HHS Public Access
Author manuscript
Cogn Psychol. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:
Cogn Psychol. 2015 August ; 80: 73–108. doi:10.1016/j.cogpsych.2015.03.004.

Revise and resubmit: How real-time parsing limitations influence grammar acquisition

Lucia Pozzan and **John C. Trueswell**
University of Pennsylvania

Abstract

We present the results from a three-day artificial language learning study on adults. The study examined whether sentence-parsing limitations, in particular, difficulties revising initial syntactic/semantic commitments during comprehension, shape learners' ability to acquire a language. Findings show that both comprehension and production of morphology pertaining to sentence argument structure are delayed when this morphology consistently appears at the end, rather than at the beginning, of sentences in otherwise identical grammatical systems. This suggests that real-time processing constraints impact acquisition; morphological cues that tend to *guide* linguistic analyses are easier to learn than cues that *revise* these analyses. Parallel performance in production and comprehension indicates that parsing constraints affect grammatical acquisition, not just real-time commitments. Properties of the linguistic system (e.g., ordering of cues within a sentence) interact with the properties of the cognitive system (cognitive control and conflict-resolution abilities) and together affect language acquisition.

Keywords

language acquisition; grammar learning; sentence processing; revision; artificial language learning

1. Introduction

In order to interpret spoken language, listeners must assign provisional structural analyses to utterances in real-time as they hear them; that is, they must rapidly categorize unfolding sound events into candidate phonemes, syllables, words, and phrases, through some kind of language parsing mechanism. Yet, for those learning a language, be they a child learning their first language, or an adult learning additional languages, language-specific rules of categorization are partially or completely unknown, even though these rules are the very same ones that ultimately permit successful interpretation. This picture is further complicated by the fact that learners' provisional structural assignments are not used solely for the purposes of interpretation, but also as input to the learning procedure itself. For

Send correspondence to: Lucia Pozzan, Ph.D., Department of Psychology & Institute of Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104, lpozzan@sas.upenn.edu, Tel: 215-898-0327, Fax: 215-573-9247.

instance, it is now well established that young children learning their first language – and adults learning a second language – will use their hypothesis about the syntactic structure of a sentence to constrain hypotheses about the meanings of unknown words and unknown morphemes within that sentence, in a process known as 'syntactic bootstrapping' (e.g., Landau & Gleitman, 1985; Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Naigles, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Snedeker & Gleitman, 2004).

It seems then that the structure building mechanism itself (i.e., the real-time parser) would play a central role in the progression of language acquisition. Yet, relatively little is known about how the challenges and limitations of real-time parsing in language learners, such as their documented difficulty revising parses (e.g., Trueswell, Sekerina, Hill & Logrip, 1999), shape acquisition trajectories, nor how the parsing process itself gets 'off the ground' in the first place in the absence of language-specific grammatical knowledge. Below we begin to explore these two interlocking issues. We assert that at the start of the learning process, the mapping of utterances onto meaning is guided by universal biases, which are gradually accompanied or supplanted by language-specific grammatical knowledge that guides parsing and interpretation more accurately (for related views in the first language acquisition literature see Fisher, Gertner, Scott, Yuan, 2010; Gertner & Fisher, 2012; Lidz, Gleitman & Gleitman, 2003; Gleitman et al. 2005; for second language acquisition, see Van Patten, 1996). We explore here a novel hypothesis within this view, that the transition from universal biases to the use of language-specific knowledge is shaped not only by the validity and reliability of language-specific cues to structure and meaning (Bates & MacWhinney, 1982; Bates & MacWhinney, 1989; MacWhinney, Bates, & Kliegl, 1984; Slobin & Bever, 1982), but also by inherent challenges associated with the real-time incremental nature of sentence processing itself, such as the difficulty of revising initial structural analyses and interpretations.[1]

1.1. Parsability and Learnability

According to the theory of syntactic bootstrapping, children discover the meanings of words not just by observing the world and keeping track of world-word contingencies, but also by taking advantage of the linguistic contexts in which words appear (Landau & Gleitman, 1985; Gleitman, 1990). Children's ability to use linguistic context during word learning is apparent early in development. For example, in her classic first studies of the learning effects of implicit syntactic analyses, Naigles (1990) showed that 25-month-olds infer aspects of a new verb's meaning from the syntactic context in which the verb appeared. Exposing children to a novel verb in a transitive sentence ("The duck is glorping the bunny") led children to believe that the novel predicate denoted a two-participant causal event rather than a one-participant non-causal event. Hearing a novel verb in an intransitive sentence ("The duck and the bunny are glorping") generated the opposite preference,

[1]This hypothesis is indebted to related views in the literature, most notably Bever (1970)'s seminal theorizing on the influence of perceptual strategies on the acquisition and representation of language. Moreover, interactions between processing and acquisition could have implications for language change and the types of natural language grammars expected to be observed in the world (see, Bever & Langendon, 1972, and more recently, Hawkins, 2004, 2012, 2014); these views are compatible and partially overlapping with our present proposal (see *General Discussion*).

Author Manuscript

Author Manuscript

Author Manuscript

indicating that children understood the novel predicate to denote a one-participant non-causal event. Numerous studies have since demonstrated similar syntactic effects on verb learning (e.g., Arunachalam & Waxman, 2010; Fisher, Hall, Rakowitz, & Gleitman, 1994; Lee & Naigles, 2008; Nappa, Wessell, McEldoon, Gleitman & Trueswell, 2009; Scott & Fisher, 2009; Yuan & Fisher, 2009; Yuan, Fisher & Snedeker, 2012), and similar use of linguistic evidence has been observed for the learning of nouns (e.g., Brown, 1957; Katz, Baker, & Macnamara, 1974; Hall, Lee, & Belanger, 2001; Smith, Jones, & Landau, 1992; Liittschwager & Markman, 1993).

To take one especially relevant example, 19-month-olds who heard the novel word "tiv" while watching a scene in which a cloth was used to wipe a block assigned it a different meaning depending on whether the novel word was used as the complement of the verb (as in (1)) or as the complement of the preposition (as in (2)).

(1) Meek the tiv

(2) Meek with the tiv

Children's looking patterns, in fact, indicate that they interpreted "tiv" to refer to the block if they heard it used as the complement of the verb (as in (1)), but to the cloth if they heard it used as the complement of the preposition (as in (2), see White, Baier, & Lidz, 2011; Lidz, Baier & White, submitted).

The ability to use structural cues to pair novel words with their intended referents is often contingent upon having some detailed language-specific knowledge. In the previous example, target-like performance depends upon the child having acquired a number of facts about English vocabulary and grammar, including the structural properties and the meaning of the preposition "with." So how do learners map structure to meanings at the beginning of the language acquisition process, when language-specific knowledge is not yet in place?

According to the theory of syntactic bootstrapping, children initially accomplish this task by taking advantage of universal, unlearned, biases to map structure to meaning (Fisher, 1996; Gleitman, 1990; Gleitman et al., 2005; Lidz, Gleitman, & Gleitman, 2003). At the beginning of the learning process, when language-specific knowledge is not yet in place, learners are thus hypothesized to rely on unlearned biases to construct partial, underspecified, structural sentence representations which are mapped to events in the world. Note that from a processing perspective, these biases can be thought of as universal parsing heuristics, which are most likely deployed in real-time by the child as each utterance is perceived[2]. As language-specific cues to structure and meaning begin to be acquired by the child, these too enter into the word learning process as additional structural constraints on interpretation.

One such universal bias (derived from the *Theta Criterion* of Chomsky, 1981) is the tendency to expect a transparent one-to-one mapping between the number of arguments in a

[2]We do not commit ourselves to the exact origin of these biases, such as whether they originate in Universal Grammar, or are better characterized as coming from a set of universal parsing/processing heuristics. One reasonable view is that the parser, as implemented in any state of development, must respect both universal and learned constraints on language representation (for discussion, see, e.g., Chomsky, 1964; Fodor, Bever, & Garrett, 1974; Pritchett, 1992). Predictions based on the account sketched in this paper do not hinge on the origin of these universal tendencies, and our manipulation and results are not meant to address this issue further.

clause (roughly, noun phrases, NPs) and the number of thematic roles (roughly, participants) in an event. Children as young as 21 months use the number of NPs in a sentence as a proxy for the number of participants, with the consequence that they misinterpret sentences in which the number of NPs does not match the number of participants. For instance, Gertner and Fisher (2012) found that younger (21-month-old) children have a tendency to misinterpret conjoined subject intransitive sentences (like the sentence used by Naigles, 1990, "The duck and the bunny are glorping") as transitive SOV sentences, in part because they do not yet know how the language specific cues of the conjunction "and" and plural agreement from "are" map onto the correct structural analysis (though see also Noble, Theakston, & Lieven, 2010, and Pozzan, Gleitman, & Trueswell, in press, for other challenges related to this structure). Children are also more willing than adults to alter the meaning of familiar verbs to match the NP contexts in which they are heard: children, but not adults, act out causative interpretations for the familiar intransitive *go*, when it is presented within a two-NP transitive context like "Noah goes the elephant to the ark" (Naigles, Fowler, & Helm, 1992; Naigles, Gleitman, & Gleitman, 1993).[3]

Children's bias to expect as many event participants as NPs in a sentence has even been found for languages in which argument omission is frequent (e.g., Mandarin, see Lee & Naigles, 2008; Kannada, see Lidz, Gleitman, & Gleitman, 2003). Lidz and colleagues examined children's performance on a sentence comprehension task in Kannada, a language in which verbal morphology is a reliable predictor of argument structure but, due to frequent argument omission, number of NPs is not. Children learning this language consistently mapped one-NP sentences to one-participant events and two-NP sentences to two-participant events, while ignoring the presence of conflicting morphological information. Differently from adults, children tended to assign the same (non-causative) interpretation to one-NP sentences, regardless of the presence (3) or absence (4) of causative morphology that perfectly predicts a two-participant causative interpretation.

(3) Kudure eer-**is**-utt-ade

Horse rise-**caus**-non-past-3sg

(4) Kudure eer-utt-ade

Horse rise-non-past-3sg

Returning to our example in (2) involving the preposition "with," the account just given predicts that a learner who has not yet acquired the meaning and structural properties of the preposition "with" might construe the novel NP "the tiv" as the argument of the verb and assign it the thematic role "patient"; relying on universal structure-to-meaning mapping biases would cause the learner not only to misinterpret the message, but also to map the novel noun onto the wrong referent (i.e., to pair it with the block in both (1) and (2)). To the extent that this is true, it becomes crucial to understand the mechanisms and evidence that learners use to substitute universal structure-to-meaning mapping biases with full-fledged

[3]Other evidence in favor of the universal, unlearned nature of this bias comes from studies investigating learners who display this bias without having been exposed to biased input. The number of noun phrases in a sentence has in fact been shown to be used as a strong predictor of argument structure in the spontaneous gesture systems (home sign) of congenitally deaf children who were not exposed to sign language (e.g., Feldman, Goldin-Meadow, & Gleitman, 1978; Goldin-Meadow, 2003).

language-specific syntactic and semantic rules, such that structures that do not conform to universal tendencies (e.g., sentences with omitted arguments) or to the majority patterns in the language (e.g., passive sentences in English), but are nonetheless licensed by the grammar of the language, can be successfully parsed and interpreted.

Arguably, the real world can provide learners with evidence that they can use to detect eventual mismatches between their interpretations and the state of affairs in the world, and update their current hypotheses accordingly (see Chang, Dell, & Bock, 2006, for an error-based acquisition model in which mismatches between the model's prediction and the actual input are used to update the model's abstract knowledge, or the *variational* model proposed by Yang, e.g., Yang, 2002, 2004, 2012, in which possible grammars are rewarded or punished according to their ability to analyze the input). For example, suppose that a learner who ignores the function and the structural properties of the preposition "with" heard it in a sentential context like (5) in which all the other vocabulary items were known, while watching a scene in which a girl was eating a cake with a spoon:

(5) The girl is eating with the spoon

Here the universal biases discussed above would not only lead to an implausible interpretation (girl-eating-spoon), but to a mismatch between the learner's hypothesis and the co-occurring event in the world. The learner could in principle use this mismatch to reject its current grammatical analysis and use properties of the observed event to inform a new hypothesis (i.e., "with" combines with NPs; the NP it combines with bears the thematic relation "instrument"). This new hypothesis can be used to parse the next utterance containing the target structure, and additional linguistic and real world evidence can be used to further refine this hypothesis.

In this paper, we propose that this learning process, which allows learners to integrate or substitute universal biases with language-specific knowledge, is influenced by the constraints and limitations of the (developing) parser. Our prediction is that language-specific cues to structure and meaning are easier to acquire and use in real time if the information that they provide can be used to prevent the parser from committing to incorrect interpretations that might need to be revised based on late arriving evidence within the same sentence. This hypothesis stems from the well-known fact that revising initial interpretative commitments is difficult for learners (for L1 acquisition in children, including languages other than English, see Choi, & Trueswell, 2010; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Omaki, & Lidz, 2014; Omaki, Davidson White, Goro, Lidz, & Phillips, 2013; Huang, Zheng, Meng, & Snedeker, 2013; Trueswell et al., 1999; Weighall, 2008; and for L2 acquisition in adults, see Pozzan, & Trueswell, 2013; Williams, Möbius, & Kim, 2001).

To illustrate in more detail, consider the hypothetical case of a learner who used the mismatch between the real world and her interpretation of (5) to infer a tentative, target-like meaning for "with." Assuming this hypothesis is available to the learner on next encounter with the word "with," she will expect it to be followed by a NP, which will be correctly interpreted as an instrument:

(6) Clean [$_{PP}$ with the cloth]

The learner's current hypothesis regarding the meaning of "with" will be reinforced in this case, provided that real world evidence supports (or at least doesn't mismatch with) this interpretation.

But what if, instead of English, our learner were acquiring one of the many world's languages in which words that convey this information appear post-nominally (e.g., as post-positions or suffixes), as in (7), rather than pre-nominally, as in (6)?

(7) Clean [$_{PP}$ the cloth with]

We predict that in this case, upon encountering the NP "cloth," the (real-time) parser will initially analyze it as the direct object of the verb "clean," and incorrectly interpret it as the patient of the predicate. In order for (7) to be interpreted correctly, this interpretation will need to be revised once "with" is heard. That is, even when a target meaning hypothesis for "with" is in place, this hypothesis can only be used to revise a preliminary structure and a partial interpretation ("clean the cloth"), once "with" is processed. But revising initial interpretations is precisely what the developing parser often fails to do; interpretations consistent with failures to revise, in fact, account on average for 50-60% of children's interpretation of temporarily ambiguous sentences (e.g., Trueswell et al, 1999). Probabilistically, then, children would be less likely to discover the intended syntactic analysis of this sentence, with the consequence that the learning of the meaning of "with" would be delayed.

Thus sentences in which disambiguating information arrives at a point in which it can no longer be easily integrated with the previous material will not only be misinterpreted by the novice learner, but might also fail to provide the learner with evidence to update their syntactic and semantic hypotheses, potentially delaying the acquisition process. An important corollary of this hypothesis is that acquisition might proceed at slower rates in languages in which function words and morphemes that carry disambiguating grammatical information become available at a point in which the parser has already committed to an incorrect interpretation. In fact, there is some suggestive evidence to support this claim: children learning Tagalog, a verb-initial language in which argument omission is pervasive but verbal morphology is a reliable predictor of argument structure, show greater sensitivity to causative morphology than children learning Kannada, a verb-final, morphology-final language, perhaps because verbal morphology can be used to guide parsing in verb-initial languages but only confirm or revise interpretative commitments in verb-final languages (Trueswell, Kaufman, Hafri, & Lidz, 2012). Similar difficulties with causative morphology in children have been reported for other verb-final languages (e.g., Turkish: Göksun, Küntay, & Naigles, 2008; Japanese: Murasugi, & Hasimoto, 2004).

Unfortunately the direct cross-linguistic comparison by Trueswell et al. (2012) is far from conclusive. First, Tagalog and Kannada have very different grammatical and morphological systems, leaving open the possibility that the superior performance shown by Tagalog-speaking children was modulated by other grammatical differences. Second, children's comprehension—not production—of morphology was surveyed, leaving open the possibility that the challenges observed in Kannada-speaking children as compared to Tagalog-speaking children reflected their difficulties revising initial interpretations rather than these difficulties impacting the acquisition of the morphology itself. In other words, the results are consistent with the alternative hypothesis that the Kannada-speaking children knew the function of causative morphology in the language, but were not able to deploy this knowledge in garden-path sentences. The more crucial question is whether such difficulties delay the learning of the meaning of morphology, and thus delay use even when revision is not required, such as in sentence production.

Since it is highly unlikely that we could compare acquisition profiles in two natural languages that differ only in the respects relevant to our research question, we opted for the next best thing: an artificial-language learning study in which learners attempt to determine how sentence structure connects to events and actions perceived in a co-present referent world (e.g., Hudson Kam, & Newport, 2005; 2009; Wonnacott, Newport, & Tanenhaus, 2008). Within this method, we can nearly perfectly match the grammars of several languages, assign them to different learners, and parametrically manipulate whether language-specific cues to structure in a language are available early in a sentence, and hence can guide interpretation, or are available only late in a sentence, and hence confirm or revise interpretations. Given the growing body of evidence showing that this methodology can be successfully employed to investigate learners' acquisition and preferences for different grammatical patterns attested in the world's languages (e.g., Culbertson, Smolensky, & Legendre, 2012; Fedzechkina, Jaeger, & Newport, 2012), it seems that artificial-language learning studies can be used as a powerful complement to cross-linguistic research to investigate preferences and biases in language acquisition in the laboratory.

The present work focuses on the performance of adult learners[4]; future work will examine children. There are however good reasons to believe that child and adult learners' grammar acquisition might be affected in similar ways by processing preferences and limitations. For instance, even native-speaker adults sometimes fail to revise initial parsing commitments (Trueswell et al., 1999; Novick, Thompson-Schill, & Trueswell, 2008), and rely on 'good enough' sentence interpretations under certain circumstances (e.g., Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Ferraro, & Bayley, 2002; Ferreira, & Patson, 2007). Moreover, the consequences of processing limitations might be exaggerated in adult second language learners, a population that has also been found to have particular difficulties revising initial interpretations (e.g., Juffs & Harrington, 1996; Juffs, 2004; Pozzan & Trueswell, 2013; Roberts & Felser, 2011; Williams et al., 2001). All of these challenges, in children and adults, may actually stem from difficulties deploying executive function / cognitive control, which is underdeveloped in children (Zelazo & Frye, 1998), and might be particularly taxed during the processing of a non-native language (Abutalebi, 2008). We will return to this issue in the general discussion.

Adults learning a second language (in this case, an artificial language) are of course likely to differ in some ways from children acquiring a first language. For instance, adults'

[4]In artificial-language learning experiments it is customary to test adult participants because, compared to children, they can be administered longer training/testing sessions and thus be taught artificial languages with larger vocabularies; moreover, adults have been shown to be more adept at learning new vocabulary from exposure to sentences in artificial-language learning studies (see Hudson Kam, & Newport, 2005).

knowledge of their native language, most notably their knowledge of language-specific grammatical tendencies and parsing procedures, is likely to be transferred to their second language (for reviews, see Gass & Selinker, 1992; Pienemann, Di Biase, Kawaguchi, & Håkansson, 2005). But, as we discuss below, comparisons of several variants of an artificial language permit us to identify such language-specific strategies, should they exist in our study.

1.2 Experimental Prospectus

In order to investigate the hypothesis that the challenges associated with parsing revision shape the way in which language learners acquire language-specific cues to meaning, we designed an experiment in which adult participants were exposed to one of four miniature-language variants during a three-day study. In all variants, the number of NPs in a sentence was not a reliable predictor of a predicate's argument structure. However, similarly to Tagalog and Kannada, morphological cues were instead reliable predictors. Specifically, one morpheme ("-ka") was always associated with an event involving a patient and an instrument while a different morpheme ("-tu") was always associated with an event involving only a patient, regardless of the presence of one or two NPs in the sentence. Thus to correctly interpret a sentence, learners needed to refrain from relying on a cue that tends to be highly predictive of argument structure across languages in general, and in the learners' L1 in particular, but is unreliable in the target language (i.e., the number of NPs in a sentence); they had to start relying instead on a cue that is absent in their L1 but perfectly predictive of argument structure in the target language (i.e., morphological cues).

The language variants differed with respect to the point at which morphological cues to argument structure would be heard in each sentence: either at the beginning (as a suffix on the first constituent) or at the end (as a suffix on the last constituent). This allowed us to test the experimental hypothesis that the characteristics of the human sentence parser (i.e., incremental processing, tendency to rapidly commit to interpretations, and difficulties revising initial interpretations) affect the time-course of grammar acquisition. We predicted that, in the absence of language-specific knowledge about the novel linguistic system (i.e., at the beginning of the study), learners will initially use the number of NPs in a sentence to infer the number of participants in an event, but will abandon this strategy with increased exposure to the language. Critically, however, the two language variants in which disambiguating morphological information became available early in each sentence should be associated with faster learning curves and higher accuracy than the two variants in which the same information was heard late in each sentence. Note that, as discussed below in *Section 2.1.2.2*, we created two early-cue and two late-cue language variants, each of which differed in whether they were similar or different to English word order; this permitted us to document (and control for) possible effects of native-language transfer.

Language exposure consisted of watching pre-recorded video interactions between two actors. One of the actors uttered a command in the language, and the other actor acted out the command. The participant's task was simply to repeat the command after watching the action being performed. Learners' comprehension and production of the language was assessed on each day in separate tests in which participants acted on spoken instructions or described perceived actions without the aid of linguistic input. At the end of the study learners were asked to indicate the meaning of the two different morphemes employed in the language.

2. Experimental Investigation

2.1 Methods

2.1.1 Participants—A total of 73 adults were recruited from the Psychology 101 pool at the University of Pennsylvania and from the community. The former received credit for their participation (4 credits), while the latter received $10 per session ($30 total).

Inclusion criteria were set to include only participants who had successfully mastered the content-word vocabulary of the language by the last day of the study (Day 3): only participants who showed evidence of having learned all twelve nouns and at least six out of the eight verbs in the language were included in the final analyses. Verbs and nouns were considered "learned" if they were used accurately on at least 75% of the cases in which they occurred in both tests of comprehension and production (see below for details of these tests). A language-background criterion was also employed. Bilingual speakers were excluded because they might present an advantage in novel word learning with respect to their monolingual peers (Kaushanskaya & Marian, 2009), and might employ mechanisms to switch from a dominant to a non-dominant language that differ in nature from those used by non-bilinguals (Costa & Santesteban, 2004).

These criteria resulted in the exclusion of 25 participants; the data from the remaining 48 participants, twelve for each of the four language variants, were included in the final analyses reported below. The 25 excluded participants were dropped for: not meeting the monolingual requirement (N = 5), failure to reach mastery of the language's content-word vocabulary (N = 8 and N = 5 in late-cue and early-cue languages, respectively), failure to return for the second or third sessions (N = 1 and N = 3, respectively), experimenter's error (N = 1), failure to follow directions (N = 1), and systematically reversing the meaning of the two morphemes during the final session (N = 1).

2.1.2 Materials and Design

2.1.2.1 Vocabulary: Every language variant contained twelve novel nouns and eight novel verbs. The referents of verbs and nouns were fixed across participants. There were no determiners in the language. Each NP was composed of two syllables and referred to either a stuffed animal or an inanimate object: "fami" (dophin), "fofo" (cow), "lepa" (frog), "naki" (pig), "poru" (bear), "tato" (horse), "kefu" (tweezers), "nunu" (clothespin), "pala" (pliers), "supo" (spoon), "sati" (sticks), "tufa" (toothbrush). All verbs were transitive and composed of two syllables. Experimental verbs, four in total, referred to actions that can be performed with an instrument or the hand: "bliki" (lift), "daxi" (turn), "glemi" (shake), "zumpi" (bounce). Filler verbs referred to actions that cannot be performed with a hand/instrument: "bobu" (blow on), "mefi" (look at), "pema" (say 'ba' to), "tiki" (stick tongue out at).

2.1.2.2 Morphology and Morpho-syntax: The languages contained two suffixes, "-ka" and "-tu", which were associated with two different versions of a given event (e.g., a human

Author Manuscript
Author Manuscript
Author Manuscript

agent causes a toy dolphin bounce up and down). The morpheme "-ka" was always used when a given event involved the use of an instrument to perform an action on a patient (e.g., the agent causes the toy dolphin to bounce up and down using a clothespin), while the morpheme ""-tu"" was always used in association with an event that involved only a patient and not an instrument (e.g., the agent causes the toy dolphin to bounce up and down with her hands).

To test the experimental hypothesis that the order with which morphological cues appear in a sentence influences the ease with which they are processed and acquired, we constructed two language variants in which morphological cues to argument structure ("-ka" and "-tu") were consistently expressed as suffixes on the verb, but, depending on the variant, the verb was either the first or the last constituent in the sentence. Both variants were thus verb-marking (more precisely, head-marking) languages, but differed with respect to the position of the verb in the sentence, and, consequently, its morphological cues. This is illustrated below using a 2-NP sentence example where the verb carrying the instrument marker "-ka" either appears clause-initially (as in I.) or clause-finally (as in II.):

I.	V_{MORPH} NP NP	Verb-Initial, Verb-Marking Language:
	Zumpi-ka	fami nunu
	Bounce-instrument	dolphin clothespin
	'Bounce the dolphin using the clothespin'	

II.	NP NP V_{MORPH}	Verb-Final, Verb-Marking Language	
	Nunu	fami	zumpi-ka
	Clothespin	dolphin	bounce-instrument
	'Bounce the dolphin using the clothespin'		

A potential shortcoming of this design was that the language variant in (I) resembles English word order more closely than the language variant in (II), thus potentially confounding the effects that cue ordering might have on acquisition with those related to L1-transfer. For this reason, we constructed two additional language variants, which are illustrated in (III) and (IV) below. In these two additional language variants, morphological cues were consistently expressed as a suffix on (inanimate) NPs; depending on the language variant, the inanimate NP was either the first or the last constituent in the sentence. Both variants were thus noun-marking (or, more precisely, dependent-marking) languages, but differed from each other with respect to the position of the verb and morphological cues in the sentence, as summarized below:

III.	NP_{MORPH} NP V	Verb-Final, Noun-Marking Language	
	Nunu-ka	fami	zumpi

Clothespin-instrument	dolphin	bounce

IV.	V NP NP_{MORPH}	Verb-Initial, Noun-Marking Language	
	Zumpi	fami	nunu-ka
	Bounce	dolphin	clothespin-instrument

In this case, the variant that most resembles English in terms of word-order (IV) is now the one in which disambiguating morphological cues appeared late in the sentence. For this second language pair, then, the predictions of the experimental hypothesis go in the opposite direction from those of facilitatory L1-transfer. The experimental hypothesis predicts (IV) to be associated with slower learning and lower accuracy rates than (III), since morphological cues appear sentence initially in the former and sentence-finally in the latter, while the opposite outcome should arise due to facilitatory L1-transfer.

In the remainder of the paper, we will investigate the extent to which comprehension and production performance is influenced by the order in which disambiguating morphological information becomes available in a sentence. For ease of exposition, we will refer to languages in which disambiguating morphological information to argument structure appears early in the sentence (i.e., variants (I) and (III) above) as **early-cue** languages and to languages in which the same information appears late in the sentence (i.e., variants (II) and (IV)) as **late-cue** languages, regardless of the position of the verb (verb-initial vs. verb-final) and the syntactic category with which the morphological cue combined (the verb vs. the NP).

2.1.2.3 Sentential Constructions: All sentences in the languages expressed commands given by an actor to a second person interlocutor. Second person subjects were always non-overt, similarly to English non-emphatic imperative clauses (e.g., Pass me the salt! Open the window!). Commands expressed by experimental sentences were to be carried out either with an instrument or the hands, depending on the morphological marker. Half of the experimental sentences contained one NP and half contained two NPs. The number of NPs in a sentence did not predict whether the action should be carried out with an instrument; only morphology did. We achieved this by employing two devices: (i) redundant NP-modification and (ii) argument omission. Redundant NP modification refers to the over-specification of a referent's attribute in a context in which the attribute is not contrastive (e.g., the attribute "blue" when the phrase "the blue cloth" is uttered in the presence of one cloth, as opposed to a context in which there are a blue and a red cloth). Adult English listeners behave in accordance with the tenets of Referential Theory (Altman & Steedman, 1988) and expect additional information about a referent to be mentioned if and only if the information necessary to establish reference. We thus expected that sentences like (8), where the NP-modifier "with the clothespin" is redundant due to the absence of a competing referent (e.g., a dolphin holding a spoon)[5], would systematically lead listeners to an incorrect (instrument) interpretation if reliable morphological information were ignored.

Argument omission refers to the omission of an obligatory argument of a given predicate (e.g., the omission of the patient NP in the sentence "wipe with the cloth"). In English, argument omission is lexically and contextually constrained in that it only occurs with a small subset of verbs (compare "I ate" with *"I hugged"), and only under specific discourse conditions, i.e., low discourse prominence (Goldberg, 2000). For this reason, we expected that sentences like (11), in which the patient NP is omitted but omission is not associated with particular discourse properties,[6] would systematically lead listeners to an incorrect interpretation if reliable morphological information were ignored. In contrast, sentences like (9) and (10) exemplify cases in which the number of NPs in the sentence is instead aligned with the number of participants/thematic relations in the event; in these cases failure to use morphological information should not be systematically associated with misinterpretations.

(8) Zumpi-tu	fami	nunu	*Non Instrument Morphology – 2 NPs*
Bounce-non-instrument	dolphin	clothespin	
Meaning: "Use your hands to bounce the dolphin, which has the clothespin"			
(9) Zumpi-ka	fami	nunu	*Instrument Morphology – 2 NPs*
Bounce-instrument	dolphin	clothespin	
Meaning: "Use the clothespin to bounce the dolphin"			
(10) Zumpi-tu	nunu		*Non Instrument Morphology – 1 NP*
Bounce-non-instrument	clothespin		
Meaning: "Use your hands to bounce the clothespin"			
(11) Zumpi-ka	nunu		*Instrument Morphology – 1 NP*
Bounce-instrument	clothespin		
Meaning: "Use the clothespin to bounce something"			

While care was taken to ensure that the grammars of verb-marking and noun-marking language variants were matched as closely as possible, the function of the morphological markers and their distribution differed slightly between them. This is due to the nature of the syntactic category that the morphological markers combined with (verbs vs. nouns): in verb-marking languages, the morpheme "-ka" indicated that the action described by the verb was to be carried out with an instrument, and the morpheme "-tu" indicated that the same action should be carried out with the hands (see Table 1 in the Appendix for the complete paradigm of the four sentence constructions in each of the four language variants). In languages in which morphological markers combined with nouns, rather than verbs, the markers functioned as case markers, and indicated the syntactic function/thematic role of the noun they combined with. Thus the morphological marker "-ka" indicated that the noun functioned as an instrument ("Nunu-ka [*clothespin-instrument*] fami [*dolphin*] zumpi [*bounce*]", i.e., 'Bounce the dolphin using the clothespin'), while the morphological marker "-tu" indicated that the noun functioned as an NP-modifier ("Nunu-tu [*clothespin-modifier*]

[5]NP modifiers were always used redundantly in the language, since it was never the case that two referents of the same type but holding different instruments appeared in the same trial.

[6]In order to make argument omission licensing as similar as possible to what is found in natural languages, the underlying regularity in the language was that the omitted patient was always the referent that was mentioned last during the naming of all the objects in the visual display that took place at the beginning of each trial, right before the command was spoken (see *Section 2.1.4*).

fami zumpi", i.e., 'Bounce the dolphin that has the clothespin'). Since "-tu" in these variants functioned as an NP modifier, it was not present on non-instrument 1-NP sentences ("Nunu zumpi", i.e., 'Bounce the clothespin'). Given that the crucial comparisons in our study involved contrasting differences in performance associated with early-cue and late-cue variants of the same language, rather than comparing between verb-marking and noun-marking languages, this difference was not expected to impact on our findings.

Experimental sentences like those above were intermixed with filler sentences, which represented half of the materials in the study (see Table 2 in the Appendix for examples of fillers in the four language variants). Fillers were included because pilot testing indicated that, when only sentences that highlighted the target morphological contrast were presented, participants became quickly aware of the manipulation and reached ceiling performance within the first training session, thus making it impossible to investigate learning curves. Filler materials consisted of actions that were not to be carried out with either an instrument or the hands, but with another body part (i.e., the mouth: "bobu" [*blow-on*]; "pema" [*say-'ba'-to*]; "tiki" [*stick-tongue-out-at*]; the eyes: "mefi" [*look-at*]). These sentences contained either two, one or no NPs, as exemplified in Table 2 in the Appendix for the four language variants[7]; filler sentences resembled experimental sentences in that the number of NPs in a sentence did not predict the number of predicate arguments.

2.1.3 Procedure

A three-day procedure was followed. Training took place on *Days 1* and *2*, together with two short testing phases, while *Day 3* was dedicated to testing. Participants were tested individually in a quiet, dimly lit room. Participants were seated in a chair positioned in front of a desk (where a podium would be placed for the "comprehension" portion of the experiment) at a comfortable distance from the wall (where videos would be projected for the training and production portions of the experiment). Words and sentences were pre-recorded by a male native speaker of English and delivered via audio speakers.

Day 1

Noun Vocabulary Training and Testing: Participants saw videos in which an actor, seen from the back, named a single object while an actress, seen from the front, pointed (correctly) to one of four props placed on each of the quadrants of a podium (see Figure 1). In each display, there were always two animals (not holding any instruments) placed diagonally from each other and two big instruments, but only one of the props was named. Participants were instructed to repeat aloud the name of each object after each pointing action. Each of the 12 objects was named four times during the training phase. Participants then received a short vocabulary comprehension test in which they saw a set of 12 novel

[7]In verb-marking languages, the morpheme -tu was associated with actions performed by hand, rather than with the absence of an instrument. This becomes clear if we compare experimental non-instrument sentences in (8) and (10) with the filler examples (see Appendix): actions performed without an instrument but with a body part different from the hands were associated with the bare form of the verb (i.e., no morphological marker), rather than the morpheme - tu. In contrast, in noun-marking languages, due to the slightly different syntactic function of the morphological markers, the marker -tu (indicating NP-modification) and the zero-marker (associated with patient NPs) occurred in both experimental and filler sentences, and thus indicated that a given action should not be associated with an instrument; the meaning of the lexical verb alone indicated whether the action should be carried out with the hands (e.g., *lift, bounce*) or a different body part (e.g., *look-at, say-'ba'-to*).

array of objects and heard each NP spoken aloud once by the actor. They were instructed to click on the correct object by clicking with the mouse on a static image of the display projected on the wall. After each response, they received non-corrective feedback, i.e., they saw a pre-recorded video showing the correct pointing action, regardless of the accuracy of their own response. Finally, participants saw 12 new videos in which the actress pointed to each object once; participants were instructed to provide the name of each object. They received non-corrective feedback, i.e., they heard a recording of the male actor naming the object, after each attempt.

Sentence Training I: Participants saw videos in which an actor, seen from the back, named all the props on the podium while an actress pointed at them. Visual displays always contained two inanimate objects (e.g., regular-sized spoon and tweezers) and two animals, each holding a miniature version of one of the regular-sized instruments (e.g., a small spoon; small tweezers). Thus for a display containing a big spoon, a frog holding a small spoon, a pair of big tweezers, and a horse holding small tweezers, the subject would hear “supo”, “supo”, “lepa”, “pala”, “pala”, “tato” accompanied by the appropriate pointing to each object.[8] The position of the first-named objects was counterbalanced across videos (e.g., lower left, upper right, etc.). This was done so that the last object named in this phase – the target omitted patient in constructions (IV) and (VIII) – was not associated to a fixed position in the display. After each prop was named, the actor gave a spoken command to the actress, who then carried out the corresponding action (for a video examples of Sentence Training items, see http://www.ircs.upenn.edu/~truesweb/tulka_link.html). Participants viewed 16 videos paired with 16 sentences (i.e., two sentences per verb). They were instructed to repeat each sentence out loud after the end of each video presentation (i.e., after the actress had carried out the action). (See General Discussion for the possible learning implications of this presentation order, i.e., hearing a sentence then seeing the action rather than the other way around.) No corrective feedback was provided. Each experimental verb was presented in two of the four possible constructions exemplified above. Each experimental construction was thus exemplified twice: two of the experimental verbs were presented in construction I and IV, while the remaining two verbs were presented in construction II and III. These constructions were paired because they do not represent minimal sentence pairs. The order of presentation was kept constant across all learners.

Short Testing: Sentence Comprehension: An inclined podium was placed in front of the participant at the beginning of this phase. At the center of the podium was a hole for a camera, which was focused on the participant’s eyes. At the beginning of each trial, an experimenter would position a prop in each of the quadrants of the podium to match the position of the same props in the video being projected on the wall (Figure 2). Each object was pointed at and named by the actors in the video. The trial began with a written English instruction prompting the participant to look at the center of the podium. The participant then heard a command in the new language and carried out an action in response to it. A second camera recorded the participants’ actions and the position of the objects on the

[8]Animals in *Sentence Training* and *Sentence Testing* videos always held small instruments. The instrument was always named before the animal. The actress never used the small instruments, which were glued to the plush animals, to carry out actions, nor did any of the subjects do so during test.

podium. No feedback was given regarding the accuracy of the participant's action. Participants viewed 16 videos (half experimental, half filler) paired with 16 sentences (i.e., two sentences per verb). Each verb was presented in the two constructions in which it was not presented during *Sentence Training I*.

Sentence Training II. Participants saw 16 videos paired with 16 commands (two sentences per verb). As before, they were instructed to repeat each sentence aloud at the end of each video presentation. Each experimental verb was presented in the two constructions in which it was not presented during *Sentence Training I*.

Short Testing: Sentence Production: Participants saw 16 videos in which the actress performed an action, but the command spoken by the actor had been replaced by a beep. The participant's task was to guess the original command. Participants' responses were recorded via a digital recorder. Each experimental verb was presented in the same two constructions in which it was presented during *Sentence Training I*. At the end of the day, each verb had appeared in each of the four possible constructions twice.

Day 2—The main differences between Day 1 and Day 2 were that (a) there was no vocabulary training on Day 2, (b) the order of the Sentence Comprehension and the Sentence Production tasks was switched, and (c) a new set of videos and corresponding sentences were employed.

Vocabulary Testing: Participants did not receive any training on the nouns of the language, but were tested on their comprehension and production. The Comprehension and Production tests were exactly like the ones on *Day 1*, but a new set of videos was used.

Sentence Training I: The training procedure was the same as the one used on Day 1, except that we used a new set of videos and sentences. As on Day 1, each experimental verb was presented in two of the four possible constructions. Two of the experimental verbs were presented in constructions I and IV, while the remaining two verbs were presented in constructions II and III; each experimental construction was thus exemplified twice. If a verb had been presented in construction I and IV, on *Day 1- Sentence Training I*, it would be presented here in construction II and III, and vice versa.

Short Testing: Sentence Production: The production procedure was the same as the one used on Day 1, but a new set of videos and sentences was used. Each experimental verb was presented in the two constructions in which it was not presented during *Day 2 - Sentence Training I*.

Sentence Training II: Once again, the training procedure was the same as the one used *on Day 1*, except that we used a new set of videos and sentences. Each experimental verb was presented in the two constructions in which it was not presented during *Day 2 - Sentence Training I*.

Short Testing: Sentence Comprehension: The comprehension procedure was the same as the one used on Day 1, but different videos and sentences were used. Each experimental verb was presented in the same two constructions as the ones from *Day 2 - Sentence Training I.*

Day 3: No training was administered on Day 3. The order of the *Sentence Production* and *Sentence Comprehension* procedures was counterbalanced across experimental lists.

Long Testing: Sentence Production: The production procedure was the same as the one used on the previous days, but 32 new videos/sentence pairs (half experimental, half filler) were created. The experimental verbs were to be combined with NPs that they had not occurred with before. Each verb was presented once in each of the four possible constructions.

Long Testing: Sentence Comprehension: The comprehension procedure was the same as the one used during the previous days, but 32 new videos and sentences (half experimental, half filler) were created. The experimental verbs were combined with NPs that they had never been presented with before. Each verb was presented once in each of the four possible constructions.

Subjects saw different videos during comprehension and production. Two lists were created, so that if a video was presented in the Production procedure in list 1, it would appear in the Comprehension procedure in list 2. The order of video presentation was counterbalanced across subjects via the creation of two additional lists in which the order of videos was reversed.

2.1.4 Coding of Actions—Trained coders watched and transcribed the video recordings of participants' actions; participants' actions were then scored independently by the first author and a trained research assistant. Inter-coder agreement was 98%; disagreements were resolved through discussion. Participants' actions were scored in terms of their accuracy for the vocabulary items (e.g., verbs and nouns) and the morphological markers independently. Verbs were scored as correct if the target action type was carried out. Morphological markers were coded as correct or incorrect depending on whether the participant's action (instrument vs. non-instrument) matched the information provided by the morphological marker ("-ka" vs. "-tu") in the sentence. Morphological accuracy was calculated independently of whether the target action was carried out in so far as an action that could be done with either an instrument or the hands was performed; for this reason, morphological errors always involved the use of an instrument in response to a non-instrument marker or vice versa. So, for example, if in response to the sentence like "zumpi-tu nunu", a participant lifted, rather than bounced, the clothespin, her response would be scored as incorrect in terms of verb accuracy but correct in terms of the morphological marker accuracy. Participants' comprehension for morphological markers and verbs are reported separately in *Sections 2.2.1.1* and *2.2.1.2*, respectively.

2.1.5 Coding of Eye Movement Data—Participants' eye movements during the sentence comprehension task were coded frame by frame from the video recordings obtained from the camera placed behind the podium. The digital camera recorded at a speed of 30 Hz, i.e. a frame every 33 ms. The video and audio annotation software ELAN (version

4.6.1.) was used to annotate the onset and offset of each experimental sentence, and the onset, offset and direction of each fixation for the duration of the sentence until 2000 ms after sentence offset. The coders were instructed to use one of six possible codes for each fixation: top left, top right, bottom left, bottom right, middle, and other. A fixation would be coded as "other" if the subject's eyes were closed, not visible, or the direction of the eye gaze could not be determined. Three research assistants coded all the videos; subsequently, 20% of the videos were re-coded by an independent coder. Average inter-coder agreement was very good (88%, Kappa = .84, $p < .001$). In cases where the two coding mismatched, the original coding was used. While eye movements were collected for the comprehension portion of the study for all three days of the study, they were coded only for Day 3. This is because of the relatively low number of items tested on Day 1 and 2 (four per condition on each day) and particularly low actout accuracy on Day 1 (~ 50% overall across languages).

2.1.6 Coding of Productions—Participants' productions were transcribed by a trained transcriber. Transcriptions were then scored independently by the first author and a trained research assistant in terms of their accuracy for vocabulary items and morphological markers. Inter-coder agreement was 96%; disagreements were resolved through discussion. Vocabulary items (verbs and NPs) were scored as correct if the target word was used. Target words with phonemic substitutions (e.g., "tufo" for "tufa"; "clemi" for "glemi") were considered correct so long as the target word was recognizable. Productions of morphological markers were coded as either correct or incorrect; incorrect productions were further coded as errors of either omissions or commission.[9] Errors of commission consisted of the production of the instrument morpheme to describe a non-instrument event or vice versa, or in the production of a morpheme that was not part of the language inventory (e.g., -to, -ku, -ta, etc.); this type of error was rare and occurred with comparable frequency in the two conditions (instrument condition: N = 13; non-instrument condition: N = 14). The majority of the errors were errors of commission (.66 of errors overall, .68 of errors in late-cue languages and .63 in early-cue languages), indicating that participants were sensitive to the morphological characteristics of the language grammars (i.e., alternation between different morphological markers), without having reached complete mastery of their grammatical function. In the analyses below errors of omission and errors of commission are collapsed together. While the overall numerical patterns (higher accuracy rates for early-cue than late-cue languages) is observed for both error types, the pattern was more pronounced for errors of commission (overall omission errors: N = 68 vs. N = 57 in late-cue languages and early-cue languages, respectively; overall commission errors: N = 144 vs. N= 98 in late-cue languages and early-cue languages, respectively). Word order errors were extremely rare (N = 19 in total) and occurred only in late-cue languages: they either involved V-initial productions when the input language was a V-final language (N = 11) or a switch in the relative position of the two NP's (N = 8). Morphological accuracy was calculated

[9]This coding was straightforward for the two verb-marking languages, since one of the two morphological markers always needed to be produced. For the two noun-marking languages, in contrast, the morphological marker -tu was only to be produced if the function of the inanimate NP was that of a NP modifier, see *Section 2.1.3.3* above. If the inanimate object was the patient of the action (i.e., for non-instrument 1-NP sentences), no morpheme should be produced. For these reasons, productions of this type were coded as correct in terms of the morphological marker if no morpheme was produced. Again, our predictions pertained to differences in early- vs. late-cue languages, thus these differences between verb- and noun-marking, which cross-cut early- vs. late-cue comparisons, cannot explain observed differences that pertain to our hypotheses.

independently of the accuracy of vocabulary items; for example, a production like “glemitu kefu” in response to a video for which the target production was “blikitu kefu” would be scored as incorrect in terms of verb accuracy but as correct in terms of morphological accuracy. Participants’ productions morphological markers and verbs are reported separately in *Sections 2.2.3.1* and *2.2.3.2*, respectively.

2.1.7 Statistical Analyses—In the results reported below, learners’ comprehension and production accuracy profiles are graphed in terms of proportions; statistical analyses are instead based on multi-level mixed effects logistic regressions; the random effect structures for our models always included by-subject and by-item random intercepts; by-item random slopes for the effect of cue order (early vs. late) were modeled whenever the latter was included as a fixed effect in the model, unless otherwise specified.[10] For the analyses that investigated differences among groups in terms of the number of subjects that reached “acquisition” criterion, chi-square analyses were performed (see end of *Sections 2.2.1.1* and *2.2.3.1* and *2.2.4*).

Statistical analyses of eye gaze patterns were performed using cluster-based permutation analysis (Maris & Oostenveld, 2007). These analyses are used to discover significant clusters of observations within a data set over time. To do this, a critical t is first calculated based on the number of degrees of freedom (for both subjects and items); the t-statistic for each time-bin in the cluster needed to exceed this value. Second, a t-value for each of the time-bins in the series is calculated. Individual bin t-values are then randomly reshuffled across the time series and the size of the maximum random cluster of t-values is calculated for each reshuffling. Random reshuffling is then repeated a large number of times (in our case, 1000 iterations); this procedure gives rise to a distribution of the maximal t-values for the randomly obtained clusters. The proportion of random partitions that result in a test statistic larger than the ones observed in the dataset is calculated. This proportion represents the p-value. If the p-value is smaller than the critical alpha-level (i.e., .05), the t-statistic for a given cluster in the data is considered to differ significantly from chance.

2.2 Results and Discussion

Results are divided into four sections below. First (in *Section 2.2.1 Actions*), we present participants’ actions in response to sentences containing instrument and non-instrument morphology. Second (in *Section 2.2.2 Eye Movements*), we analyze how participants use morphological information during online sentence processing to guide their exploration of the visual world. Third (in *Section 2.2.3 Production*), we analyze participant’s productions in response to videos depicting events carried out with or without an instrument. Finally (in Section 2.2.4 Offline Survey of Grammatical Knowledge), we report a summary of the responses of a short survey we presented at the end of the study, aimed at further

[10]In a few occasions (see *Section 2.2 Results and Discussion*), the models that included cue order as by-item random slope failed to converge; these models, were simplified by dropping the random slope. Including the by-subject random slope for the fixed effect of “day of study” resulted in convergence failure for the models; as a result, this random slope was dropped. By-item random slopes for the effect of locus of marking and by-item and by-subject random slopes for the effect of condition (instrument vs. non-instrument) also caused convergence issue in a number of models and were therefore dropped uniformly from all models.

investigating participant's knowledge of the function of morphological markers in the languages they were exposed to.

To reiterate our experimental hypotheses, we predicted that languages in which morphological cues to argument structure are available at the beginning of a sentence (and as such can be used in early processing stages to guide participants' interpretations) will be acquired more successfully and at a faster pace than languages in which the same information becomes available only at the end of the sentence (and as such can only be used to confirm or revise initial parsing commitments). We predicted that, if processing ease has an effect on language learning as a whole (e.g., learning of the morpho-syntactic properties of the language), we should see an effect of cue ordering on both comprehension and production performance; this effect should obtain in both verb-marking languages and noun-marking languages.

2.2.1 Actions

2.2.1.1 Morphological Markers: Figure 3 presents mean comprehension accuracy rates for the two morphological markers as a function of day and cue order (early vs. late-cue), collapsing across locus of marking (whether the morphology was expressed on the noun or the verb) and number of NPs. Accuracy increased linearly throughout the three days for both language variants (day (linear): *Estimate* = 1.08, *SE* = .24, $p < .001$). But, critically, the linear increase in accuracy also interacted with cue order in the predicted direction (cue order by day (linear): *Estimate* = .56, *SE* = .23, p = .01), indicating better learning curves for early-cue than for late-cue languages. Act-out accuracy for early-cue and late-cue languages did not differ on either Day 1 (*Estimate* = .11, *SE* = .26, p = .67) or Day 2 (*Estimate* = .55, *SE* = .49, p = .26) and did not interact with locus of marking on either day (Day 1: *Estimate* = .38, *SE* = .48, p = .43; Day 2: *Estimate* = −1.54, *SE* = .94, p = .10, respectively). On Day 3, however, participants who learned the early-cue variants of the grammar displayed higher accuracy than participants who had learned the late-cue variants (*Estimate* = 1.05, *SE* = .50, p = .04), regardless of the whether morphological cues were expressed on nouns or verbs (cue order by locus of marking: *Estimate* = −.50, *SE* = .99, p = .61), as shown in Figure 4.

An additional analysis was conducted to investigate whether accuracy rates were modulated by the type of morpheme ("-ka" vs. "-tu") used in the experimental sentence. In Table 3 in the Appendix, we provide average comprehension accuracy rates for the two morphemes as a function of number of NPs used in the sentence (1 vs. 2), cue order (early vs. late) and day of study. Although space limitations preclude a full discussion of these patterns, certain aspects of this table are worth noting: sentences marked with instrument morphology were associated with overall lower accuracy (Estimate = 1.76, SE = .16, p < .001); this difference did not interact with cue order (Estimate = .46, SE = .32, p = .15), indicating a comparable disadvantage for instrument morphology in both early- and late-cue language types. In contrast, this effect of morpheme type was modulated by number of NPs in a sentence (Estimate = 1.86, SE = .41, p <.001). This interaction came about because, especially in Days 1 and 2, errors occurred more often when the language-specific cue (i.e., the morphological marker) conflicted with the universal NP number cue (i.e., instrument 1-NP sentences and non-instrument 2-NP sentences). This is in line with the hypothesis that, when language-specific knowledge is not yet or only partially in place, learners have a tendency to use the number of NPs in a sentence as predictor of argument structure.

To further test the hypothesis that at the beginning of the acquisition process, language learners rely on the number of NPs in a sentence as a potent predictor of argument structure, and that this tendency decreases as language-specific knowledge is accumulated, we conducted an analysis including only those sentences for which the number of NPs mismatched with the number of thematic roles in an event (i.e., instrument 1-NP sentences and non-instrument 2-NP sentences). We asked whether learners' tendency to act-out these commands by relying (incorrectly) on the number of NPs (i.e., by not using an instrument in instrument 1-NP sentences and by using an instrument to act-out non-instrument 2-NP sentences) was modulated by day of study, and cue order.[11] Thus decreases across days would indicate that learners began to stop using the number of NPs as a predictor to argument structure and instead began to (correctly) use morphological information. As can be seen in Figure 5, learners do indeed begin with a reliance on number of NPs, in that on the first day of the study they matched the number of NPs to the number of thematic roles more often than expected by chance (50%) for both early-cue (*Estimate* = .58, *SE* = .25, p = .02) and late-cue languages (*Estimate* = .47, *SE* = .22, p = .03), consistent with the hypothesis that learners initially adopt this universal parsing heuristic. Learners' reliance on NP number decreased linearly throughout the study (day (linear): *Estimate* = −1.44, *SE* = .17, p < .001). As expected, this linear decrease in participants' reliance on number of NPs was less pronounced for late-cue languages (day (linear) by cue order: *Estimate* = .72, *SE* = .33, p = .03). Learners' initial reliance on NP number was high and did not differ across early-and late-cue languages on either Day 1 (effect of cue order: *Estimate* = −.09, *SE* = .32, p = .76) or Day 2 (*Estimate* = .94, *SE* = .68, p = .16). On Day 3, participants who learned the late-cue variants of the grammar were on average more likely than participants who had learned the early-cue variants to continue incorrectly to use NP number to predict argument structure; this difference was marginally significant (*Estimate* = 1.15, *SE* = .63, p = .07).

In this study, we were interested not only in whether our manipulation would have an effect on average accuracy rates, but also whether it would affect the number of people who could be considered "proficient" speakers of the language after a short exposure to the language system. To address this question, we first established a criterion for "acquisition" based on the child language acquisition literature (80% accuracy, see Brown, 1973) and then used this criterion to establish if participants' comprehension scores on Day 3 passed the threshold for successful acquisition. Based on this criterion, considerably more people acquired our early-cue languages (17/24 participants) as compared to our late-cue languages (10/24), (χ^2 (1) = 4.15, p = .04).

[11]Locus of marking was not included in these analyses due to data sparseness on Day 1 and Day 2. This is because only sentences in which the number of NPs mismatched with the number of thematic roles could be used in this analysis, resulting in only four sentences (and two verbs) per participant on Day 1 and Day 2. On Day 3, where double the amount of data was available, the effect of locus of marking and its interaction with cue order were not significant (locus of marking: *Estimate* = .55, *SE* = .59, p = .35, cue order by locus of marking: *Estimate* = −1.45, *SE* = 1.17, p = .22); by-item random slopes for the effect of cue order were dropped for this analysis due to convergence failure.

2.2.1.2 Verb Vocabulary: One might also expect that learners' ability to provide a target argument structure / syntactic analysis for a given sentence might result not only in faster acquisition rates for the grammar of the language (in this case, the function of the morphemes) but also in faster acquisition rates for the (untaught) meaning of vocabulary items; such an effect, however, is expected to be specific to verbs that are associated with a syntactic ambiguity, i.e., experimental verbs. In contrast, filler verbs, which never involved ambiguity and garden-path recovery, should be acquired at similar rates, independent of the cue order of the language. We examined this here by calculating action accuracy for the event type (e.g., whether, independent of correct use of intended arguments, participants carried out a correct action, such as lifting for "bliki", looking for "mefi", etc.).

Mean comprehension accuracy rates for the action vocabulary of the language as a function of day, cue order and item type (experimental vs. filler) are shown in Figure 6. On average, participants' comprehension of the action vocabulary of the language (i.e., their understanding of the meaning of the verbs) increased throughout the study for both verbs that denoted actions that could be carried out with an instrument or the hands (experimental verbs: day (linear): *Estimate* = 2.38, *SE* = .26, $p < .001$) and verbs that denoted actions that could only be carried out with a body part (filler verbs: day (linear): *Estimate* = 3.88, *SE* = .39, $p < .001$). Moreover, a significant effect of cue order emerged for experimental verbs only: while accuracy significantly increased throughout the study for both early-cue (day (linear): *Estimate* = 2.42, *SE* = .31, $p < .001$) and late-cue languages (day (linear): *Estimate* = 2.35, *SE* = .31, $p < .001$), accuracy for was overall higher for early-cue languages (cue order: *Estimate* = .72, *SE* = .36, $p = .05$). This effect was not significant for filler items (*Estimate* = .80, *SE* = .51, $p = .11$). No other effects or interactions were significant (all p's > .05).

2.2.1.3 Summary of Action Data: The act-out results indicate that the order with which morphological cues to argument structure become available to the processing system has an effect on the ease with which these cues are learned and used during sentence processing: comprehension accuracy for morphological cues to argument structure increased at faster rates and was higher on the last day of the study for early-cue than late-cue languages. Moreover, participants' comprehension of the meaning of the experimental verbs was overall more accurate for early-cue than late-cue language variants, indicating perhaps that being able to assign the correct structural analysis to the linear string had a positive impact on participant's ability to infer or remember the target meaning for the novel word. In line with this speculation, no significant effect of cue order emerged for filler verbs, for which no morphological contrast was present (i.e., the type of action to be carried out was solely determined by the meaning of the verb); this indicates that the observed cue-ordering effects are selective to cases in which morphological cues to argument structure need to be used for successful comprehension.

2.2.2 Eye Movements

2.2.2.1 Real-time Measure of Morphological Discrimination: The purpose of this analysis was to investigate learners' ability to use morphological information in real time to guide their interpretation of sentences, and examine whether it differed as a function of the order in which disambiguating morphological information became available in the sentence. We examined participants' eye gaze responses to structures with instrument and non-instrument morphology during the act-out comprehension tests. We predicted that participants acquiring early-cue languages would show greater discrimination of instrument vs. non-instrument morphology and that such discrimination patterns would appear earlier after disambiguation for early-cue than in late-cue languages, since early-cue languages do not require costly revision. These analyses are based on eye-movement patterns for correct trials only, because this allows us to investigate the extent to which the order of disambiguating cues affects online processing commitments, even for trials in which the order of morphological cues does not influence offline accuracy. All analyses are based on eye-movement patterns collected on Day 3 (*Long Testing Task: Sentence Comprehension*). We excluded the comprehension tests of Day 1 and Day 2 because of data sparseness: only eight test trials appeared on each day and off-line actions were frequently incorrect (see Figure 3 above).

In order to investigate participants' online ability to discriminate between instrument and non-instrument morphological markers and whether this ability was affected by the order with which morphological cues to meanings become available in a sentence, we examined whether instrument and non-instrument sentences were associated with different looking patterns, and how this difference was affected by cue order (late vs. early). A 'discrimination score' was calculated for each 100 ms time bin by subtracting the proportion of looks to the relevant inanimate object when non-instrument morphology was present from the proportion of looks to the same object when instrument morphology was present. For example, given the sentence pair "zumpi-ka fami nunu" ('bounce the dolphin using the clothespin') and "zumpi-tu fami nunu" ('bounce the dolphin, which is holding the clothespin'), a discrimination score was calculated by subtracting the proportion of looks to the clothespin associated with the sentence containing non-instrument morphology from the proportion of looks to the clothespin associated with the sentence containing instrument morphology. Scores close to zero thus indicate poor discrimination between the morphological markers, positive scores indicate that sentences containing instrument morphology were associated with higher proportions of looks to the inanimate object than sentences containing the non-instrument morphology, and vice versa negative scores indicated that sentences containing the marker "-ka" were associated with lower proportions of looks to the inanimate object than sentences containing non-instrument morphology.

For 2-NP sentences, target discrimination of morphology should result in discrimination scores that are positive in sign, i.e., higher proportion of looks to the inanimate object with instrument than with non-instrument morphology. This is because, in sentences marked with instrument morphology, the inanimate object is the instrument with which the action should be carried out (i.e., "Bounce the dolphin using the clothespin"); in contrast, in non-instrument sentences, the inanimate object is to be interpreted as a modifier of the direct object (i.e., "Bounce the dolphin that has the clothespin"). In this latter case, virtually no looks to the quadrant containing the inanimate object alone (i.e., the big clothespin) are expected.

In contrast, for 1-NP sentences, the expected pattern of discrimination is opposite in sign and more nuanced: looks to the inanimate object were expected in both sentences with instrument and non-instrument morphology, since in both cases the inanimate object is involved in the action. In 1-NP sentences with non-instrument morphology, the inanimate object is the intended patient in the event (e.g., "zumpi-tu nunu", 'bounce the clothespin'), and is the only relevant referent in the visual world. On the other hand, in 1-NP sentences with instrument morphology (e.g., "zumpi-ka nunu", 'bounce Ø using the clothespin'), participants' attention should be divided between the omitted patient (i.e., the last object mentioned during naming) and the inanimate object instrument (i.e., the clothespin). Accurate interpretation of 1-NP sentences, is thus expected to result in discrimination scores that are negative in sign, i.e., lower proportion of looks to the inanimate object with instrument than with non-instrument morphology.

Figure 7 plots average discrimination scores for 1- and 2-NP sentences (indicated by dashed and solid lines, respectively) over time, from the point in which both the identity and the function of the inanimate object is disambiguated[12] until 2000 ms after it, separately for early-cue (black squares) and late-cue (gray circles) languages.

For 2-NP sentences, discrimination patterns differed across time between early- and late-cue languages. For early-cue languages, discrimination scores rose significantly above zero around 700 ms after disambiguation and remained positive until the end of the time window. In contrast, for late-cue languages, discrimination scores became significantly positive only at the very end of the time-window, between 1900 and 2000 ms after disambiguation. Discrimination scores were also more pronounced for early-cue languages than late-cue ones; this difference reached statistical significance between 600 ms and 1700ms after disambiguation.

A similar picture emerged for 1-NP sentences. For early-cue languages, discrimination scores fell significantly below zero around 600 ms and remained negative until the end of the time window. For late-cue languages, discrimination scores fell significantly below zero around 1000 ms and remained negative until the end. Discrimination scores were more pronounced for early-than late-cue ones; this difference was statistically significant between 600 and 900 ms after disambiguation.

The pattern that emerges from the analysis of learners eye-movements as they process sentences in real time thus confirms and extends the results obtained from the analysis of their off-line actions: languages in which morphological cues to argument structure arrive at the beginning, rather than at the end of the sentence, enable learners to use this information faster and more effectively. Importantly, this result is not a simple reflection of the fact that disambiguation information arrives earlier in early-cue languages, and can thus be used at an earlier point, because this pattern emerges after taking into account the relative position of disambiguating information in the different language variants (by time locking our measurement to the onset of the relevant morpheme). This pattern is also not a simple

[12]The point at which both the identity and the function of the inanimate object is disambiguated is indicated for all language and sentence types by boldfacing the relevant word in the Appendix. For language variants (II)-(IV), the morphological marker signals the point in which the listener has received enough information to know both the identity and the function of the inanimate NP. For language variant (I), instead, the morphological marker expressed on the verb (e.g., "zumpi-tu") only provides information about the type of action that was to be carried out (i.e., instrument vs. non-instrument), while the identity of the inanimate object becomes available at the very end of the sentence (e.g., "nunu"). The onset of this latter NP was then taken as the onset of disambiguation.

reflection of the higher proportion of correct actions (and thus higher looks to correct objects) in early-cue languages, since the analyses were conducted on correct trials only.

2.2.2.2 Signs of Garden-Pathing in Late-Cue Languages: The previous analyses indicated that learners use morphological information that arrives early in the sentence more effectively to direct their attention to the visual scene during online sentence processing. However, these analyses do not allow us to conclude that slower rates of convergence on or away from a given referent in learners of cue-final languages reflect initial misinterpretation of its thematic role – an important prediction of our model.

To investigate this issue, we conducted a follow-up analysis on a subset of the eye-movement data. We focused on how looks to the inanimate object in (correctly acted-out) 2-NP sentences are modulated by morphological cues in the cue-final verb-marking language.[13] This language variant and structure were chosen because the inanimate NP and the morphological cues that disambiguate its thematic role in the event (i.e., instrument vs. modifier) are maximally separated in time (e.g., "kefu fami glemi-tu/-ka"), thus allowing us to distinguish more precisely between initial and final interpretations of the thematic role assigned to the inanimate object.

In order to investigate whether learners of this language variant initially interpreted the inanimate object as the instrument with which the action should be carried out and later had to revise this interpretation upon hearing the non-instrument morpheme "-tu", we identified two broad time windows: the first one, from the onset of the inanimate NP until the end of the lexical verb but before the onset of the morphological cue (approximately the first 2000 ms), and the second one, comprising the 2000 ms after the onset morphological cue. In the first time window, looks to the inanimate object should not differ between sentences marked with the instrument marker "-ka" and those marked with the non-instrument marker "-tu", since the marker has not yet been heard. More crucially, if the inanimate NP is initially interpreted as the instrument with which the action should be carried out, looks to the inanimate object should not increase significantly when the instrument marker is heard, as compared to the first portion of the sentence, before any marker is heard, but should be significantly reduced in the presence of the non-instrument maker "-tu".

Our predictions received support from the data. For "-ka" and "-tu" sentences respectively, mean proportions of looks to the instrument were .26 and .24 in the first time window and .31 and .14 in the second time window. Mixed-effects regression models based on e-logit transformed data revealed that looks to the inanimate object during the first time window (e.g., "kefu fami glemi") were comparable for sentences with instrument and non-instrument markers (*Estimate* = −.75, *SE* = .71, *p* = .31). While looks to the inanimate object increased numerically between the first time window (e.g., "kefu fami glemi") and the second time window if the instrument marker "-ka" was heard, this increase was not significant (*Estimate* = .71, *SE* = .59, *p* = .22). On the other hand, looks to the inanimate object significantly decreased in the presence of the non-instrument marker "-tu" (*Estimate* = −1.02, *SE* = .52, *p* = .05).

[13]We thank an anonymous reviewer for this suggestion.

This pattern is in line with the experimental hypothesis that learners of late-cue languages incrementally use the number of NPs in a sentence as a cue for thematic role assignment, which is then confirmed or revised by late-arriving morphological evidence.

2.2.3 Production—The results reported in the previous section suggest that languages in which morphological cues to meaning are available early, rather than late, in a sentence, are associated with higher accuracy and more rapid disambiguation in comprehension. Crucially though, we also investigated learners' production abilities so as to determine whether differences between early-vs. late-cue languages simply reflected garden-path effects in comprehension or instead reflected more general effects of cue ordering on grammar acquisition as a whole. Our prediction was that, if processing limitations have an effect on grammar acquisition generally, (a) effects of cue ordering should also emerge in tests of production and (b) individual comprehension and production performance profiles should be similar.

2.2.3.1 Morphological Markers: Figure 8 presents average production accuracy for the morphological markers as a function of day and cue order, collapsing across locus of marking. Production accuracy increased throughout the three days for both language types (day (linear): *Estimate* = 1.54, *SE* = .15, $p < .001$).[14] Critically, the linear increase in production accuracy interacted with cue order in the predicted direction (cue order by day (linear): *Estimate* = 1.20, *SE* = .27, $p < .001$), indicating a steeper learning curve for early-cue than for late-cue languages.

Production accuracy between early-cue and late-cue languages did not differ on either Day 1 (*Estimate* = −0.19, *SE* = .39, *p* = .63) or Day 2 (*Estimate* = .60, *SE* = .51, p = .23) and did not interact with locus of marking on either day (cue order by locus of marking on Day 1: *Estimate* = .19, *SE* = .77, p = .80; Day 2: *Estimate* = −1.04, *SE* = .99, p = .29). On Day 3, however, participants who learned the early-cue variants of the language displayed higher accuracy rates than participants who had learned the late-cue variants (*Estimate* = 2.28, *SE* = .88, p = .009), regardless of whether the morphological marker appeared on the verb or the NP (cue order by locus of marking: *Estimate* = .13, *SE* = 1.68, p = .94). Accuracy rates by cue order and locus of marking for the last day of the study are shown in Figure 9.

An additional analysis was conducted to investigate whether accuracy rates were modulated by the type of event being described (instrument vs. non-instrument). Event type did not have a significant effect on participants' productions: accuracy rates were comparable for events that involved the use an instrument and events that did not (*Estimate* = .19, *SE* = .14, p = .19) and production accuracy increased throughout the study in a similar manner for the two events types (event type by day (linear): *Estimate* = .41, *SE* = .25, p = .09). For the complete pattern of results, see Table 4 in the Appendix.

Substitution errors were more common in the non-instrument condition and omission errors were instead more frequent in the instrument condition (Substitutions: "-ka" used instead of

[14]The model that included by-item random slopes for the effect of cue order failed to converge, so this random slope was dropped and the final model for this analysis only included by-subject and by-item random intercepts.

"-tu": N = 126; "-tu" used instead of "-ka": N = 89, *Estimate* = −.51, *SE* = .16, p = .002; Omissions: "-ka" omitted: N = 110 vs. "-tu" omitted: N = 13, *Estimate* = 3.48, *SE* = .45, p < .001). Both types of error decreased throughout the study (day (linear): Substitutions: *Estimate* = −.78, *SE* = .14, p < .001; Omissions: *Estimate* = −1.78, *SE* = .41, p < .001), but the observed pattern of errors was not modulated by day of the study (condition by day (linear): Substitutions: *Estimate* = .33, *SE* = .28, p = .24; Omissions: *Estimate* = .57, *SE* = .81, p = .48).

As mentioned above, in this study we were interested in determining not only whether the order in which morphological cues to meaning become available to the processing system has an effect on comprehension and production performance, but whether it would affect acquisition of the grammatical system as a whole. To answer this question, we first compared early and late-cue languages in terms of the number of people whose production profiles indicated that they had "acquired" the grammatical system. For this analysis, we used the same criterion in production that we had used in comprehension (80% accuracy, see Brown, 1973). The results indicate that more people reached criterion in early-cue than late-cue languages (21/24 vs. 13/24, χ^2 (1) = 6.45, p = .01). Second, we investigated the existence of a correlation between accuracy rates across tests of production and comprehension on the last day of the study. The results of this analysis show similar profiles for production and comprehension; in both groups, accuracy rates in comprehension and production correlated significantly (r^2 = .73, p < .001 and r^2 = .43, p < .001 for late-cue and early-cue languages, respectively). More crucially, as shown in Figures 10-11, a qualitatively different pattern emerges for the two language groups: while most learners of early cue-languages display high accuracy rates for both production and comprehension, thus showing evidence of having acquired the grammatical system, a substantial number of learners of late cue-languages display low accuracy rates in both comprehension and production and thus show no evidence of having acquired the grammatical system. These results are in line with the experimental hypothesis that processing constraints and limitations affect grammatical development as a whole.

2.2.3.2. Verb Vocabulary: In order to investigate whether learners' ability to provide a target argument structure / syntactic analysis for a given sentence might result not only in faster acquisition rates for the grammar of the language (in this case, the function of the morphemes) but also in faster acquisition rates for the (untaught) meaning of vocabulary items, we calculated production accuracy for the verbs of the language (e.g., whether, independent of correct production of morphology, learners produced the correct verb, such as "bliki" for *lifting*, "mefi" for *looking*, etc.).

This prediction was not born out in the production data (see Figure 12). On average, participants' production of the verb vocabulary of the language increased linearly throughout the study both for verbs that denoted actions that could be carried out with and without an instrument (experimental items: day (linear): *Estimate* = 1.86, *SE* = .36, *p* < .001) and verbs that denoted actions that could only be carried out with a body part (filler verbs: day (linear): *Estimate* = 2.08, *SE* = .26, *p* < .001). However, there was no effect of cue order on verb production accuracy for either experimental (*Estimate* = .25, *SE* = .46, p = .58) or filler items (*Estimate* = .50, *SE* = .44, p = .25); no other effects or interactions were significant (all p's >.05).

2.3.3.3 Summary: The results from the production portion of the study confirm and extend the results that emerged from the analysis of learners' online and offline sentence comprehension performance. Although differences were not observed in verb-production, effects of cue order had the expected effects on the production of morphological markers: accuracy rates were overall higher for learners of early-cue than late-cue languages on the last day of the study, and early cue languages were associated with a higher proportion of learners whose production of morphological markers indicated mastery of the grammar of the language. Taken together, these results indicate that the order with which morphological cues to argument structure become available to the processing system has a sizeable effect on participants' ability to learn these cues and use them in real time.

2.2.4 Offline Survey on Grammatical Knowledge: The main prediction of our study was that the preferences and limitations of the human sentence parser, and in particular, difficulties revising initial interpretations, would not just affect learners' ability to use morphological cues to meaning during sentence comprehension, but the language acquisition path *tout court*. One way to investigate this is to examine production as well as comprehension, since revision of initial interpretations does not play a role in this latter case. The previous section showed an effect of cue order on production accuracy, thus lending support to our predictions. A second way to investigate this issue is to ask language learners' about the function and meaning of the morphemes and about their interpretative strategies during sentence comprehension. If cue order only had an effect on processing of temporarily ambiguous sentences and not on grammar acquisition more generally, we would not expect a difference in participants' ability to describe the meaning of the morphological cues; on the other hand, if cue order has an impact on grammar acquisition, learners of early-cue languages might display both higher performance accuracy in the processing tasks and a superior knowledge of the language grammar; the latter might translate into a better ability to describe the meaning of the morphemes. To investigate participants' knowledge of the function of morphology, we administered a post-experiment questionnaire on Day 3, which asked about the grammar of the language ("What did -*ka* mean?"; "What did -*tu* mean?"; "When did you use an instrument/your hand to carry out an action?"), and participants' learning strategies (e.g., "Were you paying attention to the ending of words?"; "Were you actively looking for a pattern?"; "When did you start noticing the pattern?").

Participants were classified on the basis of whether they could verbally describe the meaning of the morphemes (e.g., "-*ka* meant that you had to use an object to move another object, -*tu* meant that you had to do the action with your hands"). There were no differences between the two particular morphemes with respect to participants' ability to describe their meaning: participants either were able to provide the correct meaning for both or neither. There was an effect of cue ordering on participants' knowledge of the meaning of the morphology, such that more of the participants who learned an early-cue variant (.89 vs. .55, χ^2 (1) = 5.29, p = .02) provided responses indicating they could describe the function of morphology in the grammatical system (see Figure 13). No differences in participants' knowledge were found between learners of noun vs. verb-marking languages (χ^2 (1) = .15, p = .70). Most people who noticed the pattern reported that they began to do so on Day 2 (.59 and .58 in early-cue and late-cue languages) and most people reported being actively engaged in looking for a pattern (.61 in early-cue and .67 in late-cue languages; χ^2 (1) = .12, p = .73), an issue we return to in the General Discussion.

3. General Discussion

3.1 Summary and key observations

Across three days, adult participants learned a miniature language by listening and repeating commands in the language paired with short videos in which an actress acted out said commands. In all language variants studied, the number of NPs in a sentence, a cross-linguistically valid cue to argument structure, was not a reliable cue to meaning, due to frequent argument omission and redundant NP modification, whereas morphological cues appearing on verbs or NPs (depending on the language variant) perfectly predicted the number of participants/thematic roles associated with an event. Thus, in order to correctly interpret and produce sentences in the language, learners could not rely on the number of NPs, but needed to produce and attend to language-specific morphological cues to structure.

As predicted, tests of comprehension revealed that learners, independent of the language variant, displayed an initial tendency to assume that the number of NPs corresponded to the number of arguments in the sentence but over the course of exposure usually learned the correct meaning of the morphological cue (see Figure 5 above). Strikingly, the ability to learn the language-specific morphological cue to structure was systematically delayed for those individuals learning a language variant in which the morpheme arrived late in each sentence (late-cue languages) as compared to those learning a variant in which the morpheme arrived early in each sentence (early-cue languages). This was true both in real-time tests of comprehension (Figs 3-6) and production (Figs 8-9), suggesting that general knowledge of the meaning of the morphemes (“-ka” vs. “-tu”) was impaired for late-cue as compared to early-cue language variants. Also, learners of the early-cue language variants were significantly more likely to reach a successful criterion of having acquired the morphemes and were more likely to correctly report the meaning of the morphemes when their knowledge was surveyed at the end of the three-day study (Figure 13). Quite crucially, this was true regardless of whether morphological cues were expressed on NPs or verbs, indicating that the observed superior performance associated with early-cue languages cannot be accounted for in terms of facilitatory L1-transfer effects.

3.2 Evaluation of findings and implications for grammar acquisition

Our results are largely consistent with the processing account of grammar acquisition sketched in the introduction to this paper. Without prior knowledge of language-specific cues to argument structure, listeners tend to follow a simple structure-to-meaning mapping principle, in which each NP identified in a sentence is assumed to denote a separate participant / thematic role in the event denoted by the matrix verb. This pattern has been observed previously in studies of natural language acquisition (e.g., Gertner & Fisher, 2012; Lidz, Gleitman, & Gleitman, 2003; Naigles, Fowler, & Helm, 1992; Naigles, Gleitman, &

Gleitman, 1993). Regardless of the exact nature of this mapping principle (and whether it is better characterized as part of Universal Grammar or as a universal parsing and interpretation heuristic), several predictions follow based on what is known about the general limitations on real-time parsing, especially those limitations pertaining to the ability to revise initial parsing commitments.

In particular, if language-specific cues to argument structure appear late in a sentence (more precisely, if they occur after a universal parsing heuristic has had an opportunity to structure and interpret earlier aspects of the sentence) then such a language-specific cue to structure would be challenging to use and acquire. Even after forming the correct hypothesis about the meaning of such a morpheme (e.g., the meaning of "-ka"), accurate testing of this meaning hypothesis would require revising the initial parse – a parsing process that is known to be difficult for child learners and adults learning a second language. Thus such sentences would be difficult to use as evidence for confirming a correct hypothesis. This is exactly the pattern observed here: learners who were exposed to an artificial language in which a language-specific morphological cue to argument structure appeared late in sentences had greater difficulty accurately comprehending and producing this morphological cue.

Our comprehension results replicate recent findings from a cross-linguistic comparison of children acquiring one of two natural languages: children learning Kannada, a verb-final language, were found to have difficulty correctly interpreting causative verb morphology as compared to age-matched children learning Tagalog, a verb-initial language that also has causative verb morphology (Trueswell et al., 2012). However, as noted in the introduction, this cross-linguistic finding is only suggestive since comprehension, but not production, abilities were tested; moreover Kannada and Tagalog differ in many other ways, both morphologically and syntactically. By examining here both comprehension and production performance in the same group of learners, on languages that differed only in the ordering of cues, the present study provides crucial evidence indicating that the order in which cues to interpretation appear in a sentence, and can be used in real-time by the language user, affects acquisition trajectories as a whole. Given that similar effects of cue order appeared in tests of both production and comprehension, we can be more confident that adults' acquisition of the meaning of the morpheme was impaired in late-cue language variants.

Our particular account of this effect is as follows. In the absence of language-specific knowledge, the mapping of utterances onto meaning is initially guided by universal structure-to-meaning mapping principles, which are gradually abandoned and supplanted by language-specific knowledge as the learner gathers evidence about their inadequacy to process the input successfully. In order to replace these initial heuristics with language-specific knowledge, learners need to be minimally able to (a) notice a mismatch between the analysis of the input provided by their current grammar and the state of affairs in the world, (b) update their grammar to include a new hypothesis, and (c) test the new hypothesis by applying it to subsequent input. We proposed that learners' ability to test and confirm (or disconfirm) a new grammatical hypothesis will be enhanced if this hypothesis can be used to guide the parser towards building the target structure and interpretation from the start (early-cue languages) than when it can be used to confirm or revise a parsing analysis and interpretation (late-cue languages); this, in turn, should engender faster learning rates.

Consider for example the learner who has formed a tentative (and correct) hypothesis that the verbal morpheme "-tu" indicates that a given action should be carried out with the hands (i.e., non-instrument) and is now seeking confirmation of this hypothesis from additional sentences containing "-tu". In a verb-final language, this learner might subsequently encounter a sentence such as "nunu fami zumpi-tu" (clothespin dolphin bounce-non-instrument). Upon first hearing "nunu fami", these two NPs are likely to be analyzed (in real-time) as two separate NPs with different thematic roles (e.g., instrument and patient); upon hearing the inflected verb "zumpitu", our learner would need to use her fragile knowledge of "-tu" to revise this earlier interpretive commitment, and reanalyze "fami nunu" as a complex NP argument. If revision of this sort is taxing, especially for language learners, there is some possibility that the correct interpretation will never be arrived at, and hence no reinforcement will occur based on a lack of successful, coherent connection to the referent world. If on the other hand, the same person were learning a verb-initial language, they would have encountered the inflected verb at the start of the sentence "zumpi-tu fami nunu", allowing for a means to parse correctly "fami nunu" as a single NP argument, since the presence of the morpheme "-tu" should prevent the NP "nunu" to be interpreted as an instrument. This would likely lead to an easy alignment with the referential world, reinforcing the learner's hypothesized meaning of "-tu" as non-instrument.

Note that our account has within it an assumption about the learning procedure used to identify the meaning of a morpheme. We assume that learning of such morphemes is accomplished in a manner of hypothesis testing similar to what our lab has observed for the learning of the meaning of content nouns, a procedure we have dubbed *propose-but-verify* (Medina et al., 2011; Trueswell et al., 2013): Upon the first encounter with a morpheme, learners select a single most plausible meaning based on linguistic and non-linguistic evidence. Upon the next encounter, learners attempt to retrieve that meaning and test it against the current context. If consistent, this meaning is reinforced, if not, the meaning is discarded and learners attempt to identify a new meaning hypothesis. Here real-time processing constraints (and universal parsing assumptions) appear to interact with this learning process precisely in the ways expected. Yet, as we do not have direct evidence that this particular learning procedure underlies the learning of the meaning of morphemes in the present study, we acknowledge that this aspect of our account is only speculative, but consistent with the present findings. Clearly, further research examining the specific time-course of morpheme acquisition will be necessary.

3.3 Alternative Accounts of the Findings

It should be noted that a number of alternative accounts can be ruled out on the basis of the design of the current study and its results. First, the finding that early-cue languages are associated with steeper learning curves and higher accuracy rates at end state cannot be attributed to a learning advantage associated with a greater similarity between English and the verb-marking early-cue variant of the miniature language in terms of word order (V_{MORPH} NP NP vs. NP NP V_{MORPH}). In order to control for word-order similarity to English, we created early- and late-cue language variants in which the morphological markers appeared on the NP's, rather than the verb itself (NP_{MORPH} NP V vs. V NP NP_{MORPH}). The results showed an advantage for the early-cue over the late-cue variants, regardless of the lexical category on which the morpheme was expressed. This pattern cannot then be attributed to greater similarity in word order between the early-cue language variant and English: the NP-marking early-cue variant, in fact, was less similar to English in terms of word order than its late cue-variant counterpart. If anything, the most clear-cut evidence of an acquisition advantage for early-cue languages comes precisely from the NP-marking language pairs; in this case, the language that is most similar to English in terms of word order is the late-cue variant (V NP NP_{MORPH}), and it is associated with higher reliance on NP number and significantly lower accuracy rates than the early-cue variant (NP_{MORPH} NP V) (see Figure 4). While L1-transfer is an undeniable factor affecting adult second language acquisition development, our experimental manipulation allowed us to establish that facilitatory L1-tranfer effects are not the source of the observed, sizeable, learning advantage seen for the cue-initial over cue-final noun-marking variants in our study.

Second, the advantage seen for early-cue languages is unlikely to stem from morphology in clause-initial position being phonologically more salient than morphology in clause-final position. If acoustic or perceptual salience were the main determinant of differences in acquisition between the two experimental groups, late-cue languages, where morphological cues appear at the very end of the clause (as suffixes on either the verb or the NP: e.g. “Nunu fami zumpi**ka**”), should have been associated with a learning advantage over *early-cue* languages, where the morphological cues appear in the middle of the sentence (again, as suffixes on either the verb or the NP, e.g., “Zumpi**ka** fami nunu”). The available evidence, in fact, indicates that clause-final elements are associated with higher perceptual salience than clause-medial ones, due to a number of phonological phenomena occurring at the edge of phrase boundaries, among which phrase-final word lengthening (Klatt, 1976; Wightman et al., 1992; Turk & Shattuck-Hufnagel, 2007) and F0 rising (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986). This property, in turn, has important consequences for acquisition, as convincingly shown by Sundara, Demuth and Kuhl, (2011) for child learners of English: 22- and 27-month-old children were more accurate at producing and displayed higher sensitivity to the presence/absence of the third person singular marker -s when it was placed in clause-final, rather than in clause-medial position.

To the extent that perceptual salience plays a role in our study, the available evidence predicts that it should have enhanced learners’ production and comprehension performance for late-cue language variants. Since the main result of our study is that early-cue languages were associated with higher performance accuracy, we can be quite confident that perceptual salience is not the source of the observed learning advantage.

Finally, the observed advantage is also unlikely to stem from a selective memory advantage for clause-initial material. This is because recall advantages are observed both for elements that occur at the beginning (primacy effects) and at the end (recency effects) of a list, as compared to items that occur in the middle (e.g., Ebbinghaus, 1913; Murdock, 1962; Howard & Kahana, 1999). To the extent that serial-position effects play a role in our study, they do not seem to account in a straightforward manner for the *selective* early-cue (i.e., primacy) advantage seen here.

3.4 Distribution of cues to argument structure and implications for language typology

The finding that early-cue languages present a processing and learning advantage converges well with broad typological patterns found across the world's languages. As observed by Greenberg (1973) and Nichols (1992), languages in which the verb appears clause-initially tend to express syntactic relations between sentence constituents on the verb (Nichols, 1992), while V-final languages tend to express these relations on NPs via case-marking (Greenberg, 1973: Universal 41), thus indicating that early-cue languages are typologically preferred. While the ultimate explanation for such cross-linguistic covariance patterns is a much debated issue in the literature (e.g., Gibson, Piantadosi, Brink, Bergen, Lim & Saxe, 2013), the present findings suggest that certain linguistic properties might be easier to learn than others because they do not overload the (limited) cognitive resources of the language learner; in such a way, processing preferences and limitations that result from general cognitive constraints can affect language acquisition, and, by influencing what learners are more likely to acquire and pass on to the next generation, they might ultimately contribute to language change and might shape the form of language universals.

Such a typological account is largely consistent with the proposals of Hawkins (e.g., Hawkins, 2004; 2012, 2014), who suggests that processing efficiency plays an important role in shaping grammars. Most notably, Hawkins (2012) proposes the principle of "Maximize Online Processing": structures in which properties are assigned to a linguistic element X as it is processed are preferred to structures in which the property is left unassigned or, most relevant to the current discussion, *misassigned* (i.e., require revision). The present findings, and our account, provide a specific mechanism for explaining such a typological preference: initially, learners employ universal parsing heuristics; language-specific knowledge that contradicts such parsing heuristics will be especially difficult to acquire if the information tends to appear in sentence positions where universal parsing heuristics have already had an opportunity to assign a structure to the incoming material; since parsing revision is especially difficult in language learners (children and adults), it should be especially difficult to identify correct grammatical hypotheses when language specific knowledge tends to appear in such positions. If such acquisition differences result in grammatical changes over generations (not tested here), we have an acquisition/processing account of the typological theory expressed by Hawkins.

3.5. Limitations and further questions

There are of course several important issues left unresolved regarding our processing account of grammar acquisition. One potential limitation of our study, and as a consequence, of our proposed account, is that we explored here only one of two main possible communicative scenarios, namely one in which linguistic stimuli become available and need to be assigned a preliminary analysis and interpretation *before* the events to which they refer, rather than one in which linguistic input is co-present or follows the event. That is, sentence exposure for our learners always consisted of first hearing a spoken command and then seeing a video in which an actress acted out the command. In our study, the visual referent world of objects had just been presented to the learner and the objects had just been labeled, thereby setting up a relevant discourse context for the exposure sentence, but the event itself was seen only after hearing the sentence. Due to the real-time, incremental nature of the human sentence parser, sentences had to be assigned a structural analysis and interpretation, which could then be checked against the video and either confirmed or revised.

Since a crucial tenet of our account above is that difficulties associated with late-cue languages stem from documented difficulties with revision of structural and interpretative commitments, a crucial question is whether revision is generally associated with late-cue languages or whether it is only at play when sentence processing is not guided or constrained by additional information (e.g., referential, contextual, etc.). In other words, to what extent do our findings generalize to communicative circumstances where events precede or co-occur with linguistic expressions and can thus constrain and be used to guide sentence parsing? For example, suppose that learners, instead of hearing a sentence before seeing a corresponding video were first exposed to videos (e.g., a woman bouncing a dolphin holding a small clothespin) first and then to the corresponding sentence (e.g., "Nunu fami zumpi-tu"). In this scenario, could knowledge of the event drive the parser to the correct structural analysis from the start, preventing it from costly revision and thus obliterating the learning cost associated with late-cue languages? Our answer to this question can only be tentative at this point, since this possibility was not directly tested in our study.[15] Our intuition, however, is that this particular use of context could only occur under specific laboratory circumstances.

Given a simple-enough referential world, where only one event were occurring at any given time and the range of grammatical contrasts were limited, learners might indeed be able to use evidence from a scene to guide sentence parsing, with the consequence that they might never find themselves down the garden-path.

The critical issue, however, is one of generalizability to real world circumstances. In this regard, our choice to have the utterance precede the event seems to be warranted. First, although there exists some evidence that learning instances in which labeling follows, rather than precedes, exposure to referents might be associated with increased word learning performance (Ramscar et al., 2010), labeling upcoming events is quite common in language learning environments. In a study of child-directed speech, Tomasello and Kruger (1992) found that among utterances that labeled events, mothers used verbs to refer to upcoming "impending" actions about 65% of the time; mothers labeled co-occurring events about 30% of the time, with the remainder 5% of utterances labelling past events (see also Gleitman, 1990, for arguments that such a distribution would be expected).

Second, and perhaps most crucially, even in cases in which referents and events precede their linguistic description, listeners rarely know beforehand which of the many co-present referent(s) and co-occurring event(s) and aspects of events that speakers are going to talk about. This view gets support from both word learning and sentence processing studies; participants' ability to deduce the meaning of a word using exclusively evidence from a

[15]However, it should be noted that, during training, learners were asked to repeat the sentences they heard *after* seeing the videos; to the extent that sentence repetition requires the speaker to actively reconstruct the structure of the recalled sentence (e.g., Potter & Lombardi, 1990; Lombardi & Potter, 1992), learners in our study could have been in the position to use the event to correctly parse the sentences they repeated.

scene interaction is quite limited, particularly so for verbs (see Gillette et al., 1999 for an early demonstration); moreover, although listeners can use visual information to predict the structure and the content of utterances (e.g., for review see Tanenhaus & Trueswell, 2006), some have observed that this ability is hindered in more complex referential environments (Ferreira, Foucart, & Engelhardt, 2013). Finally, it is important to note also that young language learners, even as old as five years of age, show limited ability to use contextual information to guide parsing commitments (e.g., Snedeker & Trueswell, 2003; Trueswell et al., 1999) and that predictive processing requires substantial mastery of a native language (e.g., Lew-Williams & Fernald, 2010).

In sum, to the extent that contextual and referential information constrains but does not determine sentence parsing and, as a consequence, initial structural analyses and interpretation are always subject to error and subsequent revision, the proposed processing effects on acquisition are hypothesized to hold regardless of the relative ordering of linguistic input and referential information.

A second potential limitation of our study concerns the extendibility of our findings to fully-fledged natural languages. We used an experimental paradigm that involved miniature artificial languages, in which otherwise informative cues to interpretation, such as referential, pragmatic, and prosodic cues, are instead (intentionally) partially uninformative. For example, argument omission – one of the devices used in our miniature languages to disrupt the one-to-one mapping between number of NPs in a sentence and number of event participants – is lexically and contextually constrained in English, in that it only occurs with a small subset of verbs, and only under specific discourse conditions, i.e., low discourse prominence (Goldberg, 2000). In our miniature languages, in contrast, argument omission occurred with all lexical verbs, and was not contingent upon any particular discourse status, although the identity of the omitted NP was perfectly recoverable from the context (i.e., it was the last referent mentioned in the preceding discourse). Moreover, while redundant NP-modification – the second device used in our miniature study to disrupt the one-to-one mapping between number of NPs and that of event participants – is common in spoken English and is judged by adult native speakers as fully grammatical (Engelhardt, Bailey, & Ferreira, 2006), its use has also been shown to be associated with subtle linguistic cues (e.g., duration, see Engelhardt & Ferreira, 2013) that might in turn help listeners refrain from non-target interpretations; this was not the case in our miniature languages. For these reasons, future work will need to address the extent to which findings based on (adult) acquisition of highly controlled, but linguistically impoverished, artificial systems can be corroborated by evidence from natural language acquisition.

Another pressing issue left open in our investigation is whether similar results to those reported here can be observed in young children. It is conceivable that cognitive effects on sentence processing and language acquisition might be exaggerated by using artificial language systems with adult language learners, who might rely more heavily than children on domain-general problem solving skills and conscious attention to solve the language acquisition puzzle (see Bley-Vroman, 1990; 2009). In line with this criticism, more than half of the learners in our study reported being actively engaged in figuring out the grammatical regularities of the language system.

While some positive suggestive evidence for processing effects on acquisition already exists in children, in the comparison of three year olds learning verb morphology in a verb-initial vs. a verb-final language (Trueswell et al, 2012), more definitive evidence might be obtained by conducting an analogous miniature language experiment with young children. We expect similar findings would obtain, for any difficulties observed in revising parsing commitments in adults are usually found to be greater and more exaggerated in children (e.g., Choi & Trueswell, 2010; Hurewitz et al., 2000; Kidd, Stewart, & Serratrice, 2011; Trueswell et al., 1999; Weighall, 2008). Moreover, theoretical accounts of garden-path recovery would expect such continuity, especially those accounts that appeal to limitations in executive function/cognitive control. Indeed, our hypothesis stemmed from a consideration of the general constraints and limitations of the cognitive resources available to language learners. In our model, an ideal cognitive system equipped with an infinite working memory capacity, which were able to keep the processed language material within its focus of attention and effortlessly abandon an initial interpretative commitment and re-parse the string once an initial parse turned out to inadequately fit the input, is hypothesized to learn either language variant with equal accuracy. However, listeners' working memory capacity is limited and learners are notoriously bad at revising initial parsing commitments (Trueswell et al., 1999; Pozzan & Trueswell, 2013), possibly because successful revision of initial interpretations rests upon a number of shared domain-general cognitive skills (e.g., conflict resolution, inhibitory control, conflict monitoring, information updating, and shifting abilities) which are not fully developed in young children (Zelazo & Frye, 1998) and might be particularly taxed during the processing of a non-dominant non-native language in adults (e.g., Abutalebi, 2008).

Although the present study did not directly assess the hypothesis that learners' difficulties with revision stem from constraints and limitation of domain-general cognitive resources, a growing body of work supports the hypothesized link between cognitive control and sentence processing abilities. This evidence comes from both correlational and training studies.

With respect to the former, it has been shown that individual differences in (verbal and non-verbal) working memory capacity and executive functions skills positively correlate with individual differences in the ability to process complex sentences (for child learners, see Booth, MacWhinney, & Harasaki, 2000; Boyle, Lindell, & Kidd, 2013; Montgomery, Magimairaj, & O'Malley, 2008; Felser, Marinis, & Clahsen, 2003; for adult learners, see Williams, 2006; Havik, Roberts, Van Hout, Schreuder, & Haverkort, 2009; Jackson & Bobb, 2009; Hopp, 2015), revise initial misinterpretations (for children, see Qi, Fisher, & Brown-Schmidt, 2011; Woodard, Pozzan, & Trueswell, 2013; for elderly adults, see Christianson et al., 2006) and even learn a novel artificial grammar system (Kapa & Colombo, 2014). With respect to the latter, some recent studies have suggested that cognitive benefits that result from prolonged training of cognitive control/conflict resolution skills might have a beneficial effect on the processing of complex sentences, such as garden-path sentences and structures that require inhibition of preferred processing strategies in both native speakers and child second language learners (for adults, see Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; for child L2 learners, see Pozzan, Woodard, & Trueswell, 2014).

It is also important to consider how other cognitive limitations might shape grammar acquisition. For instance, successful processing of long-distance dependencies is believed to depend upon accurate storage and later re-access of relevant information within working (and long-term) memory (e.g., Gibson, 1998; Warren & Gibson, 2002). To the extent that stable individual differences (or developmental/experience differences) can be identified in such a process, they too might be expected to shape language acquisition profiles and trajectories, perhaps in a manner similar to that observed here.

Finally, examination (and testing) of other proposed universals would be of interest. Potentially relevant universals include the cross-linguistic tendency for sentential subjects to precede objects (Comrie, 1989; Dryer, 2011; Greenberg, 1963 (Universal 1); Tomlin, 1986; Whaley, 1997), bear the thematic role of agent/source (e.g., Bowerman, 1990; Dowty, 1991; Keenan, 1976; Levin & Rappaport Hovav, 1995), and for agents to be animate and patients inanimate (Comrie, 1989; Dowty, 1991; Hopper & Thompson, 1980; Langacker, 1987).

3.6 Closing comments

The present work aimed to understand how the demands of real-time parsing and interpretation impinge upon grammar acquisition. Owing to the need for greater stimulus control, the present study employed the experimental methods of multi-day acquisition of a miniature artificial language, containing a small sample of nouns and verbs and a morphological system designed to inform aspects of sentential argument structure. Measures of real-time processing commitments (eye movements during listening) accompanied by analyses of off-line comprehension responses allowed us to identify listeners' difficulty revising parsing and interpretive commitments – a pattern often observed in the processing of natural languages. Moreover, a well-documented grammar acquisition profile was also observed here: learners begin with a universally-motivated heuristic of treating each simple NP as a separate participant/thematic role in the sentence's argument structure and then transition into using language-specific knowledge for parsing and interpretation. Crucially, we show here that these two phenomena interact in ways expected if the output of real-time parsing procedures serve as input to further language acquisition (a.k.a., syntactic bootstrapping, Gleitman, 1990). If language-specific knowledge tends to be positioned serially after the parser has had a chance to analyze the input according to universal mapping biases, acquisition of this language-specific knowledge will be delayed, precisely because it will, more often than not, serve to revise rather than guide structural analyses in real-time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Alon Hafri and Yi Lin for starring in our videos and Alon Hafri for recording the experimental sentences. Thanks to Alon Hafri, Sarah Hendry, Jordan Karsch, Jacob Kaufman, Yi Lin, Kristina Woodard and Yu Qian Li for their assistance in data collection, transcription and scoring. We thank Steven Marchette for help with the cluster-based permutation analyses and Lila Gleitman for helpful comments on this paper. This work was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant 1-R01-HD-37507.

5. Appendix

Table 1

Example of the four sentential constructions for experimental verbs in the four language variants.

Language Variant	Locus of Marking	Cue Order	2 NP Sentences: Non-instrument Morphology	2 NP Sentences: Instrument Morphology	1 NP Sentences: Non-instrument Morphology	1 NP Sentences: Instrument Morphology
(I)	V	Early	Zumpi-tu fami **nunu**	Zumpi-ka fami **nunu**	Zumpi-tu **nunu**	Zumpi-ka **nunu**
(II)	V	Late	Nunu fami zumpi-**tu**	Nunu fami zumpi-**ka**	Nunu zumpi-**tu**	Nunu zumpi-**ka**
(III)	NP	Late	Zumpi fami nunu-**tu**	Zumpi fami nunu-**ka**	Zumpi nunu	Zumpi nunu-**ka**
(IV)	NP	Early	Nunu-**tu** fami zumpi	Nunu-**ka** fami zumpi	Nunu zumpi	Nunu-**ka** zumpi

Table 2

Example of the four sentential constructions for filler verbs in the four language variants.

Language Variant	Locus of Marking	Cue Order	2-NP Sentences	1-NP Sentences	0-NP Sentences
(I)	V	Early	Bobu fami nunu	Bobu nunu	Bobu
(II)	V	Late	Nunu fami bobu	Nunu bobu	Bobu
(III)	NP	Late	Bobu fami nunu-tu	Bobu nunu	Bobu
(IV)	NP	Early	Nunu-tu fami bobu	Nunu bobu	Bobu

Table 3

Comprehension accuracy rates for sentences with instrument (-ka) and non-instrument morphology (-tu) as a function of number of NPs, cue order and day of testing

NPs	Morphology	Cue Order	Day 1	Day 2	Day 3	*Subtotal*
1	Instrument	Late	.10	.46	.49	*.40*
		Early	.26	.65	.69	*.59*
		Subtotal	*.18*	*.55*	*.59*	*.70*
	Non-Instrument	Late	.80	.94	.92	*.90*
		Early	.74	.98	.98	*.94*
		Subtotal	*.77*	*.96*	*.95*	*.92*
2	Instrument	Late	.47	.73	.63	*.63*
		Early	.67	.77	.80	*.76*
		Subtotal	*.58*	*.75*	*.71*	*.70*
	Non-Instrument	Late	.63	.67	.83	*.74*
		Early	.45	.74	.93	*.77*

NPs	Morphology	Cue Order	Day 1	Day 2	Day 3	*Subtotal*
		Subtotal	*.54*	*.71*	*.88*	*.75*

Table 4

Production accuracy rates for events acted out with and without an instrument as a function of cue order and day of testing

Morpheme	Cue Order	Day 1	Day 2	Day 3	*Subtotal*
Instrument	Late	.60	.67	.76	*.70*
	Early	.49	.67	.89	*74*
	Subtotal	*.54*	*.67*	*.82*	*.72*
Non-Instrument	Late	.53	0.61	.80	*.69*
	Early	.58	0.77	.95	*.83*
	Subtotal	*.55*	*.69*	*.88*	*.76*

4. References

Abutalebi J. Neural aspects of second language representation and language control. Acta Psychologica. 2008; 128:466–478. [PubMed: 18479667]

Arunachalam S, Waxman SR. Meaning from syntax: evidence from 2-year-olds. Cognition. 2010; 114:442–446. [PubMed: 19945696]

Bates, E.; MacWhinney, B. Functionalist approaches to grammar. In: Wanner, E.; Gleitman, L., editors. Language acquisition: The state of the art. Cambridge University Press; New York: 1982. p. 173-218.

Bates, E.; MacWhinney, B. Functionalism and the competition model. In: MacWhinney, B.; Bates, E., editors. The cross-linguistic study of sentence processing. Cambridge University Press; Cambridge, UK: 1989. p. 3-73.

Beckman M, Pierrehumbert J. Intonational Structure in Japanese and English. Phonology Yearbook. 1986; III:255–309.

Bever, TG. The influence of speech performance on linguistic structure. In: Levelt, W.; Flores d'Arcais, G., editors. Advances in psycholinguistics. North-Holland: 1970. p. 21-50.

Bever, TG.; Langendoen, T. The interaction of perception and grammar in linguistic change. In: Stockwell, R.; MacCaulay, R., editors. Historical linguistics in the perspective of transformational theory. Indiana University Press; Bloomington, IN: 1972. p. 32-95.

Bley-Vroman R. The logical problem of foreign language learning. Linguistic Analysis. 1990; 20:3–49.

Bley-Vroman R. The Evolving Context of the Fundamental Difference Hypothesis. Studies in Second Language Acquisition. 2009; 31:175–198.

Booth JR, Mac Whinney B, Harasaki Y. Developmental differences in visual and auditory processing of complex sentences. Child Development. 2000; 71:981–1003. [PubMed: 11016560]

Boyle W, Lindell AK, Kidd E. Investigating the Role of Verbal Working Memory in Young Children's Sentence Comprehension. Language Learning. 2013; 63:211–242.

Bowerman M. Mapping semantic roles onto syntactic functions: Are children helped by innate linking rules? Linguistics. 1990; 28:1253–1289.

Brown R. Linguistic determinism and the part of speech. Journal of Abnormal and Social Psychology. 1957; 55:1–5.

Brown, R. A first language: The early stages. Harvard University Press; Cambridge, MA: 1973.

Chang F, Dell GS, Bock K. Becoming syntactic. Psychological Review. 2006; 113:234–272. [PubMed: 16637761]

Choi Y, Trueswell JC. Children‘s (in)ability to recover from garden-paths in a verb-final language: Evidence for developing control in sentence processing. Journal of Experimental Child Psychology. 2010; 106:41–61. [PubMed: 20163806]

Chomsky, N. Current Issues in Linguistic Theory. Mouton; The Hague: 1964.

Chomsky, N. Lectures on Government and Binding: The Pisa Lectures. 7th Edition. Foris Publications; Mouton de Gruyter; Holland: Berlin and New York: 1993. 1981

Christianson K, Williams CC, Zacks RT, Ferreira F. Younger and Older Adults' "Good-Enough" Interpretations of Garden-Path Sentences. Discourse Processes. 2006; 42:205–238. [PubMed: 17203135]

Comrie, B. Language Universals and Linguistic Typology. University of Chicago Press; Chicago: 1989.

Costa A, Santesteban M. Memory and Language Lexical access in bilingual speech production : Evidence from language switching in highly proficient bilinguals and L2 learners. Journal of Memory and Language. 2004; 50:491–511.

Culbertson J, Smolensky P, Legendre G. Learning biases predict a word order universal. Cognition. 2012; 122:306–329. [PubMed: 22208785]

Dowty D. Thematic proto-roles and argument selection. Language. 1991; 67:547–619.

Dryer, MS. Order of Subject, Object and Verb. In: Dryer, MS.; Haspelmath, M., editors. The World Atlas of Language Structures Online. Max Planck Institute for Evolutionary Anthropology; Leipzig: 2013. Available online at http://wals.info/chapter/81 [Accessed on 2014-04-17]

Ebbinghaus, H. Memory. A Contribution to Experimental Psychology. Teachers College, Columbia University; Thoemmes Press; New York: Bristol: 1999. 1913

Engelhardt P, Bailey K, Ferreira F. Do speakers and listeners observe the Gricean Maxim of Quantity? Journal of Memory and Language. 2006; 54:554–573.

Engelhardt PE, Ferreira F. Do speakers articulate over-described modifiers differently from modifiers that are required by context?: Implications for models of reference production. Language and Cognitive Processes. 2013; 11:1–11.

Fedzechkina M, Jaeger TF, Newport EL. Language learners restructure their input to facilitate efficient communication. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:17897–17902. [PubMed: 23071337]

Feldman, H.; Goldin-Meadow, S.; Gleitman, LR. Beyond Herodotus: The creation of language by linguistically deprived deaf children. In: Lock, A., editor. Action, symbol, and gesture: The emergence of language. Academic Press; New York: 1978. p. 351-414.

Felser C, Marinis T, Clahsen H. Children's processing of ambiguous sentences: A study of relative clause attachment. Language Acquisition. 2003; 11:127–163.

Ferreira F, Foucart A, Engelhardt PE. Language processing in the visual world: Effects of preview, visual complexity, and prediction. Journal of Memory and Language. 2013; 69:165–182.

Ferreira F, Patson ND. The Good Enough Approach to Language Comprehension. Language and Linguistics Compass. 2007; 1:71–83.

Ferreira F, Ferraro V, Bailey KGD. Good-enough representations in language comprehension. Current Directions in Psychological Science. 2002; 11:11–15.

Fisher C. Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. Cognitive Psychology. 1996; 31:41–81. [PubMed: 8812021]

Fisher C, Gertner Y, Scott RM, Yuan S. Syntactic bootstrapping. Wiley Interdisciplinary Reviews: Cognitive Science. 2010; 1:143–149.

Fisher C, Hall DG, Rakowitz S, Gleitman L. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. Lingua. 1994; 92:333–375.

Fodor, J.; Bever, TG.; Garrett, MF. The Psychology of Language: An Introduction to the Psycholinguistics of Generative Grammar. McGraw-Hill; New York: 1974.

Gass, S.; Selinker, L., editors. Language Transfer in Language Learning. John Benjamins; Amsterdam: 1992.

Gertner Y, Fisher C. Predicted errors in children's early sentence comprehension. Cognition. 2012; 124:85–94. [PubMed: 22525312]

Gibson E. Linguistic complexity: locality of syntactic dependencies. Cognition. 1998; 68:1–76. [PubMed: 9775516]

Gibson E, Piantadosi ST, Brink K, Bergen L, Lim E, Saxe R. A noisy-channel account of crosslinguistic word order variation. Psychological Science. 2013; 24:1079–1088. [PubMed: 23649563]

Gillette J, Gleitman H, Gleitman LR, Lederer A. Human simulations of vocabulary learning. Cognition. 1999; 73:135–176. [PubMed: 10580161]

Gleitman LR. The structural sources of verb meanings. Language Acquisition. 1990; 1:3–55.

Gleitman LR, Cassidy K, Nappa R, Papafragou A, Trueswell JC. Hard words. Language Learning and Development. 2005; 1:23–64.

Göksun T, Küntay AC, Naigles LR. Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. Journal of Child Language. 2008; 35:291–323. [PubMed: 18416861]

Goldin-Meadow, S. The Resilience of Language. Psychology Press; New York: 2003.

Greenberg, J. Some universals of grammar with particular reference to the order of meaningful elements. In: Greenberg, J., editor. Universals of human language. MIT Press; Cambridge, MA: 1963. p. 73-113.

Hall DG, Lee SC, Bélanger J. Young children's use of syntactic cues to learn proper names and count nouns. Developmental Psychology. 2001; 37:298–307. [PubMed: 11370907]

Havik E, Roberts L, Van Hout R, Schreuder R, Haverkort M. Processing subject – object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. Language Learning. 2009; 59:73–112.

Hawkins, JA. Efficiency and Complexity in Grammars. OUP; Oxford: 2004.

Hawkins, JA. Patterns of asymmetry in argument structure across languages: Some principles and puzzles. In: Suihkonen, P.; Comrie, B.; Solovyev, V., editors. Argument Structure and Grammatical Relations: A Crosslinguistic Typology. John Benjamins; Amsterdam: 2012. p. 133-150.

Hawkins, JA. Cross-Linguistic Variation and Efficiency Oxford: OUP. 2014.

Hopp H. Individual differences in the second language processing of object-subject ambiguities. Applied Psycholinguistics. 2015; 36:129–173.

Hopper PJ, Thompson SA. Transitivity in grammar and discourse. Language. 1980; 56:251–299.

Howard MW, Kahana MJ. Contextual variability and serial position effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:923–941.

Huang Y, Zheng X, Meng X, Snedeker J. Assignment of grammatical roles in the online processing of Mandarin passive sentences. Journal of Memory and Language. 2013; 69:589–606.

Hudson Kam CLH, Newport EL. Regularizing Unpredictable Variation : The Roles of Adult and Child Learners in Language Formation and Change. Language Learning. 2005; 1:151–195.

Hudson Kam CLH, Newport EL. Getting it right by getting it wrong: When learners change languages. Cognitive Psychology. 2009; 59:30–66. [PubMed: 19324332]

Hurewitz F, Brown-Schmidt S, Thorpe K, Gleitman LR, Trueswell JC. One Frog, Two Frog, Red Frog, Blue Frog: Factors Affecting Children's Syntactic Choices in Production and Comprehension. 2000; 29:597–626.

Jackson CN, Bobb SC. The processing and comprehension of wh-questions among second language speakers of German. Applied Psycholinguistics. 2009; 30:603–636. [PubMed: 20161006]

Juffs A. Representation, processing and working memory in a second language. Transactions of the Philological Society. 2004; 102:199–226.

Juffs A, Harrington M. Garden path sentences and error data in second language sentence processing research. Language Learning. 1996; 46:286–324.

Kapa LL, Colombo J. Executive function predicts artificial language learning. Journal of Memory and Language. 2014; 7:237–252.

Katz B, Baker G, MacNamara J. What's in a name? On the child's acquisition of proper and common nouns. Child Development. 1974; 45:269–273.

Kaushanskaya M, Marian V. Bilingualism reduces native-language interference during novel-word learning. Journal of Experimental Psychology. Learning, Memory, and Cognition. 2009; 35:829–35.

Keenan, EL. Towards a universal definition of "subject". In: Li, CN., editor. Subject and topic. Academic Press; New York: 1976.

Kidd E, Stewart AJ, Serratrice L. Children do not overcome lexical biases where adults do: the role of the referential scene in garden-path recovery. Journal of Child Language. 2011; 38:222–34. [PubMed: 20196901]

Klatt DH. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America. 1976; 59:1208–1221. [PubMed: 956516]

Landau, B.; Gleitman, LR. Language and experience: Evidence from the blind child. Harvard University Press; Cambridge, MA: 1985.

Langacker, R. Foundations of Cognitive Grammar, 1. Stanford University Press; Stanford: 1987.

Lee J, Naigles L. Mandarin learners use syntactic bootstrapping in verb acquisition. Cognition. 2008; 106:1028–1037. [PubMed: 17537422]

Levin, B.; Rappaport Hovav, M. Unaccusativity at the syntax–lexical semantics interface. MIT Press; Cambridge, MA: 1995.

Lew-Williams C, Fernald A. Real-time processing of gender-marked articles by native and non-native Spanish speakers. Journal of Memory and Language. 2010; 63:447–464. [PubMed: 21076648]

Lidz J, Gleitman H, Gleitman L. Understanding how input matters: Verb learning and the footprint of universal grammar. Cognition. 2003; 87:151–178. [PubMed: 12684198]

Lidz, J.; White, AS.; Baier, R. The role of incremental parsing in syntactically conditioned word learning. (submitted)

Lombardi L, Potter MC. The regeneration of syntax in short term memory. Journal of Memory and Language. 1992; 31:713–733.

Liittschwager, J.; Markman, EM. Young children's understanding of proper versus common nouns. Poster presented at the biennial meeting of the Society for Research in Child Development; New Orleans, LA. 1993.

MacWhinney B, Bates E, Kliegl R. Cue validity and sentence interpretation in English, German, and Italian. Journal of Verbal Learning and Verbal Behavior. 1984; 23:127–150.

Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. Journal of neuroscience methods. 2007; 164:177–190. [PubMed: 17517438]

Marshall PH, Werder PR. Effects of Elimination of Rehearsal on Primacy and Recency. Journal of Verbal Learning and Verbal Behavior. 1972; 11:649–653.

Medina TN, Snedeker J, Trueswell JC, Gleitman LR. How words can and cannot be learned by observation. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:9014–9019. [PubMed: 21576483]

Montgomery JW, Magimairaj BM, O'Malley MH. Role of working memory in typically developing children's complex sentence comprehension. Journal of Psycholinguistic Research. 2008; 37:331–354. [PubMed: 18521752]

Murasugi, K.; Hasimoto, T.; Kato, S. On the Acquisition of Causatives in Japanese. Paper presented at the 28th Annual Boston University Conference on Language Development; Boston, MA. 2004.

Naigles L. Children use syntax to learn verb meanings. Journal of child language. 1990; 17:357–374. [PubMed: 2380274]

Naigles LR, Fowler A, Helm A. Developmental shifts in the construction of verb meanings. Cognitive Development. 1992; 7:403–28.

Naigles, L.; Gleitman, G.; Gleitman, L. Children Acquire Word Meaning Components from Syntactic Evidence. In: Dromi, E., editor. Language and Cognition: A Developmental Perspective. Vol. 5. Ablex Publishing; New Jersey: 1993. p. 104-140.

Nappa R, Wessell A, McEldoon KL, Gleitman LR, Trueswell JC. Use of speaker's gaze and syntax in verb learning. Language Learning and Development. 2009; 5:203–234. [PubMed: 24465183]

Nichols, J. Linguistic Diversity in Space and Time. University of Chicago Press; Chicago: 1992.

Noble, C.; Theakston, AL.; Lieven, EVM. Comprehension of intransitive argument structure: The first noun as causal agent bias. Paper presented at the 35th Annual Boston University Conference on Language Development; Boston, MA. 2010.

Novick JM, Hussey E, Teubner-Rhodes S, Harbison JI, Bunting MF. Clearing the garden-path: Improving sentence processing through cognitive control training. Language Cognition and Neuroscience. 2014; 29:186–217.

Novick JM, Thompson-Schill S, Trueswell JC. Putting lexical constraints in context into the visual-world paradigm. Cognition. 2008; 107:850–903. [PubMed: 18279848]

Omaki A, Davidson White I, Goro T, Lidz J, Phillips C. No fear of commitment: Children's incremental interpretation in English and Japanese *wh*-questions. Language Learning and Development. 2014; 10:206–233.

Omaki A, Lidz J. Linking parser development to acquisition of syntactic knowledge. Language Acquisition. 2014 DOI: 10.1080/10489223.2014.943903.

Pienemann, M.; Di Biase, B.; Kawaguchi, S.; Håkansson, G. Processability, typological distance and L1 transfer. In: Pienemann, M., editor. Cross-Linguistic Aspects of Processability Theory. John Benjamins; Amsterdam: 2005. p. 85-116.

Pierrehumbert, J. Ph. D. Thesis. MIT; 1980. The Phonology and Phonetics of English Intonation.

Potter MC, Lombardi L. Regeneration in the short-term recall of sentences. Journal of Memory and Language. 1990; 29:633–654.

Pozzan, L.; Trueswell, JC. Online processing of English garden-path sentences by L2 learners: a visual world study. Poster presented at the 26th Annual CUNY Conference on Human Sentence Processing; Columbia, SC. 2013.

Pozzan, L.; Gleitman, LR.; Trueswell, JC. Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. Language Learning and Development. (in press)

Pozzan, L.; Woodard, K.; Trueswell, JC. Executive Functions Predictors of Learners' Language Processing Abilities: A Training Study. Paper presented at the 39th Annual Boston University Conference for Language Development; Boston, MA. 2014.

Pritchett, BL. Grammatical Competence and Parsing Performance. The University of Chicago Press; Chicago: 1992.

Qi, Z.; Fisher, C.; Brown-Schmidt, S. On Recovering from 'Kindergarten Path' Errors: Referential Context and Executive Functioning Influence Children's Online Ambiguity Resolution. Paper presented at the 17th Annual AMLaP Conference; Paris, France. 2011.

Roberts L, Felser C. Plausibility and recovery from garden paths in second language sentence processing. Applied Psycholinguistics. 2011; 32:299–331.

Scott RM, Fisher C. 2-year-olds use distributional cues to interpret transitivity-alternating verbs. Language and Cognitive Processes. 2009; 24:777–803. [PubMed: 20046985]

Slobin D, Bever T. Children use canonical sentence schemas: A cross-linguistic study of word order and inflections. Cognition. 1982; 12:229–265. [PubMed: 6891309]

Smith L, Jones S, Landau B. Count nouns, adjectives, and perceptual properties in novel word interpretations. Developmental Psychology. 1992; 28:273–286.

Snedeker, J.; Gleitman, LR. Why it is hard to label our concepts. In: Hall, G.; Waxman, SR., editors. Weaving a lexicon. MIT Press; Cambridge, MA: 2004. p. 257-294.

Sundara M, Demuth K, Kuhl PK. Sentence-Position Effects on Children's Perception and Production of English Third Person Singular. Journal of Speech, Language, and Hearing Research. 2011; 54:55–72.

Tomasello M, Kruger A. Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. Journal of Child Language. 1992; 19:311–333. [PubMed: 1527205]

Tomlin, RS. Basic word order: Functional principles. CroonHelm; London: 1986.

Trueswell JC, Medina TN, Hafri A, Gleitman LR. Propose but verify: Fast mapping meets cross-situational word learning. Cognitive Psychology. 2013; 66:126–156. [PubMed: 23142693]

Trueswell JC, Sekerina I, Hill NM, Logrip ML. The kindergarten-path effect: studying on-line sentence processing in young children. Cognition. 1999; 73:89–134. [PubMed: 10580160]

Trueswell, JC.; Kaufman, D.; Hafri, A.; Lidz, J. Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. In: Biller, AK.; Chung, EY.; Kimball, AE., editors. Proceedings of the 36th Annual Boston University Conference on Language Development; Somerville, MA: Cascadilla Press; 2012. p. 620-632.

Turk A, Shattuck-Hufnagel S. Phrase-final lengthening in American English. Journal of Phonetics. 2007; 35:445–472.

Van Patten, B. Input Processing and Grammar Instruction: Theory and Research. Ablex; Norwood, NJ: 1996.

Warren T, Gibson E. The influence of referential processing on sentence complexity. Cognition. 2002; 85:79–112. [PubMed: 12086714]

Weighall AR. The kindergarten-path effect revisited : children's use of context in processing structural ambiguities. Journal of Experimental Child Psychology. 2008; 99:75–95. [PubMed: 18070628]

Whaley, LJ. Introduction to Linguistic Typology: The Unity and Diversity of Language. Sage Publications; London: 1997.

White, AS.; Baier, R.; Lidz, J. Prediction and sub-categorization frame frequency in the developing parser. Paper presented at the 24th Annual CUNY Conference on Human Sentence Processing; Stanford, CA. 2011.

Wightman CW, Shattuck-Hufnagel S, Osterdorf M, Price P. Segmental durations in the vicinity of prosodic phrase boundaries. The Journal of the Acoustical Society of America. 1992; 91:1707–1717. [PubMed: 1564206]

Williams JN, Möbius P, Kim C. Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. Applied Psycholinguistics. 2001; 22:1–28.

Williams JN. Incremental interpretation in second language sentence processing. Bilingualism: Language and Cognition. 2006; 9:71–81.

Wonnacott E, Newport EL, Tanenhaus MK. Acquiring and processing verb argument structure: distributional learning in a miniature language. Cognitive Psychology. 2008; 56:165–209. [PubMed: 17662707]

Woodard, K.; Pozzan, L.; Trueswell, JC. Mapping the kindergarten-path: cognitive predictors of child sentence processing. Poster Presented at the 27th CUNY Annual Conference on Sentence Processing; Columbus, Ohio. 2014.

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

Yang C. Universal Grammar, statistics, or both. Trends in Cognitive Sciences. 2004; 8:451–456. [PubMed: 15450509]

Yang C. Computational models of syntactic acquisition. Wiley Interdisciplinary Reviews: Cognitive Science. 2012; 3:205–213.

Yuan S, Fisher C. "Really? She blicked the baby?": two-year-olds learn combinatorial facts about verbs by listening. Psychological Science. 2009; 20:619–626. [PubMed: 19476591]

Highlights

- Investigates how real-time sentence processing limitations influence grammar acquisition.
- Adults learned miniature languages with cues to structure that either guided or revised parsing commitments.
- Cues to structure that guided commitments were easier to learn than revising cues.
- Results held for comprehension and production.
- Suggests that real-time processing constraints affect acquisition trajectories.

Figure 1.
Still frame example for *Noun Vocabulary Training* videos

Figure 2.
Example of experimental set-up for *Sentence Comprehension*

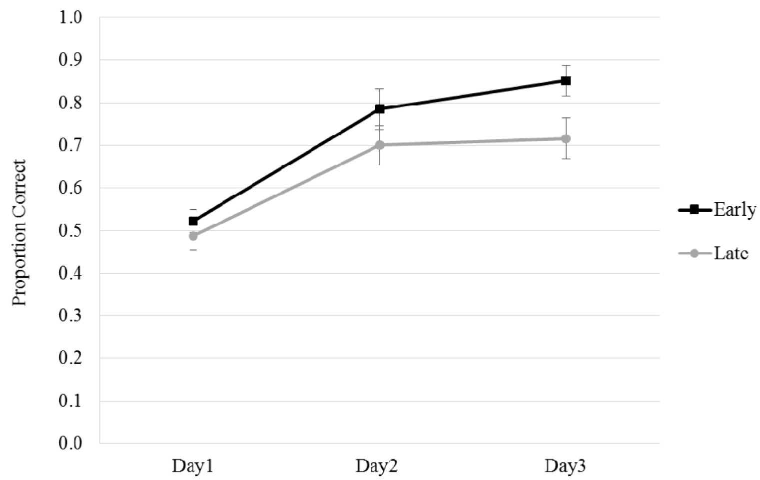

Figure 3.
Average comprehension accuracy rates for the morphological markers by Day and Cue Order. Error bars represent standard error.

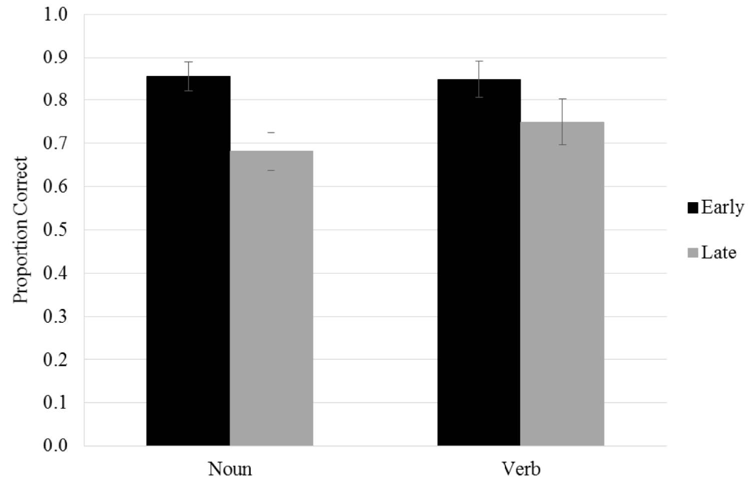

Figure 4.
Day 3: Average comprehension accuracy rates on by Locus of Marking and Cue Order. Error bars represent standard error.

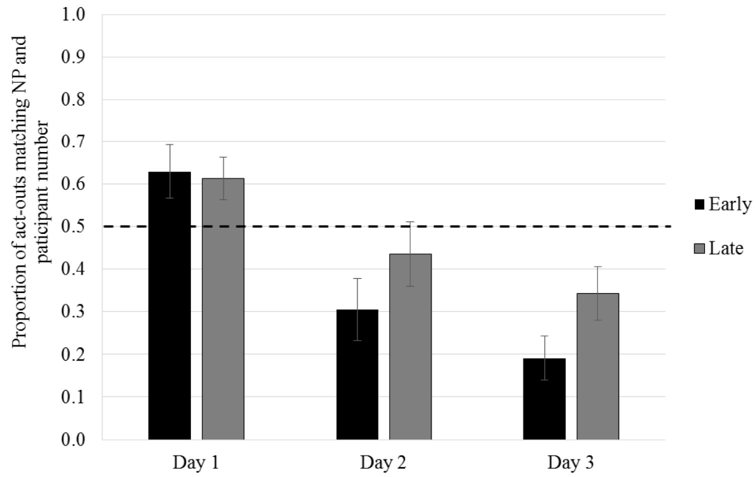

Figure 5.

Average proportions of act-outs in which the number of participants in the event matched the number of NPs by Day and Cue Order, for sentences where the number of NPs mismatched the number of event participants. Dotted line represents chance (.5). Error bars represent standard error.

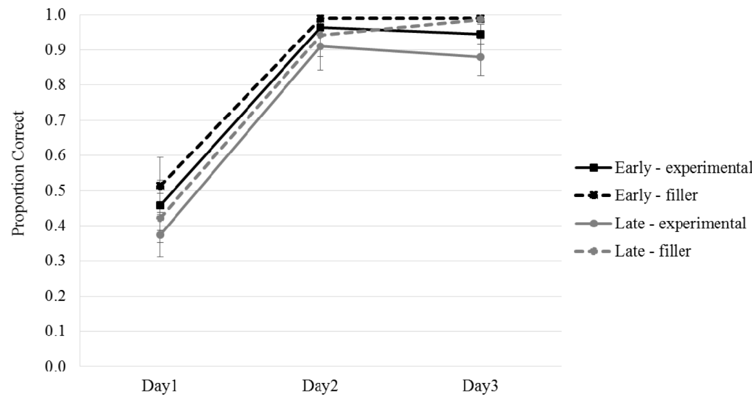

Figure 6.

Average comprehension accuracy rates for the action vocabulary by Day, Cue Order and Item-Type. Error bars represent standard error.

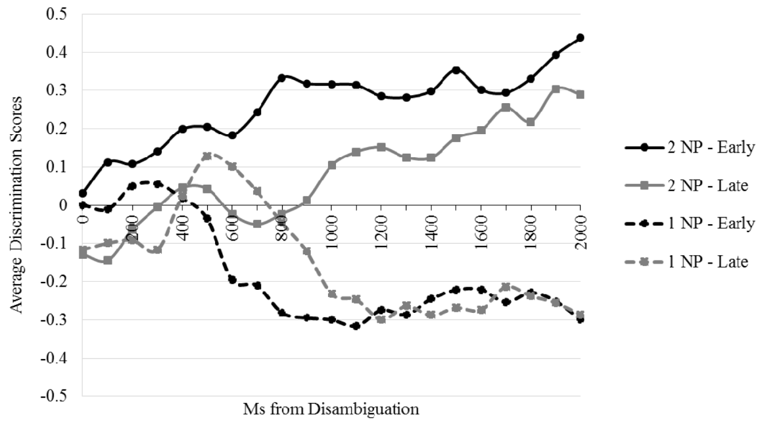

Figure 7.

Average Discrimination Scores as a function of Time (in ms.), Number of NPs in the sentence and Cue Order. Discrimination Scores are plotted from the onset of the point of disambiguation until 2000 ms. after it.

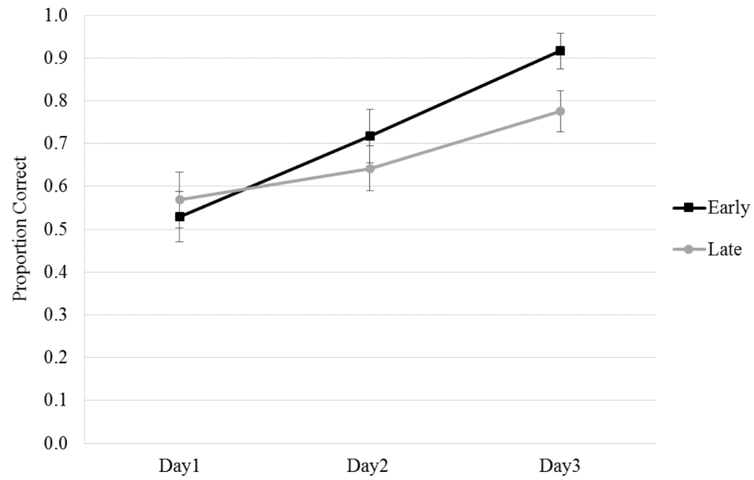

Figure 8.

Average production accuracy rates for the morphological markers by Day and Cue Order. Error bars represent standard error.

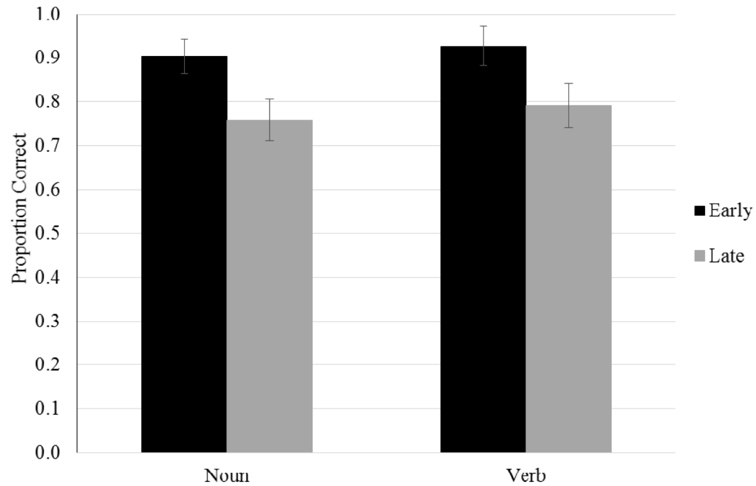

Figure 9.
Day 3: Average production accuracy rates for the morphological markers by Locus of Marking and Cue Order. Error bars represent standard error.

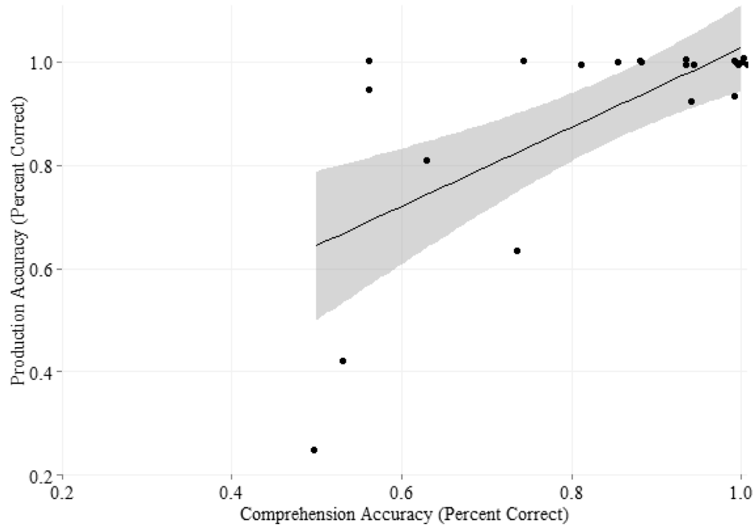

Figure 10.
Day 3: Correlation between production and comprehension accuracy rated for the morphological markers for early-cue languages.

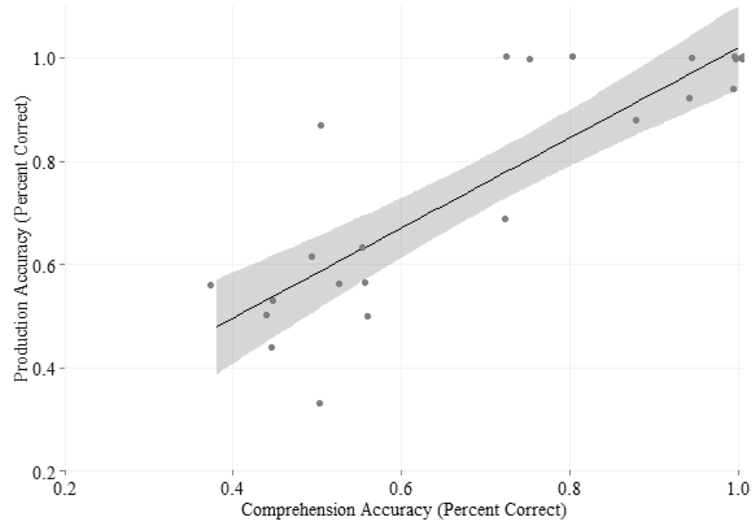

Figure 11.

Day 3: Correlation between production and comprehension accuracy rates for the morphological markers for late-cue languages.

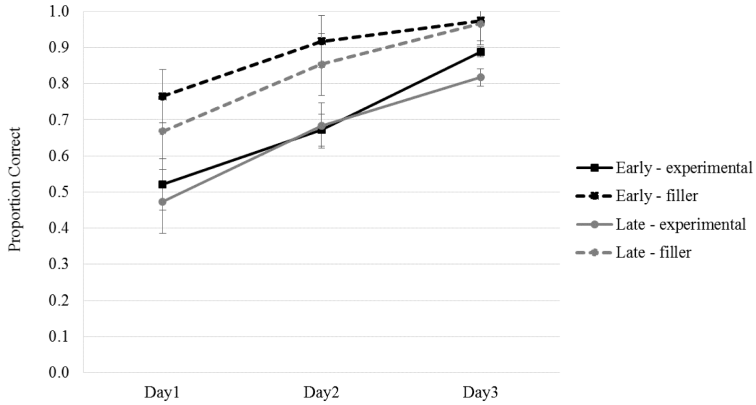

Figure 12.
Average production accracy rates for the action vocabulary by Day, Cue Order, and Item-type. Error bars represent standard error.

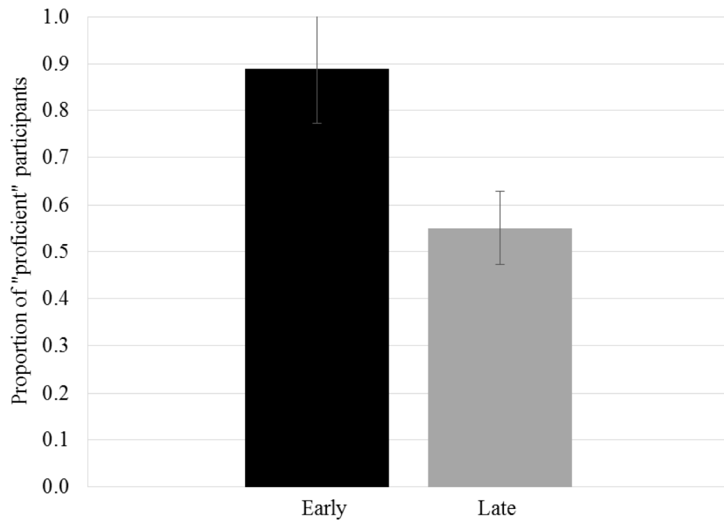

Figure 13.

Proportion of participants who were able to correctly indicate the function of morphological markers in the offline survey. Error bars represent standard error.