

How The Constraints On English Compound Production Might Be Learnt From The Linguistic Input: Evidence From 4 Connectionist Models

J.A. HAYES¹, V.A. MURPHY¹, N.DAVEY² AND P.M. SMITH¹

*Departments of Psychology¹
and Computer Science²,
University of Hertfordshire,
College Lane, Hatfield,
AL10 9AB, United Kingdom.
E-mail: J.hayes@herts.ac.uk*

Native English speakers include irregular plurals in English noun-noun compounds (e.g. *mice chaser*) more frequently than regular plurals (e.g. **rats chaser*) (Gordon, 1985). This dissociation in inflectional morphology has been argued to stem from an internal and innate morphological constraint as it is thought that the input to which English speaking children are exposed is insufficient to signal that regular plurals are prohibited in compounds but irregulars might be allowed (Marcus, Brinkmann, Clahsen, Wiese & Pinker, 1995). In addition, this dissociation in English compounds has been invoked to support the idea that regular and irregular morphology are mediated by separate cognitive systems (Pinker, 1999). The evidence of the neural network models presented here is used to support an alternative view that the constraint on English compounds can be derived from the general frequencies and patterns in which the two types of plural (regular and irregular) in conjunction with the possessive morpheme occur in the input.

1. Introduction

1.1. *Compounding in English*

Psycholinguistic research has shown that English compound words with irregular plural nouns in first position (e.g. *mice-eater*) are produced far more frequently than compound words with regular plural nouns in first position (e.g. **rats-eater*), (Gordon, 1985).

1.2. *Explanation according to the Dual Mechanism model*

The dual mechanism model (Pinker, 1999), proposes that irregular nouns and their plurals are stored as memorised pairs of words in the mental lexicon (e.g. *mouse-mice*) but that regular plurals are produced by the addition of the [-s] morpheme to the regular stem at a post lexical stage (e.g. *rat + s = rats*). Compounds are created in the lexicon. Thus as irregular plurals are stored in the lexicon they are available to be included within compound words. However, as

only the singular stems of regular nouns are stored in the lexicon the plural form is never available to be included within compound words (Marcus et al, 1995).

1.3. A Single Route Associative Memory Based Explanation of Compounding

An explanation of the constraints on English compounding based on the frequency of co-occurrence of items in the linguistic input has not been considered to date. It is argued here that children may not include the high type frequency regular [-s] plural morpheme in the middle of words such as compounds because they will always have heard it at the end of words. Thus, to include the regular plural morpheme [-s] in the middle of words contravenes a very clear pattern discernable from the input. Furthermore, frequency counts of a sample of the CHILDES (Child Language Data Exchange System) corpora (McWhinney & Snow, 1985) have shown that the plural [-s] morpheme is rarely followed by a second noun (Hayes, Murphy, Davey, Smith and Peters, 2002). Importantly, a different pattern is found with the possessive [-'s] morpheme since it is nearly always followed by a second noun. Therefore, it might be that a noun rarely follows the regular plural [-s] morpheme (i.e. patterns such as “**rat[s] chaser*” do not occur) because the pattern “*noun – morpheme[-s]- noun*” is reserved for marking possession (such as *rat’s tail*). Interestingly in other languages that do not have this competition between the plural and possessive morpheme such as Dutch (Schreuder, Neijt, van der Weide & Baayen, 1998) and French (Murphy, 2000), regular plurals are allowed within compounds. Irregular plurals may, however, appear in English compounds as they are not formed by the addition of the plural [-s] morpheme. Thus, irregulars do not compete with the possessive structure and as such may be followed by a second noun in a compound. The role of the plural [-s] morpheme as a predictor of word finality and the competitive relationship between the plural and possessive [-s] morpheme is examined here using a series of connectionist models.

2. Neural net modeling

2.1. Experiment 1.

Experiment 1, was designed to test any role that [s] might play in indicating word finality in a stream of concatenated letters. A neural network was trained on a concatenated stream of 200 sentences of child directed speech taken from CHILDES (MacWhinney & Snow, 1985). A word-ending marker was attached to each word and the words (including a word-ending marker) were

concatenated to form a stream of 3596 letters. Each letter was encoded using one of 26 random 5 bit vectors (one for each letter in the alphabet). The word-ending marker was encoded using a 27th 5-bit vector. The network was required to predict the next letter it expected to occur given the letters it had seen previously. At the beginning of a word the error was high but as more letters were presented to the network the error decreased until it was at its lowest at the end of the word. It was hypothesised that on a "next letter" prediction task of this kind, a neural network would learn that after the input [-s] there was a high probability that the next input would be a word ending marker. A simple recurrent network (SRN) was used so that at any point in time the state of the hidden units at the previous time step were used as additional input (Elman, 1990).

Results: The network's ability to learn that [-s] is a good predictor of word finality was tested using 19 unseen words that ended in [-s] and 19 unseen words that ended in other letters. The network was more accurate (i.e. the error was lower) at predicting a word ending marker after [-s] than after all other letters combined, $t(18) = -2.08$, $p = 0.05$.

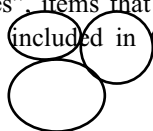
2.2. Experiment 2.

The objective of this SRN was to learn whether highly consistent patterns in the input (i.e. that a plural noun is rarely followed by another noun while a possessive noun is always followed by a second noun) can drive learning about how to manipulate plurals within noun-noun compounds. The network was required to predict the next word to occur given the words it had seen previously. Nouns, verbs, determiners and adjectives were used to make up legal English sentences. Similarly to the performance of Elman's (1990) model, this network was expected to be able to make a first order distinction between the function of the various syntactic types (learning that the words could be classified as nouns and verbs, determiners and adjectives). Furthermore, although the possessive and the plural [-s] were encoded in exactly the same manner in the input, it was predicted that the network would learn a second order distinction (a distinction that could only be learnt once the first order distinction had been learnt) that only "verbs" could appear after some [-s] morphemes and only "nouns" could appear after other [-s] morphemes. The network was trained on one group of nouns that were represented as having the properties of singulars, possessives and plurals (e.g. *hen*, *hens*, *hen's*). A second set was only represented as singulars and plurals (*coat*, *coats*), a third group was only represented as singulars and possessives (*wig*, *wig's*) and a

fourth group was represented as singulars only (*bar*). Possessives and regular plural nouns were differentiated from singular nouns because they were encoded as ending in [-s]. A set of deverbal nouns were encoded by using the localist code for the verb (e.g., *drive*) and then representing the fact that they ended in the derivational affix [-er] (e.g., *driv(e) + e r = driver*). Nouns represented as singulars were followed by verbs ending in [-s] (to represent the third person singular). Nouns represented as plurals were followed by verbs that did not end in [-s]. Possessives were followed by deverbal nouns (i.e. verbs plus the the derivational morpheme [-er]). Thus the differences between plural and possessive nouns were only represented to the network by the fact that they occurred in different sequences in the input. It was predicted that the tokens making up the 4 groups of words (1.singulars, possessives and plurals; 2.singulars and plurals; 3.singulars and possessives; 4. singulars only) would cluster together in the hidden layer representations.

The network was trained on a concatenated stream of 2000 legitimate English sentences constructed from a lexicon of 37 words. Nouns with the properties of singulars, plurals and possessives were included with equal type frequencies in the input. A sentence-ending marker was attached to each sentence and the sentences (including the sentence-ending marker) were concatenated to form a stream of 14,600 words. Each word (including the sentence-ending marker) was encoded using a 38-bit localist coding scheme. The presence or absence of [-s] at the end of a word was also explicitly coded using 2 additional input units because the focus of this model was to investigate whether items ending in [-s] (phonetically identical items) could be differentiated depending on the co-occurrence patterns in which they occurred in the input. Thus, for example, the code for the singular noun *cat* would be *localist code for cat + [-s] off*, for the plural noun *cats* it would be *localist code for cat + [-s] on and* for the possessive noun *cat's* it would be *localist code for cat + [-s] on*. The presence or absence of [-er] at the end of a word was also explicitly coded using 2 additional input units.

Results: Figure 1, shows a principle component analysis of the hidden unit representations. The dotted line superimposed on the PCA diagram shows the divide between the way nouns and verbs are represented in the hidden units. It is also apparent that singular nouns (i.e. nouns which do not end in an [-s]) are represented separately from plurals and possessives (i.e. nouns which do end in an [-s]). Most interestingly, nouns which were included in the training set as both "plurals and possessives", items that were only included as "possessives" and items which were only included in the "plural" form are all represented separately.



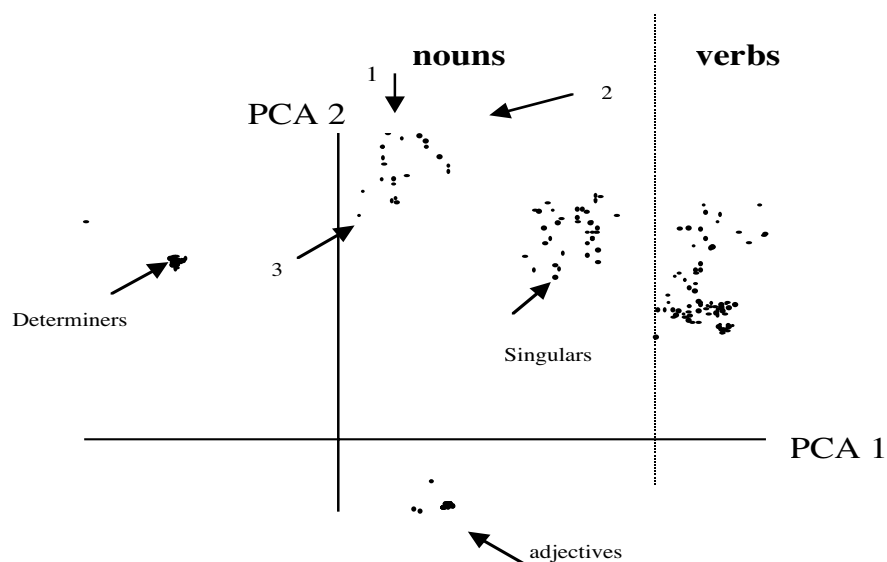


Figure 1. PCA of the hidden layer representations in Experiment 2: area 1 = plurals and possessives; area 2 = possessives; area 3 = plurals

2.3. Experiment 3

In experiment 2, the network appeared to learn that nouns that end in [-s] are different from nouns which do not end in [-s]. The network was also able to group nouns that in the training set were behaving as “plural and possessive” or as “plural” or “possessive” only. However, the network could not totally disambiguate plurals from possessives because both items were encoded in exactly the same way, the only way to separate the two types of item was to learn that some words followed some words ending in [-s] and others words followed other words ending in [-s]. In this third simulation, the same network that was used in experiment 2 was amended to include an extra input unit that encoded whether the object of the sentence in which the word occurred was either a plural or a singular noun. It was predicted that with the addition of this minimal semantic information the network would be able to disambiguate “plural” nouns from “possessive” nouns. Thus, an additional input unit was

included to encode whether the word was part of a “singular” or a “plural” sentence. Hence, although both “plural” and “possessive” words were coded as ending in [-s] only plural items were encoded as ending in [-s] and being plural, and possessive words were encoded as ending in [-s] but being singular. It was predicted that in the hidden units the plural and possessive nouns would be represented separately.

Results: Figure 2, shows a principle component analysis of the hidden unit representations. From the PCA it is evident that once again nouns and verbs are represented separately in the hidden units. With the addition of the semantic information it is now evident that singular, plural and possessive nouns are all represented separately. Interestingly, both plurals and singulars i.e., items that may be followed by a verb lie in similar positions on the x axis, while the possessives are clustering with adjectives i.e., with other items that are followed by nouns.

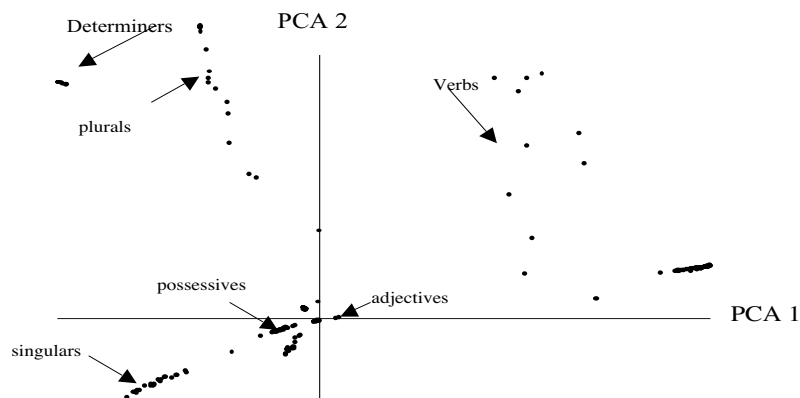


Figure 2. PCA of the hidden layer representations in Experiment 3

2.4. Experiment 4

This network was trained on a large training set of real child directed speech in which the frequencies with which the various types of morphology occurred

were not manipulated in any way. The syntactic type of each word was used as the input to the network. The network was required to predict the syntactic type of the next input it expected to receive given the syntax of the previous inputs. The models in experiments 2 and 3 were capable of learning about grammatical type from a training set in which each word token was encoded using a localist coding scheme. In these earlier models items in the training set were not explicitly coded as being representatives of a particular syntactic type (e.g. as being nouns or verbs). Instead, learning about the distinct linguistic functions that the different syntactic types perform emerged during training. However, a disadvantage of these models was that it was only possible to use a small lexicon of words because of the complexity of the learning task. The model reported here was trained on a much larger training set than our previous models. This simulation sought to reproduce the behaviour of an older child, with a much larger vocabulary, who has knowledge, though perhaps not at a metalinguistic level, of the different functions that are performed by the different syntactic types.

The frequency in which regular and irregular plurals and possessives were included in the training set was determined by the frequency in which they appeared in the child directed speech that was used as the input to the model. The performance of the network was investigated using a syntactic type prediction task in which one of three syntactic types was input (a possessive, a regular plural or an irregular plural) and the network predicted which syntactic type it expected to see next in the input stream. The difference (error) between this predicted output and the output for noun, verb, other and word ending was calculated. It was predicted that the error would be high for all items after possessives except nouns. Conversely it was predicted that there would be a high error on predicting a noun after a plural of either kind.

Results: The error on producing the target output was recorded after the network was presented with the test sequences. Many runs of the simulation were carried out but each produced almost identical results.

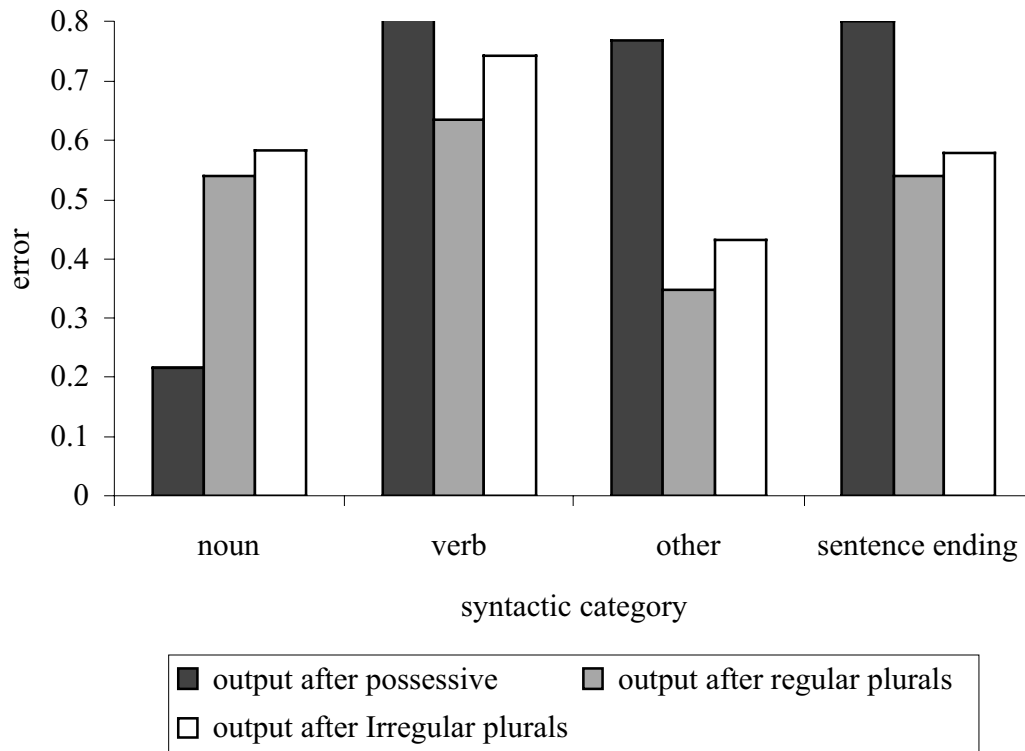


Figure 3. Error on producing nouns, verbs, other items and word endings after possessives, regular plurals and irregular plurals

Figure 3 illustrates that at a descriptive level the error on producing a singular noun after a possessive was about half as high as the error on producing a singular noun after a plural of either type. The network also learnt that the syntactic categories that make up other items and sentence-ending markers can follow plurals but not possessives.^a

^a The network, produced a high rate of error when the target output after a plural noun was a verb, despite the fact that in the input verbs followed regular plurals (25% of the time that regular plurals occurred) and irregular plurals (11% of the time that irregular plurals occurred). However, the training set contained very few verbs (6.24 % of the training set). Given that verbs were so underrepresented in the input it was unlikely that they would be predicted as the next item in a next word prediction task to any great extent.

Discussion

From experiment 1, it is evident that a neural net model trained on child directed speech was able to learn that [-s] is associated with word finality. This overwhelming pattern of [-s] at the end of words might influence language learners to omit [-s] from the middle of words. Experiment 2, showed that the model was able to learn that [-s] followed by one set of words was different from [-s] followed by a different set of words even though the [-s] was encoded in exactly the same way in the input. The same might be true for the language learner. Both the possessive [-s] and the plural [-s] sound the same phonetically but the patterns in which the two different types of morpheme appear in the input may be enough to drive learning that one type of morpheme is used in some circumstances but not in others. From experiment 3, it is evident that learning that the plural and possessive morphemes are only legal in certain sequences may be refined as the child learns that semantically the plural morpheme refers to many things while the possessive morpheme refers to one thing. In Model 4, syntactic category was explicitly encoded in the input and real child directed speech was used as input and thus the different syntactic categories were represented in the actual frequency that they occurred in real child directed speech. Under these more realistic input conditions there was still a suggestion that the network was able to recognise that the *noun –morpheme [-s]* pattern occurred in different patterns when it was plural to when it was singular. Specifically, the network showed some indication of being able to discern that nouns follow possessives but not plurals of either type and also of being able to detect that “other items” and word ending markers follow plurals of either type but not possessives. These four models taken together would seem to provide evidence for an associative account of compounding. In this associative account, the language learner notices that the [-s] morpheme tends to occur at the end rather than in the middle of a word. Furthermore, simply by exposure to the [-s] morpheme (i.e. without the plural or the possessive [-s] morpheme being explicitly labelled as being different from each other), the language learner becomes aware that the same [-s] morpheme occurs in different patterns in the input. With the addition of the absolute minimum of semantics, namely the numerical context in which the phrase is uttered, the language learner seems able to differentiate between the plural and the possessive morpheme. The possessive morpheme may be followed by a second noun but the plural morpheme may not be followed by a second noun. Thus when faced with a noun-noun compound the language user may delete the plural morpheme from the end of the first noun not because regular items of morphology are

represented in a particular manner in the brain but simply because this pattern is used to denote possession not plurality.

References

1. Marcus, G. F., Brinkmann, U., Clahsen, H., Weise, R., and Pinker, S., *Cognitive Psychology*, 29, 189 (1995).
2. Pinker, S., *Words and Rules*. London: Weidenfeld & Nicholson (1999)
3. Gordon, P. (1985). *Cognition*, 21, 73 (1985).
4. McWhinney, B. and Snow, C.E., *Journal of Child Language*, 12, 271 (1985).
5. Hayes, J.A., Murphy, V.A., Davey, N., Smith, P.M. and Peters, L. In W. D. Gray & C.D. Schunn (Eds.) *Proceedings of the 24th Annual Conference of the Cognitive science Society*. Mahwah, NJ: Lawrence Erlbaum Associates (2002).
6. Screuder, R., Neijt, A., van der Weide, F., & Baayen, R.H., *Language and Cognitive Processes*, 13, 551 (1998).
7. Murphy, V. A. *Language Learning*, 50, 153 (2000).
8. Elman, J. L. *Cognitive Science*, 14, 179 (1990).