

Modelling Atypical Syntax Processing

Michael S. C. THOMAS
School of Psychology
Birkbeck College, Malet St.,
London WC1E 7HX
m.thomas@bbk.ac.uk

Martin REDINGTON
School of Psychology
Birkbeck College, Malet St.,
London WC1E 7HX
m.redington@ucl.ac.uk

Abstract

We evaluate the inferences that can be drawn from dissociations in syntax processing identified in developmental disorders and acquired language deficits. We use an SRN to simulate empirical data from Dick et al. (2001) on the relative difficulty of comprehending different syntactic constructions under normal conditions and conditions of damage. We conclude that task constraints and internal computational constraints interact to predict patterns of difficulty. Difficulty is predicted by frequency of constructions, by the requirement of the task to focus on local vs. global sequence information, and by the ability of the system to maintain sequence information. We generate a testable prediction on the empirical pattern that should be observed under conditions of developmental damage.

1 Dissociations in language function

Behavioural dissociations in language, identified both in cases of acquired brain damage in adults and in developmental disorders, have often been used to infer the functional components of the underlying language system. Generally these attempted fractionations appeal to broad distinctions within language. However, fine-scaled dissociations have also been proposed, such as the loss of individual semantic categories or of particular linguistic features in inflecting verbs. Here, we consider the implications of developmental and acquired deficits for the nature of syntax processing.

1.1 Developmental deficits

A comparison of developmental disorders such as autism, Downs syndrome, Williams syndrome, Fragile-X syndrome, and Specific Language Impairment reveals that dissociations can occur between phonology, lexical semantics, morpho-syntax, and pragmatics. The implications of such fractionations remain controversial but will be contingent on understanding the developmental origins of language structures (Karmiloff-Smith,

1998). These processes remain to be clarified even for the normal course of development.

In the area of syntax, Fowler (1998) concluded that a consistent picture emerges. Individuals with learning disabilities are systematic in their grammatical knowledge, follow the normal course of development, and show similar orders of difficulty in acquiring constructions. However, such individuals can often handle only limited levels of syntactic complexity and therefore development seems to terminate at a lower level.

While there is great variability in linguistic function both across different disorders and within single disorders, this cannot be attributed solely to differences in ‘general cognitive functioning’ (e.g., as assessed by problem solving ability). Syntax acquisition is therefore to some extent independent of IQ. However, adults with developmental disorders who have successfully acquired syntax typically have mental ages of at least 6 or 7, an age at which typically developing children also have well-structured language. The variability in outcome has been attributed to various factors specific to language, including verbal working memory and the quality of phonological representations (Fowler, 1998; McDonald, 1997). Most notably, disorders with different cognitive abilities show similarity in syntactic acquisition. The apparent lack of deviance across heterogeneous disorders has been used to argue for a model of language acquisition that is heavily constrained by the brain that is acquiring the language (Newport, 1990).

1.2 Acquired deficits in adulthood

One of the broadest distinctions in acquired language deficits is between *Broca’s* and *Wernicke’s* aphasia. Broca’s aphasics are sometimes described as having greater deficits in grammar processing, and Wernicke’s aphasics as having greater deficits in lexical processing. The dissociation is taken to support the idea that the division between grammar and the lexicon is one of the constraints that the brain brings to language acquisition.

Dick et al. (2001) recently argued that four types of evidence undermine this claim: (1) *all* aphasics

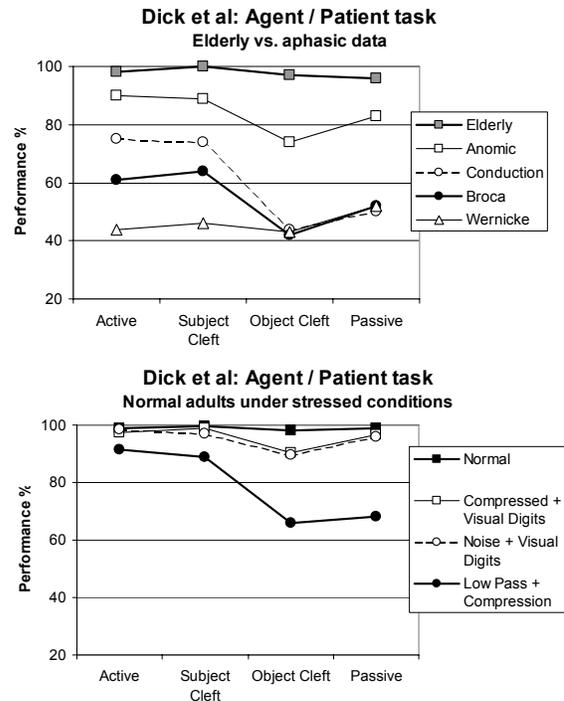
have naming deficits to some extent; (2) apparently agrammatic patients retain knowledge of grammar that can be exhibited in grammaticality judgements; (3) grammar deficits are found in many populations both with and without damage to Broca's area, the reputed seat of syntax in the brain; and (4) aphasic symptoms of language comprehension can be simulated in normal adults by placing them in stressed conditions (e.g., via manipulating the speech input or giving the subject a distracter task). Dick et al. pointed out that in syntax comprehension, the constructions most resilient in both aphasic patients and normal adults with simulated aphasia are those that are most regular or most frequent, and conversely those liable to errors are non-canonical and/or low frequency. Dick et al. (2001) illustrated these arguments in an experiment that compared comprehension of four complex syntactic structures:

- Actives (e.g., *The dog [subject] is biting the cow [object]*)
- Subject Clefts (e.g., *It is the dog [subject] that is biting the cow [object]*)
- Passives (e.g., *The cow [object] is bitten by the dog [subject]*)
- Object Clefts (e.g., *It is the cow [object] that the dog [subject] is biting*)

The latter two constructions are lower frequency, and have non-canonical word orders in which the object precedes the subject. Dick et al. tested 56 adults with different types of aphasia on a task that involved identifying the agent of spoken sentences. Patients with all types of aphasia demonstrated lower performance on Passives and Object Clefts than Actives and Subject Clefts. Moreover, normal adults given the same task but with a degraded speech signal (either speeded up, low-pass filtered, or with noise added) or in combination with a distracter task (such as remembering a set of digits) produced a similar profile of performance to the aphasics (see Figure 1).

Dick et al. (2001) argued that the common pattern of deficits could be explained by the Competition Model (MacWhinney & Bates, 1989), which proposes that the difficulty of acquiring certain aspects of language and their retention after brain damage could be explained by considering *cue validity* (the reliability of a source of information in predicting the structure of a target language) and *cue cost* (the difficulty of processing each cue). Cues high in validity and low in cost, such as Subject-Verb-Object word order in English, should be acquired more easily and be relatively spared in adult breakdown. The proposal is that for a given language, any domain-general processing system placed under sub-optimal

Figure 1. Aphasic and simulated (human) aphasic data from Dick et al. (2001)



conditions should exhibit a similar pattern of developmental or acquired deficits. Thus Dick et al. predicted that a connectionist model trained on an appropriate frequency-weighted corpus would show equivalent vulnerability of non-canonical word orders and low frequency constructions under conditions of damage. In contrast to the inferences drawn from developmental deficits, the focus here is on attributing similarities in patterns of acquired deficits to features of the problem domain rather than constraints of the language system.

2 Computational modelling

Proposals that site the explanation of behavioural data in the frequency structure of the problem domain (here, the relative frequency of the construction types) are insufficient for three reasons: (1) language comprehension is not about passive reception. The language learner must do something with the words in order to derive the meanings of sentences. It is the nature of the transformations required that crucially determines task difficulty, which statistics of language input alone cannot reveal. (2) Whatever the statistics of the environment, such information must be accessed by an implemented learning system. This system may be differentially sensitive to certain features of the input, and it may find certain transformations more computationally expensive than others, further modulating task difficulty. (3) In the context of atypical syntax processing in developmental and acquired disorders, behavioural

deficits are caused by changes in internal computational constraints. Without an implemented, parameterised learning system, we can have no understanding of how sub-optimal processing conditions generate behavioural deficits in syntax processing. To date, this issue has been relatively under-explored.

The choice of learning system is evidently of importance here. In this paper, we explore the behaviour of a connectionist network, since these systems have been widely applied to phenomena within cognitive and language development (Elman et al., 1996) and more recently to capturing both atypical development and acquired deficits in adults (Thomas & Karmiloff-Smith, 2002, 2003).

3 Simulation Design

Our starting point is a set of models of syntax acquisition proposed by Christiansen and Dale (2001). These authors employed a simple recurrent network (SRN; Elman, 1990), an architecture that is the dominant connectionist model of sequence processing in language studies and in sequence learning more generally. As is typical of current connectionist models of syntax processing, the Christiansen and Dale (henceforth C&D) model focuses on small fragments of grammar and a small vocabulary. Nevertheless, it provides a useful platform to begin considering the effects of processing constraints on syntax processing.

The following models performed a prediction task at the word level. At each time step, the network was presented with the current word and had to predict the next word in the sentence. This component of the task induces sensitivity to syntactic structures. A localist representation was used, with each input unit corresponding to a single word. The artificial corpus consisted of 54 words and included 6 nouns, 10 verbs, 5 adjectives, and 10 functions words. Nouns and verbs had inflected forms represented by separate word units (N: stem, pluralised; V: stem, past tense, progressive, 3rd person singular).

C&D investigated the effect of several cues on syntax acquisition, such as prosody, stress, and word length. *Prosody* was represented as utterance boundary information that occurred at the end of an utterance with 92% probability. The utterance boundary cue was represented by an additional input and output unit.

Distributional cues of where words appeared in various sentences, along with utterance boundary information, were available to all networks. We refer to the networks that received only these cues as the “basic” model. We also tested a second set of “multiple cue” networks that also received cues about *word length* and *stress*. *Word length* was

encoded with thermometer encoding, with one to three units being activated according to the number of syllables in the input word. In English, longer words tend to be content words. This was reflected in the vocabulary items that were selected for the grammar. *Stress* was encoded as a single unit that was activated for content words, which are stressed more heavily. The word length and stress units were present both as inputs and outputs, so that multiple cue networks had 59 input and output units to represent the words and cues.

3.1 The materials

The input corpus was a stochastic phrase structure grammar, derived from the materials used by C&D (2001). The grammar featured a range of constructions (imperatives, interrogatives and declarative statements). Frequencies were based on those observed in child-directed language. We added passives, subject and object cleft constructions to the grammar, which is illustrated in Figure 2.

Figure 2. Stochastic phrase structure grammar, including the probabilities of each construction

| | |
|------------------|--|
| S -> | Imperative [0.1] Interrogative [0.3] Declarative [0.6] |
| Declarative -> | NP V-int [0.35] NP V-tran NP active [.28] NP V-tran NP passive [0.042] subject cleft [0.014] object cleft [0.014] NP-Adj [0.1] That-NP [0.075] You-P [0.125] |
| NP-ADJ -> | NP is/are adjective |
| That-NP -> | that/those is/are NP |
| You-P -> | you are NP |
| Imperative -> | VP |
| Interrogative -> | Wh-Question [0.65] Aux-Question [0.35] |
| Wh-Question -> | where / who / what is/are NP [0.5] where / who / what do / does NP VP [0.5] |
| Aux-Question -> | do / does NP VP [0.33] do / does NP wanna VP [0.33] is / are NP adjective [0.34] |
| NP -> | a / the N-sing / N-plur |
| VP -> | V-int V-trans NP |

The four sentence types appeared with the following frequency: (Declarative) Active: 16.8%, Subject Cleft: 0.84%, Object Cleft: 0.84%, Passives: 2.52%. This gave a Passive-to-Active ratio of roughly 1:7, and ratio of OVS to SVO sentences of 1:21. Dick and Elman (2001) found that for English, the Passive-to-Active ratio ranged from 1:2 to 1:9 across corpora and that subject and object clefts appear in less than 0.05% of English sentences. They found that the relative frequency of word orders depended on whether one compares the passive OVS against transitive (SVO) or intransitive (SV) sentences and reported ratios that varied from 1:5 to 1:63 depending on corpus (spoken or written). The simulation frequencies were therefore an approximate fit, with the Subject

and Object Clefts slightly higher than in English due to the requirement to have at least a handful appear in our training corpus.

We generated a corpus of 10,000 sentences from this grammar as our training materials for the network, and a set of 100 test sentences for each of the active, passive, subject cleft and object cleft constructions.

3.2 Simulation One

The Dick et al. (2001) task consisted of presenting participants with a spoken sentence, and two pictures corresponding to the agent and patient of the sentence. The participant's task was to indicate with a binary choice which of the pictures was the agent of the sentence. For example, for sentences such as 'the dog is biting the cow', participants were asked to "press the button for the side of the animal that is doing the bad action".

Our next step was to implement this task in the model. One approach would be to train the network to output at each processing step not only the next predicted word in the sentence but also the thematic role of the current input. If the current input is a noun, this would be agent or patient. Joannis (2000) proposed just such a solution to parsing in a connectionist model of anaphor resolution. We will refer to the implementation of activating units for agent or patient (solely) on the same cycle as the relevant noun as the "Discrete" mapping problem of relating nouns to roles.

The mapping problem adds to the difficulty of the prediction task. We can assess the extent of this difficulty by measuring performance on the prediction component alone, against the metrics of two statistical models. The bigram and trigram models are statistical descriptions of the sentence set that predict the next word given the previous two or three words of context, respectively, and these were derived from the observed frequencies in the training set.

Lastly, for the purposes of this simulation, we do not distinguish between the syntactic roles of subject and object, and semantic roles of agent and patient, even though a more complex model may separate these levels and include a process that maps between them. Although these simulations conflate the syntactic and semantic categories, we use the terms agent / patient for clarity in linking to the Dick et al. empirical data.

3.2.1 Method

For Simulation 1, we added two output units to the C&D network. The network was trained to activate the first extra unit when the current input element was the subject / agent of the sentence, and to activate the second extra unit when the

object / patient of the sentence was presented. For all other inputs, the target activation of both units was zero. Thus, the number of input and output units was 55 and 57 respectively for the basic model, and 59 units and 61 units for the multiple-cue model.

The network's ability to correctly predict the next word was measured over the 55 word output units using the cosine between the target and actual output vectors. On novel sentences, a perfect network will only be able to predict the next item probabilistically. However, over many test items, this measure gives a fair view of the network's performance and we followed C&D (2001) in using this measure.

We initially chose our parameters based on those used by C&D (2001). Our learning rate was 0.1, and we trained the network for ten epochs. We performed a simple search of the parameter space for the number of hidden units to establish a "normal" condition (see Thomas & Karmiloff-Smith, 2003, for discussion of parameters defining normality). Eighty hidden units, the number used by C&D, gave adequate results for both models. This value was used to define the normal model.

We first evaluate normal performance at the end of training, then under the developmental deficit of a reduction in hidden units in the start state, and finally under the acquired deficit of a random lesion to a proportion of connection weights from the trained network.

3.2.2 Results

On the prediction component of the task, both models demonstrated better prediction ability than the bigram model, and marginally less prediction ability than the trigram model. This is in contrast to C&D's original prediction-only SRN model, which exceeded trigram model performance. It shows that the requirement to derive agent and patient roles increased the complexity of the learning problem, interfering with prediction ability.

The role-assignment component of the task was indexed by the activation of the agent and patient units *when presented with the second noun of the sentence*. At presentation of the first noun, there was no information available in the test sentences that would allow the network to distinguish between the possible interpretations of the sentence. At the second noun, the most active of the two units was assumed to drive the interpretation of the sentence and subsequent picture identification in the Dick et al. task. Therefore, the network's response was "correct" for Active and Subject Cleft sentences if the "patient" unit had the highest activation, and for Passive and Object Cleft sentences if the "agent"

Figure 3

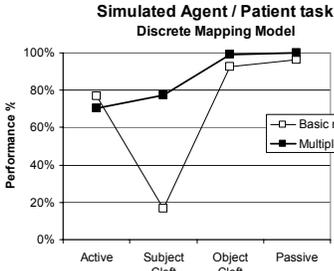


Figure 8

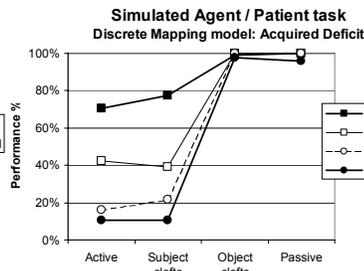


Figure 9

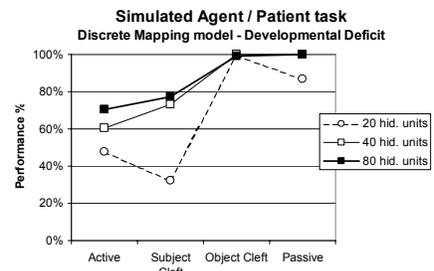


Figure 4

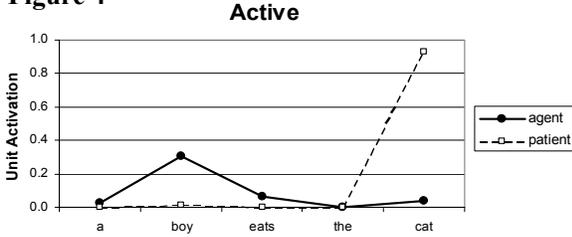


Figure 6

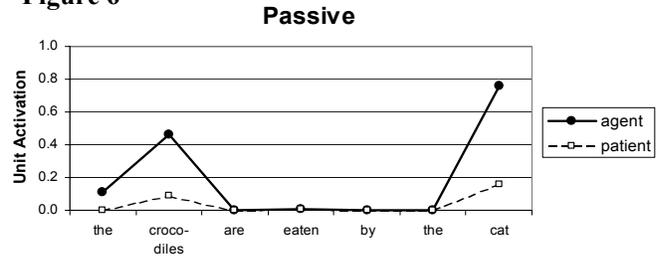


Figure 5

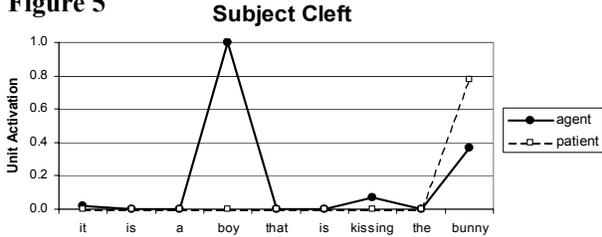
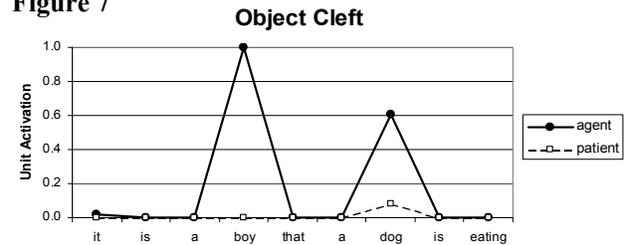


Figure 7



unit had the highest activation. The scores, measured in terms of the proportion of correct interpretations for the test sentences for each construction are shown in Figure 3.

Somewhat surprisingly, both the basic and multiple-cue models exhibited better performance on the Passive and Object Cleft sentences than on Active and Subject Cleft sentences. (These differences were statistically reliable.) The main difference between the two models was lower performance on Subject Cleft in the basic model, implying that cues to content-word status help to disambiguate the two cleft constructions.

Examining the profiles of performance for each sentence type gives some insight into the dynamics of the networks. Figures 4 to 7 show the activation of the agent and patient units for the multiple-cue model during the processing of examples of each construction, selected at random. The Subject Cleft sentence shown in Figure 5 is typical of the pattern for both Active and Subject Cleft sentences. That is, agent unit activation is close to 1.0 at the first noun, while patient unit activation is close to zero. At the second noun, the network is usually able to correctly distinguish the patient, but some agent unit activation also occurs. Therefore, using our decision criteria, the network is not always able to correctly identify the patient, and scores on Active

and Subject Cleft sentences are not perfect.

In contrast, in the example Passive and Object Cleft sentences, the network incorrectly activates the agent unit at presentation of the first noun. At this point, the network has no information that could possibly allow it to distinguish between the two different kinds of sentence, and so its response is driven by the relative frequency of the constructions. However, for the second noun (the agent), although the patient unit does show some activation, the agent unit is clearly favoured.

Generally, the advantage of the agent unit for the Passive and Object Cleft sentences is greater than the advantage of the patient unit for the Active and Subject Cleft sentences. This can be explained by a general bias in the network in favour of the agent unit. In the training set, agents (subjects) occur much more frequently than patients (objects). All of the interrogatives and imperatives only have agents, and these comprise 30% of the training sentences. Thus, paradoxically, the network suffers when attempting to produce activation on the patient unit, and this impacts on the Active and Subject Cleft performance, despite the much greater frequency of these constructions.

Figures 8 and 9 illustrate the affects of initially reducing the numbers of hidden units in the network and of lesioning connections in the

endstate. In both cases, non-optimal processing conditions exaggerated the pattern of task difficulty, with Actives and Subject Clefts failing to be learned or showing greater impairment after lesioning. Object Clefts are the most easily learnt and most robust to damage, despite their non-canonical word order and low frequency. With the task definition of responding “agent” to the second noun, this construction gains most from the prevalence of the agent status of nouns in the corpus.

This interpretation of the Dick et al. agent-identification task does not provide an adequate fit to the human data, either for normal or atypical performance. Why not? This implementation of the task requires that the network keep track of two roles at the same time and assign those roles at the correct moment. It is therefore driven by the independent probability of a noun being an agent or a patient at multiple time points through the sentence. The result is a de-emphasis of global sequence information and an emphasis on local lexical information, leading to a relative advantage of responding ‘agent’ to *any* noun.

In the Dick et al. task, the participant is asked to make a single decision based on the entire sentence, rather than continuously monitor word-by-word probabilities. Responses occurred between 2 and 4 seconds after sentence onset, with words presented at around 3 words-per-second. In the next section, we therefore provide an alternate implementation of the task based on a single categorisation decision for the whole sentence. But Simulation 1 serves as a demonstration that the statistics of the input set alone do not generate the task difficulty. It is the mappings required of the network. Moreover, we might predict that a modification of the Dick et al. study to encourage on-line monitoring of roles would alter the pattern of task difficulty. Thus, the four options might be presented as pictures (each noun twice, once as agent, once as patient), and the participants’ eye-gaze direction recorded as the sentence unfolds.

3.3 Simulation Two

An alternate implementation of the Dick et al. task is that the network should be required to make a single categorisation on the whole sentence as to whether the agent precedes the patient, or the patient precedes the agent. This implementation follows the assumption that task performance is driven by higher-level sentence-based information rather than lexically-based information. A single unit can serve to categorise the input sentence as agent-then-patient or patient-then-agent. During training, the target activation for the unit is applied continuously throughout the entire utterance. We

therefore call this the Continuous Mapping problem for sentence comprehension. Like the Discrete Mapping problem, the Continuous version has also been employed in previous connectionist models of parsing (Miikkulainen & Mayberry, 1999). (Note that Morris, Cottrell & Elman, 2000, used an implementation that combines Discrete and Continuous methods, providing a training signal that is activated when a word appears and is then maintained until the end of the sentence). The Continuous method generates a training signal for comprehension. It does not constrain on-line comprehension, which may be subject to garden-pathing and dynamic revision.

3.3.1 Method

A single output unit was trained to produce an activation of 1 for sentences with Subject-Object word order (active and subject cleft constructions), and 0 for Object-Subject word order (passives and object cleft constructions). Apart from this difference, the basic and multiple-cue models were identical in all other respects, with 55 input and output units in the basic model, and 59 units in the multiple cue model. As before, we trained the network on 10,000 sentences generated by the stochastic phrase structure grammar, and tested the trained network on sets of 100 Active, Passive, Subject Cleft and Object Cleft sentences. One hundred and twenty hidden units were required to define the ‘normal condition’ for these simulations.

3.3.2 Results

As with Simulation 1, the prediction ability of both basic and multiple-cue models suffered due to the burden imposed by the mapping task. Although the networks’ performance reliably exceeded a bigram prediction model, the trigram statistical model was slightly superior.

The network’s ability to correctly “interpret” the test sentences was measured as follows. If the semantic output unit’s activation at the time of second noun presentation was greater than 0.5, then the response was assumed to indicate that the sentence had Subject-Object word order and the agent was the first noun. If the activation was less than or equal to 0.5, then the response was assumed to indicate that the sentence had Object-Subject word order and the agent was the second noun. Although the target output for the network was consistent throughout each sentence, we selected the presentation of the second noun as our point of measurement, as this was where the network’s discrimination ability was greatest. Figure 10 depicts performance on the four constructions.

On Active, Subject Cleft, and Passive sentences the basic model showed appropriate performance,

Figure 10

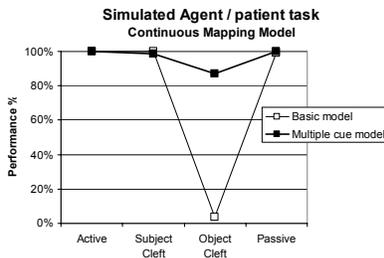


Figure 15

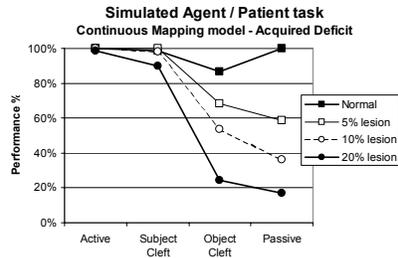


Figure 16

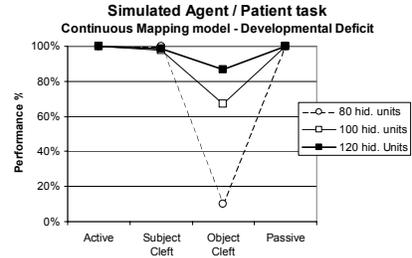


Figure 11

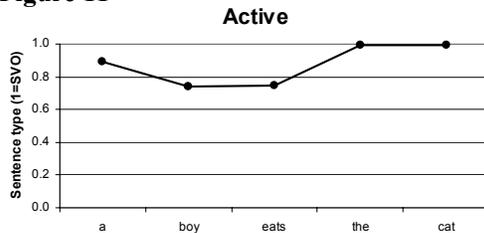


Figure 13

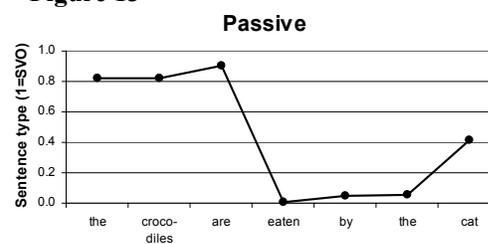


Figure 12

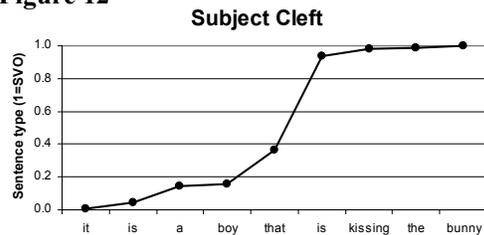
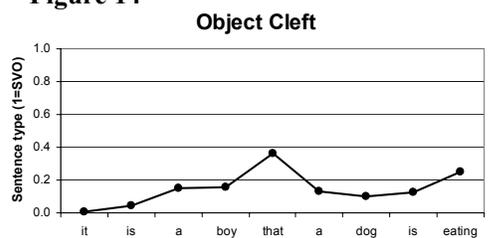


Figure 14



but it failed to correctly distinguish the Object Cleft sentences. Doubling the hidden units did not markedly alter this pattern. The multiple-cue model showed a much better fit to the human data, performing at close to ceiling for the Active, Passive and Subject Cleft constructions, and scoring in excess of 85% correct on Object Cleft constructions. The content-word cues provided in the multiple-cue model again appeared important in disambiguating the cleft constructions.

Focusing on the multiple-cue model, Figures 11-14 show the activation of the network's semantic output unit over a random sentence from each of the four test constructions. For the Active sentence, the network maintains a fairly constant high level of activation throughout the sentence. That is, it starts with the "assumption" that sentences will have a Subject-Object word order, and becomes more certain of this result (as shown by rising output activation) as the sentence proceeds.

For the Passive sentence, again, the network starts out assuming that the sentence will have the more frequent Subject-Object word order. But on seeing "eaten by", the network reverses its original diagnosis. However, the influence of this cue noticeably fades as the sentence proceeds. It persists enough that by the second noun, the network (just) manages to indicate correctly that

the sentence has Object-Subject word order.

The Cleft constructions show a very different pattern. For the Subject Clefts, the network begins with a low output value from the semantic unit. This increases slightly as the first determiner and noun are presented, but the most valuable cue arrives with the words "that is kissing". These provide a perfect indicator (in this context) that the sentence has Subject-Object word order, and the activation of the semantic unit jumps dramatically, staying near ceiling for the rest of the sentence.

Finally, examining the Object Cleft sentence, output activation again starts low and rises only modestly during presentation of the first noun. However, the presence of a second noun following immediately after the first pulls the activation back down, to correctly indicate that the sentence has Object-Subject word order. Notice that, as with the Passive sentence, as the distance increases from the cue that marks the (less common) status of the Object Cleft sentence, so the activation level of the semantic unit tends to drift back to the default of the more frequent constructions.

Figures 15 and 16 illustrate, respectively, the effects of reducing the initial numbers of hidden units in the network and of lesioning connections in the endstate. In the case of acquired damage, non-optimal processing conditions exaggerate the

pattern of task difficulty, with Passives and Object Cleft's showing greater impairment after lesioning in line with the empirical data in Figure 1. Interestingly, in the case of the developmental deficit, the pattern is subtly different. While Object Clefts show increased vulnerability, Passives are far more resilient to developmental damage.

We carried out further analysis of this difference. Using the examples in Figs. 13 and 14, the cues predicting Object-Subject order for Passives turned out to be the inflected verb 'eaten' followed by 'by', i.e., two *lexical* cues (the second redundant). For Object Clefts, the cue for Object-Subject order was *sequence-based* information: in this construction, two nouns are not separated by a verb. This is marked by the arrival of a second noun prior to a verb, that is, the words 'a' and 'dog'. While both lexical and sequence cues are low frequency by virtue of their constructions, they differ in that the Passive cue comprises lexical items unique to this construction, while the Object Cleft cue involves a particular sequence of lexical items that also appear in other other constructions. Examination of activation dynamics reveals that both low frequency cues are lost after acquired damage. However, the network with the developmental deficit retains the ability to learn the lexically-based cue that marks the Passive, but has insufficient resources to learn the sequence-based cue that marks the Object Cleft construction.

Three points are evident here. First, the model makes a strong empirical prediction that when developmental deficits are compared to acquired deficits, passive constructions will be relatively less vulnerable. This renders the model testable and therefore falsifiable. Second, the model demonstrates the differential computational requirements of tasks driven by local (lexically-based) and global (sequence-based) information in a parsing task. Third, the model reveals the distinction between acquired and developmental deficits, with compensation possible in the latter case for cues with low processing cost (see Thomas & Karmiloff-Smith, 2002, for discussion).

4 Discussion

Implemented learning models are an essential requirement to begin an exploration of the internal constraints that influence successful and atypical syntax processing. Our model necessarily makes simplifications to begin this exploration (e.g., the distribution and frequency of lexical items across constructions is not in reality uniform; cleft constructions may have different stress / prosodic cues). A precise quantitative fit to the empirical data must await models that include those factors.

However, the current model is sufficient to

demonstrate the importance of the mapping task in specifying difficulty (over and above the statistics of the input); how internal processing constraints influence performance; and how local and global information show a differential contribution to and vulnerability in sequence processing in a recurrent connectionist network.

5 Acknowledgements

This research was supported by grants from the British Academy and the Medical Research Council (G0300188) to Michael Thomas.

References

- Christiansen, M. & Dale, R. 2001. Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (p. 220-225). Mahwah, NJ: LEA.
- Dick, F. & Elman, J. 2001. The frequency of major sentence types over discourse levels: A corpus analysis. *CRL: Newsletter*, 13.
- Dick, F., Bates, E., Wulfeck, B., Aydelott, J., Dronkers, N., & Gernsbacher, M. 2001. Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, 108(3): 759-788.
- Elman, J. 1990. Finding structure in time. *Cognitive Science*, 14, 179-211.
- Elman, J., et al., (1996). *Rethinking innateness*. Cambridge, Mass.: MIT Press.
- Fowler, A. (1998). Language in mental retardation: Associations with and dissociations from general cognition. In J. Burack et al., *Handbook of Mental Retardation and Development* (p.290-333). Cambridge, UK: CUP.
- Joanisse, M. 2000. *Connectionist phonology*. Unpublished Ph.D. Dissertation, University of Southern California.
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, 2(10): 389-398.
- MacWhinney, B. & Bates, E. 1989. *The cross-linguistic study of sentence processing*. New York: CUP.
- McDonald, J. 1997. Language acquisition: The acquisition of linguistic structure in normal and special populations. *Annu. Rev. Psychol.*, 48, 215-241
- Miikkulainen, R. & Mayberry, M. 1999. Disambiguation and grammar as emergent soft constraints. In B. MacWhinney (ed.) *Emergence of Language*. Hillsdale, NJ: LEA.
- Morris, W., Cottrell, G., and Elman, J. 2000. A connectionist simulation of the empirical acquisition of grammatical relations. In S. Wermter & R. Sun (eds.), *Hybrid Neural Systems*. Heidelberg: Springer Verlag.
- Newport, E. 1990. Maturation constraints on language learning. *Cognitive Science*, 14, 11-28.
- Thomas, M.S.C. & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioural and Brain Sciences*, 25(6), 727-788.
- Thomas, M.S.C. & Karmiloff-Smith, A. 2003. Modelling language acquisition in atypical phenotypes. *Psychological Review*, 110(4), 647-682.