The Rise and Fall of English Inflectional Morphology

Patrick Juola (PATRICK.JUOLA@PSY.OX.AC.UK)[1]
Kim Plunkett (KIM.PLUNKETT@PSY.OX.AC.UK)
Department of Experimental Psychology
University of Oxford
Oxford, OX1 3UD UNITED KINGDOM

Abstract

Children acquire noun inflections before they acquire verb inflections. Noun inflections are also less affected by language disorders than verb inflections. We describe a single-system connectionist model of English noun and verb inflection which captures these facts of acquisition and atrophy, as well as other well-established developmental characteristics such as U-shaped learning and the ability to generalise to novel forms. The model makes the novel experimental prediction that irregular nouns are less affected by damage than irregular verbs, even though irregular nouns are harder to learn.

Introduction

The acquisition of past tense in English has been long studied as a general touchstone for the development of morphology and productive linguistic rules in children. The general pattern of past tense formation for English is well understood; the overwhelming majority of English verbs have a simple past tense form which can be described as the addition of one of three variants of the '-ed' suffix to a base stem. A significant minority, particularly of relatively common verbs, take a so-called "irregular" form, which may or may not be systematically related to the stem form or to the forms of other words. The developmental course is nearly as well understood; children typically begin by correctly producing a small number of both regular and irregular forms, then produce characteristically "overregularized" forms for a small but significant fraction of their verb forms. They then appear to re-learn the correct form, producing the classic "U-shaped development" (Berko, 1958; Marcus et al., 1992). English noun plurals appear to share many of the same characteristics as verb past tenses. There is a similar general rule and a small, semi-regular group of common exceptions. Brown (1973), Marcus (1995), Marchman, Plunkett, and Goodman (in press) have described broadly similar time courses for the acquisition of plural nouns, including the U-shaped curve and approximately similar overall rates of overregularization. Many of the same phonotactic features (such as voicing assimilation and/or epenthesis) are relevant to the acquisition of both nouns and verbs.

Interpretations and models of this phenomenon vary; Marcus et al. (1992) (among others) have suggested that the emergence of the overregularized forms is indicative of the development of "a mental operation implementing the *-ed*-suffixation rule posited by grammarians" [pg. 8] capturing the formation of past tense and the memorization of specific lexical exceptions to the general rule. Other researchers (Daugherty & Seidenberg, 1992; Plunkett & Marchman, 1991) have argued instead that a single connectionist network is capable of producing this pattern, and thus that a single route suffices to explain the evidence.

Here we present a single-system, feedforward, connectionist model to compare the acquisition of noun and verb inflection head-to-head. To this end, we have constructed a model that simultaneously acquires noun plurals and verb past tenses. We use this model to determine the patterns of mastery (and errors) produced by a network which only has access to phonological representations of stems and their inflected forms, and their type and token frequencies. The patterns of performance in the network are then compared directly with the acquisition data for young children. We further describe the effects of damaging this network to compare performance with the data on impaired adult performance.

The Model

Network configuration

The connectionist simulation uses a multi-layer perceptron network using backpropagation of error (Rumelhart, Hinton, & Williams, 1986). The simulation was built using the PlaNet simulator (Miyata, 1991) using 130 units for the input layer, 160 units for the output layer, and 200 units as the hidden layer. The system was trained with a learning rate (η) of 0.1 and no momentum ($\alpha = 0$). Training was performed via a pattern update schedule, where each pattern was presented individually to the network (in random order) and training performed on each pattern.

Training corpus and representation

The training data for the simulations were taken from the CELEX corpus (Baayan, Piepenbrock, & Rijn, 1993); we extracted from this database all words which were monosyllabic, which contained no "foreign" sounds in their pronunciation and for which we had evidence that they could be used as nouns or verbs. This yielded a total corpus of 2626 stems, which encompassed 3226 total inflected types (2280 nouns and 946 verbs). Of

[1]Current address: Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282 (JUOLA@MATHCS.DUQ.EDU)

these types, 26 were irregular nouns and 122 were irregular verbs. For these words, we took the corresponding token frequencies (of the stems) from the Brown corpus (Kučera & Francis, 1967) as a rough measure of token frequencies in running speech. The token frequencies of words were individually tabulated as nouns and verbs, then the function $\log_2(freq^2 + 1)$ applied to these frequencies to flatten them into something more presentable to the network. The final variance was between 1 and 21 tokens/inflected type, meaning that the most frequent words appeared just over twenty times as often as the least. The final token frequencies were heavily dominated by nouns; of the 17129 tokens in the training set, 13045 were noun tokens (204 of them irregular) and 4084 were verb tokens (997 of them irregular).

The training corpus was prepared by converting the Moby symbolic pronunciation (Ward, 1997) into a large binary vector using a modification of the PGPfone alphabet representation (Juola & Zimmermann, 1996). Each phoneme was represented as a cluster of 16 binary phonetic features including aspects such as place, manner, and height of articulation. Each word was divided into onset–nucleus–coda constituents and right-justified within a CCCVVCCC template (e.g. the word "cat" (/kAt/) would be represented by the training pattern ##k#A##t, where '#' represents an absent sound). To this 128-bit pattern, two additional bits were appended representing the syntactic form to be inflected into, either the past tense (of a verb) or the plural (of a noun). The desired outputs were a similar encoding of the phonology of the inflected form, including an optional epenthetic vowel and final consonant.

Simulations were run under two training schedules; mass training and incremental training. In mass training, all the stems and inflected forms are available to the network throughout the training process. In incremental training, training starts out with a small number of high frequency words (20 types). The training set is then gradually expanded (5% new types per epoch) to include words of decreasing frequency until the entire corpus is absorbed. These two training schedules are intended to capture the distinction between input to the child and uptake by the child (Plunkett & Marchman, 1993).

Analysis techniques

All simulations were performed with five different random starting seeds to assess the reliability of the findings. Unless otherwise stated, all results reflect the mean over these five simulations. At the end of every epoch (which recall may contain over seventeen thousand individual pattern trials), the system was evaluated to determine what, and how much, had been learned. Every output pattern was examined to determine the nearest legal phoneme in each template position and then compared with the "correct" sequence. We specifically focus here on the number of correct inflections, and on the overregularization rate, defined as

$$\left(1 - \frac{\text{overregularized types}}{(\text{overregularized types} + \text{correct types})}\right) * 100\%$$

Each of the five starting points yielded a unique weight configuration after 115 increments and was used as the basis for the lesioning experiments. Increment 115 is the earliest point at which the network had seen the entire training corpus and, as might be expected, is the point with the worst overall performance on the training corpus. Because of the high error rate under "normal" circumstances, it is reasonable to assume that it would be the most sensitive to damage and therefore an appropriate time to lesion in search of interesting error patterns.

Each of the five networks was lesioned by randomly removing individual weights while leaving the basic network topology intact. The lesioning was performed at eighteen different levels, consisting of single percentages from 1% (leaving 99% of the network weights unchanged) to 9%, and at increments of ten percent from 10% to 90%. Each network was lesioned four times at each level, yielding a total of 360 separate experimental "subjects." (All lesion levels are henceforth referred to by percentage of connections *remaining*.)

We normalized lesioned performance by calculating it as a percentage of baseline performance of the undamaged network used as a base for each subject. Because in many cases, especially for irregular nouns and verbs, the baseline performance included errors, it occasionally happened that the performance of an individual subject on an individual category would exceed the baseline, resulting in apparently paradoxical performance levels that exceed 100%. In other words, under certain circumstances, damage can actually improve performance.

We further analyzed all 360 "subjects" against a set of properties we believe to be completely psycholinguistically *unsalient*. Specifically, for each word token, we determined whether the first pronounced phoneme was from the first or second half of the alphabet, and whether it was represented in ASCII by an odd or even number (odd or even "parity"). This analysis provides us with a baseline against which to compare breakdowns in performance along linguistically salient lines.

Acquisition Results

Figure 1 shows that a single system is capable of learning both noun plural and verb past tense morphology; in fact, performance on the regular types is near-perfect after the first full epoch of training. Performance on irregular types lags significantly; performance after 200 epochs of training yielded 80% of irregular noun types and 95% of irregular verb types that were correctly inflected.

An important consideration in the simulation of U-shaped learning curves is the initial acquisition of the correct forms for some irregular forms. Because the learning of a connectionist system is strongly influenced by the type frequency of training patterns, rare forms are generally only learned well when they are made especially salient to such systems. The statistical dominance of regular forms tends to make these more influential unless irregular forms are somehow made more salient. In particular, the initial error-free period of irregular production can be problematic for connectionist

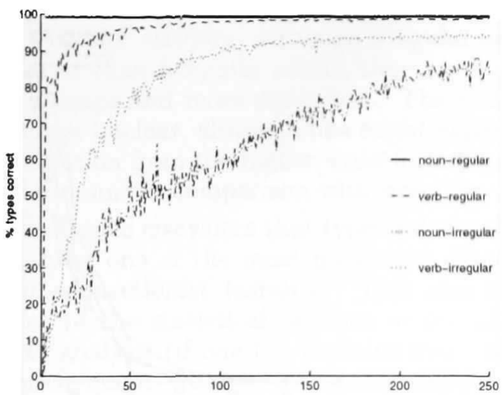

Figure 1: % of types correct by epoch (mass training).

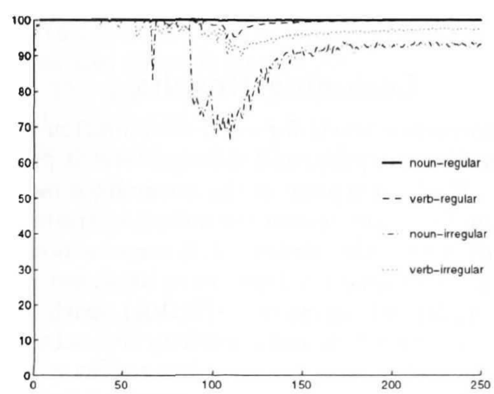

Figure 2: % of known types correct by increment.

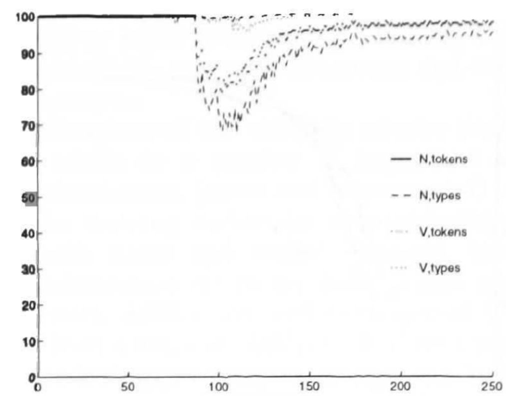

Figure 3: (1 overregularization rate) as a function of epoch (incremental training)

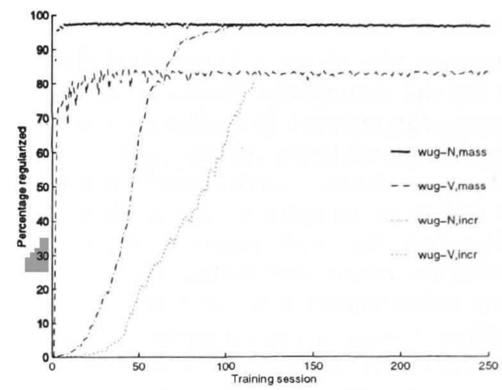

Figure 4: Regularization of nonce words by epoch or increment across training regimes.

systems (and does not, for instance, appear in figure 1).

In analysis of incremental training, it is important to keep in mind the difference between performance on words already presented to (and known by) the network, and words that will eventually be part of the training corpus but have not yet been added. Accordingly we display the mean number of *known* types correctly inflected at each increment.

As training progresses, the network learns to inflect correctly the entire training corpus. As in above, we see that performance on regular nouns eventually dominates regular verbs, as well as irregulars of all sorts, and that irregular nouns are the most difficult overall. However, figure 2 clearly demonstrates an the characteristic initial period where the networks' performances on irregular forms is perfect, (and better as a percentage of total vocabulary than that of regular forms). Overregularization performance (figure 3) shows a similar curve.

Studies of the time course and error rate of noun inflections in children are unfortuately rather rare; the best known is probably Marchman et al. (in press) and their response to the claims of Marcus (1995) about relative rates of overregularization of nouns and verbs. The Marchman et al. (in press) study is unfortunately rather limited in that they focused on the development of only five nouns and sixteen verbs, compared to the nearly 170 irregular forms modelled here. Nonetheless, our results are broadly similar in that, as found by Marchman et al. (and predicted by Marcus), overregularizations of nouns happened both earlier and more frequently than those of verbs. A significant finding is that the most frequent irregular words were remarkably resistent to overregularization; no word with a token frequency of greater than 243 (15, in our training set) was ever overregularized. Thus, moderately common words like "keep," "tell," and "let" (as well as extremely common forms like "see" and "man") were immune to overregularization, while only marginally less common words like "wife," "child," and "hold" were overregularized upon occasion.

The developmental course of generalization to novel stems is depicted in figure 4. For purposes of comparison, we also include the generalization profile under the mass training schedule of Simulation 1.

Figure 4 shows that the network is able to generalize the correct suffix to novel nouns and verbs. By the end of training over 90% of all novel nouns are inflected with the correct suffix as are over 80% of all novel verbs. Secondly, the final level of generalization of the network is independent of the training regime adopted (mass versus incremental) indicating that generalization is determined more by the training corpus than by the training method.

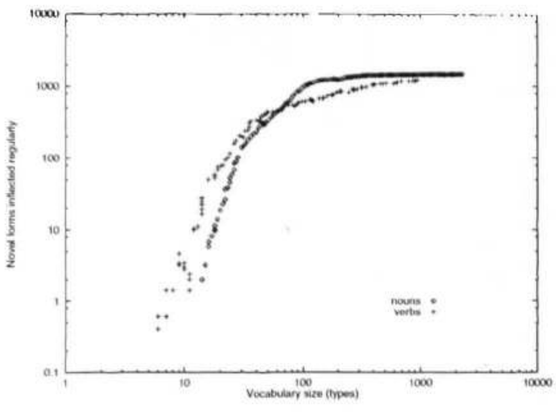

Figure 5: Evidence for critical mass effect; 'wug' rate against type rate scatter plot

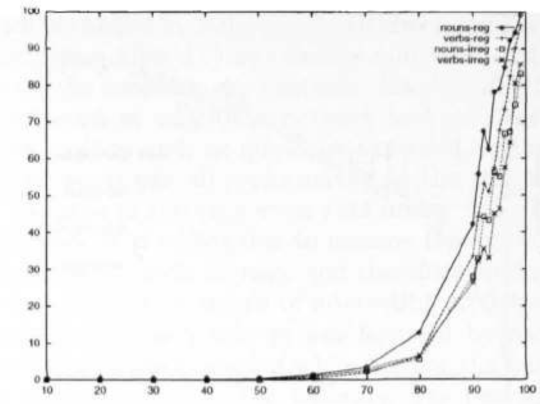

Figure 6: Normalized performance levels as a function of network remaining after lesion

Generalization also shows a pronounced developmental profile for the incremental training regime. During early training, the network is unable to add an appropriate suffix to a novel form. As the training vocabulary expands the generalization performance increases exponentially until it asymptotes at the levels indicated in figure 4. However, the developmental profile of generalization differs for nouns and verbs. The network regularizes novel verbs considerably later than novel nouns.

The profiles of development depicted in figure 4 suggest that a critical mass of nouns and verbs is required in the training set before high levels of generalization are achieved. Similarly, the sudden onset of overregularization errors depicted in figure 3 indicate a mass action effect at work. These findings are commensurate with earlier work (Plunkett & Marchman, 1993; Marchman & Bates, 1994). However, in the current simulations there is substantial delay between sudden increases in the various performance measures for nouns and verbs. The critical mass hypothesis predicts that these delays are directly related to the number of nouns and verbs in the corpus at different points in training.

Figure 5 shows the relationship between the average number of novel forms regularized and the number of types of a given syntactic category present in the training set at each point. In conjunction with figure 4, it provides support for this critical mass hypothesis in several ways. First, the curves are highly nonlinear, showing a rapid increase in regularization rate until the number of syntactic types in the training set reaches around 100 the rate of generalization increases dramatically. This is compatible with the idea of a threshhold effect. Secondly, the curves for nouns and verbs in figure 5 are nearly identical, showing that a similar process of mass action is operating for both syntactic types. The *developmental* delay between nouns and verbs shown in figure 4 reflects the composition of the input vocabulary at different stages in training. The small advantage of nouns over verbs in figure 5 reflects the relative homogeneity of noun inflections compared to verbs (recall that there are many more irregular verbs than irregular nouns).

Lesioning Results

Mean performance levels for each combination of regularity, syntactic category, and damage level is presented as figure 6. Each data point is the normalized mean performance for 20 subjects over the subset of training data corresponding to each category (e.g. regular nouns).

Individual performance level were analyzed under a four-way analyis of variance (ANOVA), with damage level, syntactic category, and regularity treated as within groups factors and base network treated as a between groups factor. The single most significant predictor of network performance was the amount of damage; greater damage resulted in lower performance ($F(3,17) = 422.093$). There were also significant main effects for regularity (regular forms were more resistant to damage; $F(3,1) = 251.328$) and for syntax (nouns were more resistant to damage; $F(3,1) = 37.357$, 1). No significant main effect of underlying network was noted. All of the above effects were significant at $p \ll 0.001$.

Significant interactions were noted between damage level and regularity, and between damage level and syntactic category (both significant at $p \ll 0.001$) ($F(3,17) = 15.238$, 4.330), although these interactions disappeared when we controlled for floor and effects.. A further interaction was observed between syntactic category and regularity at a lower but still highly significant level ($F(3,1) = 12.064$, $p < 0.003$). This interaction represents the fact that regular forms are more strongly influenced by syntactic category (regular nouns are more superior to regular verbs than irregular nouns are to irregular verbs). No significant three-way (or greater) interactions were observed.

In the 1st/2nd odd/even experiment, we found similar main effects of damage. We also found highly significant ($F(3,1) = 66.612$,1; $p \ll 0.001$) effects of first half of the alphabet superiority (which may reflect underlying distribution, as discussed below), but no significant main effect of parity or of network.

The syntax-by-regularity interaction is somewhat surprising, as we have found that irregular verbs are, in general, learned more accurately and faster than irregular nouns. The lesioning data presented here indicate an opposite trend for atrophy. Although irregular verbs are learned faster than irregular nouns, they are more sensitive to damage and more easily lost. The reasons for this are as yet unclear, although one might expect a certain contribution from the higher word-form frequencies of irregular nouns in comparison with irregular verbs.

We have argued elsewhere that type and/or token frequency can be one of the most important determining factors in connectionist learning. This also may explain some of the statistical oddities in the alphabet-half/parity analysis (if one bin contains many more, or many more frequent, words). To examine this possibility, we divided the training set into frequency quartiles (approximately 25% of types in each quartile, irrespective of syntactic category or regularity) and performed a similar 5-way ANOVA. In this reanalysis, damage remained significant ($F(4,17) = 415.847$, $p \ll 0.001$), while frequency quartile was also strongly significant ($F(4,3) = 162.442$, $p \ll 0.001$). The main effect of regularity remained significant ($F(4,1) = 312.416$, $p \ll 0.001$), while the main effect of syntactic category evaporated into non-significance.

Discounting the damage interactions (which, as above, can be attributed to floor effects), we found significant interactions between syntax and regularity ($F(4,1) = 385.872$, $p \ll 0.001$), syntax and frequency ($F(4,3) = 18.204$, $p \ll 0.001$), and regularity and frequency ($F(4,3) = 95.427$, $p \ll 0.001$). Nouns appear to be more sensitive to frequency effects than verbs, and irregulars appear to be more sensitive to frequency effects than regulars.

Because English words are not uniformly distributed throughout the alphabet, there isn't a uniform distribution of word types to the various bins. In fact, there are more nouns than verbs, and more regular than irregular forms. Similarly, there are more words in the first half of the alphabet than in the second, as well as more words of even parity than of odd. If we postulate a subset of "difficult" (i.e., hard to learn, sensitive to damage) word types, it is unlikely that they will be evenly distributed among any particular type taxonomy, with the result that the taxon with fewer difficult words would presumably perform better, perhaps significantly so. It would appear that verbs are the "difficult" forms (based on their relative inferiority with respect to nouns). Certainly, the absence of significant three-way interactions would be predicted if difficult forms were uniformly distributed within the individual categories of irregular noun, irregular verb, and so forth. However, the frequency analysis suggests that the "difficult" forms are in many cases simply the rare ones; common verb forms are happily learned (and strongly retained), while uncommon noun forms are lost. Although even when frequency is controlled for, a regular-superiority main effect is still present, arguing that, in line with common sense, irregulars may be more "difficult" than irregulars.

Psycholinguistic comparison

As far as we are aware, the simulations reported here are the first to track the developmental trajectory of a neural network trained on a realistic corpus of English nouns (2280 types) or verbs (946 types)[2]. Indeed, the model incorporated *all* the monosyllabic nouns and verbs from the Brown corpus.

The performance of the networks mimics that of children and adults in a number of important respects. First, all simulations (mass and incremental) are able to learn the training corpus to near-perfection (above 99% for both nouns and verbs). Second, the profile of overregularization errors for both nouns and verbs in the network mimics the well-documented U-shaped profile of development in children. For verb past tense forms, this result replicates and extends the findings reported for models trained on smaller vocabularies (Plunkett & Marchman, 1993, 1996). However, this is the first demonstration in which a neural network model has simulated the well-known U-shaped development for noun plurals, in particular producing a initial phase of error-free performance. This finding flatly contradicts the claim of Marcus (1995):

> 'Because irregular nouns plurals are so rare, there is unlikely to ever be a stage in which irregular plurals dominate regular plurals; hence the Rumelhart & McClelland model would probably overregularize even its earliest plurals.' (p. 450)

Our model not only produces initial error-free performance on noun plurals but does so in the context of initial error-free performance on verb past tense forms.

Third, the onset of overregularization errors on nouns tends to occur earlier than overregularization errors on verbs in the network. This result is also consistent with the data reported in Marchman et al. (in press). The data for the four children analysed in Marcus (1995) are heterogenous in this respect—Adam overregularizes nouns before he overregularizes verbs, Eve and Sarah show the reverse pattern and Abe starts overregularizing nouns and verbs around the same time[3]. However, both studies demonstrate that the rate of overregularization for nouns is greater than that for verbs. This is also true of the simulation as can be seen in figure 3. Overregularization errors are also less likely to occur on forms with a high token frequency. Again, this replicates the results of earlier work with smaller verb vocabularies and demonstrates that the frequency effect scales up to larger vocabularies and extends to noun morphology.

As ye rise, so shall ye fall. There is unfortunately rather little data on *expected* loss after injury, although (Bishop, 1997) suggests that developmental disorders hit verb acquisition harder than nouns. The network simulations agree to the extent that verbs are generally more impaired for a particular level of lesioning. Regular forms are more robust in the face of damage than irregular forms in these networks. Results from humans

[2]MacWhinney and Leinbach (1991) describe a simulation incorporating an even larger corpus of verbs and a wider range of inflections. However, their paper does not provide a detailed analysis of the developmental trajectory of the network's performance. In particular, they provide no account of early U-shaped learning.

[3]Onsets of verb overregularization for these four children are taken from Marcus et al. (1992).

(adults and children) also show this pattern of atrophy though the opposite pattern of deficit can be observed, i.e. there is a double-dissociation between regular and irregular forms. It might also be predicted that aspects of language which are difficult to learn are the most fragile. Our network generally supports this statement, with the exception that irregular nouns, which are more difficult to learn than irregular verbs, are nonetheless better preserved. We await with interest to see whether this also holds in humans.

Conclusions

We have thus presented a single-route or, more accurately, single-process model, based on an associative network, that is capable of inflecting verb stems to produce past tenses or noun stems to produce plurals. It handles both regular and irregular forms with reasonable accuracy, despite having a much larger vocabulary than many related projects e.g. Daugherty and Seidenberg (1992), Plunkett and Marchman (1991, 1993). It produces linguistically plausible generalizations, capturing aspects such as performance on nonce words. In further analyses, we have also shown that these networks provide generalization of voicing assimilation even in the absence of syntactic information, and regularized inflection of cross-categorial processes such as denominal verbs or deverbal nouns.

This system thus demonstrates that a single route to inflectional morphology is capable of producing the performance necessary for a psycholinguistically meaningful model of noun and verb inflection, even on a very limited set of information (excluding semantics entirely, for example) and provides a baseline describing what can be done in this limited domain and to which new performance can be compared, as representations and information improve.

Acknowledgements

This research was funded by ESRC grant 70 (to Kim Plunkett).

References

Baayan, H., Piepenbrock, R., & Rijn, H. van. (1993). *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia: Linguistic Data Consortium.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150-177.

Bishop, D. V. M. (1997). *Uncommon understanding : Development and disorders of language comprehension in children*. Hove, East Sussex, UK: Psychology Press.

Brown, R. (1973). *A first language : The early stages*. Cambridge, MA: Harvard University Press.

Daugherty, K., & Seidenberg, M. S. (1992). Rules of connections? the past tense revisited. In *Proceedings of the fourteenth annual conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Juola, P., & Zimmermann, P. (1996). Whole-word phonetic distances and the PGPfone alphabet. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP-96)*. Philadelphia, PA.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualisations: Revising the verb learning model. *Cognition*, *29*, 121–157.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development : A test of the critical mass hypothesis. *Journal of Child Language*, *21*(2), 331–6.

Marchman, V., Plunkett, K., & Goodman, J. (in press). Overregularization in English plural and past tense inflectional morphology. *Journal of Child Language*.

Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T., & Xu, F. (1992). Overregularisation in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(Serial No. 228).

Marcus, G. F. (1995). Children's overregularization of English plurals : A quantitative analysis. *Journal of Child Language*, *22*(2), 440–60.

Miyata, Y. (1991). *A user's guide to PlaNet version 5.6 : A tool for constructing, running, and looking into a PDP network.*

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, *38*, 43–102.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21–69.

Plunkett, K., & Marchman, V. A. (1996). Learning from a connectionist model of the English past tense. *Cognition*, *61*, 299–308.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. In (Vol. 2: Psychological and Biological Models, pp. 318–362). The MIT Press.

Ward, G. (1997). *Moby pronunciator*. 3449 Martha Ct., Arcata, CA, USA.