A Hebbian account of entrenchment and (over)-extension in language learning

Vsevolod Kapatsinski (vkapatsi@uoregon.edu)
Zara Harmon (zforough@uoregon.edu)
Department of Linguistics, 1290 University of Oregon
Eugene, OR 97403 USA

Abstract

In production, frequently used words are preferentially extended to new, though related meanings. In comprehension, frequent exposure to a word instead makes the learner confident that all of the word's legitimate uses have been experienced, resulting in an entrenched form-meaning mapping between the word and its experienced meaning(s). This results in a perception-production dissociation, where the forms speakers are most likely to map onto a novel meaning are precisely the forms that they believe can never be used that way. At first glance, this result challenges the idea of bidirectional form-meaning mappings, assumed by all current approaches to linguistic theory. In this paper, we show that bidirectional form-meaning mappings are not in fact challenged by this production-perception dissociation. We show that the production-perception dissociation is expected even if learners of the lexicon acquire simple symmetrical form-meaning associations through simple Hebbian learning.

Keywords: Hebbian learning; word learning; mental lexicon

Introduction

Extension of frequent forms to novel uses is one of the most common processes in language change, and results in the robust correlation between frequency and polysemy: upon examining a dictionary, it quickly becomes evident that it is the most frequent words that have the largest number of uses (compare *get* vs. *obtain*, Piantadosi et al., 2012; Zipf, 1949). Extension of familiar words and constructions to new uses is also one of the major mechanisms driving grammaticalization, a largely unidirectional process through which grammatical morphemes evolve out of lexical sources (Bybee, 2003, 2010). A well-studied example in English is the verb *will*, which was gradually extended from volitional lexical uses (e.g., *I will it to happen*) to grammatical future tense uses that no longer imply volition (e.g., *I will get fired for suggesting this*).

Extension can be observed not only in diachrony but also in online language use. In particular, novel extensions are frequently observed in children's use of both referential terms and verb-argument structure constructions. For example, a child may name a cow a kitty or extend the verb *giggle* to transitive use, as in *don't giggle me* (e.g. Naigles & Gelman, 1995; Pinker 1989). The words that are so overextended tend to be the frequent ones, or else ones that are highly accessible in the moment because they have just been used (Gershkoff-Stowe et al., 2006; for adults, see also Ferreira & Griffin, 2003; Burke et al., 2004). These patterns parallel the diachronic tendency for frequent words to acquire novel uses.

Crucially, a child can overextend a word in production without overextending it in comprehension. When presented with a word she over-extends in production and asked to pick out all the objects the word can refer to, the child often does not select the objects to which she over-extends the word in production as its possible referents (Naigles & Gelman, 1995).

In fact, frequency appears to have opposite effects in comprehension and production. Whereas frequent words are extended to new uses in production, frequent words are likely to be restricted to the uses in which they have been experienced. For example, Xu & Tenenbaum (2007) show that experiencing *fep* paired with a Dalmatian once leads children to think it plausible that *fep* refers to all dogs, but three *fep*-Dalmatian pairings are enough to restrict the set of referents to Dalmatians (see also Ambridge et al., 2008; Brooks & Tomasello, 1999; Theakston, 2004; Wonnacott et al., 2008, for related results with syntactic constructions). Frequent exposure to a form-meaning pairing appears to convince learners that the form always co-occurs with this meaning.

The Data

In recent work, we have confirmed the existence of this dissociation in adult learners of a miniature artificial language (Harmon & Kapatsinski, submitted; Experiment 1). In our study, participants were exposed to a language with two plural suffixes (-*dan* and -*sil*) and two diminutive suffixes (-*nem* and -*shoon*). For each participant, one suffix was more frequent than others. Each participant was tested on both comprehension and production.

Participants experienced the language through passive exposure, with nouns bearing the suffixes (e.g. *ostodan*, *zutishoon*) presented auditorily, paired with pictures of their referents. Each trial began with a picture of the referent, followed 500 ms later by the spoken word. After the offset of the spoken word, the experiment advanced to the next trial, which began 400 ms later.

Nouns bearing plural suffixes were paired with pictures of multiple large creatures (with the kind of creature determined by the stem), whereas each noun bearing a diminutive suffix was paired with pictures of a single small creature.

For half of the participants (n = 35), those in the Dan condition, the form -*dan* was more frequent than the others. For the other half, assigned to the Nem condition (n = 35), the frequent form was -*nem*. The competing -*sil* and -*shoon*

forms were always equally frequent. The unsuffixed stems constituted the singular non-diminutive form of the noun.

After exposure, participants were tested on both production and comprehension. In the production test, were presented with meanings and asked to express them. Crucially, one of the meanings was a novel one, plural diminutive (multiple small creatures). Each trial began with the presentation of the picture of a novel singular object on the computer screen. The name of the novel object was presented auditorily over headphones as in the training stage. Once the sound finished playing, the picture was removed and replaced with a display of four pictures representing four different meanings: a single object of the same type, a miniature version of the same object, multiple objects of the same type, and multiple miniature objects of the same type. Three of these pictures disappeared, leaving the participants with the one target picture to name (i.e., meaning to express). Participants were asked to generate the form for the target meaning using the stem that was presented and say the form aloud. They had five seconds to do so.

Data were analyzed using logistic mixed-effects models with maximal random-effects structure using the lme4 package (Bates et al., 2015) in R (R Core Team, 2015). Significance was assessed by comparing models with and without a predictor using a log likelihood test. Participants were significantly more likely to use a given form if it was the frequent form during exposure (Figure 1; χ^2 (1) = 21, $p <$ .0001). This was not simply an effect of semantic feature frequency, i.e. DIM.PL in Dan, where PL was frequent, was not simply interpreted as PL: the synonym of a frequent form (-*sil* in Dan and -*shoon* in Nem) was no more likely to be used to express the novel meaning than the synonym of the infrequent form (-*sil* in Nem and -*shoon* in Dan; p = .9).
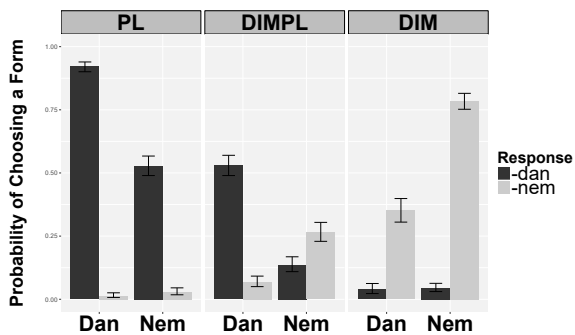

Figure 1: An illustration and results of the production test. A suffix is produced more often when it has been experienced more frequently (-*dan* in Dan and -*nem* in Nem), both to express the meaning with which it was experienced and to express the novel related meaning (DIMPL)

In the comprehension test, participants were presented with forms and asked to click on the corresponding meaning using the four-picture display briefly flashed in production. The meanings included the familiar meanings as well as the novel meaning. In this task, participants were less likely to click on the novel meaning given a form that was frequent during exposure (Figure 2; χ^2(1)= 17, p = .000037). As in production, these effects could not be accounted for by the relative frequencies of the meanings because the synonym of a frequent form was significantly more likely to be mapped onto the novel meaning than the synonym of the infrequent form (p <.001). Thus, participants are not simply more likely to click on the more familiar *meanings*, rather they are more likely to click on familiar meanings in response to the forms that have been frequently paired with them in training. For forms that have been paired with the frequent meaning less frequently, the novel meaning is preferred, despite the fact that these forms are as frequent as synonyms of infrequent forms.
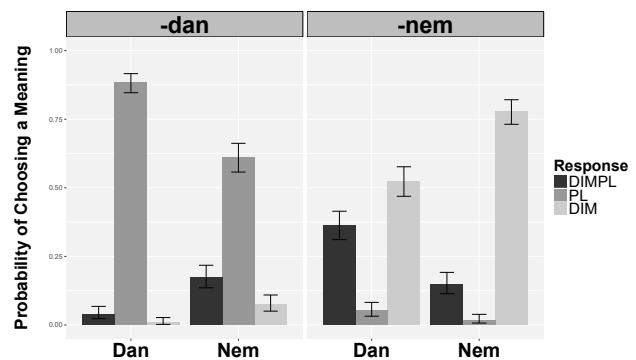

Figure 2: An illustration of the comprehension test and the corresponding results from Harmon & Kapatsinski (submitted). Responding 'DIMPL' meant clicking on multiple small creatures; 'PL' meant clicking on multiple large creatures, and 'DIM' meant clicking on a single small creature. When a suffix occurred frequently in training (-*dan* in Dan Condition and -*nem* in Nem Condition), it became less likely to be mapped onto the novel meaning, DIMPL, and more likely to be mapped onto the meaning it was paired with during exposure.

The results therefore show a production-comprehension dissociation: the forms participants were most likely to use to refer to the novel meaning in production were the forms they were least likely to map onto the novel meaning in comprehension.

Thus, frequency of a form-meaning pairing appears to have opposite effects in production and comprehension. These results therefore appear, at first glance, to be problematic for simple Hebbian models of word learning (McMurray et al., 2012; Yu & Smith, 2012) that learn symmetrical bidirectional form-meaning mappings based on form-meaning co-occurrence as well as for the notion, nearly universally accepted in linguistics (cf. Ramscar et al., 2010), that linguistic contructions, whether lexical or grammatical, are Saussurean signs – i.e., that there is a form representation that mediates the auditory-to-semantic mapping in comprehension and the semantic-to-articulatory mapping in proeuction. The aim of the present paper is to

show that, somewhat counterintuitively, the observed dissociation actually falls out of simple Hebbian learning of bidirectional form-meaning associations.

The baseline model: Frequency counter

According to Hebb (1949), neurons that fire together wire together. We assume a distinction between cues and outcomes, where outcomes follow cues. On every trial, associations between the cues present on that trial and the following outcomes strengthen. How much they strengthen is determined by the salience of the cue, the salience of the outcome, and the learning rate. During the exposure trials in Harmon & Kapatsinski (submitted), forms began 500 ms into the presentation of the referent. Therefore, we assume cues to be the semantic features of the referents (BIG, SMALL, MANY and ONE) plus a context cue, present on every trial (Pavlov, 1927; Rescorla & Wagner, 1972). This order of presentation was chosen to reflect the temporal dynamics of real-life word learning (Pereira et al., 2014). Unlike error-driven models such as Rescorla & Wagner (1972), we did not multiply the increment in association strength by prediction error. This is part of what makes the model Hebbian: it does not learn less on trials with unsurprising (or no-longer-surprising) outcomes, and would not exhibit cue competition effects such as blocking or overshadowing.

In essence, this base model is simply counting frequencies of form-meaning mappings. When it encounters a cue (meaning) followed by an outcome (form), it simply increases the weight of the link between them by a constant number, which we set to 1 in order to emphasize the model's nature as a simple frequency counter. The results do not change depending on what the number is.

Linking hypotheses

In order to connect the model's knowledge to the experimental results, we need a set of linking hypotheses connecting the weights and activations of the model to the participants' responses in the experimental tasks. We assumed that production involves activating forms given the semantic features present on that test trial and the context cue. The activation of a form is simply the sum of connection strengths from the semantic and context cues present on the test trial to that form. The choice of the form is then determined stochastically (Luce, 1963): the form is chosen in proportion to its activation value relative to the sum of all forms' activation values given the cues present. Stochastic choice implements probability matching, an empirical universal in tasks that demand repeatedly choosing between the same alternatives (Azab et al., 2016).

The linking hypothesis for comprehension is more controversial. Note that the model, like the subjects, was trained only in the meaning→form direction. However, the comprehension task required participants to choose meanings given forms, reversing the cue→outcome mappings they were trained on. Participants were extremely accurate in the comprehension task, suggesting that they were able to bring the knowledge they acquired to the task. The model must be able to do the same. We propose that the associations participants learn obey the Symmetry Principle: a cue→outcome association is as strong as the corresponding outcome→cue association (Asch & Ebenholtz, 1962; Kahana, 2002). This is another way in which the proposed model differs from models that perform using prediction error, such as Rescorla & Wagner (1972). This difference, however, is crucial for the model's ability to simulate the comprehension data.

We assume that a choice between two meanings depends on the difference in activations between the two meanings' contrasting features. For example, the probability of clicking on [small;plural] rather than [big;plural] when presented with -*dan* is proportional to the difference in association strengths between –*dan*~SMALL (=SMALL ~ -*dan*) and -*dan* ~ BIG (=BIG ~ -*dan*). The bigger this difference, the more likely participants are to click on the meaning that actually *was* paired with the form in training (Miller & Matzel, 1988).

Models between frequency and contingency

Besides the connections between the cues and outcomes present on a particular trial, there are three other sets of connections that could potentially be updated. Alternative theories of associative learning differ in their claims about whether these connections are indeed updated.

First, there are connections from the cues present on a trial to the outcomes absent from that trial. It is usually thought that these connections' weights are reduced, so that cues that are consistently paired with the *absence* of a certain outcome develop inhibitory connections to that outcome, with the subject learning the negative contingency present in the environment. Second, there are connections from the absent cues to the present outcomes. These connections are assumed *not* to be updated by Rescorla & Wagner (1972). However, van Hamme & Wasserman (1994) and Tassoni (1995) argued that – if participants know the set of cues that *could* occur on every trial – the absence of a cue could be salient. In other words, learners may notice the consistent absence of a cue on trials containing a certain outcome and develop a negative association between that cue and the outcome. Finally, one could argue that connections from absent cues to absent outcomes may also be updated, gaining strength: when a cue and an outcome are *absent together,* the learner is in a position to learn that absence of the cue predicts absence of the outcome (Tassoni, 1995). Thus, models of learning can be arranged from simplest (wiring together present cues and outcomes only) and least veridical – least able to faithfully reproduce environmental contingencies – to most complex and most veridical (updating all connections on every trial). In what follows, we examine the types of updating that are needed to capture the experimental results by independently varying whether each distinct set of connections undergoes updating. Table 1 summarizes the possible models from a simple frequency counter that updates only the connections

between present cues and present outcomes to a fully veridical contingency tracker that updates all four sets of connections (in the normative direction). We will refer to the models we evaluate with the abbreviations shown on the left sides of the table cells. For example, the Rescorla-Wagner model updates only the sets of connections in the top row and can therefore be abbreviated as (p_c).

Table 1: The four distinct sets of cue-outcome connections on every trial and whether their weights should become more positive (+) or more negative (-) in a model that is able to capture environmental contingencies veridically. The two subscripts c and o refer to cue and outcome respectively. Presence is denoted by p and absence by a.

	Outcome Present		Outcome Absent	
Cue Present	$(p_c p_o)$	+	$(p_c a_o)$	-
Cue Absent	$(a_c p_o)$	-	$(a_c a_o)$	+

Extension of frequent forms in production

Table 2 shows predicted activations of the frequent suffix, its synonym, and the two other suffixes (which are always activated equally) by the semantic features of the novel meaning (MANY and SMALL) under all logically possible models of associative learning. The left column represents the simplest possible model, a frequency counter (Bybee, 2010). Columns 2-4 represent association sets that can be added to the frequency counter in order to make contingency learning more veridical, incorporating learning of connections involving absent cues and/or outcomes. Column 5 is the model that learns only from present cues (a Hebbian version of Rescorla & Wagner, 1972). Column 6 is the full model that learns about all associations, including associations between absent cues and absent outcomes (a Hebbian version of Tassoni, 1995). Extension of frequent forms to novel meanings is predicted if the activation of the frequent form exceeds that of all other forms, including the frequent form's synonym. In other words, a preference to extend the frequent form to novel meanings is predicted whenever the largest number is in the top row.

As seen in Table 2, extension of the frequent form is predicted by increasing the weights of connections from present cues to present outcomes, as well as by decreasing the weights of connections from present cues to absent outcomes. Updating connections from absent cues (in the normative direction) acts against extension.

For the simulations reported in this table, it was assumed that an absence of a cue or outcome is noticed only half the time while its presence is always noticed. Associative learning in conditioning paradigms tends to be slower when reinforcement is signaled by the absence of a cue than when it is signaled by the presence of a cue (e.g. Wasserman et al., 1990). However, one might question whether absences are missed or ignored that often, and wonder whether noticing absences more would eliminate extension. It turns

out not to matter much: a_c does not overpower p_c even if absences are as salient as presences. All extant models of learning agree that absent stimuli are no more salient than stimuli that are actually presented and therefore all predict (over-)extension of frequent forms to related meanings in production.

Table 2: Activations of the frequent suffix, its synonym, and the other two suffixes given the novel diminutive plural meaning under alternative models.

DIM.PL→	$p_c p_o$	$p_c a_o$	$a_c p_o$	$a_c a_o$	p_c	all
Frequent	72	-42	-18	15	30	24
Synonym	24	-66	-6	21	-42	-12
Other	24	-66	-6	21	-42	-12

Entrenchment in comprehension

Table 3 reports activation differences between features that distinguish the novel meaning from the familiar meaning paired with a form in training. Because of the Symmetry Principle, the activation differences correspond to meaning→form connection weights involving the semantic features in question. For example, the activation difference between the non-diminutive and diminutive plural for –dan is the weight of the connection between –dan and BIG minus the weight of the connection between –dan and SMALL. The activation difference between the singular and plural diminutive for –nem is the weight of the connection between –nem and ONE minus the weight of the connection between –nem and MANY (cf. Miller & Matzel, 1985).

Entrenchment is observed if this difference is larger (more positive) for a frequent form compared to the 'other' forms, i.e. if the value in the top row in Table 3 is larger than the value in the bottom row.

Table 3: Comprehension effects. Each cell contains activation difference between the meaning paired with a form in training and the novel, diminutive plural, meaning. Activations of shared features of the competing meanings cancel out. Therefore, for plural suffixes this is the difference in activations between BIG and SMALL, and for diminutive suffixes it is the difference between ONE vs. MANY. Entrenchment is predicted if Frequent > Other.

Right-New	$p_c p_o$	$p_c a_o$	$a_c p_o$	$a_c a_o$	p_c	all
Frequent	36	0	36	6	36	78
Synonym	12	-12	12	-6	0	6
Other	12	12	0	0	24	24

Table 3 shows that entrenchment is favored by strengthening $p_c p_o$ connections between present cues and present outcomes, weakening $a_c p_o$ connections between absent cues and present outcomes, and strengthening $a_c a_o$ connections between absent cues and outcomes. Because updating $p_c p_o$ and $p_c a_o$ weights pull in different directions, entrenchment only occurs if absent outcomes are less salient

than present outcomes. In other words, weights of connections to absent outcomes must change less than the weights of connections to present outcomes. This appears to be a reasonable assumption (e.g. Tassoni, 1995), though not all extant models make it. For example, the Naïve Discriminative Learner (Baayen et al., 2011), which uses equilibrium equations for the Rescorla & Wagner (1972) model from Danks (2003: 116), does not show entrenchment because the learning rates for present and absent outcomes in Danks' equations are equal, a simplifying assumption (Danks, 2003: 115-116).

Conclusion

Studies of comprehension suggest that frequently encountering a form-meaning pairing convinces the learner than the form cannot be used in any other way (Braine & Brooks, 1995; Brooks & Tomasello, 1999; Regier & Gahl, 2004; Stefanowitch, 2008; Xu & Tenenbaum, 2007). Nonetheless, frequent forms are the ones most likely to be extended to new uses. Using a frequent form in a novel way seeds the process of language change because that novel use can then be picked up by others, spreading through the speech community. As the novel use diffuses through the community, it becomes conventional. Over historical time, extension of frequent forms results in the well-documented correlation between frequency of use and number of senses: in every language, it is the most frequent forms that are most polysemous (Piantadosi et al., 2012; Zipf, 1949).

Conventionalization of extensions is the primary mechanism behind the diachronic process of grammaticalization (Bybee, 2010; Heine, 2011). The importance of this diachronic process can hardly be overstated as it is the primary source of grammar: almost all grammatical morphemes, whether bound affixes or independent functors like prepositions, determiners or auxiliaries are former lexical words that have been gradually extended to new and new uses (Bybee, 2003; 2010; Christiansen & Chater, 2016).

Despite the correlation between frequency and semantic extension, the causal mechanisms behind grammaticalization remain controversial. For example, Haspelmath (1999) has argued that increases in frequency seen in grammaticalization are caused by the extension of the grammaticalizing form to new uses, which are in turn caused by semantic broadening. Bybee (2003) agrees that semantic broadening causes extension but suggests that high frequency causes semantic broadening. Like Haspelmath (1999), Heine (2011) does not allocate frequency a causal role in the process but suggests that extensions result in broadening.

In our recent experimental work, we have documented that the same speaker may extend a frequent form to a new meaning in production despite being least likely to map it onto the new meaning in comprehension. This suggests that the speaker may extend a form to a new meaning, thereby seeding the process of language change, without necessarily considering the form to be the best way to express that new meaning. Use in a new context can therefore be caused by high frequency and precede semantic broadening.

In the present paper, we have argued that this production-comprehension dissociation falls out of simple, Hebbian associative learning models, which acquire symmetrical form-meaning associations based on cue-outcome co-occurrence (Hebb, 1949; Miller & Matzel, 1985; see also McMurray et al., 2012; Yu & Smith, 2012). While such dissociations have previously been used to support the idea that form→meaning associations are distinct from meaning→form associations (Kapatsinski, 2009), the present results indicate that a single set of bidirectional associations suffices.

Remarkably, all that is required to obtain the divergence between frequency effects in production and comprehension – entrenchment of the frequent in comprehension, and extension in production – is the assumption that cue and outcome absences to be less salient than present cues and outcomes, an uncontroversial assumption (Tassoni, 1995; Wasserman et al., 1990) that is also normatively justified: almost every stimulus is absent more often than it is present, hence the presence of a stimulus is typically more informative about the contingencies in the learner's environment than its absence (McKenzie & Mikkelsen, 2007). Despite being surprising to human theorists, frequency-driven semantic extension is predicted by every associative learning theory.

References

Ambridge, B., Pine, J. M., Rowland, C. F. & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument-structure overgeneralisation errors, *Cognition, 106*(1), 87-129.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, *106*(2), 135-163.

Azab, H., Ruskin, D., & Kidd, C. (2016). Adults' guesses on probabilistic tasks reveal incremental representativeness biases. *Proceedings of the Annual Conference of the Cognitive Science Society*, *38*, 2831-36.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438-481.

Bates, D., Maechler, M., Bolker, B. M. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Braine, M. D. S. & Brooks, P. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 353-376). Hillsdale, NJ: Erlbaum.

Brooks, P., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*(4), 720-738.

Burke, M. D., Kester Locantore, J., Austin, A. A., Chae, B. (2004). Cherry pit primes Brad Pitt: Homophone priming effects on young and older adults' production of proper names. *Psychological Science, 15*(3), 164-170.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.

Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol. 2, 145-167). Mahwah, NJ: Erlbaum.

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, *47*(2), 109-121.

Ferreira, V. S., Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, *14*(1), 86-90.

Gershkoff-Stowe, L., Connell, B., & Smith, L. (2006). Priming overgeneralizations in two-and four-year-old children. *Journal of Child Language*, *33*(3), 461-486.

Harmon, Z., & Kapatsinski, V. (In preparation). Putting old tools to novel uses: The role of form accessibility in semantic extension.

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics*, *37*(6), 1043-1068.

Heine, B. (2011). Grammaticalization in African languages. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp.696-707). New York: Oxford University Press.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. Malden, MA: Wiley.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory and Cognition*, 30(6), 823-840.

Kapatsinski, V. (2009). Adversative conjunction choice in Russian (*no, da, odnako*): Semantic and syntactic influences on lexical selection. *Language Variation and Change*, *21*(2), 157-173.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review*, *70(1)*, 61-79.

McKenzie, C. R., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*(1), 33-61.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831-877.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. *The Psychology of Learning and Motivation*, *22*, 51-92.

Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *Journal of Child Language*, *22*(1), 19-46.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178-185.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280-291.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909-957.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition*, *93*(2), 147-155.

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*, 64-99. New York: Appleton-Century-Crofts.

Stefanowitsch, A. (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, *19*(3), 513-531.

Tassoni, C. J. (1995). The least mean squares network with information coding: A model of cue learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 193-204.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance limitations on grammaticality judgment tasks, *Cognitive Development, 19*(1), 15-34.

van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127-151.

Wasserman, E. A., Dorner, W. W., & Kao, S.-F. (1990). Contributions of specific cell information to judgments to interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 509-521.

Wonnacott, E., Newport, E. L., Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*(3), 165-209.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, *119*(1), 21-39.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Oxford: Addison-Wesley.