Design features of language emerge from general-purpose learning mechanisms

Padraic Monaghan (p.monaghan@lancs.ac.uk)
Centre for Research in Human Development and Learning, Department of Psychology
Lancaster University, Lancaster LA1 4YF, UK

Abstract

There are certain universal properties of language that are taken to be definitional to the concept of language itself, such as the arbitrary relationship between sounds and meanings of words. Another possibility is that these "design features" of language may instead be the expressed consequences of general purpose learning constraints within the cognitive system learning the language. To test this, generations of an inverse model learning to map between sounds and meanings of words was tested. In this model, learning to associate phonology to semantics influences the model's production of phonology from semantics, and phonological productions of one model were used as input to the next generation. Over generations of the model's learning, the language became easier to acquire, and demonstrated increased arbitrariness of mappings between phonology and semantics. The iterative modelling demonstrated that design features of natural language can spontaneously emerge in a general purpose learning system.

Keywords: Evolution; language acquisition; computational modeling; arbitrariness of the sign; phonology; semantics.

Introduction

Languages change extremely rapidly. Gray and Atkinson (2003), for instance, estimated that all the living Indo-European dialects diverged approximately 7900 years ago. In terms of cultural transmission from generation to generation, this means that language has been passed on only a few hundred times to produce such variation as that found between English, Gaelic, Greek, Italian, Lithuanian, and Hindi. There are of course multiple forces at work in determining language change (Labov, 1994), however, the fact that language has to be transmitted from one generation to the next suggests that learnability of the language is a critical selective pressure contributing to language evolution (Christiansen & Chater, 2008).

So what are these language properties that contribute to language learnability? One place to begin is the recent discussion over patterning of language universals (Christiansen, Collins, & Edelman, 2009; Scalise, Magni, & Bisetto, 2009). Evans and Levinson (2009) demonstrated that for each "language universal" proposed in the literature there is at least one extant language that violates the prevailing pattern. Instead, Evans and Levinson (2009) contend that it is language *diversity*, rather than universality, that is the critical feature of human communication to be explained. Importantly, this diversity tends to indicate statistical clusters of language properties, which are consistent with general cognitive constraints that assist in learning or transmission of language that then become embedded in language structure. Cross-linguistic consistencies (rather than absolutes) in language structure occur, then, because similar processing and learning limitations are present in all language users.

Yet, there are properties of languages that *are* universal, though these are not discussed by Evans and Levinson (2009) because they are considered definitional properties of language. These fundamental language properties, or "design features" in Hockett's (1960) terms, were listed by Greenberg (1963) as discreteness, productivity, arbitrariness, and duality of patterning. Discreteness refers to the composition of utterances in a language from smaller elements (words/morphemes or phonemes) the combination of which provides meaning. Relatedly, productivity refers to the ability to use the smaller elements of the language in multiple combinations – from a small, finite set of elements (words/morphemes) an infinite set of utterances can be generated. Arbitrariness refers to the absence of systematicity between the sounds of words and their meaning (de Saussure, 1916), and duality of patterning refers to the composition of words from smaller phonological units, where the utterance meaning is carried by the combination of words and is unrelated to the phonological composition of these words.

Though researchers such as Greenberg (1963) question the possibility of a language without each of these properties, an alternative view is instead that many of these properties could have been otherwise in language (Monaghan, Christiansen, & Fitneva, 2011). It is possible to conceive of a language where sounds of words do carry some aspect of the meaning. Instead, this paper takes as its perspective that design features are universal properties of language not because they are definitional but rather because of the constraints of our cognitive systems that mean that languages are structured to make acquisition and, consequentially, transmission easier. The current study tests a framework that demonstrates how such design features of language may have become instantiated within language structure as a consequence of general-purpose (i.e., not language-specific) learning mechanisms.

Monaghan et al. (2011) showed that one of the design features – arbitrariness of the sign – resulted in more accurate language learning by participants acquiring an artificial language, and by associative networks learning the same sound-meaning mappings. Thus, arbitrariness bestowed an advantage for learning.

However, this work stopped short of demonstrating how such arbitrariness becomes *incorporated within* the language as a consequence of learning constraints. In this paper, I present a model of cultural transmission where general purpose learning constraints that affect language

acquisition are expressed within the same model when it undertakes language production, which is then used to entrain the next generation of models. An appropriate architecture for achieving this effect of acquisition influencing production is the inverse model (Jordan & Rumelhart, 1992), which resembles the sleep-wake algorithm implemented in the stochastic inverse Helmholtz model (Hinton, Dayan, Frey, & Neal, 1995). Such models have close parallels to the learning occurring in feedforward and feedback connections in the cortex (Carpenter & Grossberg, 1987; Mumford, 1994), and have been effectively used to simulate development of features of human communication, such as segmental phonology (Plaut & Kello, 1999). These models have the advantage over previous models of iterated learning in that they permit more enriched representations to be learned and to influence performance (Kirby, 2001).

The critical feature of the modeling is that associative learning forms the basis of the model's performance, and that the model's approach to learning words' meanings from sounds influences the model's production of the words' sounds from meanings. Language structures that are easier to acquire due to general purpose learning mechanisms will be learned more accurately by the model and will, over generations of learner, become stably expressed within the words themselves. The dashed lines in Figure 1 illustrate how the learning of the spoken input to an output meaning representation can feedback to generate a spoken version for each verbal input. Hence, the model as learner can adapt its version of the language to more closely match its own internal constraints. This spoken output can then be used for the next generation of learner, and across multiple generations the language can be altered in such a way to make it more reflective of the model's learning properties and hence more easily learnable for future generations. The language can then be analysed for the key "design features" of natural languages.

The first prediction is that the expression of the model's learning constraints will result in a set of phonological representations that the model learns to map onto the meaning representations more quickly and accurately – so future generations will find the language easier to acquire. The second prediction is that "design features" of natural language will be exhibited in the representations, in particular that mappings between phonology and semantics will instantiate arbitrariness in the mapping[1]. However, based on previous studies (e.g., Monaghan et al., 2011) it is also predicted that the patterns will demonstrate systematicity at the category level, in terms of phonology-category mappings. The first simulation tests the emergent structure of phonology-semantics mappings, when the phonology is initially random, so fully arbitrary. The second simulation tests the emergent structure when the initial state of the language is systematic.

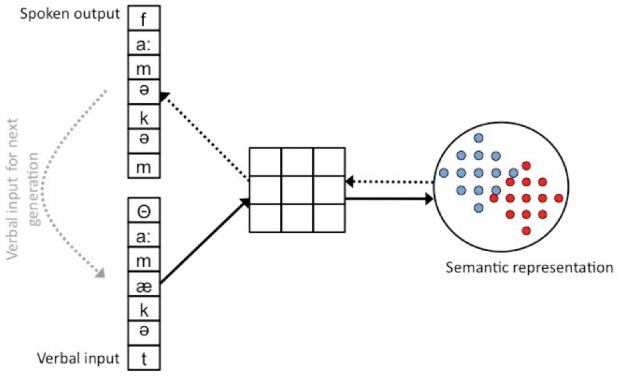

Figure 1: The model of iterated learning. Solid arrows indicate the forward model, where the model is trained to map phonology onto semantics for each word, with two categories (red/blue) centred at a different region of this multidimensional space. The dashed arrows indicate the inverse model, where the model produces phonology from semantics which forms the input to the next generation of learner.

Emergent structure from random origins

Method

Architecture

The model's architecture is shown in Figure 1. A set of 10 units each represented the phonology and the semantics for the model, and a set of 10 hidden units interconnected these representations. For the forward run of the model, to simulate language comprehension, the verbal input was connected to the hidden units which were in turn connected to the semantic units. These connections are illustrated with solid arrows. For the inverse run of the model, to simulate language production, there were connections from the semantics to the hidden, and from the hidden to the phonology to represent spoken output. The inverse model connections are illustrated in Figure 1 by dotted arrows. Weights on connections were given initial uniform random values in the range [-0.25,0.25].

Training and Testing

The model was trained on 20 patterns mapping between phonology and semantics. The phonological representations were initially constructed by randomly selecting values in the range [0,1] in 0.1 intervals for each of the 10 units in the pattern. The semantic representation remained stable throughout learning and was constructed by generating 10 exemplars of two prototypes, one centred at a value of 0.75 for each of the 10 units in the pattern, and the other centred at 0.25 for each of the 10 units. Exemplars were produced by randomly varying the prototype values by a uniform value in the range [-0.25,0.25]. There were therefore two clusters of meaning representations in the patterns to be

[1] The existence of small pockets of sound-symbolism in natural languages has little effect on the overall arbitrariness of the relationship between sound and meaning (see Monaghan et al., 2011, for discussion).

learned. In previous simulations, such semantic categories have simulated two grammatical categories of words, e.g., nouns and verbs (Monaghan et al. (2011).

The model was trained to map the phonology onto the semantics to simulate learning to comprehend language, and then learn to map the semantics back onto the phonology using the same hidden unit representations as occurred during the phonology to semantics mapping. This entailed that the model's learning constraints for acquiring the language were applied during the model's attempts to produce the language. The forward model operated as a standard three-layer backpropagation network: A pattern was selected randomly and presented at the input. The model's hidden unit representation for this pattern was recorded, and then the mean square error at the semantic representation (the difference between target and actual semantic representations) was used to adjust the weights on the forward connections using the backpropagation learning algorithm with learning rate of 0.1.

Then, the inverse model was applied to this same pattern: The target semantic representation was presented at the semantic layer of the model, and the model was required to reproduce the hidden unit representation for that pattern produced during the run of the forward model. The error at the hidden layer was propagated back to adjust weights between semantic and hidden units. Then, the model was required to reproduce the phonology for that pattern given the hidden unit representation, and once again error (between the initial input phonology and the model's actual phonological production) was propagated back to adjust weights between hidden and phonology layers.

At the end of training, the inverse model was presented with each of the semantic representations and produced a version of the phonology for these patterns that was influenced by its learning of the mapping. These new phonological patterns were used as the language for training the next iteration of the model. There were 10 iterations altogether of the model.

We varied the number of presentations of the patterns to determine an effective level of change in the language – not too much, such that the language would alter radically from one generation to the next, and not too little such that no representational change would be observed. We found that 500,000 patterns resulted in an interpretable level of representational change. With fewer presentations than this, the model produced phonological representations via the inverse model that were distant from the original phonological representations, and also that were indistinct from one another, and so extremely difficult to learn for future generations of the model.

We ran 20 versions of the model, varying the initial phonological and semantic representations, and varying the initial randomized weights on the connections between units. Each simulation run of the model was used as a separate participant in the analyses.

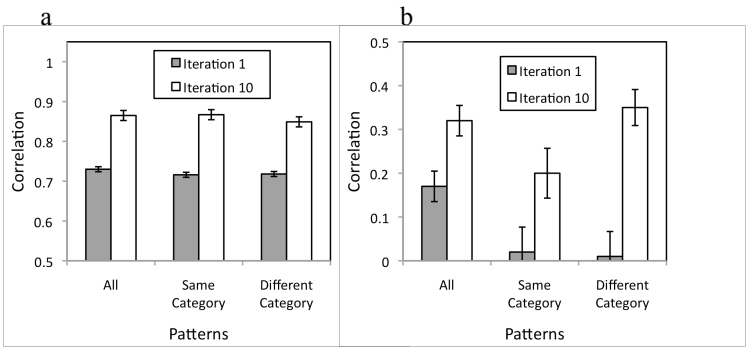

Figure 2. (a) Correlation among phonological representations for all patterns, for patterns within the same category. and across different categories, at first and tenth iteration of learning. (b) Correlation between phonology and semantic representations at first and tenth iteration for all patterns, patterns within the same category, and patterns in different categories.

Table 1: Mean error (SD in parentheses) for phonology→semantics, and semantics→phonology mappings at first and tenth iteration of training.

Iteration	phon→sem	sem→phon
1	2.40 (1.42)	116.29 (32.45)
10	1.31 (.75)	.03 (.07)

Results and Discussion

The first prediction was that over iterations of the model, the model would learn to map between phonology and meaning (comprehension task), and meaning to phonology (production task) with less error. We compared the mean square error of the model's actual productions versus the target semantic representation across the meaning layer when the phonology was inputted, and the mean square error across the phonology layer when the meaning layer was inputted, we compared the error after the first iteration to that after the tenth iteration. The error value provides a reflection of how accurately the model reproduces the semantic and phonological representations of the words. Table 1 shows the results.

For both comprehension and production tasks, there was a reduction in error over iterations, $t(19) = 3.11$, and 15.01, both $p < 0.01$, thus the generations of learning resulted in easier comprehension and production of the patterns.

To determine the changes that actually occurred to the phonology representations as a consequence of the iterated learning, we measured the cosine distance between each pair of phonological patterns and then computed the mean of these distances for each simulation run. High mean cosine values indicate that there is similarity among the patterns. We took this measure at the first iteration and at the tenth iteration. The results are shown in Figure 2a. From first to tenth iteration, the correlation increased among the phonological representations, $t(19) = 18.09$, $p < 0.001$. To determine whether this increased correlation was due to

words of the same (semantic) category becoming more similar to one another, or whether words of distinct categories were becoming more aligned, we distinguished the cosine distances among phonological representations of the same category, and those among representations of different categories. The results are also shown in Figure 2a. An ANOVA with same/different category and first/tenth iteration on mean cosine distance for each simulation run resulted in a significant main effect of same/different category, $F(1, 19) = 8.46$, $p < 0.01$, indicating that same category responses were more correlated than different category. There was also a significant main effect of iteration, $F(1, 19) = 327.40$, $p < 0.001$, reflecting the general increase in correlations from first to tenth iteration. The interaction was also significant, $F(1, 19) = 27.72$, $p < 0.001$, showing that the correlation increased more sharply for same category responses than different category responses, though the magnitude of the difference was small. Thus, the model introduced systematicity at the category level into the phonological representations resulting in easier acquisition of the patterns at the end of the set of iterations.

To determine the extent to which the mappings between phonology and semantics introduced greater or less arbitrariness in the mapping, we correlated the cosine distances between each pair of phonology representations and each pair of semantics representations. If patterns that are close together in phonology are also close together in semantics and patterns that are distant in phonology are distant in semantics then the correlation will be high, representing systematicity in the mapping. If patterns that are distinct in phonology are similar in semantics then the correlation will be low, indicating arbitrariness.

The results are shown in Figure 2b. For all patterns, there was an increase in correlation between the phonology and the semantic spaces from first to tenth iteration, $t(19) = 4.65$, $p < 0.001$. There is thus an increase in systematicity across the iterations. To determine whether this change was within each category, indicating that words of the same category were becoming increasingly systematic with respect to their meanings, or whether the change was due to systematicity across categories, we measured the correlation between distances just for words within the same category, and compared this to distances for words occurring in distinct categories.

The results are again shown in Figure 2b. An ANOVA with same/different category, and first/tenth iteration was performed. There was a marginally significant main effect of same/different category, $F(1, 19) = 3.81$, $p = 0.07$. There was a significant main effect of iteration, $F(1, 19) = 58.25$, $p < 0.001$, as correlations increased from first to tenth iteration. There was also a significant interaction, $F(1, 19) = 4.77$, $p < 0.05$, indicating that for the first iteration there was little difference in the correlation between phonology and semantics for patterns in the same versus different categories, but that after ten iterations, the correlation was substantially higher for patterns of different categories than same categories.

It may be that the development of systematicity at the category level and arbitrariness at the individual word level is an intermediate stage in the model's development to an optimal representation for the language, where this final optimal state is fully systematic. In order to rule out this possibility, the next simulation tested language change when the initial language was highly systematic.

Emergent structure from systematic origins

Method

Architecture

The architecture was identical to the first simulation.

Training and testing

The training and testing were identical to the first simulation except that the initial phonological representations were highly correlated with the semantic representations for each pattern. Each phonological representation was generated by taking the semantic representation for that pattern and varying each dimension by a random value in the range [-0.25,0.25].

a b

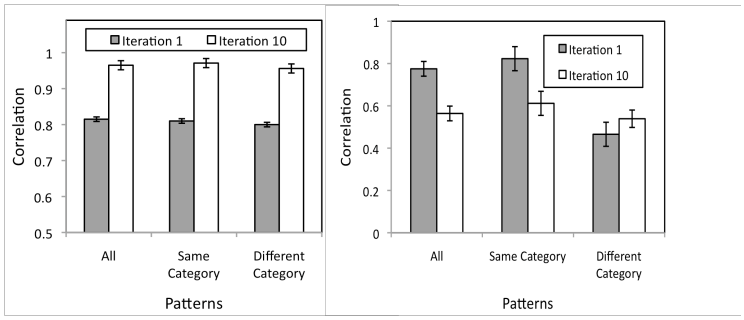

Figure 3. (a) Correlation among phonological representations for all patterns, for patterns within the same category. and across different categories, at first and tenth iteration of learning for systematic origin model. (b) Correlation between phonology and semantic representations at first and tenth iteration for all patterns, patterns within the same category, and patterns in different categories for systematic origin model.

Table 2: Mean error (SD in parentheses) for phonology->semantics, and semantics->phonology mappings at first and tenth iteration of training for the second simulation.

Iteration	phon->sem	sem->phon
1	1.06 (0.52)	18.00 (4.92)
10	15.35 (7.02)	0.003 (0.009)

Results and Discussion

In terms of the model's error, for the production task there was a significant increase in error between first and tenth iteration, $t(19) = -9.10$, $p < 0.001$, but a significant decrease for the comprehension task, $t(19) = 16.39$, $p < 0.001$ (see

Table 2). Thus, systematic mappings resulted in easier acquisition, but greater difficulties in producing discriminating output, and the model adapted the representations to meet better the production task.

The correlations among the phonological representations were compared between first and tenth iterations, and as with the initial arbitrary mapping the correlations increased, $t(19) = 31.15$, $p < 0.01$. For the same/different category comparing first and tenth iteration in an ANOVA, there was a significant main effect of same/different, $F(1, 19) = 137.99$, $p < 0.001$, with same category resulting in higher correlations. There was a significant main effect of first/tenth iteration, $F(1, 19) = 962.90$, $p < 0.001$, and a significant interaction, $F(1, 19) = 4.64$, $p < 0.05$. The difference between same/different category was greater at the tenth iteration (see Figure 3a).

For the correlations between phonology and semantic representations comparing first and tenth iteration, the correlation decreased with time, $t(19) = 6.52$, $p < 0.001$. An ANOVA with same/different category and first/tenth iteration resulted in a significant main effect of same/different, $F(1, 19) = 52.15$, $p < 0.001$, with same category resulting in higher correlation than different category. There was also a significant main effect of iteration, $F(1, 19) = 4.90$, $p < 0.05$, and a significant interaction, $F(1, 19) = 17.84$, $p < 0.001$, with a decrease in correlation between first and tenth iteration for same category but an increase in correlation for different category correlations (see Figure 3b). As with the model beginning with fully arbitrary mappings, arbitrariness increased to a greater extent for words belonging to the same category.

General Discussion

The results indicate that the model adapted phonological representations to become easier to map onto semantics. The model's general learning constraints shaped the phonology of the language to make mapping to and from semantics easier for future models to acquire. Investigating the actual changes in the phonological representations from first to tenth generation revealed that this ease of learning was accomplished through two primary changes in the representations that resembled the design features of natural language.

First, the iterations of the language increased the similarity among words of the same category in terms of their phonological representation. This accords with observations over the variety of phonological and prosodic cues that reflect grammatical categories of words cross-linguistically (Farmer, Christiansen, & Monaghan; 2006; Monaghan, Christiansen, & Chater, 2007). Indeed, for English, there are now more than 20 distinct phonological and prosodic properties that relate to grammatical category distinctions (Monaghan & Christiansen, 2008).

The inverse model presented here showed that such coherence with respect to category structure can emerge as a consequence of pressures of learning. The inverse model instantiated learning constraints into the representations themselves, and these learning constraints were expressed by reflecting the output category structure within the input phonology. In artificial language learning studies, such reflections of category structure within phonology has been shown to result in improved learning of categories (Frigo & McDonald, 1998; St Clair, Monaghan, & Ramscar, 2009), and may indeed be vital for effective acquisition of grammatical categories (Braine, 1987). The model points to the way such phonological characteristics of words can become imprinted within the language as a consequence of general-purpose learning mechanisms exerting their influence through generations of language learners.

Second, the results of the model in terms of the properties of the mapping between phonology and semantics show in addition that arbitrariness can sit alongside systematicity indicated at the category level. For words of the same semantic category there is greater distinction between individual phonological patterns in terms of the precise semantic representation that they map onto. For words of different categories, there is greater expression of words or similar sound relating to meanings that are similar. This can be interpreted in terms of the coherence among the phonological representations being tempered by the additional requirement to distinguish particular semantic representations. Thus, emerges systematicity at the category level, but arbitrariness for mapping between individual patterns.

In this respect, the iterative inverse modeling results presented here relate to learning studies of static artificial languages that map between phonological and semantic representations of words. Monaghan et al. (2011) trained associative learning models and human participants to map between phonological and semantic representations for words belonging to one of two categories. They varied the properties of the patterns in terms of whether the mappings were arbitrary or systematic between phonology and semantics, and also the extent to which additional phonological cues provided information about the general category to which the word belonged. Learning was most accurate for both the associative learning model and the behavioural results when the mapping between phonology and meaning was arbitrary, but with coherence at the category level. The iterative modeling presented here demonstrates that similar general purpose learning mechanisms imposed by requirements to associate between two sets of representations can result in an attuning of the representations themselves to approximate this structure as a consequence of constraints imposed in learning the language being expressed in production.

The model was trained with a starting language that was either fully arbitrary or largely systematic. These situations can be seen to resemble two theories of the origins of language, where words emerge either from articulatory noise (Jespersen, 1922), or from iconic or sound-symbolic forms (Ramachandran & Hubbard, 2001). In both cases, we have shown that there is an increase in accuracy of reproduction of the language across generations, and that

this is coupled with generated systematicity at the category level and greater arbitrariness in the form-meaning mappings within those categories. Future work may also permit investigation into whether the emergent pronunciations are more likely to result from iconic or noisy initial forms.

The starting point for this modeling approach was to demonstrate how learning may, over generations of learners, affect the structure of natural languages. In this respect, the modeling demonstrates that "design features" of languages may fall under the remit of the cognitive sciences in explaining how and why such properties are observable within language. Plaut and Kello (1999) demonstrated how an inverse model can account for the development of segmental phonology – a contributor to the design feature of discreteness, and we have shown here how arbitrariness of form-meaning mappings is an emergent property of constraints on learning in a similar model. Though the model learns only a small set of patterns, and consequently, the results should be treated cautiously, the observations tally closely with computational and behavioural studies on learning effectiveness from different structures of a language's vocabulary. The model presented here provides an iterative step to showing how such design features can emerge spontaneously within a learning system. Natural languages may possess "design features", then, not as necessary, definitional properties, but rather because having such structure facilitates learning, and over generations this process of learning becomes impressed within the structure of language itself.

References

Braine, M.D.S. (1987). What is learned when acquiring word classes: A step toward an acquisition theory. In MacWhinney, B. (Ed.), *Mechanisms of language acquisition*. London: Lawrence Erlbaum Associates.

Carpenter, G. A., & Grossberg, S. (1987). Neural dynamics of category learning and recognition: Attention, memory and consolidation, and amnesia. In J. Davis, R. Newburgh & E. Wegman (Eds.), *Brain structure, learning and memory*: Westview Press.

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*, 489-509.

Christiansen, M. H., Collins, C., & Edelman, S. (2009). *Language universals*. NY: Oxford University Press.

de Saussure, F. (1916). *Course in general linguistics*. New York: McGraw-Hill.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Brain and Behavioural Sciences, 32*, 429-492.

Farmer, T.A., Christiansen, M.H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203-12208.

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.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature, 426*, 435-439.

Greenberg, J. H., ed. (1963). *Universals of language*. Cambridge, MA: MIT Press.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science, 268*, 1158-1161.

Hockett, C. F. (1960). The origin of speech. *Scientific American,* 203, 89-96.

Jespersen, O. (1922). *Language: Its nature, development and origin*. London: Allen & Unwin.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16*, 307-354.

Kirby, S. (2001). Spontaneous evolution of linguistic structure - An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*, 102-110.

Labov, W. (1994, 2001), *Principles of linguistic change*, *Volume 1 Internal factors, Volume 2 Social factors*. Oxford: Blackwell.

Monaghan, P. & Christiansen, M.H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In Behrens, H. (Ed.), *Corpora in language acquisition research: History, methods, perspectives*. Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional Coherence Hypothesis: Cross-linguisitic evidence in language acquisition. *Cognitive Psychology, 55*, 259-305.

Monaghan, P., Christiansen, M. H., & Fitneva, S.A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, in press.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*. Cambridge, MA: MIT Press.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *Emergence of Language*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Ramachandran, V. S. & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3-34.

Scalise, S., Magni, E., & Bisetto, A. (Eds.). (2009). *Universals of Language Today*. Berlin: Springer.

St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*, 1317-1329.