



ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

# Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned

Inbal Arnon<sup>a</sup>, Michael Ramscar<sup>b,\*</sup><sup>a</sup> Psychology Department, University of Haifa, Haifa, Israel<sup>b</sup> Linguistics Department, Eberhard Karls University, Tübingen, Germany

## ARTICLE INFO

## Article history:

Received 26 April 2010

Revised 4 August 2011

Accepted 14 October 2011

Available online xxxx

## Keywords:

Communication

Language

Learning

Second language acquisition

Linguistic units

## ABSTRACT

Why do adult language learners typically fail to acquire second languages with native proficiency? Does prior linguistic experience influence the size of the “units” adults attend to in learning, and if so, how does this influence what gets learned? Here, we examine these questions in relation to grammatical gender, which adult learners almost invariably struggle to master. We present a model of learning that predicts that exposure to smaller units (such as nouns) before exposure to larger linguistic units (such as sentences) can critically impair learning about predictive relations *between* units: such as that between a noun and its article. This prediction is then confirmed by a study of adult participants learning grammatical gender in an artificial language. Adults learned both nouns and their articles better when they were first heard nouns used in context with their articles prior to hearing the nouns individually, compared with learners who first heard the nouns in isolation, prior to hearing them used in context. In the light of these results, we discuss the role gender appears to play in language, the importance of meaning in artificial grammar learning, and the implications of this work for the structure of L2-training.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Why is acquiring a language to native proficiency in adulthood so difficult? Numerous studies have revealed that the expertise levels of native and non-native speakers diverge across many aspects of language, including pronunciation (Moyer, 1999), morphological processing (Johnson & Newport, 1989), and the use of formulaic speech and idioms (Vanlancker-Sidits, 2003). Given the many differences between children and adults, both in terms of cognitive and neural development and in terms of the social contexts in which they learn languages, it is perhaps unsurprising that children and adults differ in their ability to learn. What is surprising, given adults’ proficiency when it comes to learning in other domains, is that children ap-

pear to learn languages far more successfully than most adults.

Over the years, a number of different approaches have been adopted in attempting to make sense of this seeming paradox. Lenneberg (1967) proposed that adults no longer have access to a biological window of opportunity for learning language, while other researchers (Kuhl, 2000; Neville & Bavelier, 2001) have highlighted maturational changes in neural plasticity, and the ways in which early neural commitment shapes consequent learning (e.g., learning the phonetic distinctions of one language changes sensitivity to non-phonemic distinctions, Werker & Tees, 1984; see also Newport, 1990). Here, we consider another difference between children and adults: experience. It is clear that adult and infant language learners differ in their experience, both in what they know about language and the world. In what follows, we examine whether the experience that adults bring to language learning may actually prove to be a hindrance when it comes to learning second languages.

\* Corresponding author.

E-mail addresses: [inbal.arnon@gmail.com](mailto:inbal.arnon@gmail.com) (I. Arnon), [michael.ramscar@uni-tuebingen.de](mailto:michael.ramscar@uni-tuebingen.de) (M. Ramscar).

### 1.1. Differences between L1 and L2 learners

Unlike children, adults come to the task of language learning with a great deal of experience with language: they know about words and grammar, and they almost certainly have knowledge of the lexical and grammatical semantics of their first language. The speech directed at adult learners also differs from that directed at children: adult-to-adult speech tends to be longer, less repetitive, and less prosodically informative (e.g., Fernald et al., 1989; Fisher & Tokura, 1996), and at least in classroom settings, it often emphasizes single vocabulary items (Doughty & Williams, 1998). Moreover, the vast majority of the adult learners whose abilities are studied are literate, and come to the task of language learning with prior knowledge and beliefs regarding the way sentences and utterances can be segmented into individual words. It is possible that the difference in the background knowledge each brings to language learning, and the different input each learns from, may in turn influence the size of the “linguistic units” children and adults are able to discriminate and use in early language learning.

Comparing the speech of young children and adult L2 learners offers some initial support for this suggestion. Early childhood speech includes multi-word utterances like ‘how-are-you’ or ‘what-is-that’ (e.g., Peters, 1983), and there is little evidence that children can “fully-analyze” these sequences in terms of their smaller lexical units. Indeed, detailed studies of 2-year olds’ speech suggest that a high proportion of the identifiable multi-word utterances they produce are relatively ‘frozen’ suggesting that children of this age have limited productive knowledge of their parts (Lieven, Pine, & Baldwin, 1997). Moreover, most of the ‘non-frozen’ utterances that children of this age produce can be derived from previous utterances using simple combinatorial operations involving the addition or substitution of a single word (Bannard, Lieven, & Tomasello, 2009; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Salomo, & Tomasello, 2009). This suggests that, at least initially, many of the lexical “units” children use and understand may not have been discriminated from the larger constructions they are encountered in (Bannard & Matthews, 2008; Goldberg, 2006; Tomasello, 2003).

In contrast, the language of adult L2 learners is often described as overly-flexible (or non-formulaic; e.g., Fillmore, 1979; Pawley & Syder, 1980; Wray, 2002). Under-segmented speech (e.g., where two units are treated as one, as in *give-it the ball*) is rare in adult L2 learners. Instead, adult learners tend to treat chunks of language as more flexible than they actually are, producing deviant collocations like ‘being taking care of’ or ‘put more attention to’ (Yorio, 1989); concomitant under- or misuse of idioms and formulaic expressions is common, and is often cited as a hallmark of non-native language use (Ellis, Sheen, Murakami, & Takashima, 2008; Granger, 1998; Wray, 2002, 2004, 2008; see Bardovi-Harlig, 2009 for a review).

Here we examine the effect that starting with smaller or larger “linguistic units” (see also Elman, 1993; Newport, 1990) has on subsequent learning by examining adults learning of an artificial grammatical gender system. Learning the agreement patterns between articles and nouns in

languages with grammatical gender is difficult for non-native speakers to master (see e.g., Harley, 1979; Scherag, Demuth, Roesler, Neville, & Roeder, 2004). If some of this difficulty is down to the units that adult learners attend to and employ, then manipulating access and exposure to these units should result in marked changes in learning. In particular, if starting with smaller units of language is what inhibits adults’ learning, then having them learn in conditions in which the article and the noun are less differentiated, and less easy to discriminate between, should facilitate learning of the relation between them.

### 1.2. Grammatical gender

Grammatical gender is a system found in many languages. It assigns all nouns (including inanimate ones) to noun classes, usually marking neighboring words for agreement (Corbett, 1991). In Hebrew, for example, verbs and adjectives are marked for gender, while in Spanish and French, articles have to agree in gender with the nouns they precede (e.g., *la pelota* – the-Feminine ball-Feminine vs. *el vaso* the-Masculine glass-Masculine). In all languages that mark gender, knowing a noun’s gender is essential for correct sentence construction.

Although children master grammatical gender relatively early (see Slobin, 1985 for cross-linguistic reports) and make few mistakes in spontaneous speech (Bassano, Maillachon, & Mottet, 2008; (Karmiloff-Smith, 1979; Slobin, 1985), L2 learners have persistent difficulty with grammatical gender even after extensive exposure (Dewaele & Véronique, 2001; Holmes & de la Batie, 1999; Rogers, 1987; Scherag et al., 2004). Native and non-native speakers also differ in their ability to use the gender information conveyed by the article in real time processing. Whereas native speakers can use article information to guide lexical access—e.g., anticipating a feminine noun following a feminine article (van Heugten & Johnson, 2011; van Heugten & Shi, 2009; Lew-Williams & Fernald, 2007a, 2007b) or slowing down if there is a gender mismatch between the article and the noun (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Grosjean, Dommergues, Cornu, Guillelmon, & Beson, 1994)—non-native speakers do not show these effects (Guillelmon & Grosjean, 2001; Lew-Williams & Fernald, 2010; Scherag et al., 2004).

These findings suggest that native speakers treat the article and the noun as a more cohesive unit than do non-native speakers, allowing them to both select the correct article in production, and to use it to facilitate recognition in comprehension (see also, Clahsen & Muysken, 1986; Clahsen & Felser, 2006; Hawkins & Chan, 1997). Indeed, researchers from a variety of theoretical backgrounds have suggested that children initially treat the article and the noun as a *single* unit, rather than two separable ones, as an adult might (Carroll, 1939, 1989; Chevrot, Dugua, & Fayol, 2008; MacWhinney, 1978). This would be a natural consequence of the way children encounter nouns, which is most often in the company of articles (especially in gender-marking languages, Mariscal, 2009).

Numerous findings support this observation. For instance, children’s early use of articles tends to be quite lexically specific. Instead of using articles with all nouns,

children will initially only use a given article with a given noun (e.g., producing only the definite article with one noun and only the indefinite with another, Mariscal, 2008; Pine & Lieven, 1997). Similarly, the patterns of liaison acquisition in French support the idea that articles and nouns are initially learned as a single unit: children often make mis-segmentation errors, incorrectly treating the liaison consonant as part of the noun (Chevrot et al., 2008; Dugua, Spinell, Chevrot, & Fayol, 2009).

Adults, on the other hand, appear far less likely to treat the article and the noun as a single unit. As a result of their experience with their first language, adult L2 learners often know that nouns and articles are separate entities. Further, the way older learners often encounter nouns and articles, particularly in a classroom setting, may emphasize their independence (Doughty & Williams, 1998). For instance, while none of children's early language input is written, adults are likely to learn from written input in which the distinction between the article and the noun is made explicit. Finally, while there is evidence that while adults can use cognitive control to selectively attend to particular aspects of the input, children appear to largely lack this facility (Ramscar & Gitcho, 2007; Thompson-Schill, Ramscar, & Chrysikou, 2009). In other words, adults not only know that articles and nouns are separate, but they can also 'choose' to focus their attention on one or the other.<sup>1</sup>

### 1.3. Why might granularity matter in learning?

Why might starting with noun-labels make it harder to learn about the relations between nouns and articles? A possible answer lies in formal learning theories, which see the acquisition of knowledge as being driven by discrepancies between what is expected based on prior experience and what is encountered in the environment (Rescorla & Wagner, 1972). In these theories, prior learning can **block** (Kamin, 1969) the learning of later information by rendering it redundant. Thus learning noun-labels before article + noun sequences might affect the learning of articles if a learner's knowledge about the relationship between a noun and its meaning marginalizes the informative contribution of an article. For example, if an English-speaker learning the French word for cat has already learned to fully expect a label ("chat") given a semantic context (chat/cat), then there will be little room for a definite article ("le") to add new information to this relationship: what the learner already knows will interfere with learning the article.

The basic information principles that govern learning theory (Rescorla & Wagner, 1972; Gallistel, 2003), and which give rise to blocking effects can be illustrated by first considering the effect of background rates on learning. In a classic study of this, Rescorla (1968) conditioned several groups of rats to expect mild shocks following tones. The association rate between tones and shocks was held constant, while the **background rate** of tones varied (Fig. 1).

<sup>1</sup> A tendency to focus on noun-labels at the expense of gender information may be exaggerated in learners whose first language does not have gender categories, leading to an effect of L1 on L2 learning of grammatical gender (Marinova-Todd, 1994; Sabourin, Stowe, & de Haan, 2006).

Rescorla (1968) found that the degree to which rats conditioned to the relationship between tones and shocks depended on the background rate of tones: the higher the background rate, the less the rat conditioned to the pairing. This is because the **informativity** of the tone – that is, the degree to which it predicts an upcoming shock – depends *both* on how often it is coupled with a shock, and how often it simply occurs on its own.

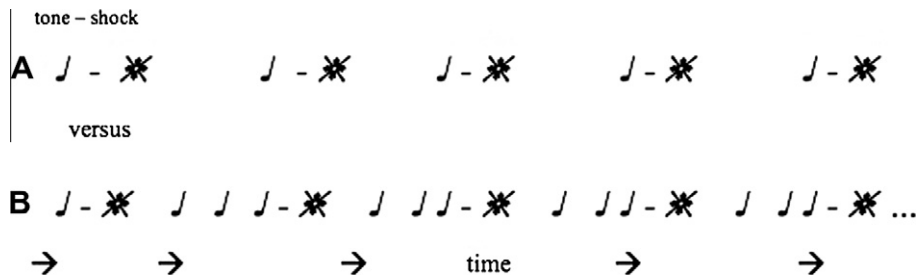
It should be noted that this has important implications for the way learning is understood (Rescorla, 1988): given that there was no change in the 'association rate' between A and B – only the background rate varied – the difference in learning between the two groups can only be explained in terms of what was learned on the "no shock" trials. In other words, the rats' learning was driven by the non-occurrence of an expected event (**prediction error**). If the difference in learning in the two groups of rats is driven by the *non*-occurrence of expected events, it also follows that learning cannot simply be a process of tracking the co-occurrences of cues and events (such as rewards or punishment).

However, there is still more to learning than simply counting successful and unsuccessful predictions. Not only is learning predictive, it is also competitive. **Blocking** (Kamin, 1969) is a simple statistical consequence that follows from the idea that learning serves to reduce uncertainty about future events: once a learner has learned to fully predict an outcome given a cue or set of cues, learning about additional cues is unnecessary (the information they provide is redundant). Informally, one can think of uncertainty as a resource, which is depleted as successful predictions lead to more certain expectations. When a set of cues already fully predicts an event, there will be little uncertainty available to drive the learning of other cues to that event. Cues can thus be seen as *competing* with one another for the predictive value.

Accordingly, if a rat has already learned that it will be shocked when it hears a tone, then when a light is paired with the tone, the rat will fail to learn to value the light as an additional predictive cue. Because the tone is already fully informative about the upcoming shock, the light is redundant, and prior learning about the tone will block subsequent learning about the light. Such results indicate that rats do not learn simple "associations" between stimuli and responses. Rather, they learn about the degree to which cues are **informative** about events (Dickinson, 1980; Rescorla, 1988; Gallistel, 2003).

This idea – that rats learn to predict, and in some sense 'understand' the world around them by monitoring their predictions – can be formalized in a relatively straightforward manner (e.g., Rescorla & Wagner, 1972; McLaren & Mackintosh, 2000; Pearce & Hall, 1980; Gallistel, 2003).<sup>2</sup> Moreover, the idea of discrepancy monitoring that moti-

<sup>2</sup> Although these models are usually described as "associative," because they explain phenomena that traditionally labelled "associative learning" in the literature, computationally, these models are actually driven by discriminative principles. Discriminative learning doesn't simply involve 'tracking associations,' as many psychologists assume: instead these models actively discriminate against cues that generate error in order to maximize successful predictions (see Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010, for a tutorial).



**Fig. 1.** Schematic of the different training schedules in Rescorla (1968). **A** represents a condition in which the background rate is 0 (the ‘background rate’ is simply the number of uncoupled tones that are observed), whereas **B** represents a condition in which the background rate is twice that of the rate of co-occurrence between tones and shocks.

vates the many mathematical models of animal learning appears to have a correspondingly straightforward neural implementation, and there is good neurobiological evidence for these mechanisms in humans (e.g., Fiorillo, Tobler, & Schultz, 2003; Hollerman & Schultz, 1998; Holroyd & Coles, 2002; Schultz & Dickinson, 2000; Waelti, Dickinson, & Schultz, 2001; Montague, Hyman, & Cohen, 2004; Schultz, 2006).

Further, research suggests that language learning may be sensitive to error in learning in the way that these models suggest: error-driven models and analyses have been used to predict and explain the specific difficulties L2 learners have in learning temporal reference in new languages (Ellis & Sagarra, 2010), as well as temporal sequencing effects in word learning in children and adults (Ramscar et al., 2010; Ramscar, Dye, Witten, Klein, 2012; Ramscar, Dye, Popick & O'Donnell-McCarthy, 2011); error patterns in the acquisition of morphosyntax (Ramscar & Yarlett, 2007; Ramscar, Dye, & Yarlett, 2009; see also St. Clair, Monaghan & Ramscar, 2009); and reading performance across a wide-variety of measures in adults (Baayen, 2010, 2011; Baayen, Milin, Filipovic Durdevic, Hendrix, & Marelli, 2011; see also Colunga, Smith, & Gasser, 2009; Smith, Colunga, & Yoshida, 2010).

#### 1.4. Blocking and grammatical gender learning

As we noted above, the basic principles that govern discrimination learning suggest a reason why grammatical gender is difficult for adults to learn. If adults learn about smaller linguistic units such as noun-labels in isolation, they will learn to associate, say, the informative features of a given noun's semantics (such as the objects associated with that noun) with a particular noun label. This will serve to concentrate their knowledge about the noun label on its semantic features, which may subsequently impair learning about the relation between the *article* and the noun. Indeed, if a noun label is already fully predicted by the presence of an object (or semantic cues that lead to the expectation of that object), then an article will provide little to no information to a learner (Ramscar et al., 2010); in this case, the object – noun-label association will effectively block learning of the article.

If, however, the association between semantics, articles and noun-labels are not learned in isolation, then learners will have to learn to discriminate articles from nouns,

which – given appropriate experience – will happen as a result of *cue-competition* (a central component of discriminative learning models). Two factors will influence this process:

1. Each time a noun-label occurs with an object (but without its article, or with a different article), the association between the noun-label and the object will strengthen at the expense of the article. Since sets of objects will tend to occur more faithfully with noun-labels than articles (given that in most languages, more than one article can appear with any given noun), any given noun's association with those objects will come to strengthen at the article's expense.
2. Just as importantly, every time an article occurs *without* a particular noun-label or set of objects, its association with them will weaken. Since articles occur ‘promiscuously’ with many nouns and objects, a given article's association with a particular noun-label or set of objects will thus weaken relative to the noun-label, such that objects will become more strongly associated with nouns than with articles.

Crucially, however, because articles will be informative about upcoming nouns as language unfolds in time, they will remain associated with objects and nouns as a result of initial learning (Ramscar et al., 2010; Rescorla & Wagner, 1972).

Blocking – and learning and information theory more generally – thus offers an explanation for many of the empirical findings described above: It may be, for instance, that young language learners – who are learning to discriminate linguistic “segments” and meanings at the same time – initially produce “under-analyzed” segments because they initially form associative relationships between an object and an article–noun sequence, and that these only gradually give way to more flexible usage as associativity between the object and article diminishes as a result of further discrimination learning. Conversely, it may be that adult language learners – attending to smaller linguistic units – fail to match nouns and articles correctly, and are overly ‘flexible’ in their usage because they never form associations between objects and articles (which in turn may cause them to pair nouns and articles incorrectly as well).

Learning theory thus provides testable predictions in this case: First, it ought to be the case that learning smaller

semantic units in isolation should lead to blocking in the way we have described, because it leads to a weaker association between the article and the object; and second, it follows that it should be possible to enhance adult learning by ‘forcing’ adults to learn from larger units.

## 2. Does granularity matter?

To systematically examine the effect of unit size on learning, we created an auditorily presented novel language and contrasted the effect of initially exposing adult learners to article–noun sequences – in which the boundaries between articles and nouns were unclear – with that of initially presenting them with the noun-labels as identifiable units. (Here, we use the term “noun-label” to refer to a noun appearing without an article.)

Learners were divided into two groups. In the **sequence-first** group, learners were first exposed to article + noun sequences in whole sentences, and then to noun-labels. In the **noun-label-first** group, learners were first exposed to noun-labels and then to full sentences. By the end of the experiment, both groups had received exactly the same input, but in different orders. This allowed us to manipulate the size of the initial units learners were exposed to while keeping frequency of exposure constant.

### 2.1. The artificial language

The artificial language comprised fourteen novel labels for familiar concrete objects (e.g., piano-*slindot*), two articles (*sem* and *bol*) and a carrier phrase (*os fernal en*, see Appendix for full list of items). The nouns were divided into two “classes,” and each noun only appeared with one article. There were no additional semantic or phonological cues to class membership. Articles always followed the carrier phrase and preceded nouns. An example of a full sentence in the language is given in (1).

(1)	<i>Os-ferpal-en</i> Carrier phrase	<i>bol</i> article 1	<i>slindot</i> “piano”
-----	---------------------------------------	-------------------------	---------------------------

All noun-labels were two syllables long. The objects were matched for familiarity, and for frequency and age-of-acquisition of the English word (using the Bristol norms, [Stadthagen-Gonzalez & Davis, 2006](#)).

A male speaker recorded the carrier phrase, the articles, and the nouns separately. These were concatenated to create the full sentences. One recorded token of each noun, each article and the carrier phrase was used throughout the experiment to ensure that the nouns had the same prosody in full sentences and in isolation, and that the articles had the same acoustic features as the nouns. The duration of the two articles was kept identical to ensure that neither had any acoustic prominence.

In addition to the experimental items, a second block of phrases was constructed. This “distracter block” comprised the same carrier phrase, seven different nouns and two additional articles (*tid* and *gob*). In contrast to the experimental items, the mapping between the articles and the

nouns was not consistent, and nouns could appear with either article.

### 2.2. Formal learning model

To formally examine the effects of exposure to different unit sizes on learning, we simulated acquisition of the language in both sequence-first and noun-label-first conditions using the Rescorla–Wagner model ([Danks, 2003](#); [Rescorla & Wagner, 1972](#)). Rescorla–Wagner is perhaps the most widely used learning rule in psychology, and has been applied to numerous learning effects in both animals and humans ([Rudy, 1974](#); [Gluck & Bower, 1988a, 1988b](#); [Miller, Barnet, & Grahame, 1995](#); [Siegel & Allan, 1996](#); [Danks, 2003](#)). While the model does not account for all the phenomena observed in ‘associative’ learning, it provides the most accessible and most widely used formalization of the basic principles of learning theory (i.e., cue competition and blocking), its basic mechanisms are supported by a wealth of neurobiological evidence (e.g., [Schultz, Dayan, & Montague, 1997](#); [Fiorillo et al., 2003](#); [Hollerman & Schultz, 1998](#); [Holroyd & Coles, 2002](#); [Schultz & Dickinson, 2000](#); [Waelti et al., 2001](#); [Montague et al., 2004](#); [Schultz, 2006](#); [Daw & Shohamy, 2008](#)) and it is sufficiently detailed to allow a straightforward testing of the analysis we present here. It should be noted, however, that our analysis of blocking effects is consistent with a wide range of learning models that implement cue competition in discrimination learning (e.g., [Gallistel, 2003](#); [Gallistel & Gibbon, 2000](#); [McLaren & Mackintosh, 2000](#); [Pearce & Hall, 1980](#); [Rosenblatt, 1959](#); see also [Danks, 2003](#); [Dayan & Daw, 2008](#)).

The Rescorla–Wagner model simulates changes in the associative strengths between individual cues and an outcome as the result of discrete learning trials. If the presence of a cue or outcome  $X$  at time  $t$  is defined as present ( $X, t$ ), and its absence as absent ( $X, t$ ), then the predictive value  $V$  of a cue  $C_i$  to outcome  $O$  after a learning event at time  $t + 1$  can be specified as:

$$V_i^{t+1} = V_i^t + \Delta V_i^t \quad (1)$$

and changes ( $\Delta$ ) to the predictive value of  $V_i^t$  are defined as:

$$\Delta V_i^t = \begin{cases} 0 & \text{if } ABSENT(C_i, t) \\ \alpha_i \beta_1 \left( \lambda - \sum_{PRESENT(C_j, t)} V_j \right) & \text{if } PRESENT(C_j, t) \& PRESENT(O, t) \\ \alpha_i \beta_2 \left( 0 - \sum_{PRESENT(C_j, t)} V_j \right) & \text{if } PRESENT(C_j, t) \& ABSENT(O, t) \end{cases} \quad (2)$$

Learning is thus determined by a discrepancy function in which  $\lambda$  is the total value of a predicted event  $O$  (the maximum amount of associative strength that the event can support) and  $V_j$  is the predictive value for  $O$  given the cues  $C_j$  present at time  $t$ , while the rate of change ( $\Delta$ ) at  $t$  is determined by two factors: the overall learning rate  $\beta$  (where  $0 \leq \beta \leq 1$ ), and the individual saliency of cues,  $\alpha$  (where  $0 \leq \alpha \leq 1$ ).

When cues present on a trial are positively supported – i.e., where a predicted outcome occurs – the Rescorla–

Wagner learning rule will produce a negatively accelerated learning curve (the result of events becoming better predicted, which reduces the discrepancy between what is expected and what is observed) and asymptotic learning over repeated trials (as events become fully predicted). When cues present on a trial produce **error** – i.e., when a predicted outcome fails to occur –  $\lambda$  (the value of the expected outcome) takes a value of zero because it didn't occur. In these cases, the discrepancy function will produce a negative value, resulting in a reduction in the associative strength between the erroneous cues and the absent outcome. Because error reduces the predictive value of cues, and because the total amount of value a given outcome can support is finite, this process causes cues to compete with one another for relevance in learning, leading to patterns of learning that usually differ greatly from those that would arise if learned values simply reflected correlations between cues and outcomes (a common misconstrual of learning; Rescorla, 1988).

### 2.2.1. Implementation

Two models were constructed to simulate either sequence-first or noun-label-first training. Because we were interested in how learning affects the relative value of cues, the  $\alpha$  parameter was set to 1 in these simulations, eliminating its influence. The other values were set at  $\lambda = 100\%$  and  $\beta = 0.2$  (a learning-rate that typically predicts human learning with a good degree of accuracy, see e.g. Ramscar et al., 2010; Ramscar et al., 2011). Accordingly, the only free parameter in all of the simulations was the learning rate ( $\beta$ ), which was fixed in advance and held constant.

The models assume that language learners will seek to associate semantics with acoustic events that unfold in time (Ramscar et al., 2010), and were thus configured to reflect the conditions under which this could take place in the training environment, in which visual stimuli were present throughout each training trial, whereas acoustic stimuli occurred sequentially in each trial (Fig. 2). Accordingly, cues representing the object depictions and the different presentation contexts (i.e., the presence or absence of the gesturing figure in the learning environment) were

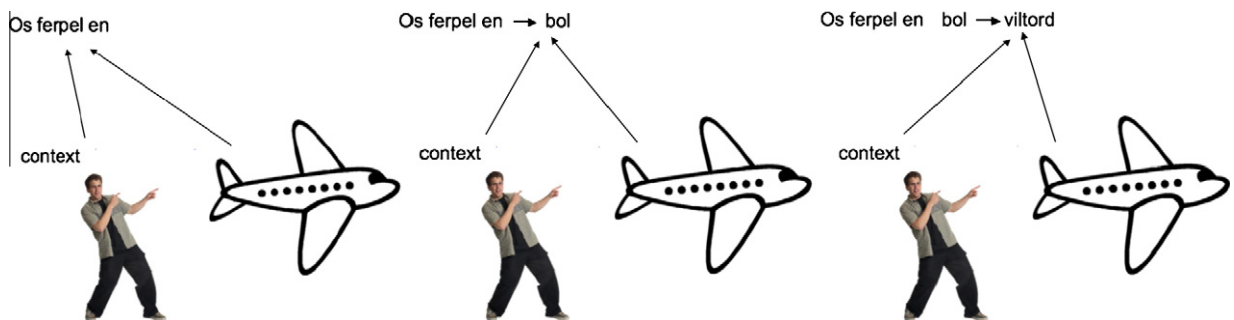
available as cues to initial acoustic stimuli, and both earlier acoustic stimuli and the object depictions and context were available as cues to later acoustic stimuli. Note that our simulation thus closely mirrors the training our human participants will receive.

Specifically, in the **sequence-first** condition, each object depiction and trial context were initially available as cues to the carrier phrase “os ferpal en” (left). The carrier phrase then served as an additional cue to the article (“bol,” center), which in turn served as an additional cue to the noun-label, “viltord” (right). That is: a visual object stimulus, the trial context and the carrier phrase competed as cues to each determiner, and then an object stimulus, the trial context and a determiner competed as cues to each noun-label (see Gureckis & Love, 2010, for a similar architecture, along with evidence of its ability to simulate human performance in sequence learning). In the **noun-first** condition, the situation was simpler: each object depiction and the trial context served as the cues to each noun label.

The models were then trained on the same number of trials in two blocks, to simulate the training human participants would receive. The noun-label-first model was trained on 70 noun-label trials (each noun-label was repeated five times) and then 70 full sentences (each noun in a sentence five times). In the sequence-first model, the order of these training blocks was simply reversed.

Note that in the sequence training trials, all of the cues start with equal values (0) and incrementally strengthen and weaken with each trial, competing for relevance as training proceeds. However, when sequence training follows noun training in the noun-label-first condition, the object depiction and context cue will have already been associated with the noun-label, and this will (probabilistically) block learning of the article. For example, if planes (Fig. 2) have already been learned as cues to “viltord” prior to the sequence learning trials, the article “bol” will be less informative about “viltord” (right panel), simply because “viltord” will already be strongly expected to occur given only the presence of the plane on screen.

To offer a simple analogy for how blocking and informativity work here: if you do not know what you are having for dinner, an open jar of pasta sauce might be an



**Fig. 2.** The cue structure of sequence-training trials modeled in the simulation. The visual stimuli are present throughout each trial, while the aural stimuli unfold in time (the text representation of this has been included as an illustration). Initially, each noun depiction (in this instance the plane) and each trial context (the gesturing man) are available as cues to the carrier phrase “os ferpal en” (left). The carrier phrase then serves as an additional cue to the article (“bol,” center), which in turn can serve as an additional cue to the noun, “viltord” (right). In the sequence-training trials, these cues will incrementally strengthen and weaken, competing for relevance as training proceeds. However, if the plane has already been learned as a cue to “viltord” prior to sequence learning, this will block the learning of “bol” as a cue to “viltord” (right panel), because “viltord” will be fully anticipated given the presence of the plane.

informative cue to what is cooking, and is thus something you might take note of; on the other hand, if you already know that pasta is on the menu, the same jar of sauce will add little to what you already know, and it is likely that you will pay far less attention to it as a consequence.

An examination of what the models had learned at the end of training confirmed our analysis of the advantages of sequence-first training: In the sequence-first model, the value of the association between the articles and the noun-labels was *five times* greater than that learned by the noun-first model (10.6% versus 2%; see Fig. 3). This was the case, even though the summed total of the predictive association from the articles and noun stimuli to the noun-labels was the same in both conditions (i.e., 100%).

Intriguingly, the results of these simulations point to a possible benefit of learning to associate gendered articles with noun-labels. For speakers of gendered languages who learn associations between articles and nouns, different articles may serve to convey information about which noun is likely to follow, which may in turn facilitate processing. Having a stronger association between the article and the noun will enable the speaker to use this to narrow down the set of items that can follow it, a useful thing given the large number of nouns in language.

Nouns form the richest, most diverse part of speech in many languages, meaning that formally, nouns carry more information than other parts of speech. Indeed, at any point in speech where context suggests a noun is likely to occur (for instance, at the point an article occurs) uncertainty about exactly which element will follow will be at its greatest. In English, one of the most likely places for disfluencies to occur is at the article preceding a noun, and the more complex a given noun is, the more likely it is that disfluencies will occur (Clark & Wasow, 1998). Nouns are also the most common sites for incorrect lexical retrieval and a host of other speech errors (Vigliocco, Antonini, & Garrett, 1997). These processing difficulties have been shown to

correlate with *entropy*, a formal, information-theoretic measure of the uncertainty about which linguistic element can be expected to follow in a given context (McDonald & Shillcock, 2003). Since the value of this uncertainty is a function of the size of the set of possible candidates (Shannon, 1948), anything that serves to *reduce* that set will serve to reduce the absolute level of entropy, and thus ought to facilitate language processing, for both speaker and listener alike.

In gendered languages, the gendered determiners associated with particular sets of nouns may serve precisely this purpose. To give an informal example, in German, almost all alcoholic drinks are in the masculine class, except beer, which is neuter. Hence, in the context of being offered a drink in a bar, simply hearing which article is used will be informative to a listener (see also Kopcke & Zubin, 1984). This idea – that learning a gendered article might facilitate noun processing – is consistent with more general principles of encoding specificity, and the benefits specific contextual cues provide in searches of memory (Tulving & Thompson, 1973).

### 3. Behavioral experiment

Since the modeling results formally supported our analysis of the benefits of learning from larger units in the artificial language, we then assessed people's learning under the same conditions.

#### 3.1. Participants

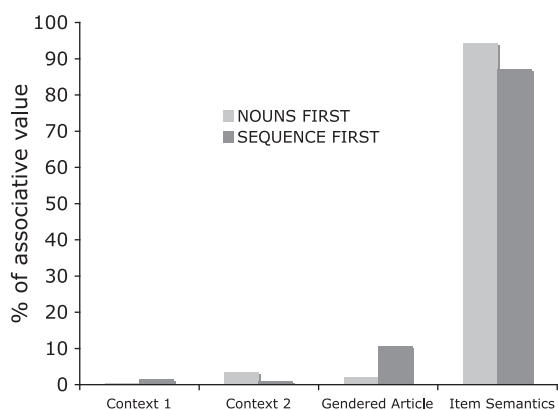
Thirty-two native English-speaking undergraduate students at Stanford University participated (18 females, 14 males, mean age 20 years and three months). None of the participants were speakers of a gender marking language.

#### 3.2. Procedure

The experiment was divided into two phases: learning trials and test trials. Participants were told that they would be tested on the novel language and were asked to repeat the sounds they heard to enhance learning. The experiment lasted 25 min (20 min of training and 5 min of testing). Training and testing sessions took place in a lab room, and were video-taped. Forced-choice responses and reaction times were collected using a response box.

The artificial language was taught to our participants by presenting them with pictures of objects on screen along with an accompanying “description” voiced in the artificial language. In training, participants were exposed to two types of stimuli: noun-labels and full sentences (carrier-phrase + article + noun) that were presented in separate blocks of trials. In noun-label trials, a picture of the named object was presented on screen alone; in full-sentence trials, a picture of the named object was presented on screen along with a picture of a male gesturing to the object. Object size was held constant across all trials.

Participants in the sequence-first condition heard a block of sentences followed by a block of noun-labels, while participants in the label-first condition heard a block



**Fig. 3.** Predictive associations learned from the item semantics, contexts and gendered-article to each noun-label by the two models. (Context 1 = gesturer present; context 2 = gesturer absent; the y-axis shows the percentage of the total predictive value that the model has assigned to each of the potential cues to each noun-label). In the nouns-first condition, the degree to which the initial learning of the object–semantics to noun–label association *blocked* later learning of the article to noun–label association is clearly visible.

of noun-labels followed by a block of sentences. The only difference between the two conditions was the order of the blocks. Following the two learning blocks, participants in both learning conditions were exposed to a distracter block of 35 sentences (accompanied by pictures of the objects). The distracter block was introduced to control for any recency effects in testing, and ensured that the last block before testing was identical in the two learning conditions.

### 3.2.1. Training trials

Stimulus presentation was timed; objects appearing with full-sentences stayed on the screen for 3500 ms and objects appearing with noun-labels stayed on the screen for 2000 ms. Consistent with the simulations, participants in both learning conditions were exposed to the same number of noun-labels (each noun-label was repeated five times, with a total of 70 labels) and full sentences (each noun in a sentence five times, with a total of 70 sentences).

### 3.2.2. Test trials

Test trials were identical in the two learning conditions. Participants completed a forced-choice task and then a production task. In the forced-choice task, participants saw a picture, heard two sentences and had to indicate which sentence was the correct one in the language. They were told that only one sentence was correct.

Half of the forced-choice trials tested knowledge of the article + noun pairing. On these trials, the incorrect sentence had the right noun-label but the wrong article (e.g. participants saw a piano and heard: \**Os-ferpal-en sem slindot* versus *Os-ferpal-en bol slindot*). The other half of the trials tested knowledge of the noun-labels. On these trials, the incorrect sentence had the right article but the wrong noun-label (see piano and hear: \**Os-ferpal-en bol viltord* versus *Os-ferpal-en bol slindot*). Each object was presented once in an article trial and once in a noun trial, yielding 28 forced-choice trials (half testing article + noun pairing and half testing noun knowledge). Order of presentation was randomized for each participant.

In the production task, participants saw a picture and had to produce a full sentence to describe it. They were encouraged to produce full sentences even if they were unsure about all the parts. There were 14 production trials (one for each object). Order of presentation was randomized for each participant. Responses were coded for accuracy by a research assistant blind to the study goals (reliability with coding by author 1 was high,  $\kappa = .95$ ). Nouns and articles were coded as correct if they did not differ from the target in more than one sound (*slipdot* for *slindot*, and *vol* for *bol* were coded as correct). An article + noun sequence was coded as correct only if both the article and the noun were correctly produced. The carrier-phrase was coded for accuracy on a scale from 1 to 3 (1-fully accurate, 2-partially accurate, 3-not accurate).

### 3.3. Results

As predicted, participants in the sequence-first condition showed better learning of the article + noun pairing. A mixed-effect regression model with trial type and learn-

ing condition as fixed effects, and subject and item as random effects, revealed a main effect of learning condition that was not qualified by a significant interaction: participants in the sequence-first condition were more accurate overall (80% vs. 71% correct,  $B = .44$  (SE = .21),  $p < .05$ ). They were better at selecting both the correct article (61% vs. 54%) and the correct noun-label (98% vs. 92%). Not surprisingly, given the difficulty of grammatical gender, both sets of participants selected the correct noun-label more often than they selected the correct article (95% vs. 57.5% correct,  $B = 2.72$  (SE = .26)  $p < .001$ ; see Fig. 4). However, post hoc tests revealed that while participants in the sequence-first condition were significantly above chance (61%) in choosing the correct article  $t(15) = 3.55$ ,  $p = .003$ , participants in the noun-label-first condition were not  $t(15) = .81$ ,  $p > .4$ .

Participants in the sequence-first condition also made correct responses faster than participants in the noun-label-first condition (990 ms. compared to 1188 ms.). This was true when selecting the correct determiner (1194 vs. 1436) and when selecting the correct noun (854 vs. 1032). The effect of learning condition on reaction times was significant across items,  $F_2(1, 13) = 6.57$ ,  $p = .01$ , but not across subjects,  $F_1(1, 30) = 1.63$ ,  $p = .2$ .

The overall accuracy rates in the production task were not high, which is not entirely surprising given the short exposure time (20 min) and the number of noun-labels taught (14). Despite this, the production results showed the same pattern as for comprehension (Fig. 5). Participants in the sequence-first condition were more likely to produce a correct article + noun sequence (40% of the time) than were participants in the label-first condition (24% of the time),  $B = .85$  (SE = .33),  $p < .05$ . Again, this is consistent with our hypothesis that the gender of determiners is informative about nouns, and that this information serves to benefit processing (McDonald & Shillcock, 2003).

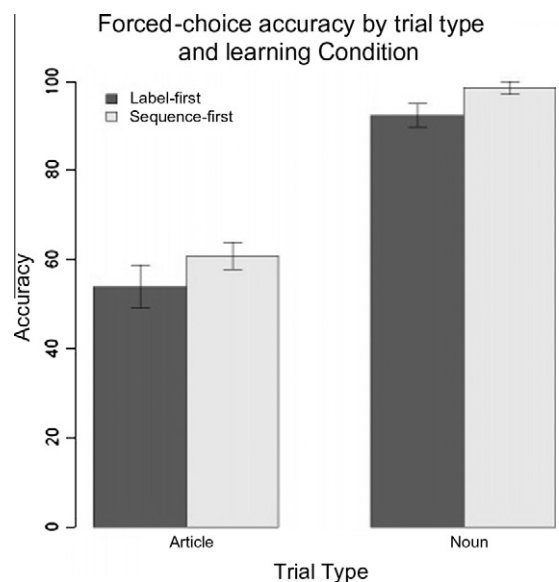


Fig. 4. Forced choice accuracy by test type and training condition.



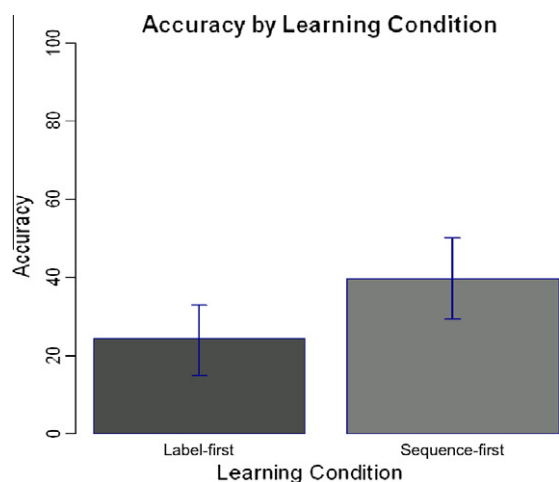


Fig. 5. Production accuracy for the article–noun pairs by learning condition with 95% confidence intervals.

Finally, consistent with the results of the simulations, both groups learned to produce the carrier-phrase with the same degree of proficiency,  $t(30) = -1.08$ ,  $p > .2$ . Thus, as expected, when the structure of the information available to learners was held constant – as was the case for the carrier phrase – no differences in learning performance were detectable between the groups. However, when the structure of information was manipulated – by encouraging blocking – participants in the sequence-first condition learned the association between the articles and the nouns better than participants in the noun-label-first condition.

#### 4. Discussion

Our artificial grammatical gender system was learned better when both our human participants and the learning models started with “less segmented” input. Participants in the sequence-first condition were both more likely to choose the sentence with the correct article in a forced-choice task and more likely to produce the appropriate article for a given noun in a production task. This occurred even though participants in both conditions had been exposed to exactly the same training items exactly the same number of times.

One of our most striking results was that the sequence-first condition not only allowed participants to better learn the gendered articles, it also facilitated learning of the *noun-labels*. This finding is consistent with our suggestion that learning the gender of articles could benefit language processing by reducing uncertainty about upcoming nouns. This finding is intriguing, because it has often been argued that grammatical gender systems serve no useful linguistic function at all (Twain, 1880; see Kilarski, 2007, for a review). Indeed, the supposed pointlessness of grammatical gender, taken together with the difficulties that L2 learners have with trying to master these systems, prompted the developmental psychologist Michael Maratsos (1979) to say of the German gender system:

“The presence of such systems in a human cognitive system constitutes by itself excellent testimony to the occasional nonsensibility of the species. Not only was this system devised by humans, but generation after generation of children peaceably relearns it.”

Nouns are by far the most diverse – and hence most unpredictable – part of speech in most languages, and they are thus a very likely location for processing errors (McDonald & Shillcock, 2003). It makes sense, therefore, that by making nouns more predictable, gendered articles ought to help ameliorate this problem, and thus facilitate the processing of nouns (Dahan et al., 2000; Lew-Williams & Fernald, 2009; see also Frigo & McDonald, 1998). This would suggest that grammatical gender might have a more functional role in language processing than has often been supposed (Kilarski, 2007).<sup>3</sup>

Given that the results of both our experiment, and a growing body of evidence from the processing of gendered languages, suggests that grammatical gender has a functional part to play in language processing, this raises the question of why adults fail to master this functionality in the way children do. While we did not test children’s language learning (we have not shown that children start from larger units, but see Bannard & Matthews, 2008), we noted many reasons to believe that adults are more likely than children to start with smaller linguistic units, and to focus on noun-labels in learning. Accordingly, the poor performance of our label-first participants may offer some explanation for why adults struggle to learn grammatical gender, and for why the representations adults *do* learn are shallow and hard to access in real time (Clahsen & Felser, 2006). In short, starting from noun-labels should hinder learning about the relation between articles and nouns because learning segments individually comes at the cost of blocking later learning about the relations between segments.

When noun-labels and articles co-occur in speech, our simulations suggest that child learners will only gradually discriminate them from one another, as their experience with language and the mapping between form and meaning develops over time. This fits nicely with usage-based models of language, which posit that grammatical relations emerge from a gradual process of abstraction over stored utterances – including multi-word ones (Bod, 2009; Bybee, 1998; Tomasello, 2003). Much like a second language learner, infants cannot immediately discriminate word boundaries in speech. But unlike that learner, they do not even know those boundaries exist, which will lead them to initially use larger, less well-discriminated sequences of language in learning.

In keeping with this suggestion, there is evidence that infants are initially sensitive to the acoustic correlates of major prosodic units (namely clauses) early on, but take longer to detect smaller prosodic units like phrases or

<sup>3</sup> If gender systems have functional benefits for noun processing, this raises the question of whether the *loss* of a gender system (as occurred in the development of Modern English; Curzan, 2003) would in turn make the nouns in that language more difficult to process compared to those of a gendered language, or whether the role previously played by gender might be taken up by some other part of the grammar.

words in running speech (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). By two months, they remember speech better when it is packaged in a well-formed prosodic clause and are capable of remembering multi-word utterances (Mandel, Jusczyk, & Kemler Nelson, 1994; Mandel, Kemler Nelson, & Jusczyk, 1996). We can also find multi-word units in their early productions; e.g., young children produce under-segmented utterances like ‘give-it the ball’ where give-it is treated as a single unit (Peters, 1983). Older children also attend to sequences (Bannard & Matthews, 2008): they are better at repeating higher frequency four-word sequences, even when the frequency of individual words is matched (e.g., ‘a drink of milk’ compared to ‘a drink of tea’), indicating they have memory traces of such units. Finally, production of irregular plurals is facilitated in familiar sequences (e.g., *teeth* in *brush your teeth*; Arnon & Clark, 2011), suggesting that children utilize larger, less discriminated units in production. Indeed, as our simulations suggest, children’s learning of “grammatical relations” (like those between articles and nouns) may well result from the process of “analyzing” and segmenting larger sequences in the course of discrimination learning.

Whereas children may have no choice other than to gradually learn to discriminate the various elements of languages, adult L2 learners will have prior experience of another language, and this should enable them to focus on what they perceive to be the “meaning-carrying” units (such as the noun-labels) of a new language. To the extent that they do this, it may take away from their ability to learn the kind of relationship between articles and nouns that child learners acquire. This tendency may also be exaggerated if a learner’s first language does not have gender categories, leading to an effect of L1 on L2 learning of grammatical gender (Sabourin et al., 2006). Thus although in natural language learning situations, the difference between adult and child learning may not be as straightforward as in our experiment, because learning – and therefore, blocking – are probabilistic, it follows that to the extent that the distribution of linguistic experience for L2 learners and children differs, we would expect to see corresponding differences in what they learn.

While this account of the problems facing adult learners differs in certain respects from previous suggestions about the differences in adult and infant learning capacity (Elman, 1993; Newport, 1990), it is consistent with an important insight in those proposals: namely, that adults learn from the input in ways that are different from children, and these differences in how they learn ultimately affect the nature of *what* they learn (see also Hudson Kam & Newport, 2009; Singleton & Newport, 2004). Traditionally, the problem of L2 learning has been seen as a task of explaining what it is that children lose (see e.g., Lenenberg, 1967). However, our results suggest another perspective: that L2 learning deficits are a by-product of adult gains in knowledge and cognitive capacities, and not the opposite. The trouble with adult learners may be that they are just a little too ‘clever’ for their own good (Ramscar & Gitcho, 2007; Thompson-Schill et al., 2009).

Thus, for example, adult learners have prior knowledge about the kinds of cues that are informative in their first

languages, which may help when cues overlap between languages, but not when they fail to overlap. Indeed, where information is encoded differently across languages, learning of L2 cues may be blocked by previously learning of L1 cues. For instance, the temporal positioning of events can be marked in English in multiple ways, among them the use of verbal morphology (*walked*) and temporal adverbs (*yesterday*). Chinese, by contrast, doesn’t make use of verbal morphology to mark tense, but uses temporal adverbials instead. In terms of cue competition, native speakers of Chinese have learned to use temporal adverbials – but not verbal morphology – as a cue to temporal positioning. As a result, they have a harder time (compared to English speakers), when learning verbal cues for temporal positioning in an artificial language (Ellis & Sagarra, 2010).

#### 4.1. The role of semantics in our models of learning

The points above are particularly relevant for understanding why the results of our training experiment differ in subtle but important ways from the many findings that have shown how learners are able to detect and learn simple co-occurrence information in artificial language learning (e.g., Saffran, Aslin, & Newport, 1996). Participants in our experiments did not learn the relationship between the articles and the nouns equally well in both conditions—even though they had access to the same overall sequential co-occurrence information in the strings in our artificial language—because unlike many other studies of statistical learning (especially those concerned with word segmentation, e.g., Saffran et al., 1996), the current study had a semantic as well as a sequential component: learners were exposed to pairings of sounds *and* objects, rather than just sounds alone.

Using the artificial language to talk about ‘the world’ altered the distributional structure of the input, introducing co-occurrence patterns between linguistic units and objects, as well as between the linguistic units themselves. This is important because the nature of phonemic perception means that when adults learn patterns in linear strings of phonemes (i.e., in a standard artificial grammar learning paradigm), the way they learn is necessarily non-competitive (Ramscar et al., 2010). The value of a phoneme as a cue to another phoneme in a sequence will increase when the expected phoneme follows it, and decrease when the expected phoneme fails to appear. Because competition between cues becomes attenuated when the temporal relations between them vary (Amundson & Miller, 2008), when phonemes appear sequentially, there is no cue competition (Ramscar et al., 2010): the value lost by a phonemic cue after an error cannot be appropriated by a competing cue, because in a sequence of phonemes, there are no competing cues (Ramscar et al., 2010). In the absence of cue competition, the cue value of a phoneme in an artificial grammar will simply come to represent the proportion of successful predictions it has made relative to the proportion of unsuccessful predictions. Accordingly, its value will track the frequency with which phonemes co-occur, approximating the conditional probability of one phoneme given another (Ramscar et al., 2010; see also Cheng, 1997; Wasserman, Elek, Chatlosh, & Baker, 1993),

and this is why participants acquire a good understanding of the transitional probabilities between phonemes in the training sequences in such tasks (see Saffran, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Saffran et al., 1996).

Because learning in an artificial grammar task where there is only one available cue (e.g., transitional probabilities) is not competitive, introducing a semantic component qualitatively changes the learning environment. The structure of information in the task will now encourage competitive learning (which is inhibited by the information structure in a standard artificial grammar learning paradigm; Ramscar et al., 2010). When cues compete for predictive value, what people learn about them changes. Competition results in a situation where the learned value of a cue is determined not only by the statistical structure of the environment, but also by the structure of a learner's knowledge about the environment. The value of a cue is thus determined relative to the system of cues a learner has already acquired to help them understand the environment, rather than independently (Ramscar et al., 2010). It is this 'systematic' nature of competitive learning that gives rise to effects such as blocking (Kamin, 1969; Rescorla & Wagner, 1972).

This point is best illustrated computationally, by reference to our simulations. If semantics are eliminated in the sequence-first condition, the learned value of an article → noun-label association is 100%, as one might expect in a traditional artificial grammar learning task. When semantics are introduced, however, the learned value of this association is reduced by competition to 10.6%, and when training on noun-labels alone precedes sequence learning, it reduces to a mere 2%. Given an appropriate information structure, it is entirely conceivable that it could be blocked altogether. As this illustrates, the way in which learning is structured has a considerable impact on what gets learned.

Interestingly, it seems that the emphasis in cognitive science on simple non-competitive learning tasks such as those studied in "artificial grammar learning" – in which participants learn only the transitional probabilities between individual phoneme cues and subsequent individual phonemic events – has inadvertently contributed to the widespread misconception of learning as being limited to simple probability learning (see also Rescorla, 1988). While participants can and will learn transitional probabilities when that is the only information available to them, when semantic and contextual information are included in the input distribution, they will learn to value the cues most relevant to the task at hand – namely, the semantics (Ramscar et al., 2010). Our findings suggest that once the semantic dimension of language is taken into account, the way linguistic information is learned, and *what* gets learned as a result, can shift dramatically.

To return to natural language, while it might be (very fairly) objected that the kind of blocked design we used to train adults here is not at all similar to the environments in which children actually acquire the ability to use language, what is encouraging is that once we constrained the structure and sequencing of the information available to our adult participants, they learned exactly as learning and information theory predicted they should have done.

This adds some support to the idea that in the richer, far more statistically varied environments that children encounter language, their learning is driven by the same principles. This is gratifying because, to the extent that we assume that the mind is a computational device (and there seems to be no better alternative conception), and to the extent that learning and information theory represent our best efforts to describe the laws of computation, having cognitive theories that are compatible with learning and information theory ought to be the goal of cognitive science.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to Michael Ramscar. We are grateful to Melody Dye, Eve Clark, Daniel Jurafsky and Daniel Yarlett for discussion of these ideas.

## Appendix A. Items used in the experiments

### A.1. Articles

1. Sem
2. Bol

### A.2. Nouns

Word	Meaning
Etkot	key
Fertsot	sock
Geesoo	hat
Gorok	bike
Hekloo	bath
Hertin	iron
Jatree	pan
Panjol	television
Perdip	house
Pikroo	car
Slindot	piano
Sodap	spoon
Toonbot	clock
Viltord	plane

## References

- Amundson, J. C., & Miller, R. R. (2008). CS-US temporal relations in blocking. *Learning and Behavior*, 36(2), 92–103.
- Arnon, I., & Clark, E. V. (2011). When 'on your feet' is better than 'feet': Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7, 107–129.
- Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, 5, 436–461.
- Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11, 295–328.
- Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438–482.
- Bannard, C., & Matthews, D. E. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19, 241–248.
- Bannard, C., Lieven, E., & Tomasello, T. (2009). Modeling Children's Early Grammatical Knowledge. *Proceedings of the National Academy of Sciences*, 106, 41 17284–17289.

- Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, 59, 755–795.
- Bassano, D., Maillolochon, I., & Mottet, Y. (2008). Noun grammaticalization and determiner use in French children's speech: A gradual development with prosodic and lexical influences. *Journal of Child Language*, 35, 403–438.
- Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*, 33, 752–793.
- Bybee, Joan L. (1998). The emergent lexicon. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Proceedings of the Chicago Linguistic Society 34: The Panels* (pp. 421–435). Chicago: Chicago Linguistic Society.
- Carroll, J. B. (1939). Determining and numerating adjectives in children's speech. *Child Development*, 10, 215–229.
- Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39, 535–594.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367–405.
- Chevrot, J. P., Dugua, C., & Fayol, M. (2008). Liaison acquisition, word segmentation and construction in French: A usage-based account. *Journal of Child Language*, 1, 40.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42.
- Clahsen, H., & Muysken, P. (1986). The availability of universal grammar to adult and child learners – A study of the acquisition of German word order. *Second Language Research*, 2(2), 93–119.
- St. Clair, M., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, 33, 1317–1329.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201–242.
- Colunga, E., Smith, L. B., & Gasser, M. (2009). Correlation and prediction in children's word learning: Cross-linguistic evidence and simulations. *Language and Cognition*, 1(2), 197–217.
- Corbett, G. G. (1991). *Gender*. Cambridge, England: Cambridge University Press.
- Curzan, A. (2003). *Gender Shifts in the History of English*. Cambridge: Cambridge University Press.
- Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42(4), 465–480.
- Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, 47, 109–121.
- Daw, N., & Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Social Cognition*, 26(5), 593–620.
- Dewaele, J. M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Bilingualism: Language and Cognition*, 4, 275–297.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.
- Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
- Dayan, P., & Daw, N. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive Affective and Behavioral Neuroscience*, 8, 429–453.
- Dugua, C., Spinell, E., Chevrot, J.-P., & Fayol, M. (2009). Usage-based account of the acquisition of liaison: Evidence from sensitivity to the singular/plural orientation of nouns. *Journal of Experimental Child Psychology*, 102, 342–350.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36, 353–371.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Ellis, N. C., & Sagarra, N. (2010). The Bounds of Adult Language Acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, 32, 552–580.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501.
- Fillmore, C. S. (1979). On fluency. In J. Fillmore, D. Kempler, & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). London: Academic Press.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299, 1898–1902.
- Fisher, C., & Tokura, H. (1996). Acoustic Cues to Grammatical Structure in Infant-Directed Speech: Cross-Linguistic Evidence. *Child Development*, 67, 3192–3218.
- Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39, 218–245.
- Gallistel, C. R. (2003). Conditioning from an information processing perspective. *Behavioural Processes*, 62, 89–101.
- Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review*, 107, 289–344.
- Gluck, M. A., & Bower, G. H. (1988a). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117(3), 227–247.
- Gluck, M. A., & Bower, G. H. (1988b). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166–195.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology* (pp. 145–160). Oxford: Clarendon Press.
- Grosjean, F., Dommergues, J. Y., Cornu, E., Guillemon, D., & Besson, C. (1994). The gender marking effect in spoken word recognition. *Perception & Psychophysics*, 56, 590–598.
- Guillemon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, 29, 503–511.
- Gureckis, T. M., & Love, B. C. (2010). Direct associations or internal transformations? Exploring the mechanisms underlying sequential learning behavior. *Cognitive Science*, 34, 10–50.
- Harley, B. (1979). French gender 'rules' in the speech of English-dominant, French-dominant and monolingual French-speaking children. *Working Papers in Bilingualism*, 19, 129–156.
- Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of universal grammar in second language acquisition. *Second Language Research*, 13, 187–226.
- Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269–286.
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1, 304–309.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.
- Holmes, V. M., & Dejean de la Batie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics*, 20, 479–506.
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59, 30–66.
- van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language*, 28, 87–100.
- Van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, 12, 419–425.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G. K., Kennedy, L. J., Woodward, A. L., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behaviour*. New York: Appleton-Century-Crofts.
- Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language*. Cambridge, England: Cambridge University Press.
- Kilarski, M. (2007). On grammatical gender as an arbitrary and redundant category. In Douglas, Kilbee (Ed.), *History of linguistics 2005: Selected papers from the 10th international conference on the history of language sciences (ICHOLS X)* (pp. 24–36). Amsterdam: John Benjamins.
- Kopcke, K.-M., & Zubin, D. A. (1984). Sechs Prinzipien für die Genuzuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, 93, 26–50.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, 97(22), 11850–11857.
- Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.
- Lew-Williams, C., & Fernald, A. (2007a). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18, 193–198.

- Lew-Williams, C., & Fernald, A. (2007b). How first and second language learners use predictive cues in online sentence interpretation in Spanish and English. In H. Caunt-Nulton, S. Kulatilake, & I. Woo (Eds.), *Proceedings of the 31st annual Boston university conference on language development* (Vol. 2, pp. 382–393). Somerville, MA: Cascadilla Press.
- Lew-Williams, C., & Fernald, A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ? In J. Chandlee, M. Franchini, S. Lord, & G. Rheiner (Eds.), *Proceedings of the 33rd annual Boston university conference on language development* (Vol. 1, pp. 290–301). Somerville, MA: Cascadilla Press.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447–464.
- Lieven, E., Pine, J., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–220.
- Lieven, E. V. M., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30, 2.
- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20, 481–507.
- MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, 43(1/2), 1–123.
- Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants to organize and remember speech information? *Cognition*, 53, 155–180.
- Mandel, D. R., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, 11, 181–196.
- Maratsos, M. P. (1979). How to get from words to sentences. In D. Aaronson & R. Rieber (Eds.), *Perspectives in psycholinguistics*. Hillsdale, NJ: Erlbaum.
- Marinova-Todd, S. H. (1994). The critical period in second language acquisition: The case of gender. Unpublished thesis, York University.
- Mariscal, S. (2008). Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language*, 35, 1–29.
- McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science*, 14, 648–652.
- McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, 28, 211–246.
- Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, 431, 760–767.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21, 81–108.
- Neville, H. J., & Bavelier, D. (2001). Variability of developmental plasticity. In J. McClelland & R. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Newport, E. L. (1990). Maturation constraints on language learning. *Cognitive Science*, 14, 11–28.
- Pawley, A., & Syder, F. H. (1980). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. Schmidt (Eds.), *Communicative competence* (pp. 191–225). London: Longmans.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pine, J. M., & Lieven, E. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18, 123–138.
- Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Science*, 11(7), 274–279.
- Ramscar, M., Dye, M., & Yarlett, D. (2009). No representation without taxation: The costs and benefits of learning to conceptualize the environment. In *Proceedings of the Analogy 09 Conference, Sofia, Bulgaria*.
- Ramscar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The Enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE*, 6, e22501. doi:10.1371/journal.pone.0022501.
- Ramscar, M., Dye, M., Witten, J., & Klein, J. (2012). Dual routes to cognitive flexibility: learning and response conflict resolution in the dimensional change card sort task. *Child Development*.
- Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31, 927–960.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The feature-label-order effect in symbolic learning. *Cognitive Science*.
- Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative & Physiological Psychology*, 66, 1–5.
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151–160.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rogers, M. (1987). Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics*, 8, 48–74.
- Rosenblatt, F. (1959). Two theorems of statistical separability in the Perceptron. In V. Blake & A. M., Uttley (Eds.), *Mechanisation of thought processes: Proceedings of a symposium held at the National Physical Laboratory* (pp. 419–456). London: Her Majesty's Stationery Office.
- Rudy, J. W. (1974). Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning & Verbal Behavior*, 13, 282–296.
- Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1–29.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Scherag, A., Demuth, L., Roesler, F., Neville, H. J., & Roeder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93, B97–B108.
- Schultz, W., & Dickinson, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience*, 23, 473–500.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.
- Schultz, W., Dayan, P., & Montague, R.-R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423. 623–656.
- Siegel, S. G., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review*, 3(3), 314–321.
- Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American sign language from inconsistent input. *Cognitive Psychology*, 49, 370–407.
- Slobin, D. I. (1985). *The crosslinguistic study of language acquisition* (Vol. 1). The Data Erlbaum.
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Cued attention and children's novel noun generalizations. *Cognitive Science*, 34, 1287–1314.
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49, 249–267.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods*, 68, 598–605.
- Thompson-Schill, S., Ramscar, M., & Chryssikou, M. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, 8(5), 259–263.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352–373.
- Twain, M. (1880). The awful German language (Appendix D). In M. Twain (Ed.), *A tramp abroad* (pp. 390–402). Hartford, CT: American Publishing.
- Vanlancker-Sidits, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics*, 24, 45–57.
- Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8(4), 314–317.

- Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*, 43–48.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: The role of probability in judgments of response–outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 174–188.
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866–1878.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 249–268). Amsterdam: John Benjamins.
- Wray, A. (2008). The puzzle of language learning: From child's play to 'linguaphobia'. *Language Teaching*, *41*, 253–271.
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the life-span* (pp. 55–72). Cambridge: Cambridge University Press.