languages

MDPI

Article

Multimodal Cue Competition in Adults' Novel Verb Generalization

Bhuvana Narasimhan

Department of Linguistics, University of Colorado Boulder, Hellems 290, 295 UCB, Boulder, CO 80309, USA; narasimb@colorado.edu; Tel.: +1-303-492-8456

Academic Editor: Usha Lakshmanan
Received: 8 December 2016; Accepted: 16 March 2017; Published: 21 March 2017

Abstract: In addition to identifying the referents of novel words, language learners also have to learn to generalize newly acquired words to the appropriate range of referents. Here we ask: what is the relative importance of visual, auditory, and linguistic information in influencing how adult learners generalize newly acquired verbs to novel contexts? In our study, participants learned two novel verbs associated with distinct auditory, visual, and linguistic cues. Then they labeled unfamiliar events in which each cue was either presented in isolation or placed in conflict with other cues. Participants' production of the verb associated with each cue when in conflict with other cues was assessed relative to their baseline tendency to produce the verb associated with each cue presented in isolation. Findings show that visual cues dominate over linguistic and auditory cues in influencing participants' verb extension patterns. In contrast, participants are rarely influenced by auditory or linguistic cues when they are placed in conflict with the other cue types. Our findings suggest that any account of word learning needs to factor in the dynamics of how multimodal cues interact to drive attention during word extension.

Keywords: multimodal; cue competition; word learning; verb; generalization; language acquisition

1. Introduction

One of the central questions in psycholinguistics has to do with how word meaning is acquired and represented by adult and child language learners. Much of the word learning literature focuses on the challenge of referential indeterminacy faced by learners in their first encounters with novel words. Learners have to solve the problem of identifying the referents of unfamiliar words used in rich environments that contain many potential referents. For instance, a novel word could refer to an object (e.g., a rabbit), an object part (the rabbit's ear), a property (fluffy), what it is doing (running), and so on [1]. In order to home in on the appropriate referent, learners may attend to the speaker's focus of attention, the perceptual or functional characteristics of the referent, as well as the linguistic context in which the word is embedded, among others [2–6].

As documented in the language acquisition literature, both adults and children exhibit 'fast-mapping' skills': the ability to rapidly home in on the broad meaning of a novel word based on a few incidental exposures [7,8]. However, the challenge of acquiring word meaning extends well beyond the problem of identifying referents during initial encounters with a word. Learners also have to generalize the newly acquired word to the appropriate range of referents in the target language, a task that requires a grasp of the underlying dimensions and features of word meaning that influence categorization patterns [9]. For instance, acquiring the central meaning of a motion verb such as *enter* entails learning that the verb can be extended to any situation that involves the spontaneous change of location of a 'Figure' object resulting in its inclusion within the boundaries of a 'Ground' object. Learning to use *enter* appropriately also involves disregarding other co-occurring features of motion

events such as the manner of motion (e.g., whether the 'Figure' object rolls, slides, or spins while changing location). Thus, in order to be able to use a verb flexibly in different situations, child and adult language learners "need to understand which specific aspect of the action is invariant for the verb and which aspects can vary across different situations in which the verb is used" [10].

The refinement of word meaning beyond the initial stage of 'fast-mapping' is a protracted process during which learners gradually learn to extend words to the appropriate range of referents [11–15]. The patterns of word extension are influenced by a number of factors including the word's frequency of occurrence, the breadth and make-up of the category associated with the word in the target language, and the number of competing words with overlapping category boundaries, among others [16]. Since language is learned within complex multisensory environments, different types of information—perceptual as well as linguistic—are recruited when learners attempt to home in on what a word means and how it can be extended to different contexts of use [6,17]. For instance, adult learners are sensitive to fine-grained differences in frequencies of co-occurrence between spoken words and visual referents in the environment [18]. In addition, learners can exploit linguistic cues in inferring the meanings of novel verbs, including the words that co-occur with the novel word, viz. distributional cues [19,20]. Distributional cues may also aid the learner when deciding whether to extend the use of a word to novel verb arguments [20] or to new referents. To illustrate, the noun *apple* often co-occurs with the verb *eat* whereas the noun *ball* co-occurs with verbs such as *bounce* or *throw*. Thus, hearing the verb *eat* used in the presence of a novel object that is round and red (as in: *Can I eat it*?) may encourage learners to label the novel object as *apple* even though it physically resembles a ball.

However, in spite of remarkable advances in research on the mechanisms of word learning, we know relatively little about how we integrate information from multiple sensory modalities—such as the visual, tactile, auditory, taste, and olfactory modalities—during word learning and generalization. Here we ask: how do different types of perceptual cues (e.g., visual, auditory) interact with linguistic cues (e.g., co-occurring words in the speaker's utterance) in influencing learners' choice of a word to label a novel object or action?[1] A well-established methodology to investigate how learners attend to different cues in language learning involves the use of tasks that present learners with cues that are in competition with each other [21,22]. Prior research shows that infants attend to social cues (the gaze direction of an adult participant) as well as salience cues (the relative movement of images on a screen) [23]. Adult learners recruit both cross-situational regularities in multimodal cues as well as prior knowledge about linguistic context in acquiring word meaning; further, they are influenced equally by both sources of information when the two types of evidence are in conflict [24].

However, whereas prior research focuses on how learners use multiple cues to identify the appropriate referent of a novel word, the present study investigates how word learners recruit information from different modalities when extending the use of an already learned word to an unfamiliar referent. For instance, if a learner is confronted with an action and has to label it based on either visual information (the action *looks* like breaking) or conflicting auditory information (it *sounds* like tearing), which verb do they select: *tear* or *break*? Alternatively, if learners encounter auditory information (a tearing sound) together with conflicting linguistic information (e.g., a question such as *what did he do to the stick?*), do they choose the verb *tear* based on the auditory cue or do they prefer the verb *break* based on the greater frequency with which the verb *break* co-occurs with *stick*?

Any account of word learning and generalization needs to provide an account of how multimodal cues conspire or compete to drive attention across contexts. Thus, our study focuses on the following

---

[1] Both non-linguistic sounds (e.g., non-speech sounds such as rattling sounds) and linguistic strings in spoken language (e.g., words that co-occur with unfamiliar verbs that are in the process of being acquired) consist of information in the auditory modality. For the purposes of the current study, we distinguish between the two types of information using the terms 'auditory' cues for non-speech sounds and 'linguistic' cues for lexical distributional cues without implying that speech constitutes a special perceptual modality distinct from general audition.

question: what is the relative importance of visual, auditory, and linguistic information when learners generalize the use of a newly acquired verb to a novel situation? Psychophysical studies investigating the issue of sensory dominance in multisensory interactions suggest that the brain does not rely equally on different sources of information: the visual sensory modality can overshadow the auditory sense in adults [25]. In these studies, participants make one response to auditory targets, another response to visual targets, and both responses to concurrently presented auditory and visual targets. Although participants respond accurately to unimodal targets, they often fail to respond to auditory targets on the bimodal target trials [26]. The visual dominance effect also prevails in studies that employ more complex stimuli such as line drawings, photos, and familiar sounds [27,28]. In contrast, auditory dominance has been shown to prevail in 6-year-old children tested in a similar paradigm involving speeded responses to audiovisual stimuli [29]. If the patterns of sensory dominance found in these studies also extend to word learning contexts, we would predict that adult learners would privilege visual information over auditory and linguistic information in contexts where they are asked to generalize words to novel referents.

However, psychophysical studies on multisensory processing often rely on relatively simple stimuli (e.g., tones and flashing lights), and even those studies that do employ more complex stimuli do not investigate the larger linguistic context in which words are typically encountered, viz. co-occurring words or phrases in the same utterance. Moreover, the tasks employed involve the detection or discrimination of the source modality [30] or the identification of target concepts embedded in rapidly presented pictures and/or sounds [28]. However, most typical language learning situations involve meaningful words embedded in conversational discourse between interlocutors. Further, auditory, visual, and linguistic information are not matched in salience and do not always occur simultaneously. Rather, the onset of a sound that co-occurs with an action may occur at different stages during the unfolding of the action (e.g., the crashing sound accompanying the breaking of a glass happens at the end of the causal event that leads an agent to bring about material disintegration in the glass). Nor does the utterance containing the verb that describes an action typically overlap in time with the action [31,32].

In order to emulate typical communicative situations involving language use, the present study engages participants in an interactive task during a training phase in which puppets teach participants two novel verbs that are associated with distinctive actions, sounds, and co-occurring lexical items. The actions involve an agent pulling/rolling and releasing a toy part such that it returns to its original position with a characteristic sound. The co-occurring lexical items consist of adjunct phrases (the adverb *carefully* or the prepositional phrase *with his/her hand*). Adverbs are known to facilitate verb learning in certain contexts [33] and the specific adjunct phrases chosen for this study could be applied to describe the actions associated with either of the two novel verbs (the pulling or rolling actions were both performed 'carefully' and 'with the hand'). Thus, the semantic content of the adjuncts could not be used to distinguish the two verbs. The syntactic frame in which the two verbs were used is also identical, as are the core arguments themselves. Thus, the only respect in which the linguistic context of the two verbs differs is in the lexical items used in the adjunct phrases. This manipulation allowed us to examine the relative contributions of two kinds of auditory stimuli that differed minimally in terms of whether they were speech sounds (words, phrases) or non-speech sounds (twangy/rattly sounds).

After participants in our study learn the two novel verbs, they participate in a forced-choice task in two different experimental conditions during the test phase. In the 'unimodal baseline' condition, participants are presented with each type of cue in isolation and are asked to produce the associated verb. This establishes a baseline for how strongly each type of cue is associated with the corresponding verb after the learning phase. In the 'multimodal cue competition' condition, participants encounter novel situations in which they are presented with unfamiliar scenarios in which visual, linguistic, and auditory cues are placed in conflict. Difference scores are then calculated by subtracting participants' accuracy scores in producing the verb associated with each type of cue in the 'multimodal cue competition' trials from participants' accuracy in producing the verb associated

with the same type of cue in the corresponding 'unimodal baseline' trials.[2] Comparing difference scores for auditory, linguistic, and visual cues allows us to calculate the relative effectiveness of each type of cue in eliciting the associated verb in the presence of competing cues, relative to their baseline accuracy in associating those verbs with the appropriate cue. We can thus establish a hierarchy representing the relative weighting of auditory, visual, and linguistic cues in influencing learners' novel verb generalization.

Although it is important to examine how learners recruit multiple cues in learning words from any part of speech, we focus here on verb learning in adults. Verbs have relational meaning and an argument structure that is projected at the sentence level. Thus, acquiring the relational meanings of verbs plays a central role in acquiring the grammar of a language [10]. Second, investigating novel verb generalization strategies in adults will help in establishing multimodal cue weighting strategies in word learners with relatively stable cognitive systems. This allows us to establish a basis for subsequent comparative studies of similar strategies in child learners undergoing developmental change as well as adult second language learners with different first language backgrounds.

2. Materials and Methods

2.1. Participants

Sixteen adult speakers of English (11 females, five males) participated in the study, which was approved by the Institutional Review Board at the University of Colorado Boulder (Boulder, CO, USA) (IRB Protocol #15-0165, approved 7 April 2015, renewal approved 28 March 2016). A number of speakers reported fluency in other languages including Arabic, French, Spanish, Portuguese, and Malay. Informed consent was obtained from all participants, who received cash or course credit for participation.

2.2. Materials

The stimuli included video clips and unfamiliar 'Martian toys' that were constructed to enable participants to learn novel verbs associated with novel actions. Four different versions of the toy were constructed, consisting of bucket-shaped objects with a spring attached to one side and a cylindrical object affixed to the base. Hand puppets ('Marsie' and 'Lionie') were used to create training and testing videos. The video clips depicted the two puppets enacting novel actions performed on each of the four toys. The novel actions were labeled by two novel verbs: *wug* and *meek*. Each verb was associated with a cluster of three distinct cues. For instance, the verb *wug* was associated with a puppet enacting a pulling–releasing action on the toy's spring (a visual cue), a twangy sound accompanying the return of the spring to its original position (an auditory cue), and a prepositional adjunct phrase (*with his/her hand*) (a linguistic cue). The second verb, *meek*, was associated with a puppet enacting a rolling–releasing action on the cylindrical part affixed to the toy's base (a visual cue), a rattly sound accompanying the return of the cylinder to its original position (an auditory cue), and an adjunct adverbial (*carefully*) (a linguistic cue). The training scenarios do not present the three types of cues concurrently. Rather, the linguistic cue (the adjunct phrase describing the manner in which the action is performed) is embedded in a question about the puppet's action that is posed either before or after the caused motion event is enacted by the puppet. The auditory cue consists of a sound that overlaps with the result phase of the caused motion event initiated by the puppet: the releasing of the spring or the cylinder. All cues occurred with equal frequency and each cue was associated with only one of the verbs.

---

[2] In the 'unimodal baseline' condition, the term 'accuracy' refers to participants' production of the appropriate verb when presented with the linguistic cue, the visual cue, or the auditory cue that was associated with that verb in the training phase in isolation. In the 'multimodal cue competition' condition, the term 'accuracy' refers to the production of the appropriate verb when presented with the associated linguistic cue, visual cue, or auditory cue, when that cue is presented as the 'odd man out' cue in the presence of two conflicting cues (that are both associated with a different verb).

Both puppets took turns in enacting the novel action associated with each verb on all four toys which varied in color and other, small accoutrements (small clips, blocks affixed to the toys). The characteristic sounds associated with each novel verb contained minor variations. There was also variation in the forms of the verbs *wug* and *meek.* Both verbs were used in the progressive (e.g., *I am meeking the toy with my hand*) or in their bare form (e.g., *Do you want to meek the toy*?). Their direct object arguments occurred as either full noun phrases or pronouns (e.g., *I am wugging the toy/it carefully*). The arguments were also omitted if licensed by prior discourse context (e.g., in questions: *Where is she meeking ( ) with her hand?* or in statements: *She is wugging ( ) carefully*). The subject arguments referring to the puppets were either pronouns (e.g., *I, he, she*) or obligatorily null in complement clauses (e.g., *Do you want to meek the toy?*).

Different sets of audio files were superimposed on the same set of videos using iMovie (Apple Inc., Cupertino, CA, USA) and embedded in PowerPoint files (Microsoft Corporation, Redmond, WA, USA) to create a total of sixteen different versions of the stimuli. Four versions of the training phase were created. In two versions, the verb *wug* was associated with a pulling–releasing action, a twangy sound, and an adjunct prepositional phrase (*with his/her hand*), while the verb *meek* was associated with a rolling–releasing action, a rattly sound, and an adjunct adverbial (*carefully*). In the other two versions, the mapping between verbs and cue constellations was reversed. For each of these verb–action–sound–adjunct phrase constellations, the order in which the verbs were introduced during training was counterbalanced. Each of the four training phase videos was then combined with a common set of stimuli used in comprehension training, comprehension assessment, and production assessment. Each of the resulting four training sequences was then combined with four versions of a common set of test trials that varied in how the test stimuli were presented. The test stimuli were presented in two blocks that counterbalanced the order in which the linguistic and perceptual cues were presented during the test phase (i.e., they varied in whether the video clip containing the question with the adjunct phrase was played before or after the video clip containing action and/or sound was played). The order of the stimuli within each block was also counterbalanced.

The test stimuli consisted of twelve pairs of video clips. Six out of the twelve pairs of clips were constructed to assess participants' ability to associate each type of cue with the appropriate verb when the cues would be presented in isolation—the 'unimodal baseline' condition. In the 'visual cue only' clips, the participant would be shown only the action (e.g., the puppet pulling and releasing the toy's spring), but hear no accompanying sound (e.g., no twangy sound) and no adjunct phrase in the question asked by the puppet (*What is he doing*?). In the 'auditory cue only' clips, they would hear only the sound (e.g., a twangy sound), but see no action (a black screen would cover the video of the action) and no adjunct phrase would be used in the question asked by the puppet (*What is he doing*?). In the 'linguistic cue only' clips, they would hear the adjunct phrase in the question asked by the puppet (e.g., *What is he doing* ***with his hand***?), but hear no sound (e.g., no twangy sound) and see no action (a black screen would cover the video of the action).

The remaining six pairs of clips were designed to assess participants' preference for each type of 'odd man out' cue (auditory, visual, or linguistic) when pitted against two cues of different types in the 'multimodal cue competition' condition. Thus, in the 'visual odd man out cue' clips, the participant would be shown the *action* associated with one verb (e.g., the puppet rolling and releasing the cylindrical part of the toy) and would hear the *sound* (e.g., a twangy sound) and an *adjunct phrase* in the question asked by the puppet (e.g., *What is he doing* ***with his/her hand***?) that was associated with the other verb. In the 'auditory odd man out cue' clips, the participant would hear the *sound* associated with one verb (e.g., a rattly sound), and would see the *action* (e.g., the puppet rolling and releasing the toy's spring) and hear an *adjunct phrase* in the question asked by the puppet (e.g., *What is he doing* ***with his/her hand***?) associated with the other verb. In the 'linguistic odd man out cue' clips, the participant would hear an *adjunct phrase* in the question asked by the puppet that was associated with one verb (e.g., *What is he doing* ***carefully***?), and would hear the *sound* (e.g., a twangy sound) and see the *action* (e.g., the puppet rolling and releasing the toy's spring) associated with the other verb.

2.3. Procedure

2.3.1. Training Phase

Based on methodology developed in prior research [34], our study involved two phases (see Table 1). In the learning phase, participants learned two nonce verbs: *wug* and *meek*. Participants viewed video clips of two puppets producing each verb while demonstrating the associated novel action on novel toy objects on a computer screen. Participants were asked to reproduce the novel actions associated with the nonce verbs by performing the actions themselves on the novel toys. Then they watched eight video clips of the two puppets naming and demonstrating the actions associated with each verb. Finally, participants' knowledge of each of the two verbs was assessed in a comprehension and a production task.

Table 1. Overview of the different phases of the experiment.

Experimental Phase	Procedure
Training	Two puppets interacted with the participant on the computer. The puppets demonstrated two novel actions using novel toys and labeled them with nonce verbs (*wug*, *meek*). The participant also performed each action on the actual toy and labeled the action. Then the participant watched eight pairs of video clips in which a puppet described each of the two actions and the other puppet enacted it.
Comprehension assessment	The participant was shown four pairs of video clips. In each pair, a puppet asked the participant to point to the action corresponding to *wug* or *meek*. Then two video clips played side-by-side and feedback was provided after the participant provided a pointing response.
Production assessment	The participant was shown eight pairs of video clips. In one video clip in each pair, a puppet asked the participant to label the action performed by the other puppet (e.g., *What is she doing carefully/with her hand, wugging or meeking?*) The second video clip showed a puppet demonstrating the action and the participant produced a label (*wugging* or *meeking)*. Feedback was provided about the accuracy of the participant's response.
Test	The participant was shown 16 pairs of video clips, 12 of which consisted of the novel test stimuli. In each pair, a puppet asked the participant to label the action performed by the other puppet (e.g., *What is she doing (carefully/with her hand), wugging or meeking?*). A second video clip showed a puppet demonstrating the action (unless the visual information was hidden with a black screen in the 'unimodal baseline' condition) and the participant produced a label (*wugging* or *meeking)*. Six of the 12 clips were 'unimodal baseline' test stimuli providing only visual information about the action, only the sound, or only the linguistic information. The remaining six 'multimodal cue competition' test stimuli consisted of one 'odd man out' cue associated with one nonce verb that was placed in conflict with two other types of cues associated with the other verb. Four 'control' clips showed familiar actions from the training session to measure post-training accuracy. These clips contained linguistic, auditory, and visual cues that were congruent (i.e., all the cues were associated with the same verb). No feedback was provided about the accuracy of the participant's response in the test phase.

2.3.2. Test Phase

During the 'multimodal cue competition' condition in the test phase, participants were presented with hybrid scenarios involving conflicting cues and asked to pick one of the two nonce verbs to describe the scenario. For example, the participant might be asked to label a pulling–releasing action associated with the verb *wug* but hear a twangy sound together with the adjunct phrase (*with his/her hand*) associated with the verb *meek.* The participants' verb choice indicates their preferred

basis for categorizing the referent action. For instance, if the participant responded with the verb *wug*, it would suggest that they relied on the visual cue to a greater extent than either the auditory or the linguistic cue.

Since we employed a relatively naturalistic word learning task, we did not design the different types of cues to be equally salient. They also differed in their relative times of presentation. However, in order to accurately assess the relative weighting of each type of cue independently of differences in their inherent salience and timing as well as in individual learners' preferences, we included a 'unimodal baseline' condition during the test phase. In the 'unimodal baseline' trials, participants selected one of the two nonce verbs to describe scenarios in which each cue is presented in isolation (i.e., the action, the sound, or the adjunct phrase associated with each verb is presented in isolation to the participant). Participants' accuracy in producing the appropriate verb when each type of cue is presented in isolation provides a baseline measure of the relative strength of association of each of these cues with the verb when the cues are not placed in conflict with each other. Difference scores are calculated by subtracting participants' accuracy scores in the 'unimodal baseline' trials from their accuracy scores in the corresponding 'multimodal cue competition' trials. These scores provide a measure of participants' tendency to produce the verb associated with the linguistic, auditory, or visual ('odd man out') cue when each type of cue is placed in competition with the two other cues, after taking into account participants' baseline accuracy in producing the appropriate verb when presented with each type of cue in isolation.

For instance, if visual cues are highly weighted relative to the other cues, the frequency with which participants will produce the verb associated with the visual cue in the presence of conflicting auditory and linguistic cues will not differ greatly relative to when the visual cue is presented in isolation. In contrast, if visual cues receive little weight, the frequency with which participants will produce the verb associated with the visual cue is likely to be greatly reduced in the presence of competing auditory and linguistic cues, relative to when the visual cue is presented in isolation. Comparing the decrement in the production of the verb associated with each type of cue when placed in conflict relative to when it is presented in isolation allows us to establish participants' preference hierarchy for different sources of information during the process of verb generalization.

3. Results

The participants' responses were coded for (i) comprehension accuracy during the training phase (the number of correct pointing responses in the four comprehension trials); (ii) production accuracy during the training phase (the number of correctly produced nonce verbs in the eight production trials); (iii) production accuracy in the control clips during the testing phase (the number of correctly produced nonce verbs in the four control trials); (iv) production accuracy in the six test trials in the 'unimodal baseline' condition (the number of correctly produced nonce verbs when an auditory, linguistic, or visual cue was presented in isolation); and (v) production accuracy in the six test trials in the 'multimodal cue competition' condition (the number of correctly produced nonce verbs when an auditory, linguistic, or visual 'odd man out' cue associated with one verb was presented concurrently with two cues associated with the other verb).

3.1. Learning the Nonce Verbs in the Training Phase

Participants learned the nonce verbs *wug* and *meek* with a high level of accuracy after just a few minutes of exposure in an interactive task in which they watched puppets label novel actions using the nonce verbs, performed the novel actions corresponding to the nonce verbs, and produced the verbs to label the corresponding actions. Reponses were 100% accurate in four comprehension and eight production assessment trials conducted during the training phase, and in the four control trials presented during the test phase. Thus, when participants receive congruent linguistic, auditory, and visual cues, they are able to associate the constellation of multimodal cues with the appropriate verb 100% of the time.

3.2. Ranking Multimodal Cues in the Test Phase

Since we used a forced choice response procedure, we elicited binary responses in the test trials. These responses consisted of the appropriate verb associated with a specific cue, either presented in isolation or as an 'odd man out' cue (scored as '1'), or they consisted of the inappropriate verb (scored as '0'). Recall that, in the 'multimodal cue competition' condition, the 'inappropriate' verb was the verb associated with the two cues that were placed in conflict with the 'odd man out' cue. Figure 1 shows the mean proportion of verbs associated with each type of cue in the 'multimodal cue competition' condition (white bars) and in the 'unimodal baseline' condition (grey bars). Participants frequently produced the nonce verb associated with a visual cue, whether it occurred in isolation or it was placed in competition with conflicting auditory and linguistic cues. In contrast, participants were far less likely to produce verbs associated with either linguistic or auditory cues when they competed with conflicting cues compared to when these cues occurred in isolation.

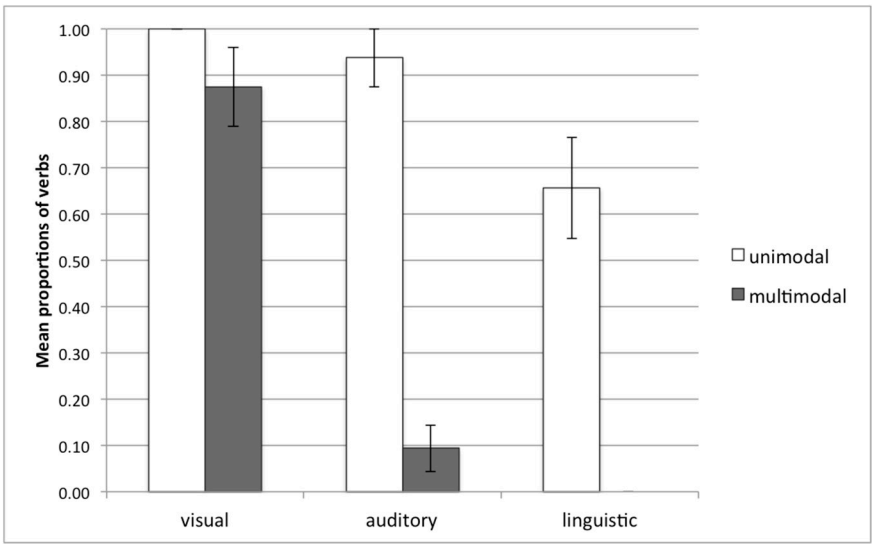

Figure 1. Mean proportions of verbs associated with visual, auditory, and linguistic cues in the 'unimodal baseline' condition (white bars) versus the 'multimodal cue competition' condition (grey bars). Error bars represent standard error of the mean.

As discussed above, difference scores were computed by subtracting the participants' accuracy scores for each type of cue in trials where the cue was presented in conflict with two other cues ('multimodal cue competition' trials) from their accuracy scores for the same type of cue in corresponding trials where they were presented in isolation ('unimodal baseline' trials). The scores thus represent the decrement in the likelihood of producing a verb associated with a cue of a particular type when it was placed in competition with the other cues versus when it occurred on its own.

The difference scores were analyzed using mixed effects logistic regression with treatment coding for fixed effects and the participant and items as random effects. The predictor variable was the Cue Type (linguistic, visual, auditory) and the outcome variable was difference scores (coded as '1' or '0'). A number of control variables were included to control for the effects of 'nuisance' variables including Group (the particular cluster of cues that was associated with the verb during the training phase), Version (the order in which the linguistic versus perceptual cues were presented during the test phase), and Test Item Order (the order of presentation of test stimuli during the test phase).

As Table 2 shows, none of the control variables had a significant effect on participants' responses, a finding that was confirmed by likelihood ratio tests. On the other hand, the significant coefficients for the Cue Type variable as well as likelihood ratio tests (comparing the full model with a model that omitted the Cue Type variable) reveal that the type of cue—visual, linguistic, or auditory—significantly

influences participants' verb generalization tendencies. Both linguistic and auditory cues differed significantly from visual cues in their ability to elicit the associated verb in the cue competition condition, even when differences in their relative strength of association with the verb in isolation were taken into account (linguistic cue: $\beta = 3.09$, z value = 3.75, $p < 0.001$; auditory cue: $\beta = 4.30$, z value = 4.49, $p < 0.001$). The linguistic and auditory cues were directly compared by releveling the cue type variable, which revealed that there is no significant difference between auditory versus linguistic cues in their ability to elicit the corresponding verb in the cue competition condition ($\beta = 1.20$, z value = 1.80, $p = 0.07$).

Table 2. Effects of Stimulus Type, Group, Version, and Test Item Order on differences in participants' accuracy in producing the verb associated with different cue types in the 'multimodal cue competition' condition versus the 'unimodal baseline' condition.

	Estimate	**Standard Error**	***z* Value**	***p*-Value**
Intercept	−2.45	0.92	−2.66	0.007808 **
Cue Type: linguistic	3.09	0.83	3.75	0.000177 ***
Cue Type: auditory	4.30	0.96	4.49	0.000007 ***
Group: 2	−0.64	0.71	−0.91	0.364638
Version: Order B	0.94	0.72	1.31	0.189696
Test Item Order: forward	0.00	0.69	0.00	0.999715

***: $p < 0.001$; **: $p < 0.01$.

4. Discussion

The current study informs accounts of word learning by providing empirical data on how information from different modalities is recruited during the process of word generalization. Participants often produce the verb associated with the visual cue even in the presence of two conflicting cues that pull for a different verb. On the other hand, both linguistic and auditory cues suffer substantial decrements in the cue competition condition. Our findings demonstrate visual dominance effects in adults in a novel verb generalization task that is interactive, involves meaningful, complex stimuli, and imposes no time constraints on participants' responses. These results parallel similar findings in multisensory processing tasks that require participants to provide speeded responses to auditory or visual cues provided using simple, non-meaningful stimuli [25] as well as more complex stimuli [28].

It could be argued, however, that our findings can be explained in terms of participants' differential weighting of event components rather than a preference for a particular sensory modality. In our study, the visual cue depicts a combination of the 'cause' (the agent pulls and releases the spring/cylinder on toy) and 'result' (spring/cylinder moves back to its original position) components of the event. In contrast, the auditory cue corresponds only to the 'result' event component (the twangy/rattly sound is produced as a 'result' of the agent's action); and the linguistic cue conveys only the 'means' (with the hand) or 'manner' (carefully) of the agent's action. Thus, participants' preference to rely on the visual cue could be interpreted instead as a preference to rely on a cue that depicts the 'cause + result' components of the event versus the result alone (the auditory cue) or the manner/means component (the linguistic cue). However, some prior research suggests that an explanation of our findings in terms of event components may not be sufficient [35,36]. For instance, in one study, adults learned a novel verb for a novel event, after which they viewed additional novel events in which either the 'action', the 'result', or the 'instrument' components of the original actions had been altered [35]. Adult learners were unwilling to extend the novel verb to label an event in which the 'result' component of the event had been changed from the training events. Changes in the 'action' component had a moderate effect on participants' willingness to use the novel verb while changes in the 'instrument' had the weakest effect. While this study differs from ours in important respects, it does suggest that adult learners consider the 'result' to be a more central component of an event

relative to either the 'action' or the 'instrument'. If so, one would predict that participants in our study would place greater reliance on the auditory cue (since it corresponds to the result component in our stimuli) than on the linguistic cue (which provides information about the means/manner component). However, we do not observe this pattern in our data. Further research is required that systematically disentangles the effects of event structure from modality-specific preferences.

Further research is also needed to determine how the mechanisms that give rise to visual dominance in the verb generalization task relate to the ones that operate in producing the 'Colavita visual dominance effect' in perceptual tasks [25]. For instance, does participants' preference for visual cues arise from attenuated processing of auditory and linguistic cues when the different types of cues are placed in conflict? Or does the visual bias arise from participants' decision processes during the process of verb generalization [37]? For instance, it is possible that the dominance of visual cues over linguistic cues in the current study arises from the use of adjunct phrases that are not considered to be central to verb meaning. Participants' preference for visual cues might be reduced or even reversed if linguistic cues that are more relevant to verb meaning are employed, such as noun phrases that are core arguments of the verb. Additionally, the novel verbs could be consistently presented in transitive frames with an overt object (*She is wugging the toy carefully*) rather than in transitive frames presented either with or without an overt object (*She is wugging (the toy) carefully*). Exposure to transitive frames that consistently label the entity involved in the result subevent may increase the overall salience of both the linguistic cues as well as the nonlinguistic sounds associated with the result component of the event. On the other hand, if visual dominance effects arise largely from the attenuated processing of (linguistic and nonlinguistic) auditory cues when competing with visual cues, then visual cues may continue to prevail over linguistic cues, even if the latter are more central to verb meaning.

Additional studies are also required to investigate whether the ranking of modality-specific effects obtained in the current study can be modulated using other tasks or stimuli. For instance, it is possible that increasing the relative frequency of the multimodal trials or the salience and timing of visual, auditory, and linguistic cues will modulate the size and direction of the visual dominance effect [26]. The type of words used (nouns versus verbs) as well as factors such as motivation or communicative goals might also play a role in influencing modality-specific effects on word generalization. Finally, age is an important factor that modulates the relative importance of visual and auditory cues [29,37]. As discussed in the introduction, prior research shows auditory dominance in young children in tasks involving multisensory processing. Whether they would also exhibit auditory dominance in a verb generalization task is an interesting question that needs to be investigated in further research exploring the role of sensory information in word meaning and use [38].

Acknowledgments: I gratefully acknowledge the contributions of the research assistants of the Language, Development, and Cognition Lab (University of Colorado Boulder, Boulder, CO, USA), including Fanyin Chen, Patricia Davidson, and Madison Wagner who assisted in the design and running of the experiment. Thanks also to Pui Fong Kan for valuable feedback, and to Norielle Adricula, Caroline Good, Sean Kelly, and Jayne Williamson-Lee for their assistance in the lab. Additionally, I would like to thank the two anonymous reviewers of this paper for their valuable feedback.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Quine, W.V.O. *Word and Object*; MIT Press: Cambridge, MA, USA, 1960.
2. Clark, E.V. What's in a word: On the child's acquisition of semantics in his first language. In *Cognitive Development and the Acquisition of Language*; Moore, T.E., Ed.; Academic Press: New York, NY, USA, 1973; pp. 65–110.
3. Nelson, K. Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* **1973**, *38*, 1–135. [CrossRef]
4. Gentner, D. What looks like a jiggy but acts like a zimbo?: A study of early word meaning using artificial objects. *Pap. Rep. Child Lang. Dev.* **1978**, *15*, 1–6.
5. Rescorla, L. Overextension in early language development. *J. Child Lang.* **1980**, *7*, 321–335. [CrossRef] [PubMed]

6. Hollich, G.; Hirsh-Pasek, K.; Golinkoff, R. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monogr. Soc. Res. Child Dev.* **2000**, *65*, 1–135. [CrossRef]
7. Markson, L.; Bloom, P. Evidence against a dedicated system for word learning in children. *Nature* **1997**, *385*, 813–815. [CrossRef] [PubMed]
8. Vlach, H.A.; Sandhofer, C.M. Fast Mapping Across Time: Memory Processes Support Children's Retention of Learned Words. *Front. Dev. Psychol.* **2012**, *46*, 1–8. [CrossRef] [PubMed]
9. Murphy, G. L. *The Big Book of Concepts*; MIT Press: Cambridge, MA, USA, 2002.
10. Imai, M. Children's use of argument structure, meta-knowledge of the lexicon, and extra-linguistic contextual cues in inferring meanings of novel verbs. In Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, Keihanna, Japan, 28–30 July, 2008; Mueller, S., Ed.; CSLI Publications: Stanford, CA, USA, 2008; pp. 399–416.
11. Pye, C.; Loeb, D.F.; Pao, Y.Y. The Acquisition of Breaking and Cutting. In Proceedings of the Twenty-seventh Annual Child Language Research Forum, Stanford, CA, USA, 7–9 April 1995; Clark, E.V., Ed.; CSLI Publications: Stanford, CA, USA, 1996; pp. 227–236.
12. Ameel, E.; Malt, B. C.; Storms, G. Object naming and later lexical development: From baby bottle to beer bottle. *J. Mem. Lang.* **2008**, *58*, 262–285. [CrossRef]
13. Carey, S. Beyond fast mapping. *Lang. Learn. Dev.* **2010**, *6*, 184–205. [CrossRef] [PubMed]
14. Saji, N.; Imai, M.; Saalbach, H.; Zhang, Y.; Shu, H.; Okada, H. Word learning does not end at fast-mapping: Evolution of verb meanings through reorganization of an entire semantic domain. *Cognition* **2011**, *118*, 45–61. [CrossRef] [PubMed]
15. Wagner, K.; Dobkins, K.; Barner, D. Slow mapping: Color word learning as a gradual inductive process. *Cognition* **2013**, *127*, 307–317. [CrossRef] [PubMed]
16. Bowerman, M. Why can't you "open" a nut or "break" a cooked noodle? Learning covert object categories in action word meanings. In *Building Object Categories in Developmental Time*; Gershkoff-Stowe, L., Rakison, D.H., Eds.; Erlbaum: Mahwah, NJ, USA, 2005; pp. 209–243.
17. Wellsby, M.; Pexman, P. Developing embodied cognition: Insights from children's concepts and language processing. *Front. Cogn. Sci.* **2014**, *5*, 506. [CrossRef] [PubMed]
18. Vouloumanos, A. Fine-grained sensitivity to statistical information in adult word learning. *Cognition* **2008**, *107*, 729–742. [CrossRef] [PubMed]
19. Scott, R.M.; Fisher, C. 2-year-olds use distributional cues to interpret transitivity-alternating verbs. *Lang. Cogn. Process.* **2009**, *24*, 777–803. [CrossRef] [PubMed]
20. Twomey, K.; Chang, F.; Ambridge, B. Lexical distributional cues, but not situational cues, are readily used to learn abstract locative verb-structure associations. *Cognition* **2016**, *153*, 124–139. [CrossRef] [PubMed]
21. Bates, E.; MacWhinney, B. Competition, variation, and language learning. In *Mechanisms of Language Learning*; MacWhinney, B., Ed.; Erlbaum: Hillsdale, NJ, USA, 1989; pp. 157–193.
22. Ellis, N.C. Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Appl. Linguist.* **2006**, *27*, 164–194. [CrossRef]
23. Houston-Price, C.; Plunkett, K.; Duffy, H. The use of social and salience cues in early word learning. *J. Exp. Child Psychol.* **2006**, *95*, 27–55. [CrossRef] [PubMed]
24. Koehne, J.; Crocker, M. The interplay of cross-situational word learning and sentence-level constraints. *Cogn. Sci.* **2015**, *39*, 849–889. [CrossRef] [PubMed]
25. Colavita, F.B. Human sensory dominance. *Percept. Psychophys.* **1974**, *16*, 409–412. [CrossRef]
26. Spence, C.; Parise, C.; Chen, Y.C. The Colavita visual dominance effect. In *The Neural Bases of Multisensory Processes*; Murray, M.M., Wallace, M.T., Eds.; CRC Press: Boca Raton, FL, USA, 2012; pp. 529–556.
27. Sinnett, S.; Spence, C.; Soto-Faraco, S. Visual dominance and attention: The Colavita effect revisited. *Percept. Psychophys.* **2007**, *69*, 673–686. [CrossRef] [PubMed]
28. Stubblefield, A.; Jacobs, L.A.; Kim, Y.; Goolkasian, P. Colavita Dominance Effect Revisited: The Effect of Semantic Congruity. *Atten. Percept. Psychophys.* **2013**, *75*, 1827–1839. [CrossRef] [PubMed]
29. Nava, E.; Pavani, F. Changes in sensory dominance during childhood: Converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Dev.* **2013**, *84*, 604–616. [CrossRef] [PubMed]
30. Spence, C. Explaining the Colavita visual dominance effect. *Prog. Brain Res.* **2009**, *176*, 245–258. [PubMed]

31. Lederer, A.; Gleitman, H.; Gleitman, L. Verbs of a feather flock together: Semantic information in the structure of maternal speech. In *Beyond Names for Things: Young Children's Acquisition of Verbs*; Tomasello, M., Merriman, W.E., Eds.; Erlbaum: Hillsdale, NJ, USA, 1995; pp. 277–297.
32. Tomasello, M.; Kruger, A. Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *J. Child Lang.* **1992**, *19*, 311–333. [CrossRef] [PubMed]
33. Syrett, K.; Arunachalam, S.; Waxman, S.R. Slowly but surely: Adverbs support verb learning in 2-year-olds. *Lang. Learn. Dev.* **2014**, *10*, 263–278. [CrossRef] [PubMed]
34. Narasimhan, B.; Cheng, F.; Davidson, P.; Kan, P.F.; Wagner, M. The Influence of Visual, Auditory, and Linguistic Cues On Children's Novel Verb Generalization. In *Perspectives on the Architecture and Acquisition of Syntax: Essays in honour of R.Amritavalli*; Sengupta, G., Sircar, S., Raman, G., Balusu, R., Eds.; Springer Nature: Berlin, Germany, in press.
35. Behrend, D.A. The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Dev.* **1990**, *61*, 681–696. [CrossRef] [PubMed]
36. Bunger, A.; Lidz, J. Constrained flexibility in the extension of novel causative verbs. In Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, USA, 10–12 February 2006; Antić, Z., Chang, C.B, Cibelli, E., Hong, J., Houser, M.J., Sandy, C.S., Toosarvandani, M., Yao, Y., Eds.; Berkeley Linguistics Society: Berkeley, CA, USA, 2006; pp. 479–490.
37. Robinson, C.W.; Sloutsky, V.M. When audition dominates vision: Evidence from cross-modal statistical learning. *Exp. Psychol.* **2013**, *60*, 113–121. [CrossRef] [PubMed]
38. Maouene, J.; Sethuraman, N.; Laakso, A.; Maouene, M. The body region correlates of concrete and abstract verbs in early child language. *Cogn. Brain Behav. Interdiscip. J.* **2011**, *25*, 449–484.