measures of semantic fit that have been exploited extensively in studies of N400."

Most notably, Osterhout, Holcomb and Swinney (1994) presented participants with reduced-relative-clause sentences in which the verbs varied in transitivity-bias, creating four conditions: intransitive-only verbs (e.g., *The doctor decided the prescription had changed*), intransitive-biased verbs (e.g., *The doctor remembered the prescription had changed*), transitive-biased verbs (e.g., *The doctor forgot the prescription had changed*) and transitive-only verbs (e.g., *The doctor followed the prescription had changed*). The logic here is that the bias of the verb dictates the likelihood of the post-verbal NP being interpreted as a direct-object of the verb (if used transitively) or the subject of the complement clause (if used intransitively) (e.g., Trueswell, Tanenhaus, & Kello, 1993). A P600 was evoked at the disambiguating word (*was*) in the transitive-biased condition relative to the intransitive-biased condition and crucially, a significantly larger P600 was evoked for the transitive-only condition. Thus, a verb's bias *can* modulate the amplitude of the P600 when two syntactically *ambiguous* sentences are compared (i.e., intransitive-biased vs. transitive-biased), and when a syntactically-ambiguous sentence is compared against an outright syntactic-violation (i.e., transitive-biased vs. transitive-only). Although this indicates that P600 can be sensitive to fine-grained syntactic manipulations, the question that is of more interest to this thesis is whether a verb's bias can also modulate the P600 evoked by two sentences that *both contain a syntactic violation* (i.e, verb-argument structure violations). The results of *Study 3* suggest this may not be the case. Nevertheless, given that other studies have demonstrated that the amplitude of the P600 is modulated by the probability and salience of subject-verb number agreement violations (e.g., Coulson et al., 1998), it is still realistic to believe that the P600 evoked by overgeneralization errors can be modulated by the frequency of a verb. For now, though, one must be cautious with regard to the suitability of the ERP paradigm for examining this issue. A number of methodological considerations for future research are considered in the manuscript for *Study 3* (*Section 4*).

A discussion of methodological implications would not be complete without briefly considering methodological issues of *Study 1* and *Study 2*. *Study 1* used a priming-production paradigm to demonstrate 3-4-year-old children's use of pre-emption and entrenchment to restrict use of *un-* prefixation. *Study 2* used a more traditional elicited-production paradigm to demonstrate 3- to 4- year-old children's use of phonological fit to restrict use of regular past-tense marking. The findings are methodologically significant because previous developmental investigations of these effects have used judgment paradigms, which detected these effects (with respect to the relevant domain of overgeneralization) for only children aged 9-10, despite testing children as young as 5-6. Thus, compared to the judgment paradigm, the production paradigm appears to be more sensitive to young children's use of pre-emption, entrenchment and phonological fit.

On the other hand, *Study 1* demonstrated that the judgment paradigm – but not the production paradigm – was sensitive to 5-6-year-olds' use of semantic fit, perhaps reflecting on the judgment paradigm's suitability for detecting such fine-grained semantic effects. Overall, the findings of the thesis demonstrate the importance of selecting a paradigm that is most suitable for examining the effect of interest and the age-group of interest. At the same time, the field should work towards developing experimental paradigms so that they are optimally sensitive to the full-range of restriction mechanisms used in the retreat from error, and which can potentially be applied to children even younger than those studied in this thesis. Any theory of retreat from overgeneralization should aim to test its predictions using both behavioural measures (i.e., production and judgment paradigms) and more implicit measures of sentence-processing such as ERP and eye-tracking.

5.3. Directions for Future Research

The findings of the thesis demonstrate that any successful account of the retreat from overgeneralization error must yield probabilistic effects of pre-emption, entrenchment and semantic/ phonological fit. The major challenge facing future research in this field is to specify the relative contribution of each of these effects and how this might change under different conditions.

5.3.1. Relative contributions of pre-emption and entrenchment

Section 5.2.1 has already alluded to the fact that under some circumstances, entrenchment plays a more robust role than pre-emption (e.g., at the argument-structure level) and in other circumstances, pre-emption plays a more robust role (e.g., at the morphological-level and possibly at later stages of development). The *FIT* account (e.g., Ambridge & Lieven, 2011) (*Section 1.4.4; Section 5.2.3*) can potentially explain this by proposing that frequency effects of 'pre-empting' and 'entrenching' constructions fall out of the same statistical-learning mechanism (i.e., item-in-construction-frequency) but that the relative weighting of these constructions varies according to their *relevance* to the intended message. In order to evaluate the predictions that arise from this account, it is necessary to outline this proposal in a little more detail.

The first assumption is that all constructions compete to convey the intended message, with the relative effects of each construction determined by the frequency with which the relevant verb is experienced in the construction (*item-in-construction-frequency*). Entrenchment effects fall naturally out of this process because any construction that is not the error-construction receives activation at the expense of the error-construction. For example, **The clown laughed Mummy* is blocked by the frequency with which *laugh* is used in the periphrastic-causative [*The clown made Mummy laugh*], the intransitive [*Mummy laughed*], the appropriate fully-lexically-specified construction [*Laugh*!], etc. Note that all constructions, including 'pre-empting' constructions (i.e., those considered most-nearly-synonymous to the error-construction) exert their own frequency effect under the notion of *item-in-construction-frequency* effects. However, in the domain of inflectional morphology and verb argument-structure, the 'pre-emption' effect yielded by *item-in-construction-frequency* is simply a subset of the overall entrenchment effect, unless an additional factor is posited that adds weight to the pre-emptive-construction. An additional factor is needed to account for circumstances in which effects of pre-emption are independent from entrenchment effects (this is almost certainly the case in the domain of English past-tense; *sle[pt]* for **sleep[ed]; not sleep[s/ing]* for **sleep[ed]*).

For this reason, *FIT* posits that activation of constructions is determined not only by *item-in-construction-frequency*, but also by the *relevance* of the construction to the intended message. Thus, at least in domains of overgeneralization where there exists a near-perfect alternative-phrasing to the error-construction (e.g., *slept* for **sleeped*; *empty* for **unfilled*), pre-emptive constructions outcompete more distant 'entrenching' constructions because the former receive overwhelmingly higher activation based on near-perfect relevance*:* the effect of *item-in-construction-frequency* will be more apparent for the (relevant) pre-emptive construction than (less-relevant) entrenching-constructions. In other domains, such as transitive-causative overgeneralizations, where the pre-empting construction is only moderately relevant, the weighting of each construction's *item-in-construction-frequency* effect will be more evenly spread between the (moderately-relevant) pre-emptive construction and the (slightly-less relevant) entrenching-constructions. The key point is that the account posits no clear divide between pre-emption and entrenchment effects. The relative weighting of a 'pre-empting' or 'entrenching' construction's *item-in-construction frequency* effect is determined by the construction's relevance to the intended message.

Indeed, recent evidence suggests that the relevance of a construction to the intended message may even have a *moderating* role on the influence of frequency effects, at least under extreme circumstances where there is no alternative-phrasing to a novel generalization. That is, *item-in-construction frequency* may exert an effect only if the construction meets at least some relevance-threshold. Robenalt and Goldberg (2015) demonstrated that adults preferred novel uses of a low-frequency verb to its higher-frequency equivalent, but only for sentences for which there exists a competing alternative-phrasing (e.g., **The magician vanished/disappeared the rabbit*; cf. *The magician made the rabbit vanish/ disappear*). Participants had no preference for novel uses of low vs. high frequency verbs when there was no competing alternative-phrasing to the sentence (e.g., ?*The chief will chuckle/laugh you back to your deskjob*). The *FIT* account explains these effects by positing that all competing-constructions – 'pre-empting' and 'entrenching' – fall on a gradient of relevance to the intended message and that as this relevance declines, less weight is given to a construction's *item-in-construction-frequency*; when a construction is completely irrelevant to the intended message (as is the case with any other construction that competes to convey the message in *The chief will laugh you back to your deskjob*), its *item-in-construction-frequency* exerts no influence. In other words, a construction must be at least *somewhat* relevant in order for *item-in-construction-frequency* to exert an influence.

Future research should investigate the extent to which *item-in-construction frequency* effects can be mediated by the relevance of a construction. For example, holding verb-use constant, the *FIT* account predicts that a novel generalization with a near-perfectly relevant competing construction will be dispreferred to a novel generalization with only a moderately-relevant competing construction; a novel generalization with only a moderately-relevant competing construction will be dispreferred to a novel generalization with only a minimally-relevant competing construction. A profitable paradigm to investigate these predictions is Artificial Grammar Learning (AGL) (e.g., Casenhiser & Goldberg, 2005; Perek & Goldberg, 2015; Wonnacott et al., 2008; Wonnacott, 2011) which avoids confounds that arise when these effects are examined in natural language corpora. For example, Casenhiser and Goldberg (2005), successfully taught 5-to-7-year-old children a novel [PERSON] [LOCATION] [NOVEL-VERB] construction that was associated with semantics of 'appearance' (children's understanding of the construction semantics was assessed using a forced-choice pointing paradigm). Future studies can use this paradigm to compare the frequency effects of different novel constructions that vary in their relevance to the intended message (e.g., describing an appearance scene).

The research should also investigate whether the balance between relevance and *item-in-construction frequency* changes across development. Such investigations may provide further insight into why very young children (say, at 2-to-4-years-old) tend to be very conservative in their generalizations (e.g., Akhtar, 1999; Tomasello, 2003). If, as posited under constructivist accounts, these children have less adult-like knowledge of the semantic properties associated with a construction, they will likely be less proficient at gleaning the relevance of different competing-constructions, thus relying more on *item-in-construction-frequencies* (i.e., lexical-conservatism) and less on the relevance of a construction (which can be hypothesised to motivate creative generalizations).

Developmental investigations may also provide insight into why, with age, children become more creative with their generalizations. Once children have developed knowledge of the semantic properties associated with competing-constructions that is sufficient to glean varying levels of relevance, children may begin to override *item-in-construction-frequencies* in favour of relevance (in a similar manner as demonstrated by adults in Robenalt & Goldberg, 2015). As has been demonstrated in the present thesis however (and discussed in the next sub-section, *5.3.2*), young children by no means have *adult-like* knowledge of constructions' semantic (or phonological) properties, and thus, their creativity will lead to occasional *over*-generalization of verbs into semantically (or phonologically) unsuitable constructions.

5.3.2. Relative contribution of fit

The central assumption of the *FIT* account is that overgeneralization errors are a consequence of a lexical item being used in a construction-slot with which it is not compatible. The thesis has supported this assumption by demonstrating that young children's linguistic generalizations are restricted by the extent to which the phonological or semantic properties of a verb 'fit' with the properties associated with those of the [VERB] slot in a construction. However, less is known about how this ability develops. The *FIT* account makes two developmental predictions that should be investigated by future research. First, the account predicts that children are more likely to produce overgeneralization errors when they have yet to fully acquire the semantic properties of the relevant verb. Gropen, Pinker, Hollander and Goldberg (1991b) have provided preliminary evidence for this prediction. Children (aged 2-3; 4-5; 5-6) who misunderstood the meaning of the verb *fill* (assessed using a forced-choice pointing paradigm where children picked between a picture of a ¾ full cup [incorrect] and completely-full cup [correct]) were significantly more likely to erroneously use the verb in a *contents-locative* (e.g., **Fill water into the glass*) than a container-locative (*Fill the glass with water*). The finding is consistent with the notion that children use constructions with semantically-inappropriate verbs when they misunderstand the semantics of the verb. However, future research should investigate this possibility with a broader set of verbs as well as within other domains of overgeneralization.

The second important prediction of the *FIT* account is that overgeneralization errors occur when a child has not yet developed adult-like knowledge of the properties associated with a [VERB] slot in a construction and consequently uses an inappropriate verb in that slot. Future studies should investigate this prediction using AGL paradigms such as the one introduced by Casenhiser and Goldberg (2005) (see *Section 5.3.1*). *FIT* predicts that children who do not understand a novel construction's semantics, are more likely to overgeneralize semantically-unsuitable verbs into this construction. Future studies should also investigate whether children are more likely to generalize to novel constructions with a complex set of semantic properties (similar to the notion of Whorf's cryptotype) than another novel construction with relatively simple set of semantic properties. If this were the case, it would be consistent with the assumption that *Study 1*'s findings, where a semantic fit effect was observed in 5-6-year-old children but not 3-4-year-old children, can be attributed to younger children's difficulty with learning the complex semantic properties associated with verbs that use *un-* prefixation.

Another prediction made by the *FIT* account is that generalization of a verb to a particular construction will be more likely if the [VERB] slot of the construction has high heterogeneity (i.e., if the range of verbs that have previously appeared in that slot are not particularly similar on any dimension). The assumption goes against the more traditional approach that the likelihood of generalization is determined by a construction's type-frequency (i.e., the number of different verbs that have appeared in that slot; e.g., Bybee, 2001) and instead posits that type-frequency effects are potentially a proxy for a slot's heterogeneity. Again, the AGL paradigm is well suited to investigate these effects, such that one can pit type-frequency versus homogeneity-of-the-slot. The *FIT* account would predict that a low-type-frequency construction can nevertheless see high levels of generalization, as long as verbs that have previously been used in that slot are not very similar to one another. On the other hand, a high-type frequency construction which has been used with verbs that share a narrow set of semantic properties should see relatively low levels of generalization.

5.3.3. Other considerations

Future research should also consider individual differences in the retreat from overgeneralization. Motivation for such research comes from a recent study by Kidd and

Arciuli (2016). Children's (age 6-8) comprehension of passives and relative clauses was predicted by individual performance on a test of statistical learning ability which used non-linguistic visual stimuli, thus indicating that the statistical learning mechanism used to support syntax acquisition is domain-general, not language-specific. A similar test of statistical-learning ability could be applied to investigate whether children's performance on this task predicts the likelihood of their producing an overgeneralization (e.g., those who score poorly may produce more errors). If the task is found to predict overgeneralizations made by young children but less-so with older children, this may demonstrate that statistical-learning is more strongly implicated early in the retreat from error, with a more prominent role for other factors like relevance and fit emerging with development.

5.3.4. Practical Applications

The thesis has outlined the theoretical importance of investigating younger children's restrictions of overgeneralization errors and has argued that future research should investigate the relative contributions of different statistical and semantic cues to generalization. It is now important to consider the practical application of this research.

The research can benefit children with specific language impairments (SLI), who constitute a significant minority of the child population (3-7%; Conti-Ramsden, Botting & Faragher, 2001). Children with SLI fail to acquire language normally despite no cognitive or perceptual impairments, and treatments for SLI and other language disorders are informed by research on typically developing children. Identifying how typically developing children use different statistical and semantic cues from the input to retreat from overgeneralization errors can help to inform speech and language therapists about the input required to successfully retreat from errors. From here, training interventions can be designed to ensure that SLI children are exposed to the input that is needed to restrict errors.

On a related note, it is also important to disseminate research findings to stakeholders such as teachers and parents, especially since it is often reported that they lack the information needed to monitor a child's language development. As well as arranging traditional dissemination activities (such as visiting schools and nurseries to present research), it will be useful to harness new technologies (e.g., smart-phone applications) which can make it easier for these stakeholders to monitor and assess children's language development. Indeed, more effective dissemination to stakeholders can help earlier detection of atypical language use and thus improve the chances of SLI children receiving early training interventions (which is crucial to prevent further setbacks in language development).

5.4. Final conclusions

The thesis began by outlining the need to develop experimental paradigms for the purpose of testing young children's retreat from overgeneralization, and to investigate what these paradigms tell us about this process. The findings of the thesis show that children as young as 3-4-years-old probabilistically restrict generalization of verbs to linguistic constructions based on the verb's frequency in other contexts (entrenchment), the frequency of other formulations that convey the same message (pre-emption), and the extent to which the properties of the verb overlap with properties of the construction (fit) (although the latter effect was observed only in terms of phonological fit [*Study 2*], with an effect of semantic fit emerging by 5-6-years-old [*Study 1*]). The conclusion reached was that these findings are best interpreted in terms of an interactive, probabilistic learning mechanism where constructions compete to convey the speaker's intended message based on relevance, statistical entrenchment of the lexical-item, and fit between the properties of the item and those of the construction slot in which it is used.

The thesis also highlights the importance for researchers in this field to seek converging evidence across a variety of experimental paradigms before reaching final conclusions. The task for future theoretical and experimental work is to specify the relative contribution of pre-emption, entrenchment and fit in the retreat from error, and how this might change across development.

Overall, the research summarised in this thesis indicates that even young children's generalizations are sensitive to the linguistic input (i.e., statistical regularities and generalized semantic or phonological patterns of use) and are not well explained by a system of abstract rules that act on discrete categories, whether this applies to syntactic categories (e.g., add "*-ed*" to any instance of the category "VERB") or semantic verb classes (e.g., a narrow-range rule that acts invariably on any verb that is part of an 'alternating' verb-class).

6. References

Abbot-Smith K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23,* 275-290.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90*, 119-161.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*, 41–61.

Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language, 26,* 339-356.

Ambridge, B. (2010). Children's judgments of regular and irregular novel past tense forms: New data on the dual- versus single-route debate. *Developmental Psychology 46,* 1497-1504.

Ambridge, B. (2011). Paradigms for assessing children's knowledge of syntax and morphology. In: E. Hoff, E (ed.), *Guide to Research Methods in Child Language*. (pp. 113-132). London: Blackwell-Wiley.

Ambridge, B. (2012a). How Do Children Restrict Their Linguistic Generalizations?: An (Un-) Grammaticality Judgement Study. *Cognitive Science, 37*, 508-543.

Ambridge, B. (2012b). Assessing Grammatical Knowledge (with reference to the graded grammaticality judgement paradigm) in E. Hoff (ed.). *Research Methods in Child Language: A Practical Guide* (pp 113 - 132). London: Blackwell-Wiley.

Ambridge, B. (2013). How Do Children Restrict Their Linguistic Generalizations?: an (un-) grammaticality judgement study. *Cognitive Science, 37*, 508-543.

Ambridge, B. A., & Blything, R.P. (2015). A Connectionist Model of the Retreat from Verb-argument structure overgeneralizations. *Journal of Child Language, 43, 1-32.*

Ambridge, B., Freudenthal, D., Pine, J. M., Mills, R., Clark, V., & Rowland, C. F. (2009). Unlearning un-prefixation errors. In A. Howes, D. Peebles, & R. Cooper (Eds.), 9th *International Conference on Cognitive Modelling* (pp. 158-163). Manchester: ICCM.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

Ambridge, B., & Lieven, E. V. M. (2015). A Constructivist account of child language acquisition. In B. MacWhinney and W. O'Grady (Eds). *Handbook of Language Emergence* (pp. 478-510). Hoboken, NJ: Wiley Blackwell.

Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2014). Child language acquisition: Why universal grammar doesn't help. *Language, 90,* 54-89. doi:10.1353/lan.2014.0051

Ambridge, B., Pine, J. M., & Rowland, C. F. (2011). Children use *verb-semantics* to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics, 22,* 303–323.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition, 123,* 260–279.

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb-semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language, 8,* 45-66.

Ambridge, B, Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science, 4,* 47–62.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (*entrenchment*) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition, 106,* 87–129.

Ambridge, B., Pine, J. M., Rowland, C. F., Freudenthal, D. & Chang, F. (2014). The retreat from dative overgeneralization errors: Semantics, statistics or both? *Language, Cognition and Neuroscience, 29,* 218-243.

Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R. L., & Clark, V. (2009). A semantics-based approach to the "no negative evidence" problem. *Cognitive Science, 33,* 1301-1316.

Audacity Team (2015). Audacity®. Version 2.1.1. Audio editor and recorder. Available from: http://audacityteam.org/

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Enquiry, 10,* 533-581.

Baker, M. (2001). *The Atoms of Language*. New York: Basic Books.

Bannard, C., Lieven, E., & Tomasello,M. (2009). Modelling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 17284–17289.

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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. arXiv:1506.04967 [stat]. Retrieved from http://arxiv.org/abs/1506.04967

Bates D., Maechler M., Bolker B., & Walker S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-9, https://CRAN.R-project.org/package=lme4.

Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (157-193). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bates D, Maechler M, & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-3.

Berent, I., Pinker, S., & Shimron, J. (2002). The nature of Regularity and Irregularity: Evidence from Hebrew Nominal Inflection. *Journal of Psycholinguistic Research, 31*, 459-502.

Bidgood, A., Ambridge, B., Pine, J. M. & Rowland, C. F. (2014). The retreat from locative overgeneralisation errors: A novel verb grammaticality judgment study. *PLoS ONE 9(5)*: e97634. doi:10.1371/journal.pone.0097634.

Blything, R.P., Ambridge B., & Lieven E.V.M. (2014). Children Use Statistics and Semantics in the Retreat from Overgeneralization. *PLoS ONE 9(10)*: e110009. doi:10.1371/journal.pone.0110009

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In Wanner, E., Gleitman, L. (Eds.) *Language acquisition: The state of the art* (pp. 319-346). New York: Academic Press,

Bowerman, M. (1988). The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford, England: Blackwell.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics, 28*, 1251–1289.

Bowerman, M. (1996). *Argument Structure and Learnability: Is a Solution in Sight?* Paper presented at the Proceedings of the twenty-second annual meeting of the Berkeley Linguistics Society, Berkeley, California.

Bowerman, M. & Croft, W. (2008). The acquisition of the English causative alternation. In Melissa Bowerman & Penelope Brown (eds.) *Crosslinguistic perspectives on argument*. York: Basic Books.

Braine, M. D. S. (1992).What sort of innate structure is needed to 'bootstrap' into syntax? *Cognition*, *45*, 77–100.

Braine, M. D. S., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 352–376). Hillsdale, NJ: Erlbaum.

Brooks, P.J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*, 720–738.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development*, *70*, 1325–1337.

Brooks, P.J. & Zizak, O. (2002). Does pre-emption help children learn verb transitivity? *Journal of Child Language, 29*, 759–81.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes 10*, 425-455.

Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. . *Language, 59*, 251-270.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science*, *8*, 500–508.

Childers, J.B. & Tomasello,M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, *37*, 739–48.

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

Chomsky, N. (1965), Aspects of the Theory of Syntax, MIT Press, Cambridge, MA.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Foris: Dordrecht.

Chouinard, M.M., & Clark, E.V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language, 30*, 637–669.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*, 767–811.

Clark, E. V., Carpenter, K. L., & Deutsch, W. (1995). Reference states and reversals: Undoing actions with verbs. *Journal of Child Language, 22*, 633–662.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 741-748.

Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21-58.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Demetras, M.J., Post, K.N. and Snow, C.E. (1986). Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language*, *13*, 275-92.

Dodson, K. & Tomasello, M. (1998). Acquiring the transitive construction in English: the role of animacy and pronouns. *Journal of Child Language, 25*, 555-74.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). *Cognition*, *82*, 259–78.

Friederici, A. D., & Frisch, S. (2000).Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*, 476-507.

Frisch, S., Hahne, A., & Friederici, A.D. (2004).Word category and verb-argument structure information in the dynamics of parsing. *Cognition*, *91*, 191-219.

Gertner, Y., Fisher, C. & Eisengart, J. (2006). Learning words and rules: abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*, 684–91.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics, 13*, 327–356.

Gouvêa, A.C., Phillips, C., Kazanina, N. & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*, 149-188.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language, 18*, 115–151.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in english. *Language, 65*, 203–257.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci. 12,* 556–568.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38,* 1531-1549.

Hartshorne, J. K., & Ullman, M. (2006). Why girls say "holded" more than boys. *Developmental Science, 9,* 21-32.

Ibbotson, P. (2013). The role of semantics, *pre-emption* and skew in linguistic distributions: the case of the un-construction. *Frontiers in Psychology, 4*, DOI 10.3389/fpsyg.2013.00989.

Kaan, E., Harris, A., Gibson, G., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15,* 159–201.

Kidd, E. & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development, 87,* 184-193.

Kielar, A., Meltzer-Asscher, A., & Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia, 50,* 3320-3337.

Kotz, S. A. & Friederici, A. D. (2003). Electrophysiology of normal and pathological language processing. *Journal of Neurolinguistics, 16,* 43–58.

Kutas, M. & Hillyard, S. A. (1980). "Reading senseless sentences: Brain potentials reflect semantic incongruity". *Science, 207,* 203–208.

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621-647.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version:2.0-32. http://cran.rproject.org/web/ packages/ lmerTest /index.html

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). Stanford, CA: CSLI.

Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24,* 187–219.

Luke, S.G. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*. Advance Online Publication. doi:10.3758/s13428-016-0809-y

Li, P., & Macwhinney, B. (1996) Cryptotype, overgeneralization, and competition: a connectionist model of the learning of English reversive prefixes. *Connection Science*, *8*, 3-30.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., & Friston, K. (2011). EEG and MEG data analysis in SPM8. *Comput Intell Neurosci*, 852-961.

Luck, S.J. (2014). *An Introduction to the Event-Related Potential Technique, Second edition*. Cambridge, MA: The MIT Press.

MacWhinney, B. (2000) The CHILDES project: Tools for analyzing talk. Lawrence Erlbaum Associates. *Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B, (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31*, 883-914.

Maratsos, M. (2000). More overregularisations after all: New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language, 27*, 183–212.

Maslen, R. J. C., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language and Hearing Research, 47*, 1319-1333.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57*, 1-182.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*, 465–472.

McNeil, D. (1966). Capacity for language acquisition. *Volta Review, 68*, 17-33.

Miwa, K., Libben, G., Dijkstra, T., & Baayen, R. H. (2014). The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *Quarterly Journal of Experimental Psychology, 67*, 79-113.

Noble, C. H., Rowland, C. F. & Pine, J. M. (2011). Comprehension of argument structure and semantic roles: evidence from infants and the forced-choice pointing paradigm. *Cognitive Science, 35*, 963–982.

Ogden, D. C. & Ellis, N. C. (2014) Distributional and semantic properties of verb–argument constructions in corpora of child and parent language. Manuscript in preparation.

Perek, F. & Goldberg, A. (2015). Generalizing beyond the input: the functions of the constructions matter. *Journal of Memory and Language, 84*, 108-127.

Pine, J. M. & Lieven, E. V. M. (1993). Reanalyzing rote-learned phrases – individual differences in the transition to multi-word speech. *Journal of Child Language*, *20,* 551–71.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition* (pp. 339–441). Hillsdale, NJ: Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT.

Pinker, S. (1995). Language acquisition. In L.R. Gleitman & M. Liberman (Eds.), *An Invitation to Cognitive Science* (pp. 135-182), Cambridge, MA: MIT Press.

Pinker, S. (1999). *Words and Rules: The ingredients of Language*. New York: Basic Books.

Pinker, S. (2013). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA/London: MIT Press.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*, 456-463.

Prasada, S., & Pinker, S. (1993). Generalisation of Regular and Irregular Morphological Patterns. *Language and Cognitive Processes, 8*, 1-56.

Pye, C. (1990). The Acquisition of Ergative Languages, *Linguistics*, *28*, 1291-1330.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L. & Holcomb, P. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*, 413-438.

Osterhout, L., Holcomb, P.J., Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exper. Psychol., Learn., Mem., Cogn. 20*, 786–803.

Quene, H., & Van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425.

R Development Core Team (2013). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology, 45,* 45-94.

Robenalt, C. & Goldberg, A. E. (2015). Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well backstroke or swim children to shore. *Cognitive Linguistics. 26, 467-504.*

Rosler, F., Putz, P., Friederici, A., & Hahne, A. (1993). Event-Related Brain Potentials While Encountering Semantic and Syntactic Constraint Violations. *Journal of Cognitive Neuroscience, 5,* 345–362.

Rowland, C.F., Chang, F., Ambridge, B., Pine, J.M., & Lieven, E.V.M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition, 125,* 49–63.

Roncaglia-Denissen, M.P., Schmidt-Kassow, M., & Kotz, S.A. (2013). Speech Rhythm Facilitates Syntactic Ambiguity Resolution: ERP Evidence. *PLoS ONE, 8,* e56000. doi:10.1371/journal.pone.0056000

Rosler, F., Putz, P., Friederici, A., & Hahne, A. (1993). Event-Related Brain Potentials While Encountering Semantic and Syntactic Constraint Violations. *Journal of Cognitive Neuroscience, 5,* 345–362.

Rumelhart, D.E. & McClelland, J. (1986). *On learning the past tense of English verbs. In Parallel Distributed Processing*. Cambridge, Mass: MIT Press.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analysis reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain & Language, 137,* 29-39.

Schutter, D.J.L.G., Leitner, C., Kenemans, J.L., & van Honk, J. (2006). Electrophysiological correlates of corticosubcortical interaction: A cross-frequency spectral EEG analysis. *Clinical Neurophysiology, 117,* 381–387.

Theakston, A.L. (2004). The role of *entrenchment* in children's and adults' performance on grammaticality judgement tasks. *Cognitive Development, 19,* 15-34.

Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2005). Beyond formalities: the case of language acquisition. *The Linguistic Review*, *22*, 183–97.

Tomasello, M. & Abbot-Smith, K. (2002). A tale of two theories: response to Fisher. *Cognition*, *83*, 207–214.

Tomasello, M. & Brooks, P. J. (1998). Young children’s earliest transitive and intransitive constructions. *Cognitive Linguistics,* 9, 379–395.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.

Wlotko, E. W., & Federmeier, K .D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message level meaning. *NeuroImage*, *62*, 356-366.

Wonnacott, E., Newport, E.L., & Tanenhaus, M.K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165–209.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory & Language*, *65*, 1–14.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language*, *66,* 458-478.

Wurm, L. H. & Fisicaro S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37-48.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*, 1176-1190.

7. Appendices

Appendix 1. ***Table A2.1.*** **CHILDES Frequency Counts of Each Verb.** Verbs in bold indicate that the verb has been registered less than 10 times in the CHILDES database. Also note that verbs which were heard by children were also produced at a similar frequency, indicating that spontaneous production of verbs were reflective of the frequency at which they were heard.

Verb	Input (1,678,227 utterances)	Children (854,696 utterances)	Verb	Input (1,678,227 utterances)	Children (854,696 utterances)
bandage	**2**	**2**	allow	107	20
buckle	43	12	ask	2605	361
button	139	67	believe	530	112
chain	**1**	**3**	bend	235	75
cork	**2**	**0**	close	1693	982
crumple	**2**	**1**	come	31683	7205
delete	**3**	**0**	**embarrass**	**6**	**0**
do	59495	13776	fill	292	76
fasten	84	17	freeze	48	14
hook	161	41	give	10156	2572
lace	**3**	**0**	go	58170	23377
latch	**7**	**0**	lift	427	115
leash	**0**	**0**	loosen	27	13
lock	214	157	open	3709	2336
mask	**2**	**3**	press	492	185
pack	145	35	pull	2977	849
reel	**0**	**0**	put	38227	13281
roll	941	275	**release**	**7**	**0**
screw	243	119	**remove**	**26**	**0**
snap	177	70	sit	8349	3082
tie	685	367	squeeze	235	95
veil	**0**	**0**	stand	1858	768
wrap	242	81	straighten	85	11
zip	240	70	tighten	69	25

Appendix 2. ***Table A2.2.*** **Practice and Test Sentences for Production Study**

Practice Sentences (Production)

Prime Sentences	**Target Sentences**
Bart pinned the picture to the wall and then he unpinned it	Bart dressed the dog and then he…
Marge twisted the wire and then she untwisted it	Homer crossed his legs and then he…
Homer plugged in the toy and then he unplugged it	Lisa covered the ball and then she uncovered it

Test Sentences (Production)

Verb Set A

Prime Sentences	**Target Sentences**
Bart chained the dog to a post and then he unchained it	**Bart embarrassed everyone and then he...
Bart laced his shoes and then he unlaced them	**Bart pulled the cord and then he...
Bart masked the cat and then he unmasked it	**Homer asked a question and then he...
Homer buckled his belt and then he unbuckled it	**Homer loosened his tie and then he...
Homer did his tie and then he undid it	**Homer stood on the box and then he...
Homer fastened his seatbelt and then he unfastened it	**Lisa believed in unicorns and then she...
Homer packed his case and then he unpacked it	**Lisa froze the ice lolly and then she...
Homer snapped the lego bricks together and then he unsnapped them	**Lisa opened the box and then she...
Homer wrapped the present and then he unwrapped it	**Lisa squeezed the sponge and then she...
Lisa bandaged her arm and then she unbandaged it	**Marge allowed Bart some chocolate and then she...
Lisa tied her shoelaces and then she untied them	**Marge closed the door and then she...
Marge deleted the email and then she undeleted it	**Marge released the bees and then she...
Bart clenched his fist and then he unclenched it	*Bart buttoned his shirt and then he...
Bart clogged the sink and then he unclogged it	*Bart hooked the picture on the wall and then he...
Bart linked the railway tracks and then he unlinked them	*Homer corked the bottle and then he...
Homer bolted the door and then he unbolted it	*Homer latched the gate and then he...
Homer clipped the papers together and then he unclipped them	*Homer veiled the bride and then he...
Homer tangled the strings and then he untangled them	*Lisa leashed the dog and then she...
Lisa braided her hair and then she unbraided it	*Lisa locked the door and then she...
Lisa curled her eyelashes and then she uncurled them	*Lisa rolled up the newspaper and then she...
Lisa strapped on her watch and then she unstrapped it	*Marge crumpled the paper and then she...
Marge coiled the rope and then she uncoiled it	*Marge reeled the cotton and then she...
Marge folded her arms and then she unfolded them	*Marge screwed the top on the container and then she...
Marge loaded the basket and then she unloaded it	*Marge zipped her coat and then she…

Verb Set B

Prime Sentences	Target Sentences
Bart buttoned his shirt and then he unbuttoned it	*Homer fastened his seatbelt and then he...
Bart hooked the picture on the wall and then unhooked it	*Lisa bandaged her arm and then she...
Homer corked the bottle and then he uncorked it	*Bart masked the cat and then he...
Homer latched the gate and then he unlatched it	*Homer wrapped the present and then he...
Homer veiled the bride and then he unveiled it	*Homer snapped the lego bricks together and then he...
Lisa leashed the dog and then she unleashed it	*Lisa tied her shoelaces and then she...
Lisa locked the door and then she unlocked it	*Bart laced his shoes and then he...
Lisa rolled up the newspaper and then she unrolled it	*Marge deleted the email and then she...
Marge crumpled the paper and then she uncrumpled it	*Homer buckled his belt and then he...
Marge reeled the cotton and then she unreeled it	*Homer did his tie and then he...
Marge screwed the top on the container and then she unscrewed it	*Bart chained the dog to a post and then he...
Marge zipped her coat and then she unzipped it	*Homer packed his case and then he...
Bart clenched his fist and then he unclinched it	**Homer came home and then he...
Bart clogged the sink and then he unclogged it	**Marge pressed the lever and then she...
Bart linked the railway tracks and then he unlinked them	**Bart filled the balloon and then he...
Homer bolted the door and then he unbolted it	**Homer sat on the dog and then he...
Homer clipped the papers together and then he unclipped them	**Homer tightened the screws and then he...
Homer tangled the strings and then he untangled them	**Marge put the book on the table and then she...
Lisa braided her hair and then she unbraided it	**Bart went to the hospital and then he...
Lisa curled her eyelashes and then she uncurled them	**Homer lifted his arms and then he...
Lisa strapped on her watch and then she unstrapped it	**Marge removed the television and then she...
Marge coiled the rope and then she uncoiled it	**Marge gave Bart a cookie and then she...
Marge folded her arms and then she unfolded them	**Homer bent the metal bar and then he...
Marge loaded the basket and then she unloaded it	**Marge straightened the picture and then she...

*= Target Sentence containing "un" verb
**= Target Sentence containing "zero" verb

Appendix 3. *Table A2.3.* Practice and Test Sentences for Judgment Study

Practice Sentences (Judgments)

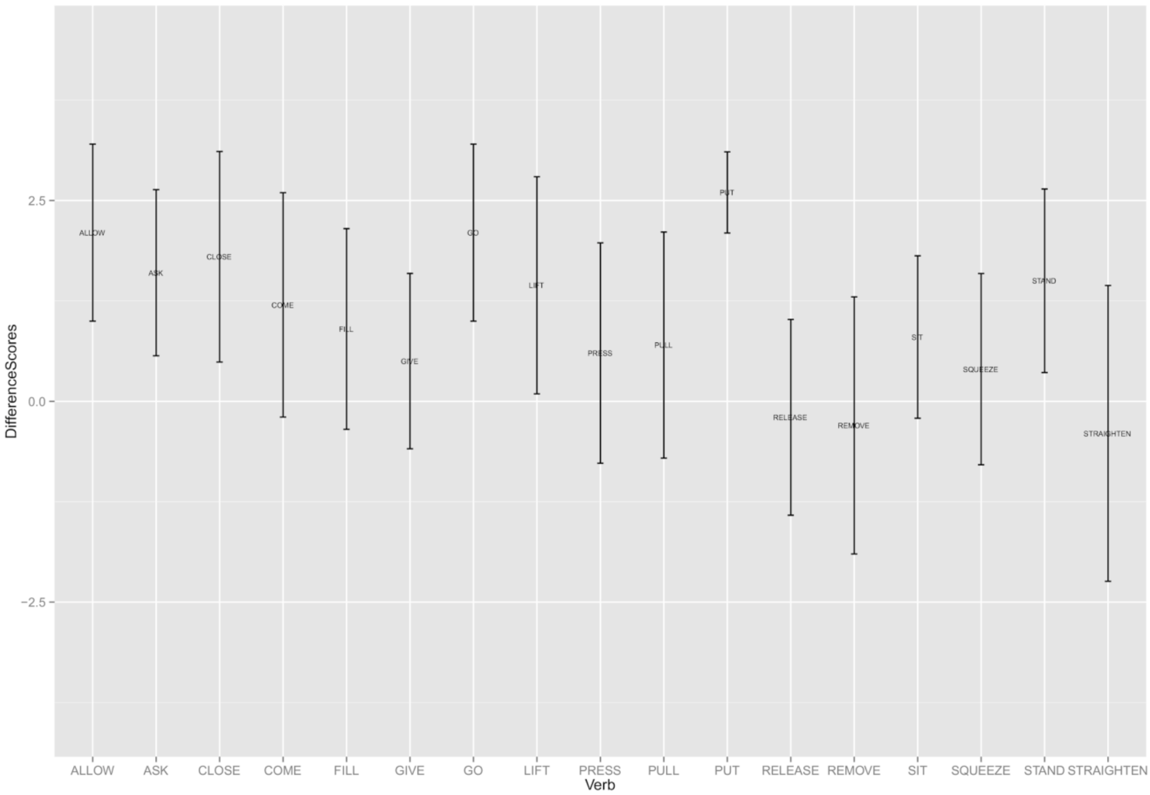
1) Lisa broke the cup (5/5)
2) Lisa breakded the cup (1/5)
3) Bart spilled soup onto his shirt (suggested score 4/5 as *spilt* is preferred in British English)
4) Homer eated the ice cream (suggested score 2/4 as more acceptable than *breakded*)
5) Homer drap water into the cup (suggested score 1/5 – over-irregularization of *drip*)
6) Bart sticked the stickers onto his shirt (suggested score 2-3/5; better than *breakded/drap*)
7) Lisa spreaded butter onto the bread (suggested score 4/5 as *spread* is the correct form, though many find *spreaded* acceptable)

Test Sentences (Judgments)

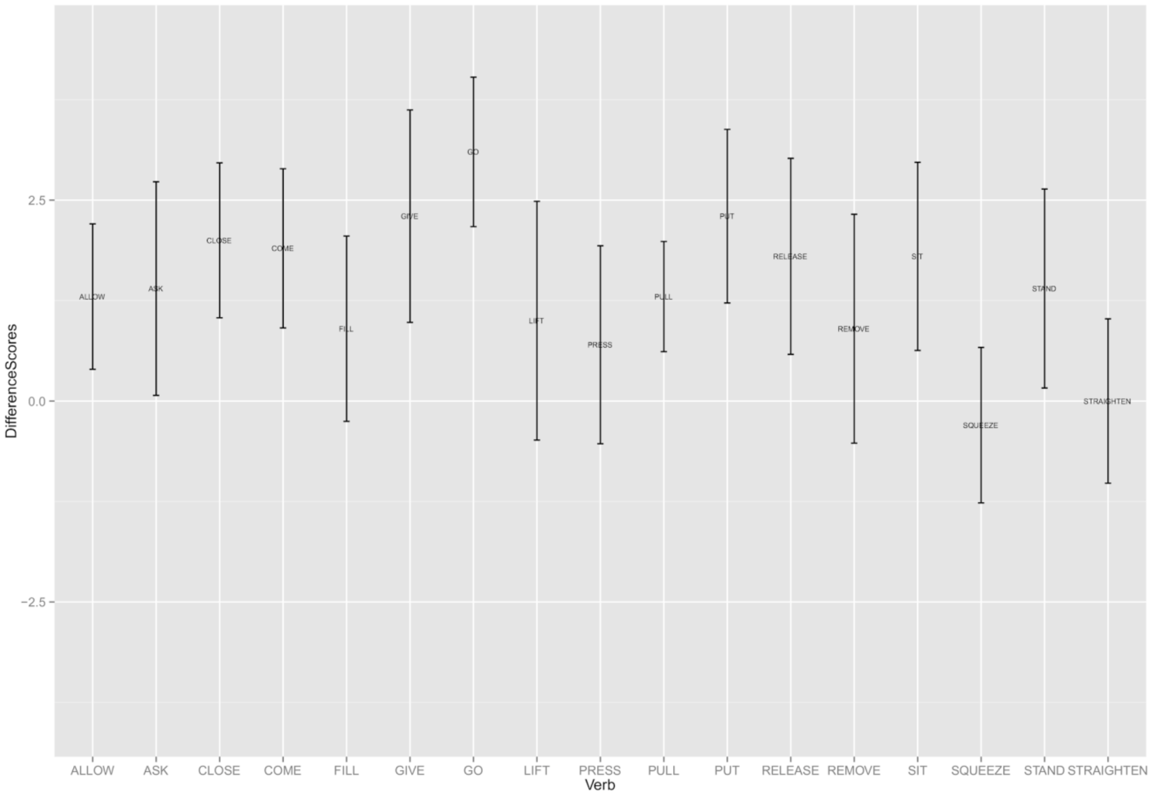
Verb Set A	Verb Set B
**Bart (un)embarrassed everyone	*Homer (un)fastened his seatbelt
**Bart (un)pulled the cord	*Lisa (un)bandaged her arm
**Homer (un)asked a question	*Bart (un)masked the cat
**Homer (un)loosened his tie	*Homer (un)wrapped the present
**Homer (un)stood on the box	*Homer (un)snapped the Lego bricks together/apart
**Lisa (un)believed in unicorns	*Lisa (un)tied her shoelaces
**Lisa (un)froze the ice lolly	*Bart (un)laced his shoes
**Lisa (un)opened the box	*Marge (un)deleted the email
**Lisa (un)squeezed the sponge	*Homer (un)buckled his belt
**Marge (un)allowed Bart some chocolate	*Homer (un)did his tie
**Marge (un)closed the door	*Bart (un)chained the dog to/from a post
**Marge (un)released the bees	*Homer (un)packed his case
*Bart (un)buttoned his shirt	**Homer (un)came home
*Bart (un)hooked the picture on/from the wall	**Marge (un)pressed the lever
*Homer (un)corked the bottle	**Bart (un)filled the balloon
*Homer (un)latched the gate	**Homer (un)sat on the dog
*Homer (un)veiled the bride	**Homer (un)tightened the screws
*Lisa (un)leashed the dog	**Marge (un)put the book on/off the table
*Lisa (un)locked the door	**Bart (un)went to the hospital
*Lisa (un)rolled (up) the newspaper	**Homer (un)lifted his arms
*Marge (un)crumpled the paper	**Marge (un)removed the television
*Marge (un)reeled the cotton	**Marge (un)gave Bart a cookie
*Marge (un)screwed the top onto/from the container	**Homer (un)bent the metal bar
*Marge (un)zipped her coat	**Marge (un)straightened the picture

*= Sentence containing "un" verb
**= Sentence containing "zero" verb

Appendix 4. ***Figure A2.1.*** **Mean Difference Scores for 3-4 Year Olds.** Mean difference scores were calculated by subtracting the mean rating for each verb's *un-* form from the mean rating for each verb's bare form. If mean difference scores for verbs that do not take *un-* (i.e. "zero" verbs – defined by whether or not they had appeared in *un* –form in BNC) fell below the value of zero then we assert that the child did not understand the meaning of the verb; using this rationale, 3-4 year old children rated only three "zero" verbs as more grammatical than their bare form equivalent (*release, remove, straighten*) and thus we can be confident that test verbs used in the current study were suitable for use with these children.



Appendix 5. ***Figure A2.2.*** **Mean Difference Scores for 5-6 Year Olds.** Mean difference scores were calculated by subtracting the mean rating for each verb's *un-* form from the mean rating for each verb's bare form. If mean difference scores for verbs that do not take *un-* (i.e. "zero" verbs) fell below the value of zero then we assert that the child did not understand the meaning of the verb. Five-to-six year old children rated one "zero" verbs as more grammatical than its bare form equivalent (*squeeze*). Thus, we can be confident that test verbs used in the current study were suitable for use with this age-group.



Appendix 6. *Table A3.1*. Core set of 40 novel verbs adapted from Albright and Hayes (2003)

	Verb forms				Model Prediction (similarity to...)		
No.	Verb.	Present	Regular Past	Irregular Past	Reg.	Irreg.	Verb Island
1	bize	bÈYz	bÈYzd	bÈoz	0.53	0.02	Regular&Irregular
2	dize	dÈYz	dÈYzd	dÈoz	0.55	0.04	Regular&Irregular
3	drice	drÈYs	drÈYst	drÈos	0.42	0.05	Regular&Irregular
4	flidge	flÈIJ	flÈIJd	flÈÃJ	0.4	0.03	Regular&Irregular
5	fro	frÈo	frÈod	frÈu	0.62	0.07	Regular&Irregular
6	gare	gÈer	gÈerd	gÈor	0.65	0.05	Regular&Irregular
7	glip	glÈIp	glÈIpt	glÈÃp	0.51	0.01	Regular&Irregular
8	rife	rÈYf	rÈYft	rÈof	0.28	0.06	Regular&Irregular
9	stin	stÈIn	stÈInd	stÈÃn	0.58	0.1	Regular&Irregular
10	stip	stÈIp	stÈIpt	stÈÃp	0.53	0.05	Regular&Irregular
11	blig	blÈIg	blÈIgd	blÈÃg	0.4	0.03	Irregular Only
12	chake	CÈek	CÈekt	CÈUk	0.46	0.04	Irregular Only
13	drit	drÈIt	drÈIt«d	drÈIt	0.35	0.03	Irregular Only
14	fleep	flÈip	flÈipt	flÈEpt	0.43	0.07	Irregular Only
15	gleed	glÈid	glÈid«d	glÈid	0.31	0.02	Irregular Only
16	glit	glÈIt	glÈIt«d	glÈIt	0.37	0.06	Irregular Only
17	queed	kwÈid	kwÈid«d	kwÈEd	0.29	0.03	Irregular Only
18	plim	plÈIm	plÈImd	plÈÃm	0.56	0.03	Irregular Only
19	skride	skrÈYd	skrÈYd«d	skrÈod	0.23	0.11	Irregular Only
20	spling	splÈIN	splÈINd	splÈÃN	0.37	0.19	Irregular Only
21	teep	tÈip	tÈipt	tÈEpt	0.45	0.04	Neither
22	gude	gÈud	gÈud«d	gÈud	0.34	0.02	Neither
23	nung	nÈÃN	nÈÃNd	nÈQN	0.56	0.01	Neither
24	pank	pÈQNk	pÈQNkt	pÈÃNk	0.55	0	Neither
25	preak	prÈik	prÈikt	prÈok	0.51	0.01	Neither
26	rask	rÈQsk	rÈQskt	rÈÃsk	0.6	0	Neither
27	shilk	SÈIlk	SÈIlkt	SÈQlk	0.31	0.03	Neither
28	tark	tÈark	tÈarkt	tÈork	0.46	0	Neither
29	trisk	trÈIsk	trÈIskt	trÈIsk	0.59	0.01	Neither
30	nold	nÈold	nÈold«d	nÈold	0.43	0.01	Neither
31	blafe	blÈef	blÈeft	blÈEft	0.31	0	Regular Only
32	bredge	brÈEJ	brÈEJd	brÈoJ	0.52	0	Regular Only
33	chool	CÈul	CÈuld	CÈol	0.69	0.02	Regular Only
34	dape	dÈep	dÈept	dÈQpt	0.43	0	Regular Only
35	ghez	gÈEz	gÈEzd	gÈaz	0.54	0.01	Regular Only
36	nace	nÈes	nÈest	nÈos	0.44	0.01	Regular Only
37	spack	spÈQk	spÈQkt	spÈÃk	0.51	0	Regular Only
38	stire	stÈYr	stÈYrd	stÈor	0.66	0.02	Regular Only
39	tesh	tÈES	tÈESt	tÈaS	0.44	0	Regular Only
40	whiss	wÈIs	wÈIst	wÈÃs	0.47	0.02	Regular Only

Appendix 7. *Table A4.1.* Table of Test Sentences for Study 3

	Transitive Structure; High Frequency Verb	Transitive Structure; Low Frequency Verb	Intransitive Structure; High Frequency Verb	Intransitive Structure; Low Frequency Verb
1	*On Monday Jack laughed the girl in the kitchen*	*On Monday Jack giggled the girl in the kitchen*	*On Monday Jack laughed in the kitchen*	*On Monday Jack giggled in the kitchen*
2	*On Tuesday Jake laughed the girl in the garden*	*On Tuesday Jake giggled the girl in the garden*	*On Tuesday Jake laughed in the garden*	*On Tuesday Jake giggled in the garden*
3	*On Wednesday Luke laughed the girl in the kitchen*	*On Wednesday Luke giggled the girl in the kitchen*	*On Wednesday Luke laughed in the kitchen*	*On Wednesday Luke giggled in the kitchen*
4	*On Thursday Mike laughed the girl in the garden*	*On Thursday Mike giggled the girl in the garden*	*On Thursday Mike laughed in the garden*	*On Thursday Mike giggled in the garden*
5	*On Friday Matt laughed the girl in the kitchen*	*On Friday Matt giggled the girl in the kitchen*	*On Friday Matt laughed in the kitchen*	*On Friday Matt giggled in the kitchen*
6	*On Monday Rob laughed the girl in the garden*	*On Monday Rob giggled the girl in the garden*	*On Monday Rob laughed in the garden*	*On Monday Rob giggled in the garden*
7	*On Tuesday Bob laughed the girl in the kitchen*	*On Tuesday Bob giggled the girl in the kitchen*	*On Tuesday Bob laughed in the kitchen*	*On Tuesday Bob giggled in the kitchen*
8	*On Wednesday Mark laughed the girl in the garden*	*On Wednesday Mark giggled the girl in the garden*	*On Wednesday Mark laughed in the garden*	*On Wednesday Mark giggled in the garden*
9	*On Thursday Rick laughed the girl in the kitchen*	*On Thursday Rick giggled the girl in the kitchen*	*On Thursday Rick laughed in the kitchen*	*On Thursday Rick giggled in the kitchen*
10	*On Friday Scott laughed the girl in the garden*	*On Friday Scott giggled the girl in the garden*	*On Friday Scott laughed in the garden*	*On Friday Scott giggled in the garden*
11	*On Monday Jack fell the girl onto the floor*	*On Monday Jack tumbled the girl onto the floor*	*On Monday Jack fell onto the floor*	*On Monday Jack tumbled onto the floor*
12	*On Tuesday Jake fell the girl down the stairs*	*On Tuesday Jake tumbled the girl down the stairs*	*On Tuesday Jake fell down the stairs*	*On Tuesday Jake tumbled down the stairs*
13	*On Wednesday Luke fell the girl onto the floor*	*On Wednesday Luke tumbled the girl onto the floor*	*On Wednesday Luke fell onto the floor*	*On Wednesday Luke tumbled onto the floor*
14	*On Thursday Mike fell the girl down the stairs*	*On Thursday Mike tumbled the girl down the stairs*	*On Thursday Mike fell down the stairs*	*On Thursday Mike tumbled down the stairs*
15	*On Friday Matt fell the girl onto the floor*	*On Friday Matt tumbled the girl onto the floor*	*On Friday Matt fell onto the floor*	*On Friday Matt tumbled onto the floor*
16	*On Monday Rob fell the girl down the stairs*	*On Monday Rob tumbled the girl down the stairs*	*On Monday Rob fell down the stairs*	*On Monday Rob tumbled down the stairs*
17	*On Tuesday Bob fell the girl onto the floor*	*On Tuesday Bob tumbled the girl onto the floor*	*On Tuesday Bob fell onto the floor*	*On Tuesday Bob tumbled onto the floor*
18	*On Wednesday Mark fell the girl down the stairs*	*On Wednesday Mark tumbled the girl down the stairs*	*On Wednesday Mark fell down the stairs*	*On Wednesday Mark tumbled down the stairs*
19	*On Thursday Rick fell the girl onto the floor*	*On Thursday Rick tumbled the girl onto the floor*	*On Thursday Rick fell onto the floor*	*On Thursday Rick tumbled onto the floor*
20	*On Friday Scott fell the girl down the stairs*	*On Friday Scott tumbled the girl down the stairs*	*On Friday Scott fell down the stairs*	*On Friday Scott tumbled down the stairs*
21	*On Monday Jack disappeared the girl at the picnic*	*On Monday Jack vanished the girl at the picnic*	*On Monday Jack disappeared at the picnic*	*On Monday Jack vanished at the picnic*
22	*On Tuesday Jake disappeared the girl as if by magic*	*On Tuesday Jake vanished the girl as if by magic*	*On Tuesday Jake disappeared as if by magic*	*On Tuesday Jake vanished as if by magic*
23	*On Wednesday Luke disappeared the girl at the picnic*	*On Wednesday Luke vanished the girl at the picnic*	*On Wednesday Luke disappeared at the picnic*	*On Wednesday Luke vanished at the picnic*
24	*On Thursday Mike disappeared the girl as if by magic*	*On Thursday Mike vanished the girl as if by magic*	*On Thursday Mike disappeared as if by magic*	*On Thursday Mike vanished as if by magic*
25	*On Friday Matt disappeared the girl at the picnic*	*On Friday Matt vanished the girl at the picnic*	*On Friday Matt disappeared at the picnic*	*On Friday Matt vanished at the picnic*
26	*On Monday Rob disappeared the girl as if by magic*	*On Monday Rob vanished the girl as if by magic*	*On Monday Rob disappeared as if by magic*	*On Monday Rob vanished as if by magic*
27	*On Tuesday Bob disappeared the girl at the picnic*	*On Tuesday Bob vanished the girl at the picnic*	*On Tuesday Bob disappeared at the picnic*	*On Tuesday Bob vanished at the picnic*
28	*On Wednesday Mark disappeared the girl as if by magic*	*On Wednesday Mark vanished the girl as if by magic*	*On Wednesday Mark disappeared as if by magic*	*On Wednesday Mark vanished as if by magic*
29	*On Thursday Rick disappeared the girl at the picnic*	*On Thursday Rick vanished the girl at the picnic*	*On Thursday Rick disappeared at the picnic*	*On Thursday Rick vanished at the picnic*
30	*On Friday Scott disappeared the girl as if by magic*	*On Friday Scott vanished the girl as if by magic*	*On Friday Scott disappeared as if by magic*	*On Friday Scott vanished as if by magic*
31	*On Monday Jack smiled the girl in the hall*	*On Monday Jack grinned the girl in the hall*	*On Monday Jack smiled in the hall*	*On Monday Jack grinned in the hall*
32	*On Tuesday Jake smiled the girl by the seaside*	*On Tuesday Jake grinned the girl by the seaside*	*On Tuesday Jake smiled by the seaside*	*On Tuesday Jake grinned by the seaside*
33	*On Wednesday Luke smiled the girl in the hall*	*On Wednesday Luke grinned the girl in the hall*	*On Wednesday Luke smiled in the hall*	*On Wednesday Luke grinned in the hall*
34	*On Thursday Mike smiled the girl by the seaside*	*On Thursday Mike grinned the girl by the seaside*	*On Thursday Mike smiled by the seaside*	*On Thursday Mike grinned by the seaside*
35	*On Friday Matt smiled the girl in the hall*	*On Friday Matt grinned the girl in the hall*	*On Friday Matt smiled in the hall*	*On Friday Matt grinned in the hall*
36	*On Monday Rob smiled the girl by the seaside*	*On Monday Rob grinned the girl by the seaside*	*On Monday Rob smiled by the seaside*	*On Monday Rob grinned by the seaside*
37	*On Tuesday Bob smiled the girl in the hall*	*On Tuesday Bob grinned the girl in the hall*	*On Tuesday Bob smiled in the hall*	*On Tuesday Bob grinned in the hall*
38	*On Wednesday Mark smiled the girl by the seaside*	*On Wednesday Mark grinned the girl by the seaside*	*On Wednesday Mark smiled by the seaside*	*On Wednesday Mark grinned by the seaside*
39	*On Thursday Rick smiled the girl in the hall*	*On Thursday Rick grinned the girl in the hall*	*On Thursday Rick smiled in the hall*	*On Thursday Rick grinned in the hall*
40	*On Friday Scott smiled the girl by the seaside*	*On Friday Scott grinned the girl by the seaside*	*On Friday Scott smiled by the seaside*	*On Friday Scott grinned by the seaside*