Language at three timescales: The role of real-time processes in language development and evolution

Bob McMurray

Dept. of Psychological and Brain Sciences, Dept. of Communication Sciences and Delta Center, University of Iowa

Abstract

Evolutionary developmental systems (Evo-Devo) theory stresses that selection pressures operate on entire developmental systems rather than just genes. This paper extends this approach to language evolution arguing that selection pressure may operate on two quasi-independent timescales. First, children clearly must acquire language successfully (as acknowledged in traditional Evo-Devo accounts) and evolution must equip them with the tools to do so. Second, while this is developing they must also communicate with others in the moment using partially developed knowledge. These pressures may require different solutions and their combination may underlie the evolution of complex mechanisms for language development and processing. I present two case studies to illustrate how the demands of both real-time communication and language acquisition may be subtly different (and interact). The first case study examines infant directed speech (IDS). A recent view is that IDS underwent cultural to statistical learning mechanisms that infants use to acquire the speech categories of their language. However, recent data suggest is it may not have evolved to enhance development, but rather to serve a more real-time communicative function. The second case study examines the argument for seemingly specialized mechanisms for learning word meanings (e.g., fast-mapping). Both behavioral and computational work suggest that learning may be much slower, and served by general purpose mechanisms like associative learning. Fast-mapping, then, may be a real-time process meant to serve immediate communication, not learning, by augmenting incomplete vocabulary knowledge with constraints from the current context. Together, these studies suggest that evolutionary accounts consider selection pressure arising from both real-time communicative demands and from the need for accurate language development.

Any theory of language evolution must address [at least] two questions (c.f., Bickerton, 2007). First, is the question of *selection pressure:* what functional demands led to language evolution? This emphasizes the function of language or of properties of the language system. The second question is *evolutionary implementation*: how did the language system change due to this pressure, how are these changes inherited, and what are their consequences for further evolution? These questions cannot be easily divorced. Evolved capacities at one generation (richer learning mechanism) create opportunities for selection pressure at the next. These questions—the *why* and *how* of language—are difficult to answer for language

Address Correspondence to Bob McMurray, Dept. of Psychology, E11 SSH, University of Iowa, Iowa City, IA 52242, bob-mcmurray@uiowa.edu, 319-335-2408.

as a whole, but more tractable for specific aspects of language like systematic phonology, referential word use, or combinatoric syntax.

Developmental science offers insight into the implementation of language evolution. By understanding *how* people acquire language within their lifespan, we can understand what is inherited across generations and the mechanisms of inheritance. Indeed, much of language development embraces such a soft synthesis, identifying functional properties of human language like referential communication (Waxman & Gelman, 2009), symbolic rules (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), and social/pragmatic inference (Tomasello, Carpenter, Call, Behne, & Moll, 2005), and making evolutionary arguments for why those capacities must be inborn (Hauser, Chomsky, & Fitch, 2002; Jackendoff, 1999; Tomasello, 2003).

Thus, traditional evolutionary theory often offers a straightforward explanation: evolutionary pressure selected genes for language and related cognitive functions, which gave rise to the relevant innate capacities or neural modules over development (Arbib, 2003; Hauser et al., 2002; Pinker & Bloom, 1990; Spelke & Kinzler, 2007). These genes, modules or capacities need not be specific to language (Jackendoff, 1999; Pinker & Bloom, 1990), but the emphasis on innate capacities argues for a unidirectional account in which selection pressure leads to genes for the neural innovations needed for language.

In contrast, *developmental systems theories* (Gottlieb, 2007; Johnston & Edwards, 2002; Oyama, 2000; Spencer et al., 2009) offer richer developmental accounts that raise new possibilities for language evolution. Development is the product of ongoing, *bidirectional* interactions between genes, proteins, cells, neural structures, behavior and the environment (Figure 1). Genes do not produce capacities directly, but respond in cascades to the chemical environment (created by regulatory genes and the cellular enrivonment), to lead in complex ways to phenotypes (Dediu & Christiansen, this issue).

Developmental systems offers multiple routes to inheritance. Stable sources of inheritance include genes, the prenatal environment (Francis, Szegda, Campbell, Martin, & Insel, 2003), and caregiver behavior (Francis, Diorio, Liu, & Meaney, 1999). In language, a number of mechanisms of extended inheritance may operate. Children inherit their language, which itself evolves to support learnability or communicative efficiency (Christiansen & Chater, 2008; Kirby, Dowman, & Griffiths, 2007) (Wedel, this issue). Most children inherit prenatal experience with sound that shapes early perceptual abilities (e.g., Nazzi, Bertoncini, & Mehler, 1998). Finally, mathematical regularities in how large networks (e.g., of words) interact can give rise to-syntax (Ferrer i Cancho, Riordan, & Bollobás, 2005). These offer avenues beyond genes on which variation and selection can evolve richer communicative behavior.

This is consistent with **Evolutionary Developmental Systems (*Evo-Devo*)** theory (Lickliter & Honeycutt, 2003; Oyama, 2000) which argues that evolution selects for not just genes, but for the whole developmental system that underlies capacities like language. This offers new ideas for how evolution may be implemented as part of a developmental system. This allows one to pinpoint mechanisms of variation, selection and inheritance across different levels of

the language system. For example, many accounts of language acquistion agree language acquisition harnesses statistical regularities in the environment (Elman, 1990; Hsu & Chater, 2010; Saffran & Thiessen, 2007). These regularities derive from the structure of the language, which in turn is shaped by cultural evolution. In an Evo-Devo framining, this suggests co-evolution of the language (the environment or input) and the learner (Christiansen & Chater, 2008; De Boer, 2005; Kirby et al., 2007).

Even as development can illuminate *how* language evolved, a proper understanding of the developmental system also offers us insight into it the question of *why*—the selection pressures responsible for such change. A focus on accurate langauge *development* as a goal of evolution too limited. Children and caregivers are motivated not just to acquire language over the long timescales of learning and development, but they also have immediate needs to use langauge in the moment to achieve communicative goals. These real-time processes have functional goals and demands—selection pressures—that are unique from the demands on development. We must properly characterize both to understand development and evolution.

I illustrate this with two case studies. The first examines the role of infant directed speech (IDS) in the development of speech categorization; the second examines word learning. Both cases are framed around well-known learning mechanisms as a core to development and evolution, but highlight the need to understand real-time processes operating in the child and the caregiver to properly characterize the developmental and selection pressures operating on those learning systems.

Case Study 1: Infant Directed Speech and the Acquisition of Phonetic Categories

One of the earliest in achievements in language acquisition is tuning infants' perceptual abilities to the phonology of their language. English-learning children, for example, need to discriminate /r/ and /l/; Japanese-learning babies do not. This is thought to occur during the first year (Werker & Curtin, 2005) when infants know few words and possess poor speech production skills. Consequently, many researchers have argued for some form of perceptual learning.

One promising account is ***distributional learning*** (de Boer & Kuhl, 2003; Maye, Werker, & Gerken, 2003). Phonological categories like voicing (which distinguishes /b,d,g/ from /p,t,k/) are distinguished by cues like Voice Onset Time (VOT, the time difference between the release of the consonant and the onset of voicing). The statistical structure of this cue across many utterances shows clusters (Figure 2) reflecting categories. If infants track these distributions, they could extract the categories of their language, and laboratory studies suggest they can do this rapidly (e.g., Maye et al., 2003). Further, distributional learning has been shown across domains(Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004), and species (Pons, 2006). Thus evolution could have harnessed an existing learning mechanism in a new way for language development.

However, distributional learning is not a perfect solution. Acoustic cues are highly variable due to talker differences, speaking rate, and coarticulation (McMurray & Jongman, 2011).

This variability creates overlap in the distributions of speech sounds, making it difficult to unambiguously categorize a given sound. Could evolved cultural practices help? Cross-linguistically, many caregivers use Infant Directed Speech (IDS). IDS affects every level of speech including syntax, word choice, prosody and segmental cues. If IDS systematically changes the statistics of the input, caregivers could structure these statistics to help infants overcome variability. Kuhl et al. (1997) support this with measurements of vowel formant frequencies for mothers speaking in IDS and Adult Directed Speech (ADS). Mothers' vowel spaces in IDS were stretched with greater separation (in F1×F2 space) than in ADS. Liu, Kuhl, and Tsao (2003) further showed that mothers with more separation have infants who discriminate speech better.

These findings suggest an Evo-Devo account where cultural practices co-evolve with learners (e.g., De Boer, 2005). Learning is based on domain general and evolutionarily conserved processes like distributional learning. However, IDS is a product of cultural evolution: it is transmitted across generations (or between individuals within a generation), and can support this difficult learning problem. Thus, the difficulty in acquiring phonological categories acts as a form of selection pressure on the way that parents speak (see also, Christiansen & Chater, 2008), selection pressure for better developmental outcomes.

However, recent research challenges this account, suggesting the effects observed by Kuhl et al. (1997) could also derive from caregivers slowing down, using more stressed words, or as a consequence of non-speech facial movements like smiling (Benders, 2013; Cristia & Seidl, 2013; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013).

A recent study from my lab (McMurray, Kovack-Lesh, et al., 2013) examined these issues in two ways. First, we examined a different cue, VOT. As VOT is a temporal cue, it is sensitive to speech rate. If IDS effects reflect speaking rate or prosody, VOTs for both voiced (/b,d,g/) and voiceless sounds (/p,t,k/) should increase, without enhancing the contrast. This was what was observed. VOTs lengthened in IDS for all sounds (Figure 3A; see Englund, 2005), and after accounting for speaking rate, the effect of IDS vanished (Figure 3B). Second, prior analyses of vowels focused on their mean locations in F1×F2 space; however IDS may also increase their variance may, impeding learning. Our results supported this with marked increases in variability (Figure 3C,D). Consequently when individual tokens were used to train a logistic regression to classify vowels, the classifiers trained on lower variance ADS performed slightly better.

Thus, the effect of IDS on segmental cues may be largely a by-product of other more global changes like speaking rate, prosody, or affect. This is consistent with other studies (Benders, 2013; Cristia & Seidl, 2013; Lam & Kitimura, 2011) which point to more general changes in segmental cues that do not specifically support learning.

This challenges the idea that IDS evolved in part to support perceptual development. So why did it evolve? One possibility is that IDS is useful for other developmental outcomes such as learning words, prosody or syntax. Alternatively, a developmental benefits may not be the primary function of IDS. Instead, many of the changes in IDS may derive from more

immediate demands. IDS helps modulate infants' affect, arousal and attention. Parents are sensitive to such things, and adjust their speech in the moment, responding to real-time cues from infants (Smith & Trainor, 2008). IDS may also better communicate such affective or intentional information to infants (Bryant & Barrett, 2007). Thus, IDS may be about the immediate needs of communicating with an infant that has not mastered language—any segmental changes come along for the ride. Weighing immediate selection pressure on an individual's behavior (e.g., struggles communicating with a baby), against long-term developmental outcomes, the real-time approach seems more plausible. That doesn't undermine the role of cultural evolution in shaping IDS; rather it suggests additional selection pressures on it.

If IDS did not evolve to support speech category acquisition, this renews the question of how distributional learning copes acoustic variability to find phonetic boundaries. What if it doesn't need to? Data like figure like 3C,D suggests there may be no such boundaries even under optimal learning conditions. There is just too much variability (and even more in IDS) (see also, Bion, Miyazawa, Kikuchi, & Mazuka, 2013). Indeed, McMurray and Jongman (2011) showed that even in a 24-cues space, there are no boundaries that discriminate fricatives as well as adult listeners. However, the addition of real-time compensation processes (on top of the statistical structure) can achieve listener-like performance. This puts limits on how much distributional learning alone must accomplish. The function of development is not just to find boundaries—children must acquire real-time processes for coping with variation. Consequently, any evolutionary account must go beyond learning examine how this real-time skill (and the developmental processes that give rise to it) evolved (see also, Christiansen & Chater, in press).

Case Study 2: Word Learning

A second problem faced by the developing child is learning to associate words and meanings. This problem is challenging. In any naming situation, there are many possible meanings for a novel word: the available objects, their properties, a talker's intentions and so forth. A common view is that children have biases or inferential strategies that constrain the search (Bloom & Markson, 1998; Golinkoff, Mervis, & Hirsh-Pasek, 1994). For example, when children encounter known objects (e.g., a plate and a bowl) with a novel object (a spork), they assume a new name refers to the novel object, a strategy termed fast-mapping by mutual exclusivity.

Little work on language evolution has considered vocabulary acquisition. However, common references to innate or human-specific biases or inferential abilities that solve the ambiguity problem (Tomasello et al., 2005; Waxman, 2003), suggest a soft evolutionary account: the demands of learning words led to the evolution of such strategies; or conversely the evolution of such abilities (for other purposes, like social interaction) enabled larger vocabularies. Such arguments seem to disqualify evolutionarily older mechanisms like associative learning (Nazzi & Bertoncini, 2003; Waxman & Gelman, 2009).

This again puts the functional goals—selection pressure—on developmental needs. However, as in the prior case study children must also understand language in the moment,

and such needs may put pressure on the system for different solution. For example, Horst and Samuelson (2008) (see also, Bion, Borovsky, & Fernald, 2013)found that children perform well in typical fast-mapping tasks, but are at chance five minutes later when these supposedly learned words are tested again. Thus, the product of fast-mapping is fleeting— possibly an inference to be used in the moment—and not synonymous with learning. This suggests a novel evolutionary account. The functional demands on in-the-moment behavior differ from those on learning. What makes for decisive behavior in the moment may not be good for learning (and vice versa). Instead, learning could rely on evolutionarily conserved mechanisms like associative learning, even as these simple mechanisms are buttressed by more sophisticated real-time inferences.

The different demands on real-time and learning-time processes (and the way they interact) are well illustrated by a recent computational model (Figure 4; McMurray, Horst, & Samuelson, 2012; McMurray, Zhao, Kucker, & Samuelson, 2013) built on simple, domain general processes. To capture real-time processing, we used a competition algorithm, similar to mechanisms for making complex "inferences" in domains like categorization, visual search, and speech perception (Spivey, 2007). Competition occurs over associations that link auditory and visual representations. These links are formed using a simple form of associative learning which builds gradually over multiple repetitions of a word.

The network is trained by activating one word and several objects (simulating a cluttered scene). Competition forces the model to settle on one object, and a small amount of learning occurs throughout this process. While on any one trial there is no information to indicate the word's referent, *across trials* the correct referent is more likely to co-occur with a word than other objects. Associative learning, by strengthening links between co-occurring words and objects, can thus acquire the correct mappings (Yu & Smith, 2007).

This simple architecture offers a unifying explanation for a host of developmental phenomena including cross-situational learning, fast-mapping by mutual exclusivity, taxonomic advantages, and changes in speed of processing familiar words. Two of these phenomenon illustrate how considering both real-time and developmental-time functional goals change our understanding of development, and hence of the evolution of word learning.

Figure 4B shows a network's accuracy over training. The network was trained on 35 words with a high degree of ambiguity (on each trial ~25 objects was present). It was tested with a 3AFC task (one word with its referent and two foils), and with an analysis of its associations to determine its underlying knowledge. 3AFC performance far exceeded competence (weight analysis): at 50,000 epochs, the network "knew" two words, but performed correctly for almost all 35. Real-time competition allowed the network to leverage partial knowledge —constrained by the objects in the scene—to perform accurately. Most real-world situations offer similar constraints for partially learned words, creating a similar situation.

Slow, gradual learning may be advantageous since most words have multiple meanings which are contextually determined. If a learner commits to a mapping on the basis of only one or two exposures it could be erroneous and difficult to change; slow learning offers more

statistical certainty as to the correct mappings based on more evidence. However, if learning is slow, what does the child during the extended period of uncertainty about many words. Real-time processing constrained by context may allow learners to act (e.g., respond to request) even as learning is slow. This combination of timescales may represent a novel evolutionary solution to competing selection pressures: learning should be slow to achieve accuracy, but real-time processes enable good behavior while words are learned.

Figure 4C shows model performance on a fast-mapping task similar to Horst and Samuelson (2008). Early in development the model is at chance at fast-mapping, even as it can recognize familiar words. Fast-mapping quickly reaches ceiling, much like children, although retention does not exceed chance until later. Thus, the model roughly fits the pattern of children's performance. However, this suggests that fast-mapping is a real-time ability, not a mechanism of learning—the model can use mutual exclusivity to identify a novel word's referent, but does not retain it.

However, fast-mapping does not solely derive from real-time processing. It is not built into the competition—the model cannot do it early in training. An analysis of the associations reveals that fast-mapping is a *product of learning.* The model starts with small random connections among all units. Over training, connections to familiar words and objects are gradually pruned (except correct associations). However, connections between units that are never active are not, and remain in their initial [random] state. Consequently, even after learning there are many small random connections between unlearned words and objects (Figure 4D). These offer many pathways by which a novel auditory unit can activate a novel visual unit, but few paths by which it can activate familiar units (those connections are pruned). Simple learning mechanisms set up the network to perform this intelligent in the moment behavior.

This model suggests a novel evolutionary account of word learning. First, it suggests two distinct selection pressures: 1) the child must accurately learn the meanings of many words with multiple contextually determined meanings; and 2) the child must accurately use words in the moment with only partial knowledge. Our model suggests that slow learning coupled to real-time processing offers a novel solution to this problem.

Second, classic analyses argued for specialized *learning* mechanisms—endowed by evolution—to solve the problem of ambiguity. Our division of timescales suggests that much of this can be offloaded to real-time processing. This in turn may enable evolutionarily older learning mechanisms like association. Given this, the major challenge left for associative learning is to overcome is acquisition of large numbers of associations and generalizing to new stimuli. These have both been shown to some extent by animals who can learn and generalize dozens or even hundreds of words in similar ways to humans (Kaminski, Call, & Fischer, 2004; Wasserman, Brooks, & McMurray, 2015), though no animal has yet acquired a vocabulary on the scale of humans. Moreover, complex real-time inferences may also not require specialized mechanisms. We've suggested here that mutual exclusivity is built in part on general-purpose competition mechanisms, and novelty and salience can mimic effects mutual exclusivity (Horst, Samuelson, Kucker, & McMurray, 2011). Thus, real-time processes may also be evolutionarily conserved[1], and not surprisingly other species have

been shown to fast-map by mutual exclusivity (Kaminski et al., 2004), even as they do not show retention (Griebel & Oller, 2012).

Given this, evolution may have harnessed both learning mechanisms and real-time inference mechanisms from other domains and species. What may be uniquely evolved for humans is the way they are assembled, and the types of information children use as input to such mechanisms. Children, for example, may be able to flexibly weigh social cues with lexical knowledge and use them to form associations in a way that is more difficult for other species. This combination of processes at two timescales may be a unique response to the competing selection pressure for learning a set of symbolic mappings far larger and more flexible than seen in any other species' communications, while at the same time requiring children to "fake it" and behave accurately in the moment while they did so.

General Discussion

These case studies illustrate that both evolution and development have dual functional goals. Child must acquire the sound categories or words of their language. But while this extended acquisition is unfolding, children and their caregivers must also be able to communicate in the moment.

In speech perception, I propose that IDS evolved (via cultural evolution) more in response to these real-time communicative demands than to the goal of enhancing developmental outcomes (at least in speech). This work suggests that statistics of the input may not be enough for account for perception; listeners need real-time processing to actively account for variability in the speech signal (McMurray & Jongman, 2011). In word learning, empirical results on fast-mapping (Bion, Borovsky, et al., 2013; Horst & Samuelson, 2008) and our computational model suggests that learning is simple and slow (to achieve the most accurate mappings), but real-time processing allows the child to act intelligently on the basis of partially learned information. Similarly fast-mapping is not a mechanism of learning, but rather, a product of learning that enables rapid inferences about new words in the moment. Again we see the importance of real-time processing in buttressing an only partially developed communicative system.

These case studies examine simple domain general learning mechanisms (statistical and/or associative learning). However, they argue that in isolation, such learning mechanisms are insufficient to understand language behavior. Rather, learning must be understood in the context of how learned information is used in the moment to understand and use language. This does not argue for an isomorphism between learning and real-time processing (e.g., language acquisition as learning to process: Christiansen & Chater, in press). Instead, there are independent selection pressures for what constitutes good processing in the face of uncertainty, and what constitutes good learning over development. This argument is also distinct from generativist approaches which distinguish learning and processing in a performance/competence framework. In contrast, real-time performance has a life of its

---

[1]That is not to say that children only bring simple mechanisms to problem. Processes like social inference (Diesendruck & Markson, 2001) may also help children rapidly identify referents; even as their results are gradually embedded in associative networks. However, such processes cannot be treated as evolutionary primitives–they also develop via complex routes.

own. It can, in some cases, exceed competence, for example, when a child uses real-time compensation to categorize a phoneme despite poorly defined boundaries, or uses information in the moment to identify the referent of a novel word. These "performance" phenomena are important in their own right as targets of both developmental and evolutionary investigation and are quasi-independent of learning.

A two-year-old hearing the phrase: "look at the chicken" must immediately determine that the intended referent is a small white bird, not the food they more typically associate it with. They should also do some learning to modify their representation of *chicken*. Over the next few years, they must also flexibly learn that chicken can refer to food and animals, that it can be used as an insult, or to refer to a game played with muscle cars. This may require substantially more input than what can be obtained in one barnyard encounter (despite the child's excellent performance there). This argues for independent processes for fast accurate inference, coupled to slow gradual learning to sort out this complexity.

In many cases, the processes that "live" at both timescales can be fairly simple and domain general, things like distributional learning, associative learning, and real-time competition. Such mechanisms are likely to be evolutionarily conserved. But this seems to run counter the obvious fact that no other species appears to have anything like language. This is underscored by an intriguing discrepancy in comparative work on language. The famous ape language studies noticeably failed to teach apes anything resembling the complexity of language (e.g., Gardner & Gardner, 1984). However, studies individual properties of language often show animals can succeed at tasks like rule-learning (Murphy, Mondragón, & Murphy, 2008), categorization (Wasserman et al., 2015), large vocabularies (Kaminski et al., 2004), rapid inference (Griebel & Oller, 2012) or even recursion (Gentner, Fenn, Margoliash, & Nusbaum, 2006; though see van Heijningen, de Visser, Zuidema, & ten Cate, 2009). There are undoubtedly differences in how non-humans achieve these abilities, and some may have arrived by convergent evolution. However, this suggests many of the precursors to language may be rooted in our evolutionary heritage.

So what makes human language different? The animal work and computational models of such simple mechanisms suggests there unlikely to be an evolutionary silver bullet, a critical ability that was necessary for language evolution. Instead, an Evo-Devo approach may offer the most clarity by offering a framework for thinking about how multiple components—likely built on simple general purpose mechanisms—can come together, and be tuned over development in a complex, inherited and evolved environment. Our work suggests a novel way to think about how such components are brought together. Even within a single subsystem of language (e.g., word learning, speech perception) distinct selection pressures for real-time processing and learning, may lead to quasi-independent solutions to both problems. It is this interaction (cascaded across many different selection pressures, on many different aspects of language and communication) on which the uniqueness of language may rest. But when we consider such interaction among processes, surprisingly simple—and evolutionarily shared—mechanisms may suffice to solve complex developmental problems.

Acknowledgements

References

Arbib MA. The evolving mirror system: a neural basis for language readiness. Studies in the Evolution of Language. 2003; 3:182–200.

Benders, T. Nature's distributional-learning experiment. The University of Amsterdam, Amsterdam, The Netherlands; 2013. (Ph.D.)

Bickerton D. Language evolution: A brief guide for linguists. Lingua. 2007; 117(3):510–526.

Bion RAH, Borovsky A, Fernald A. Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. Cognition. 2013; 126(1):39–53. [PubMed: 23063233]

Bion RAH, Miyazawa K, Kikuchi H, Mazuka R. Learning Phonemic Vowel Length from Naturalistic Recordings of Japanese Infant-Directed Speech. PLoS ONE. 2013; 8(2):e51594. [PubMed: 23437036]

Bloom P, Markson L. Capacities underlying word learning. Trends in Cognitive Sciences. 1998; 2(2): 67–73. [PubMed: 21227068]

Bryant GA, Barrett HC. Recognizing Intentions in Infant-Directed Speech: Evidence for Universals. Psychological Science. 2007; 18(8):746–751. [PubMed: 17680948]

Christiansen MH, Chater N. Language as shaped by the brain. Behavioral and Brain Sciences. 2008; 31(05):489–509. [PubMed: 18826669]

Christiansen MH, Chater N. The Now-or-Never Bottleneck: A Fundamental Constraint on Language. Behavioral and Brain Sciences. in press.

Cristia A, Seidl A. The hyperarticulation hypothesis of infant-directed speech. Journal of Child Language, FirstView. 2013:1–22.

De Boer, B. Infant directed speech and the evolution of language. In: Tallerman, M., editor. Evolutionary Prerequisites for Language. Oxford University Press; Oxford, UK: 2005. p. 100-121.

de Boer B, Kuhl PK. Investigating the role of infant-directed speech with a computer model. Auditory Research Letters On-Line (ARLO). 2003; 4:129–134.

Dediu D, Christiansen MH. Language evolution: constraints and opportunities from modern genetics. Topics in Cognitive Science. this issue.

Diesendruck G, Markson L. Children's avoidance of lexical overlap: a pragmatic account. Developmental Psychology. 2001; 37:630–641. [PubMed: 11552759]

Elman JL. Finding structure in time. Cognitive Science. 1990; 14:179.

Englund KT. Voice onset time in infant directed speech over the first six months. First Language. 2005; 25(2):219–234.

Ferrer i Cancho R, Riordan O, Bollobás B. The consequences of Zipf's law for syntax and symbolic reference. 2005; 272

Francis DD, Diorio J, Liu D, Meaney MJ. Nongenomic Transmission Across Generations of Maternal Behavior and Stress Responses in the Rat. Science. 1999; 286(5442):1155–1158. [PubMed: 10550053]

Francis DD, Szegda K, Campbell G, Martin WD, Insel TR. Epigenetic sources of behavioral differences in mice. Nat Neurosci. 2003; 6(5):445–446. [PubMed: 12665797]

Gardner RA, Gardner BT. A vocabulary test for chimpanzees (Pan troglodytes). Journal of Comparative Psychology. 1984; 98(4):381–404. [PubMed: 6509904]

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. Nature. 2006; 440(7088):1204–1207. [PubMed: 16641998]

Golinkoff RM, Mervis CB, Hirsh-Pasek K. Early object labels: The case for a developmental lexical principles framework. Journal of Child Language. 1994; 21:125–155. [PubMed: 8006089]

Gottlieb G. Probabilistic epigenesis. Developmental Science. 2007; 10(1):1–11. [PubMed: 17181692]

Griebel U, Oller DK. Vocabulary Learning in a Yorkshire Terrier: Slow mapping of spoken words. PLoS ONE. 2012; 7(2)

Guenther FH, Nieto-Castanon A, Ghosh S, Tourville J. Representation of Sound Categories in Auditory Cortical Maps. Journal of Speech Language and Hearing Research. 2004; 47:46–57.

Hauser MD, Chomsky N, Fitch WT. The faculty of language: what is it, who has it, and how did it evolve? Science. 2002; 298(5598):1569–1579. [PubMed: 12446899]

Horst JS, Samuelson L. Fast mapping but poor retention in 24-month-old infants. Infancy. 2008; 13(2): 128–157.

Horst JS, Samuelson LK, Kucker SC, McMurray B. What's new? Children prefer novelty in referent selection. Cognition. 2011; 118(2):234–244. [PubMed: 21092945]

Hsu AS, Chater N. The logical problem of language acquisition: A probabilistic perspective. Cognitive Science. 2010; 34(6):972–1016. [PubMed: 21564242]

Jackendoff R. Possible stages in the evolution of the language capacity. Trends in Cognitive Sciences. 1999; 3(7):272–279. [PubMed: 10377542]

Johnston TD, Edwards L. Genes, Interactions, and the Development of Behavior. Psychological Review. 2002; 109(1):26–34. [PubMed: 11863039]

Kaminski J, Call J, Fischer J. Word learning in a domestic dog: Evidence for "Fast Mapping". Science. 2004; 304(5677):1682–1683. [PubMed: 15192233]

Kirby S, Dowman M, Griffiths TL. Innateness and culture in the evolution of language. Proceedings of the National Academy of Sciences. 2007; 104(12):5241–5245.

Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Lacerda F. Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. Science. 1997; 277(5326):684–686. [PubMed: 9235890]

Lam C, Kitimura C. Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. Developmental Science. 2011; 15(2):212–221. [PubMed: 22356177]

Lickliter R, Honeycutt H. Developmental Dynamics: Toward a Biologically Plausible Evolutionary Psychology. Psychological Bulletin. 2003; 129(6):819–835. [PubMed: 14599279]

Liu H-M, Kuhl PK, Tsao F-M. An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science. 2003; 6(3):F1–F10.

Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants. Science. 1999; 283(5398):77–80. [PubMed: 9872745]

Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition. 2003; 82:101–111.

McMurray B, Horst JS, Samuelson L. Word learning emerges from the interaction of online referent selection and slow associative learning. Psychological Review. 2012; 119(4):831–877. [PubMed: 23088341]

McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review. 2011; 118(2):219–246. [PubMed: 21417542]

McMurray B, Kovack-Lesh K, Goodwin D, McEchron WD. Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? Cognition. 2013; 129:362–378. [PubMed: 23973465]

McMurray, B.; Zhao, LB.; Kucker, SC.; Samuelson, LK. Probing the limits of associative learning: generalization and the statistics of words and referents. Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence. In: Gogate, L.; Hollich, G., editors. IGI Global; Hershey, PA: 2013. p. 49-80.

Murphy RA, Mondragón E, Murphy VA. Rule Learning by Rats. Science. 2008; 319(5871):1849–1851. [PubMed: 18369151]

Nazzi T, Bertoncini J. Before and after the vocabulary spurt: two modes of word acquisition? Developmental Science. 2003; 6(2):136–142.

Nazzi T, Bertoncini J, Mehler J. Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. Journal of Experimental Psychology: Human Perception and Performance. 1998; 24(3):756–766. [PubMed: 9627414]

Oyama, S. The ontogeny of information: Developmental systems and evolution. 2nd. Duke University Press; Durham, NC, US: 2000. rev. and expanded

Pinker S, Bloom P. Natural language and natural selection. Behavioral and Brain Sciences. 1990; 13(04):707–727.

Pons F. The effects of distributional learning on rats' sensitivity to phonetic information. Journal of Experimental Psychology: Animal Behavior Processes. 2006; 32(1):97–101. [PubMed: 16435970]

Saffran, JR.; Thiessen, ED. Domain-general learning capacities. In: Hoff, E.; Shatz, M., editors. Handbook of Language Development. Blackwell; Cambridge, UK: 2007. p. 68-86.

Smith NA, Trainor LJ. Infant-Directed Speech Is Modulated by Infant Feedback. Infancy. 2008; 13(4): 410–420.

Spelke ES, Kinzler KD. Core knowledge. Developmental Science. 2007; 10(1):89–96. [PubMed: 17181705]

Spencer J, Blumberg M, McMurray B, Robinson SR, Samuelson L, Tomblin JB. Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. Child Development Perspectives. 2009; 3(2):79–87. [PubMed: 19784383]

Spivey, MJ. The continuity of mind. Oxford University Press; New York: 2007.

Tomasello M. On the different origins of symbols and grammar. Studies in the Evolution of Language. 2003; 3:94–110.

Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences. 2005; 28(05):675–691. [PubMed: 16262930]

van Heijningen CAA, de Visser J, Zuidema W, ten Cate C. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proceedings of the National Academy of Sciences. 2009; 106(48):20538–20543.

Wasserman EA, Brooks DI, McMurray B. Pigeons acquire multiple categories in parallel via associative learning: A parallel to human word learning? Cognition. 2015; 136:99–122. [PubMed: 25497520]

Waxman, SR. Early category and concept development: Making sense of the blooming, buzzing confusion. The Oxford University Press; New York, NY: 2003. Links between object categorization and naming: Origins and emergence in human infants; p. 213-241.

Waxman SR, Gelman S. Early word-learning entails reference, not merely associations. Trends in Cognitive Sciences. 2009; 13(6):258–263. [PubMed: 19447670]

Werker JF, Curtin S. PRIMIR: A Developmental Framework of Infant Speech Processing. Language Learning and Development. 2005; 1(2):197–234.

Yu C, Smith LB. Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. Psychological Science. 2007; 18(5):414–420. [PubMed: 17576281]

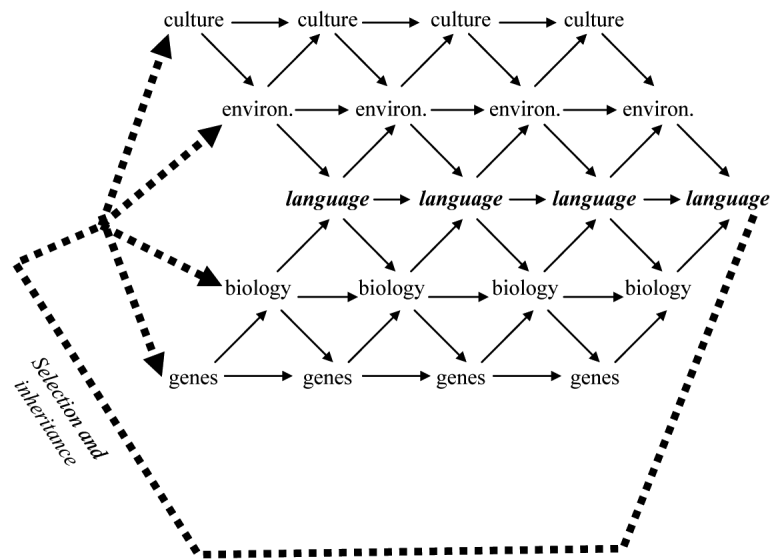**Figure 1.**
In Evo-Devo accounts, language develops as bidirectional and ongoing interactions between genes, biology (which includes a range of processes not shown like gene expression, epigenetics, protein transcription, brain organization), cognitive/language processes, the environment and broader culture. Inheritance is possible at each level, and selection operates to tune the entire developmental system across generations.
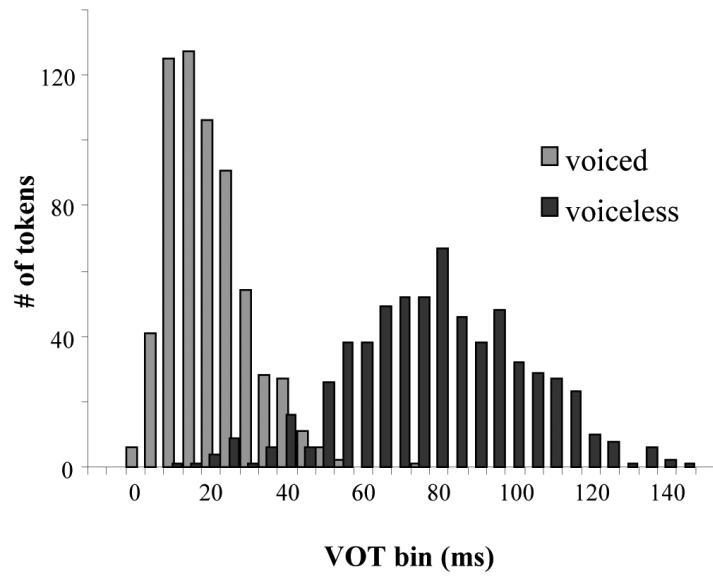
Figure 2.
Frequency distribution of VOTs in English (adult directed speech, from McMurray et al., 2013)
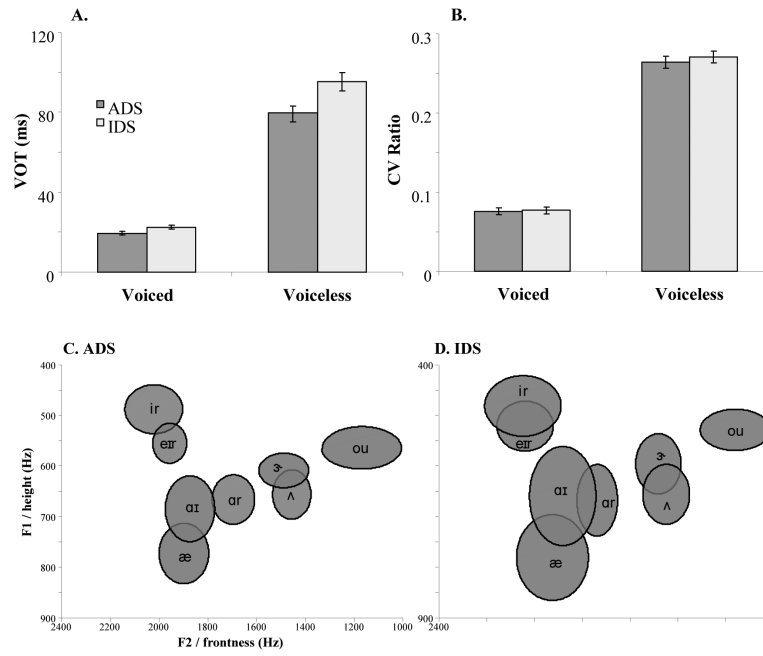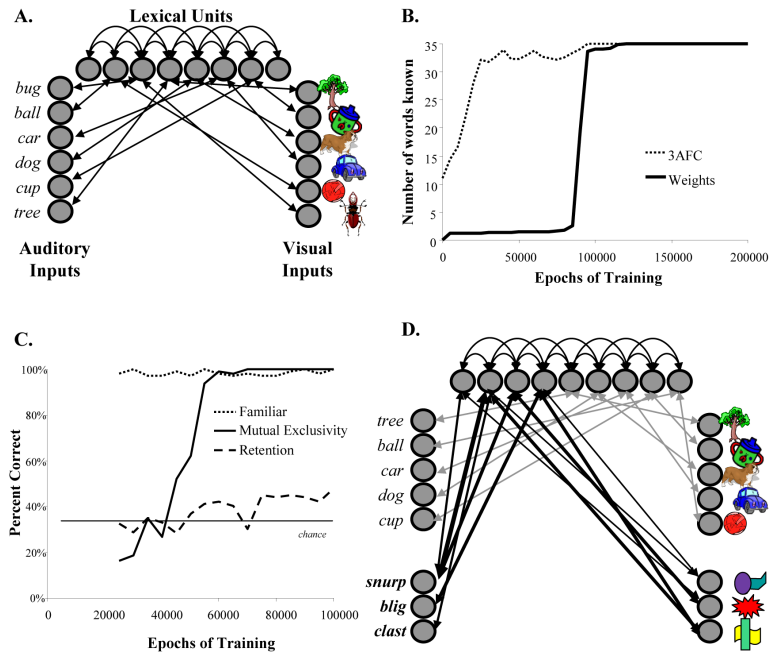
Figure 3.
Results of McMurray et al., (2013). A) In IDS VOTs for both voiced and voiceless sounds lengthen – the distance between voiced and voiceless sounds does not increase. B) When speaking rate is taken into account (with a ratio of VOT to vowel length), effects of IDS disappear. C) Vowel measurements in ADS – ellipsis are centered at the mean F1 × F2 location for each vowel, and the size of the ellipse corresponds to one SD. D) Variance increases dramatically in IDS.

Figure 4.

The McMurray, Horst and Samuelson (2012) Dynamic Associative Model. A) Structure of the model. Inputs correspond to words and objects, linked to a layer of lexical units. Activation flows from inputs to the lexicon, and feeds back to inputs as it settles over time; connections are gradually tuned via associative learning over several trials to link words to their referents. B) Results of a single network trained with a high degree of referential ambiguity. Shown is the number of words known when the model is tested with a 3- or 10-alternative forced choice task, and by an analysis of the association weights. C) Performance in a version of Horst and Samuelson (2008) task over training. The model was then tested in a series of trials with two familiar objects and a novel object. Familiar trials test the model's ability to identify one of the familiar referents from this array; on Mutual Exclusivity trials, the auditory stimulus was a novel, untrained word. D) Representation of the association weights for fast-mapping. The lower three words and objects are untrained. Because they have not been encountered, they retain random connections between them, but not between novel words/object and familiar ones. This allows activation to flow from the novel word to the novel object on a fast-mapping trial.