Why Form-Meaning Mappings are not Entirely Arbitrary in Language

Padraic Monaghan (pjm21@york.ac.uk)
Department of Psychology, University of York
York, YO10 5DD, UK
Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology, Cornell University
Ithaca, NY 14853, USA

Abstract

We discuss two tasks that the child must solve in learning their[1] language: identifying the particular meaning of the word being spoken, and determining the general category to which the word belongs. We present a series of simple language learning models solving these two tasks. We show that learning the precise meaning of the word is more easily accomplished if there are arbitrary mappings between the spoken form of words and their meanings, when these words are presented with contextual information. We also show that learning general categories is best achieved when there is systematicity in the mappings between forms and categories. We present corpus analyses of English and French indicating that there is both arbitrariness and systematicity in language, and suggest that this co-habitation is a design feature of natural languages that benefits learning.

Introduction

Since Saussure (1916), the relationship between the sound and the meaning of words has been regarded as largely arbitrary. Indeed, the arbitrariness of form-meaning mappings has long been highlighted as a design feature of human language (e.g., Hockett, 1960). Recent support for arbitrariness has come from computational simulations by Gasser (2004), who demonstrated that, for large vocabularies, there is a learning advantage for arbitrary form-meaning relationships. Because systematic pairings of forms and meanings require strong constraints on the space of possible pairings (e.g., a particular onset phoneme is restricted to only co-occur with a particular facet of meaning) it is only possible to encode efficiently a relatively small number of words. In contrast, arbitrary mappings between form and meaning impose fewer constraints and therefore permit the learning of a large and extendable vocabulary, which is the hallmark of human language.

Though the general picture is one of arbitrariness between the individual phonological form of a word and its meaning (see Tamariz, 2005 for an estimate of the correlation), some systematic mappings do exist in natural language; for example, expressives in Japanese and Tamil show evidence of a systematic form-meaning mapping (Gasser, Sethuraman, & Hockema, in press). In English certain groups of words display similar sound symbolism – such as, *-ash* which tends to occur at the end of words indicating the

application of (destructive) force to something:*bash, clash, thrash, trash, slash, mash, dash*, etc. – the psychological reality of which has been confirmed through priming experiments (Bergen, 2004). Moreover, as we show below, systematic mappings also exist between words and their lexical categories, and these go well beyond the effects of morphological affixing. In this paper we present a simple model of word learning in order to investigate the circumstances under which systematic form-meaning mappings may be advantageous for language learning.

The Child's Dual Task

The context of the utterance of a word (e.g., situational, gestural, verbal co-occurrence) provides a great deal of information about the general ballpark meaning of the word (Tomasello, 2003). Given this contextual information, then, would it be more conducive to learning the pairing between the spoken form of a word and its precise meaning if there is a correlation between the spoken representations and the output representations, or if there is no, or little, correlation? We hypothesise that if each word within a general region of semantic space was very different in its spoken form to other words then this would benefit the learning of the mapping – the learner has more individual sources of information to exploit in determining the mapping. If there is a correlation, then precise and subtle differences between the spoken forms of words have to be attended to in order to identify the exact intended meaning.

As an example, imagine the situation where a child is observing animals milling around in a farmyard. In English, the child hears the words *"cow", "sheep", "goat",* and *"chicken"*, and is required to form a mapping between each word and each animal. The words in another language, SystemEnglish for these same animal referents are *"bim"*, *"bin", "bing",* and *"pim"*. We predict that the child learning SystemEnglish is going to find the task significantly more difficult, partly because subtle differences between words have to be attended to, and partly because such differences may be over-ruled by the noise present in the auditory environment, which may obliterate the distinctions entirely. Indeed, although 12-month-old children can distinguish between minimal pairs of sounds such as *"bin"* and *"bim"*, they are unable to associate these terms with distinct referents at the age of 14 months. However, children can perform this association when the words are more phonologically distinct (Werker, Fennell, Corcoran, & Stager, 2002).

---

[1] Throughout this paper, we employ the epicene *they*: "**A person cannot help their birth**". Thackeray, W.M. (1848). *Vanity fair*. London: Punch.

Hence, we propose that one contributory factor towards the arbitrariness of the form-meaning relationship for words is the effect of learnability of such pairings under *noisy conditions* and when *contextual information* is present.

However, the child has also to learn the structure of the language in addition to its lexical items. Hence, developing knowledge about the general category to which a word belongs as quickly as possible is going to be useful for developing knowledge about the structure of the language and further supplementing the contextual information available to assist in identifying the particular word. As a shorthand for general categories to which words belong, we use the word's lexical category.

In this paper, we explore four hypotheses about the role of arbitrary and systematic mappings for language learning.

1. In the first study, we employ a neural network model to investigate our hypothesis that if contextual information is present, words can be learned more quickly under noisy conditions when the mapping between phonology and meaning is arbitrary.
2. Our second study tested the hypothesis that the general category to which a word belongs will be learned more quickly if mappings between words and categories are systematic.
3. In the third study, we explored a prediction derived from Studies 1 and 2: There should be arbitrariness between spoken word forms and particular meanings of words, and systematicity between spoken forms and categories of words evident in natural language.
4. Finally, the fourth study examined the hypothesis that this combination will be optimal for a system learning both individual meanings and categories. Specifically, Study 4 tests the combined effect of arbitrariness and systematicity on learning in a neural network.

Study 1

In this simulation study we investigated the hypothesis that if context is provided, neural networks will learn faster under noisy conditions when the mappings between word forms and meanings are arbitrary.

Method

Architecture The model is shown in Figure 1, in a screenshot of the PDP++ software (O'Reilly, Dawson, & McClelland, 1995). The model is comprised of four layers of units. There are two input layers. The "context" input layer contains four units, and provides an indication of the general category of the pattern: one unit for each category. The "phonological" input layer contains 20 units across which a pattern is presented. These two input layers are connected to a hidden layer that contains 40 units. The hidden layer is connected to a "semantic" output layer, containing 20 units.

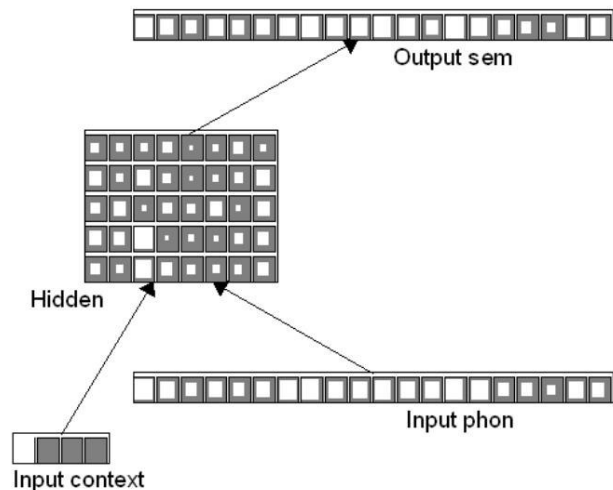The model is trained on four sets of 10 patterns, each set

Figure 1: The neural network model of learning mappings between phonological and semantic representations, with the presence of contextual information.

of 10 belonging to one of four categories. The patterns were designed to represent words with proximal meanings within a category. We constructed these sets of category patterns by generating a prototypical 20-dimensional pattern for each of the four categories with randomly selected values in the range [.2, .8] in intervals of .1. Then, each pattern within a set is created by randomly changing 20% of the values of this prototype. Consequently, patterns within a category were correlated with one another, $r = .70$.

In the systematic simulations, the input representations are correlated with the output representations for each pattern (mean $r = .92$). These were generated by taking the output representation and randomly changing 20% of the dimensional values by an amount of -0.2, -0.1, 0.1, or 0.2, though values were capped with minimum value of 0 and maximum value of 1. Figure 1 shows an instance of a systematic input-output pairing, with activity of units represented in terms of area of shading. In the arbitrary simulations, the same set of input representations were randomly paired with the output representations such that the correlation between input and output representations was reduced (mean $r = .21$).

Training and testing During training, each pattern was presented in the phonological input layer along with one context unit activated according to which category the word belonged. In Figure 1, the word is in category 1, and so the first unit is active in the context layer. The model was trained using the backpropagation learning algorithm, with learning rate .1 and momentum .9. In each block of training, all 40 patterns were presented in a randomised order. Training was stopped once the model's production for each pattern was closer to the target pattern than it was to any other pattern in the training set, by computing a correlation between the output representation and the target representation for each pattern in the training set. We term

this the *identification task*. The dependent variable was how many blocks of training were required to achieve this target. If the model had not reached criterion by 5000 blocks of training, the simulation was halted.

The model was trained with noise added across the input phonological representations. The noise distribution was uniform with mean 0 and variance .1.

In order to ascertain that it was the general context that was crucial in learning the arbitrary mappings, we repeated the simulation but omitted the context input layer from the simulations.

We repeated each simulation 20 times, and the results were compared in a within-subjects design.

Results and Discussion

An ANOVA on time taken to learn the identification task, with presence/absence of contextual information as between-subjects factor and arbitrary/systematic language as within subjects factor, indicated a significant main effect of context with context making mappings easier to learn, $F(1, 38) = 39.33$, $MSe = 1,615,861.34$, $p < .001$, but no overall significant effect of arbitrary/systematic mapping, $F < 1$. As hypothesized, there was a significant interaction between context presence and arbitrary/systematic language, $F(1, 38) = 60.64$, $MSe = 780,486.96$, $p < .001$.

Post hoc tests revealed that when the context layer was present, the systematic language required more training than the arbitrary language (2476 and 1094 blocks of training, respectively), $t(19) = 5.66$, $p < .001$. When the context layer was excised, the model with the systematic mappings learned more quickly than the arbitrary mappings, 2721 and 4416 blocks of training, respectively, $t(19) = -5.46$, $p < .001$. Contextual information made almost no difference to learning the systematic mappings, $t < 1$, but made a large difference to learning the arbitrary mappings, $t(38) = 15.91$, $p < .001$.

A further advantage of arbitrary mappings between forms and meanings is that, once learned, identifying the particular pattern should be less prone to impairment by noise in the environment. This is because within each category there is a larger distance between patterns in the arbitrary condition than the systematic condition. After training the model, we tested its vulnerability to noise by testing it on the set of 40 patterns when uniform distributions of noise with mean 0 and variance .2 and .5 were added to the input phonological representations

For the .2 noise level, we found that the arbitrary mapping model reproduced 31.30 from 40 patterns correctly, whereas the systematic mapping model produced a mean 26.20 from 40 patterns, which was significantly less, $t(19) = -6.90$, $p < .001$. For the .5 noise level, the accuracy was 16.10 and 12.15 for arbitrary and systematic mapping models, respectively, $t(19) = -3.72$, $p < .005$.

When provided with contextual information indicating the general category of a word, learning arbitrary mappings between phonological forms and semantics was facilitated relative to systematic mappings.

Study 2

The identification of the word is not the only task facing the child learning their language. In addition, the child must learn the general category to which the word belongs. For learning individual items, arbitrary mappings have been shown to be more beneficial to learning. For learning general categories, however, we hypothesise that systematic mappings will be more effective. This is because the model does not have to identify the precise characteristics of the word, but only the general region of space that the word inhabits. We tested this in Study 2.

Method

Architecture We used the same architecture as in Study 1, except without the context layer.

Training and testing The same input and output pairs were used for training as in Study 1. For testing, however, we computed the distance from the model's output to the nearest prototypical representation for the category to which the output pattern belonged. We stopped training when the model produced an output closest to the prototypical category representation for all 40 patterns. We term this the *categorization task*. The dependent variable was once again the number of blocks each simulation took to reach the training criterion, and we repeated each of the arbitrary and systematic mapping simulations twenty times.

Results and Discussion

The model presented with systematic mappings learned to solve the categorization task after a mean of 5 blocks, whereas the model learning the arbitrary mappings took a mean of 1217 blocks of training, which was significantly longer, $t(19) = -4.69$, $p < .001$. The model learning the systematic mappings was significantly faster at solving the categorization task than the identification task: comparing the systematic simulations of Study 1 and the current simulation, $t(38) = 8.621$, $p < .001$. Comparing the results of Studies 1 and 2 indicated that solving the categorization task when no context was present and the identification task when context was present made little difference for the models learning the arbitrary mappings, $t(38) < 1$.

Hence, our simulations have indicated that systematicitiy in the mapping between form and category significantly benefits learning the general category to which a word belongs. However, learning the precise representation, rather than the general category is better performed by a model with arbitrary mappings when supported by contextual information. Given that language learning requires not only the formation of mappings between particular spoken forms and meanings, but also the learning of categories of words, for the purpose of syntactic processing, we suggest that learning will be best accomplished if the language contains some degree of systematicity between the phonological forms of words and their general category, but also arbitrariness between the phonological form and the individual meaning. An optimal

system of communication is likely to incorporate both properties. In Study 3 we investigate whether natural languages reflect these hypothesised properties.

Study 3

Natural languages contain phonological information about the grammatical category of a word, where grammatical category is one approach to considering general groupings of meanings of words (Kelly, 1992; Monaghan, Chater, & Christiansen, 2005; Onnis & Christiansen, 2005). So, where in the word are we likely to find arbitrariness in terms of meaning and systematicity in terms of category? Speech processing is a fast, online, sequential process, consequently there is pressure on the beginning of a word to be distinct from other words, so that it can be uniquely identified as quickly as possible (Hawkins & Cutler, 1988; Lindell, Nicholls, & Castles, 2003). Hence, placing phonological information shared between many different words at the beginning of the word would impede the process of identification. Placing this shared information towards the end of the word (reflecting the language-universal preference for suffixing over affixing, Hawkins & Cutler, 1988) would provide systematicity for the categorization task without impairing the identification task. We hypothesized that words in a natural language would contain more grammatical category information reflected in phonology at the end of the word, but not at the beginning. We assessed this through corpus analyses, focusing on the distinction between nouns and verbs, as in previous studies (Kelly, 1992; Onnis & Christiansen, 2005). We also assessed whether the effect is cross-linguistic by analyzing both English and French.

Method

Corpus preparation We took the 5000 most frequent nouns and verbs from the English CELEX database (Baayen, Pipenbrock, & Gulikers, 1995) that were classified unambiguously in terms of grammatical category. For French, we took the 5000 most frequent unambiguous nouns and verbs from the BRULEX database (Content, Mousty, & Radeau, 1990). Previous studies have focused on the noun/verb distinction in order to estimate the potential phonological information present in the lexicon for reflecting category (Kelly, 1992; Onnis & Christiansen, 2005), and we follow their lead here. There were 3,818 nouns and 1,182 verbs in the English corpus, and 3,657 nouns and 1,343 verbs in French.

Corpus analysis To investigate the relationship between the phonology at the beginning of the word and grammatical category, we took as a cue the onset and nucleus of the first syllable (so for the word *penguin*, we used /pɛ/, and for the word *ant*, we used /æ/). For the end of the word cue, we took the nucleus and coda of the last syllable of the word (for *penguin*, we used /ɪn/, and for *ant*, we used /ænt/). We chose the first and last few phonemes as participants have

been found to be sensitive to the first few letters of words for grammatical categorization (Arciuli & Cupples, in press a) and the first and last two letters of words have been shown to reflect stress patterns that in turn reflect grammatical category (Arciuli & Cupples, in press b). There were 536 distinct word beginnings, and 564 endings for English, and 455 beginnings and 167 endings for French. The cues were entered into a discriminant analysis to determine how effectively the beginnings or endings of words related to the noun/verb distinction. As a baseline, we randomly reassigned the grammatical category labels to the words, and reran the analyses.

Results and Discussion

For English, the discriminant analysis on the beginning cues resulted in 62.0% correct classifications compared to 50.3% for the baseline. The ending cues correctly classified 81.9% of the words compared to 50.1% for the baseline. Both analyses were significant, though the ending cues were an order of magnitude greater in terms of effect size, $\chi^2 = 365.49$ for beginning cues and $\chi^2 = 1,914.29$ for ending cues, both $p < .001$.

For French, the beginning cues resulted in 58.5% correct classification compared to 49.4% for the random baseline, $\chi^2 = 486.31$. For the ending cues, performance was again more accurate, with 89.8% correct classification compared to the 50.0% baseline, $\chi^2 = 3,055.70$.

The cues we have used in these analyses highlight the useful phonological information present in languages for determining grammatical category (Kelly, 1992). The use of the first two and last two phonemes for each word reflects previous studies that have used the phonological form of the entire word (Monaghan et al., 2005), or just the first or last phoneme for categorization (Onnis & Christiansen, 2005). The current studies contain morphological information, but repeated analyses on monomorphemic nouns and verbs resulted in similar effects, indicating that morphological markers to grammatical category are only a part of the contribution of phonological properties of words related to grammatical category. For both the English and French analyses we found phonological information well above chance levels for both beginnings and endings.

Of particular interest in the current study, in both English and French the beginning of words provides more information about the identity of the word – providing more distinctiveness to assist in the identification of the unique word, yet the second half of the word is where greater systematicity can be observed between phonological forms and general category for words. In the next study, we examined whether models learning to map between patterns that were partially arbitrary and partially systematic were beneficial for learning compared to systems that were entirely arbitrary or entirely systematic.

Study 4

This study tested the hypothesis that a combination of systematic and arbitrary mappings will be optimal for a

system learning both individual meanings and categories of words.

Method

Architecture The model was the same as that used in Studies 1 and 2.

Training and testing The model was trained and tested for the identification task, and the categorization task. There were three conditions of mapping between phonological and semantic representations. We used the arbitrary and the systematic mappings as in Studies 1 and 2, as well as a third condition that we term the half-arbitrary mapping. In this condition, there was little correlation between the first 10 input and output units, but there was a correlation between the second set of 10 input and output units. In the arbitrary mapping condition, all the input unit representations were randomized, whereas in the half-arbitrary mapping condition, this randomization was performed only for ten of the input units. In the identification task, the context layer was active, but was inactive for the categorization task. All other aspects of the simulations were identical to Studies 1 and 2.
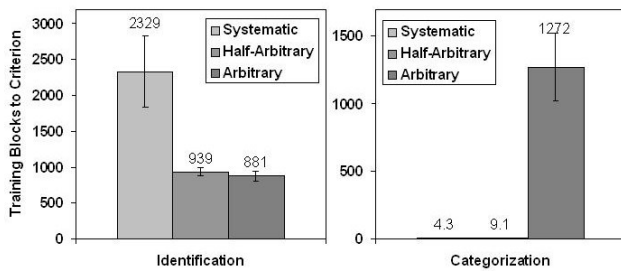

Figure 2: Blocks of training to criterion for the identification task, and the categorization task for the systematic, half-arbitrary, and arbitrary simulations in Study 4.

Results and Discussion

Figure 2 shows the results of the simulations for learning the arbitrary, half-arbitrary, and systematic patterns. For the identification task – where the model has to identify the precise pattern – a one-way ANOVA was significant, $F(2, 38) = 29.32$, $MSe = 458,281.29$, $p < .001$. Post hoc tests revealed a significant difference between the arbitrary and systematic condition and the half-arbitrary and the systematic condition (both $p < .001$), but no significant difference between the arbitrary and half-arbitrary condition ($p > .5$). For the representations of input and output patterns that we used in these simulations, the presence of some degree of arbitrariness was sufficient to produce good performance for the identification task. The arbitrary and half-arbitrary networks were comparably good, and both beneficial for learning over the networks learning the systematic pairings.

For the categorization task – where the model was tested on the closeness of the model's production to the nearest

prototype category – there was a significant difference between the three conditions, $F(2, 38) = 26.18$, $MSe = 407,461.76$, $p < .001$. Post hoc tests revealed a significant difference between the systematic and the arbitrary conditions and the half-arbitrary and the arbitrary conditions (both p < .001), and also a small but significant difference between the systematic and half-arbitrary conditions (p < .001). Both the half-arbitrary and the systematic mappings demonstrated a large advantage over the arbitrary mappings for learning the general category of the word. Without any systematicity in the mappings, learning the general category of the pattern was as difficult as learning the precise identity of the pattern.

General Discussion

Study 1 confirmed the advantage for learning arbitrary mappings in a connectionist model found by Gasser (2004). However, our explanation for the effect is different. Gasser (2004) concentrated on the additional degrees of freedom available for forming arbitrary mappings that enabled better performance when many patterns had to be learned by the model. Yet, the effect was only observed when the patterns were densely populated in the representational space. We have shown that the presence of noise to the input representations, or the environment of the learner together with general contextual information about the word resulted in a robust effect, even in a sparsely-populated representational space.

Study 2 considered the additional task facing the child: to learn not only the precise identity of the word but also the general category to which the word belongs. We found that when the models were assessed in terms of their ability to reflect the general category of the pattern, then systematic mappings were more effective.

The corpus analyses in Study 3 were performed to test first whether information about general categories could be observed in the phonological form of words, and second whether this general information was located at a particular position in the word. We hypothesized that it was likely to occur toward the end of the word, given the sequential constraints on speech perception to identify the word uniquely as quickly as possible. The corpus analyses of English and French demonstrated a similar pattern in both languages: namely, that there is information about the noun/verb distinction present at both the beginning and the end of the word, but that this is substantially more informative about category at the end of the word. Hence, the farmyard is populated by cows, ducks, and lambs rather than scow, sduck, and slamb. The corpus analyses are undoubtedly related to inflections tending to occur at the end of the word, but these make up a small proportion of the cues we used in the analyses (there were 564 endings for English and 167 for French). Relatedly, Onnis and Christiansen (2005) found that categorization based on inflections was no more effective than categorization based merely on single phonemes.

The final set of simulations in Study 4 incorporated the language structure suggested by the corpus analyses into the input-output patterns. When each pattern was a combination of systematic and arbitrary mappings – the half-arbitrary condition – learning was as effective for the identification task as when the patterns were entirely arbitrary. For the categorization task, the half-arbitrary condition produced a huge advantage over the arbitrary condition, and learning was only slightly slower than in the entirely systematic condition.

The Saussurian notion that language is arbitrary has been an influential view for almost a century. The simulations presented here indicate that there is a learnability advantage to this arbitrariness when information about general context is available to the learner. This is the natural situation for language learning – children are exposed to words spoken in phrases, rather than entirely in isolation. Furthermore, the words they hear are accompanied by gestures and expressions, and occur in the presence of objects in the environment (Tomasello, 2003). A host of cues are thus available to the child to narrow down the possible meanings of the word.

Yet the Saussurian notion does not reflect entirely the nature of the form-meaning relationship present in natural languages. The systematicity of the mapping between the forms of words and the general categories to which they belong – evidenced here by the relationship between phonological properties and grammatical categories – indicates that sound-symbolism occupies more than merely small pockets of the lexicon (Gasser et al., in press).

The child faces two tasks in learning their language: to identify the precise word, and to learn the word's category. These two processes are served by complementary and co-present features of the phonology of the word. In Hockett's (1960) terms, language has been designed to incorporate both arbitrariness and systematicity in the language.

Acknowledgments

References

Arciuli, J. & Cupples, L. (in press a). Would you rather 'embert a cudsert' or 'cudsert an embert'? How spelling patterns at the beginning of English disyllables can cue grammatical category.

Arciuli, J. & Cupples, L. (in press b). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*.

Baayen, R.H., Pipenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bergen, B.K. (2004). The psychological reality of phonaesthemes. *Language, 80*, 290-311.

Content, A., Mousty, P., & Radeau, M. (1990). BRULEX: Une base de donnés lexicales informatisé pour le français écrit et parlé. *Anné Psychologique*, *90*, 551-566.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Cognitive Science Society Conference* (pp.434-439). Hillsdale, NJ: LEA.

Gasser, M., Sethuraman, N., & Hockema, S. (in press). Iconicity in expressives: An empirical investigation. In S. Rice and J. Newman (Eds.), *Experimental and empirical methods*. Stanford, CA: CSLI Publications.

Hawkins, J.A. & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J.A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.

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

Kelly, M.H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*, 349-364.

Lindell, A., Nicholls, M., & Castles, A.E. (2003). The effect of orthographic uniqueness and deviation points on lexical decisions: Evidence from unilateral and bilateral-redundant presentations. *Quarterly Journal of Experimental Psychology, 56A*, 287–307.

Onnis, L. & Christiansen, M. (2005). New beginnings and happy endings: Psychological plausibility in computational models of language acquisition. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society,* (pp.1678-1683). Hillsdale, NJ: Lawrence Erlbaum.

O'Reilly, R.C., Dawson, C.K. & McClelland, J.L. (1995). *PDP++ neural network simulator software*. Carnegie Mellon University.

Saussure, F.d. (1916). *Cours de linguistique générale*. Paris: Payot.

Tamariz, M. (2005). *Exploring the adaptive structure of the mental lexicon*. Unpublished PhD thesis, The University of Edinburgh, Edinburgh.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Boston, MA: Harvard University Press.

Werker, J.F., Fennell, C.T., Corcoran, K., & Stager, C.L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*, 1-30.