
A CONNECTIONIST MODEL OF DEPENDENCY LENGTH MINIMISATION

UNDERGRADUATE DISSERTATION

Ashna Ahmad
Trinity College, Cambridge, CB2 1TQ
aa2155@cam.ac.uk

March 14, 2022

ABSTRACT

Psycholinguistic approaches to the apparent privileging of short constituents before longer, more complex constituents in English syntax often link this phenomenon to the inherent ease of processing short constituents before long ones, rather than the reverse. This ‘dependency length minimisation’ strategy is a standard way to motivate many length-based structural choices in syntax, including the well-known phenomenon of heavy-NP shift. One corpus study, however, found that the positive correlation between dependency length and sentence-final positioning only holds for certain dependency types. Here we use a connectionist model to simulate human processing of four different dependency types and test whether it develops a significant preference for shifting long dependents to the end of the sentence, as well as whether this preference is contingent on dependency type. Results from the model’s output support the corpus analysis. Sentence-final dependencies do tend to be the longest dependency in the produced sentences, but not when they are direct object dependents, and in fact the opposite trend is observed in this case. An explanation is proposed for why this does not constitute conclusive evidence that long/heavy direct object dependencies pose less of a processing challenge than dependencies of other types.

1 Introduction

1.1 Dependency length minimisation

Dependency length minimisation (DLM) is the positioning of any syntactic head/dependent pair (verb/subject, verb/object, noun/adjective, and so on) such that the distance between the elements in the pair is as short as possible. DLM is a demonstrably significant syntactic feature across languages: in 37 major languages, overall dependency length per sentence (as a function of sentence length) is consistently lower than random baselines [1]. The relative difficulty of processing longer dependencies is a common prediction of language processing models [2], and has also been supported by reading studies [3]. Therefore, DLM is primarily considered a processing facilitation strategy, adopted to reduce working memory burdens and enable more efficient sentence construction.

A canonical example of DLM is heavy-NP shift, a phenomenon first described by Ross [4] and extensively studied in syntax ever since. The following sentences provide an example (they are represented with displaCy visualisations [5] to illustrate their dependencies):

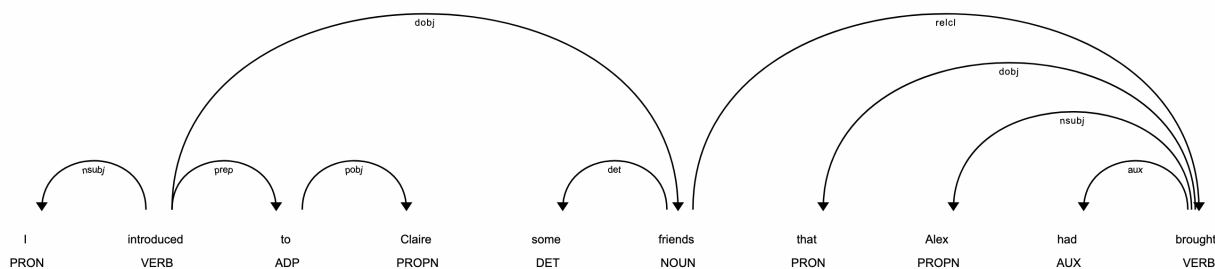


Figure 1: Sentence with heavy-NP shift

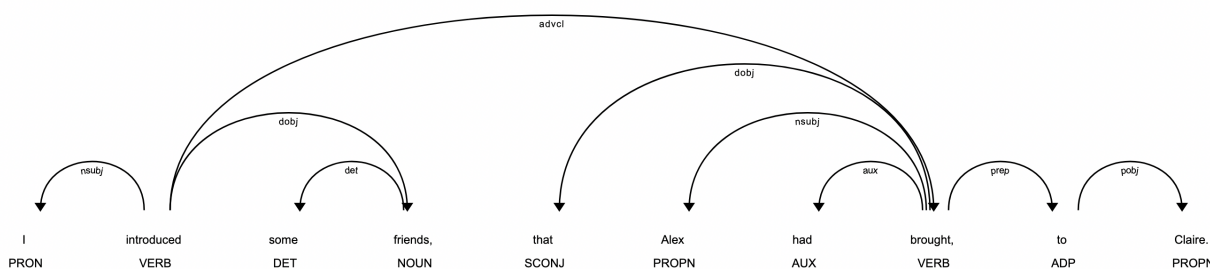


Figure 2: Sentence without heavy-NP shift (the heavy dependency is in the middle of the sentence)

In this example, the relative clause in the NP ('some friends that Alex had brought') lengthens the direct object dependency between 'friends [that]' and 'brought'. This causes a usually nonstandard word order, *verb* >_L *indirect object* >_L *direct object*, to be acceptable in the interest of dependency length minimisation. Figure 1 shows that shifting the NP containing the relative clause to the end of the sentence removes the sentence's longest (i.e. heaviest) dependency, thus facilitating processing. Other observations supporting the DLM strategy include Hawkins' Heaviness Serialisation Principle (HSP) [6], which is a description of the nominal modifiers with a greater tendency to be placed rightward relative to their head. According to this idea, relative clauses exhibit the strongest rightward-movement tendencies, followed by genitives, adjectives, and lastly demonstratives and numerals. The examples in Figures 1 and 2 could be reframed according to the HSP: the rightward movement of the relative clause 'that Alex had brought' serves to place the heaviest constituent at the end of the sentence. Although Hawkins does not frame this constraint in terms of dependency length, it makes identical predictions to DLM: the modifier types with stronger rightward-movement tendencies under the HSP also cause longer dependencies.

Sentences with long dependencies in the middle - i.e. without minimal dependency length - have varying degrees of unacceptability. While most English speakers would agree that 'I introduced to Claire some friends' is unacceptable while 'I introduced some friends to Claire' is not, there may be some disagreement over the sentences in Figures 1 and 2, especially if the dependency is even slightly shortened ('I introduced some friends Alex had brought to Claire' has ambiguous grammaticality). The literature generally holds that DLM becomes increasingly imperative as the relative length of the dependency increases. A series of recall-based production experiments [7] found that for both noun phrases and prepositional phrases, participants' preference for shifting a phrase to a sentence-final position increased in line with the relative (but not absolute) length of the phrase. Another production study [8] reported similar findings, with participants preferring to shift longer dependents to sentence-final positions. However, this study also suggested that verbs with higher 'shifting disposition' are more likely to induce speakers to shift heavy constituents. These include any verbs which are frequently found in non-adjacent positions with their complements (i.e. verbs which take sentential complements, such as indirect statements).

We can see that a variety of factors inform speakers' preferences regarding DLM, and psycholinguistic (rather than purely syntactic) exploration of the phenomenon is still relatively new. Therefore, there is still much scope for adding more nuance and detail to psycholinguistic accounts of long dependency processing. A very recent study [9] did so by exploring whether the observed pattern of longer dependencies shifting to a sentence-final position holds across all dependency types. The researchers applied a distributional regression model to evaluate the interaction between

dependency type (among other factors) and the DLM-driven shifting of longer dependencies. The data used to build the regression model was from the Universal Dependencies (UD) database [10] and consisted of six dependency types involving major lexical categories. These dependency types were verb/subject (the *car is red*), verb/object (she *gave me a raise*), verb/oblique (*give the toys to the children*), verb/adverbial modifier (*genetically modified* food), noun/adjective (*large house*), and noun/nominal modifier (the *office of the boss*). Curiously, this corpus study only found the expected link between dependency length and sentence-final dependent shift in the case of obliques and nominal modifiers. The figure below, reprinted under open access from their paper [9], shows the results:

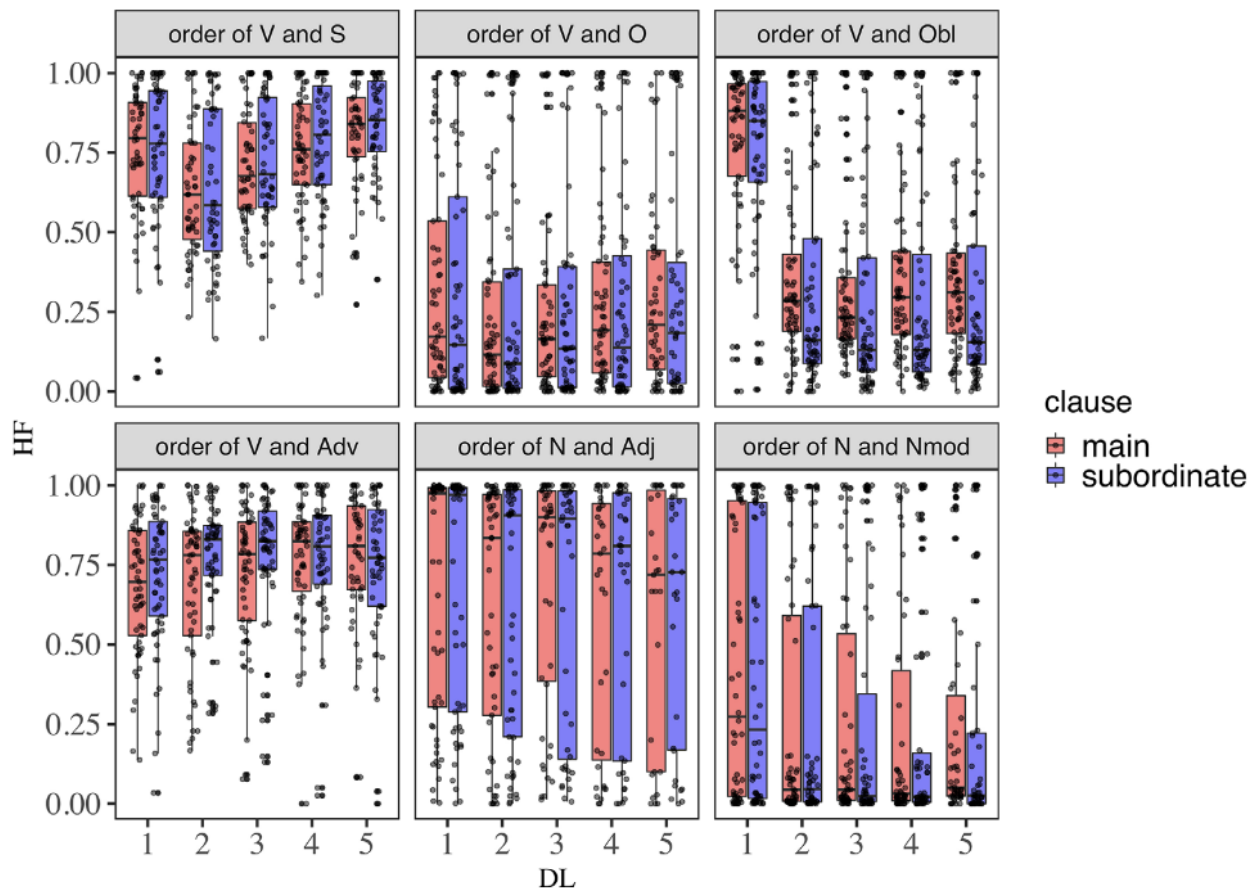


Figure 3: Boxplot showing the results of Jing et al.'s corpus analysis

Here, the x-axis shows dependency length (absolute) and the y-axis shows the rate of head-finality (a single-variable proxy for word order) in the raw data from the UD corpus. The only significant changes in word order in line with an increase in dependency length appear to come from oblique and nominal modifier dependencies. In other words, the UD corpus seems to show that the prevalence of DLM-driven shifts in word order is contingent on dependency type. In the study that follows, we train a connectionist model to produce sentences with varying dependency types of different lengths, to verify whether it supports the prediction that only certain dependency types cause syntactic shifts motivated by DLM. To simplify the modelling process, we train and test the model only on English sentences, whereas the UD corpus analysis included many more languages; however, since DLM effects (including shifts of long dependencies to either periphery of a sentence) have been observed to hold cross-linguistically [11], variation across languages should not be a significant variable.

1.2 Connectionist modelling

Connectionist modelling is the experimental technique of using artificial neural networks to simulate human processing. Simple recurrent networks (SRNs) [12] are used to explicitly represent a learning process involving modifiable connections, i.e. the adjustment of the weights which control activation strengths. This adjustment allows the

connections that make up a neural network to be continuously updated in order to increase output accuracy; this process occurs via backpropagation, an algorithm based on the difference between correct and actual network output [13]. Activation in connectionist networks takes place in distributed patterns across groups of processing units. These usually consist of an input ('previous word') layer (i.e. the last word in the sentence that was generated by the network), an output ('next word') layer, and several intermediate 'hidden layers' which receive the activation from the previous word as input from the network's 'context units', thus learning to predict the next word based on a combination of previous word information and whatever other information was fed into the network. Connectionist modelling has been a successful strategy for modelling many phenomena in human language processing, including the cognitive processes involved in the acquisition of English past-tense morphology [14] and in reading aloud [15], among other things. While not every feature of a connectionist network is neurally plausible, the gradual and local nature of the weight updates which occur during backpropagation mimics the processes underpinning neural plasticity ([16], [17]). With that in mind, there are three main advantages of building a connectionist model to simulate the effect of dependency type on dependency length minimisation:

A processing account. There may be semantic variables which influence human judgements regarding the necessity of DLM. Since corpora such as the UD database are collections of sentences produced by humans under natural, non-experimental conditions, other variables - most notably of the semantic variety - may influence results from corpus analyses. When using a connectionist model, however, we can control how much semantic information the model receives as input in order to isolate the desired independent variables. In this case, we are specifically interested in whether certain dependency types pose an increased processing challenge to the human brain when long and trigger stronger DLM tendencies as a result, since DLM is usually described as a processing facilitation strategy. We can therefore provide limited non-syntactic information to the model.

A 'brain systems' account. Chang [16] describes corpus-based studies as an 'empirical approach' and connectionist modelling as a 'brain systems' approach. This is because statistical trends in corpora provide information about the effect of "statistical relationships in the linguistic environment" on linguistic output, whereas connectionist modelling emphasises the role of neural systems in language production. Using a SRN to build a largely cognitively plausible model of language processing - e.g. by using backpropagation as an analogue to plasticity - allows for exploration of the neuropsychological constraints on sentence production. We hope to add to Jing et al.'s insights on the relationship between DLM and dependency type by approaching the question from a neuropsychological, processing-oriented perspective, rather than a purely statistical perspective.

Incrementality. There is overwhelming evidence that sentence production is an incremental process [18]. DLM might appear to run contrary to incrementality because it requires advance planning of a sentence to take into account the heaviness of a constituent. However, there is a way to reconcile incremental sentence planning with DLM effects within a connectionist framework, which will be covered in the next section.

The neural network architecture which we use in this study has successfully captured cross-linguistic variation in DLM (specifically heavy-NP shift) patterns in the past. Chang [11] used the architecture to study the processing biases influencing word order in English and Japanese. The model learnt an artificial input language based on subsets of English and Japanese, where sentences either contained no heavy dependencies, a heavy patient, or a heavy recipient (where the length condition for heaviness was provided by relative clauses). Training examples were evenly split between long-before-short and short-before-long orders, allowing for any prevalence of either order to be ascribed to the preferences developed by the model. It emerged that the model did develop the specific DLM preferences for each language, and this was attributed to the model picking up intra-sentence statistical regularities which mainly influence the activation of the model's event-semantics units (particularly for the Japanese input language). The fact that this model learnt language-specific DLM preferences without any mechanism to control dependency length hardcoded into the model was used to support the claim that "grammar is shaped by processing" [19], rather than via a computationalist method based on explicit rules. However, since the structural preferences of the model were developed through a learning process, wherein statistical tendencies in the input informed which structures outcompeted others for activation, this study shows that some preferences regarding dependency length can be acquired purely through learning as opposed to satisfying any formal ease-of-processing requirements.

By using this same architecture, we hope to build on the successful record of the dual-path model at simulating DLM effects, and test whether the model can also capture the effect of dependency type on DLM which was observed in Jing et al.'s corpus study. If it can indeed capture this effect, this will serve as an indication that the contingency of DLM effects on dependency type can be accounted for by appealing to neural language processing mechanisms (specifically learning-based lexical competition for activation, which is discussed further in the following section).

2 The model

2.1 Architecture

The connectionist architecture used in this experiment is an adaptation of the dual-path model ([12], [20]). The original Python implementation of the model was developed by Chara Tsoukala and can be found at https://github.com/xtsoukala/dual_path.

Below is a diagram representing the dual-path architecture [20]:

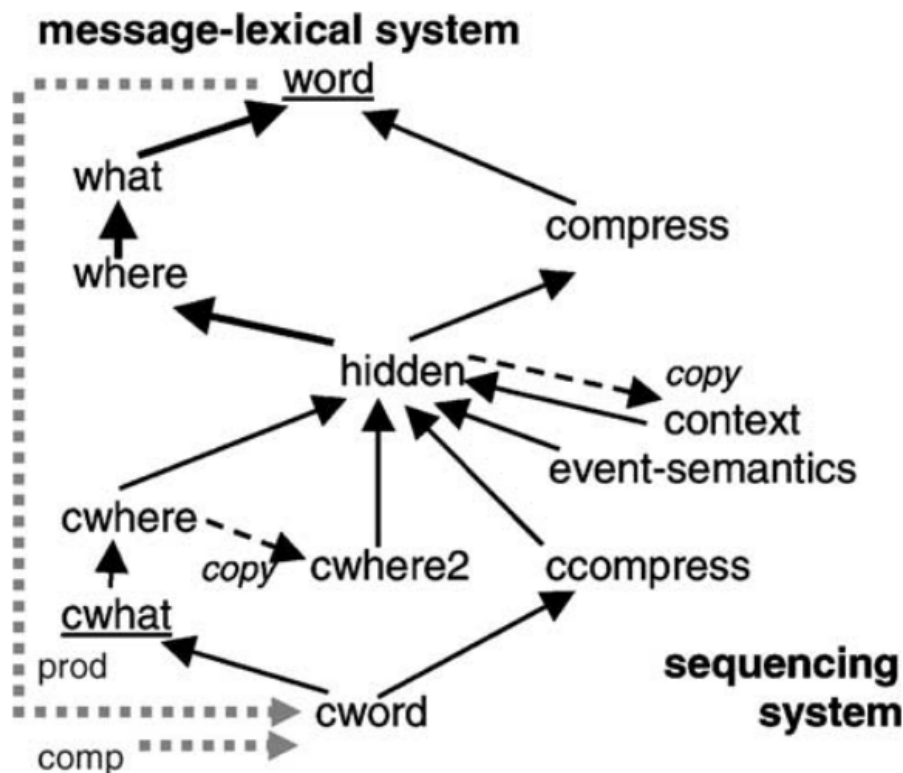


Figure 4: Diagram of the dual-path model architecture

In this diagram, 'prod' represents the flow of information from the previous word's activations, while 'comp' represents other information fed into the network. The top-to-bottom ordering of the layers, meanwhile, represents the order of the spreading activation (backpropagation was applied in the reverse direction).

The term 'dual-path' refers to the two systems which together comprise the model architecture: the message-lexical system and the sequencing system. Both systems receive information about the previous word in the current sentence (where 'cwhere' stands for the semantic role of the previous word, e.g. 'agent' or 'patient', and 'cwhat' for the word itself). They also receive a 'message', which is selected from the list of messages fed into the network (the list is available at <https://github.com/4944-iib/dissertation/blob/main/input/structures.csv>) and provides a target sentence for the model to plan out. In turn, both systems influence the final output layer. Between the input and output layers are two 'compress' layers (consisting of 12 units each in our experiment) and a recurrent 'hidden' layer (30 units).

The sequencing system is the SRN which deals with information about syntactic categories and their distributions. It is structured as follows:

- It has a feed-forward architecture consisting of the previous word layer, the hidden layers, and finally the output layer.

- The previous word layer activates hidden unit representations based on the activation strengths corresponding to the previous word, in order to help predict the next word.
- The hidden units use the binding between a word ('what') and its semantic role ('where'), the outputs from the context and compress units, the number of semantic roles in the target sentence (encoded in the 'event-semantics' units), and a copy of the previously filled roles in the sentence ('cwhere2') to generate the new role-word bindings which will correspond to the next word in the sentence.
- These new bindings are sent to the compress units, causing words in the output layer to be activated. The word with the highest activation is selected as the next word in the sentence. A sentence is generated by repeating this process. In the training stage, these generated sentences are compared with the target sentences from the input in order to enable backpropagation, and hence learning.
- Context units contain a copy of the hidden layer activations of the previous word, preventing the model from losing sight of longer-distance syntactic relations.
- Compress units force the model to carry out syntactic clustering according to distributional regularities in the input (in this case, in the message), leading to correct reproduction of English syntax; for more detail see [21] and [22].

While the model is learning to produce sentences, the interaction of these network layers leads to the generation of a sentence from a message (by passing the message sequentially through the network until each of the new words has been generated). The message, however, is only one part of the input pair in the training stage; there is also the target sentence (from the training set). The aim of the model in the learning stage is to produce a sentence as close as possible to the item from the training set. In our experiment, the training and test sets for the model were randomly generated by the model (2000 training sentences and 1004 test sentences) by matching lexical items to their semantic roles in the list of messages provided to the model. In the learning stage, the generated sentence is then compared with the target sentence from the training set and the resulting error (i.e. the difference between the target and the output) is backpropagated through the system. This leads to the output sentences increasing in grammaticality with every training epoch.

After training, the model is tested on the 1004-item test set. During testing, the model receives the (randomly-generated) test set as input, where each item in the set is structured like the 'message' example in section 2.2 (the next section describes how we generated our results in more detail). This process produced the results of our experiment.

While the sequencing system concerns itself with activating words in the output layer based on syntactic information from the message and the previous word(s), the message-lexical system deals with semantic role information. The meaning system also receives a target sentence from the test set and the message corresponding to that sentence. However, in this case, the semantic role of each word (given in the message) is bound to the word in question. For instance, if 'the man' is the agent of a sentence, 'where' units corresponding to 'agent' and 'what' units corresponding to '[the] man' are activated before (rather than during) the start of sentence generation. Once again, the message 'tells' the system to activate these units.

The system then proceeds as follows:

- Rather than feeding role-word bindings from the hidden layers to the output layer, there are separate role ('where') and concept ('what') layers. The weights of the connections between these layers change quickly (on the assumption that, as mentioned above, role-word bindings are already known pre-production).
- 'What' units connect to output units (connection weights are learned gradually)
- The hidden layer connects to the 'where' units, enabling the model to learn to activate semantic roles at particular positions, so that the same sentence structure can be filled/reproduced with different lexical items when necessary.

In our model, all layers used the tanh activation function, apart from the previous word layer and the output layer, both of which used softmax. Since both systems affect the activation of words in the output layer, by the time a single word has the highest activation and is therefore selected, its high activation should be a result of its fulfilment of both syntactic criteria (i.e. it is the right part of speech) and semantic criteria (i.e. it is capable of fulfilling the semantic role assigned to it by the message - so an inanimate noun would not be coactivated with an 'agent' role).

This specific neural network architecture is very appropriate for connectionist language modelling due to two key features which have also been hypothesised to exist in human language processing. The first of these, as discussed earlier, is incrementality. This architecture accounts for incrementality through the 'reverse message' system, which maps 'what' units back onto 'where' units. This allows previous words in the sentence (which will have been selected in the output layer, then activated in the previous word layer as the network moves onto the next word) to retain their

roles. For instance, if a sentence is being generated according to a message containing just one ‘patient’ role, once a word with a ‘patient’ role has been activated in the previous word layer, the reverse message prevents the network from producing another ‘patient’ word.

When humans produce sentences, their memory of producing a word corresponding to a certain role influences their subsequent structural choices. Given that sentences are planned incrementally, human speakers may produce a word with a certain role before planning the rest of the sentence, and subsequently modify the sentence structure on-line in order to avoid restating the sentence. For example, if they produce the word with a ‘patient’ semantic role first, they are likelier to go on to produce a passive sentence. [11] The reverse message allows the dual-path network to mimic this human-like incremental planning strategy. This is particularly relevant when investigating DLM, since it seems nontrivial to reconcile an incremental sentence planning mechanism with a phenomenon which seems to require advance knowledge of dependency length. In fact, previous studies using the dual-path architecture have proven that it is possible to reconcile realistic modelling of heavy-NP shift with incrementality. Using this architecture, it is possible to accurately model the DLM preferences of not only English, a verb-initial language where DLM leads to the privileging of short constituents before long ones, but also of Japanese, where verb-finality causes DLM to work in the opposite direction and privilege long constituents before short ones. [11]

The second key feature is the idea of units which compete for activation. Many proponents of connectionism in psycholinguistics have suggested that a significant feature of human language production is competition between the different lexical items which could potentially be selected to complete an utterance ([15], [23]). The dual-path architecture incorporates lexical competition via the assignment of multiple sentence structures to each message. For example, a sentence containing an agent, patient, and recipient can be written in one of two ways: with the patient preceding the recipient and the inclusion of the preposition ‘to’ in between, or with the patient directly following the recipient (‘I gave the book to the girl’ vs. ‘I gave the girl the book’). These two sentences could be generated by the same message, since the semantic roles within them - ‘agent’, ‘patient’, ‘recipient’ - are the same. However, due to these two possible sentence structures, when the network is reading this message, multiple ‘where’ units will become active after the verb has been generated. This leads to competition between the word corresponding to the patient role and the word corresponding to the recipient role (between ‘girl’ and ‘book’ in the example). This competition is critical for our experiment, since when a message is paired with a DLM and a non-DLM sentence structure, the ‘winning’ structure in the competition will provide us with information about the circumstances under which the units which correspond to a DLM strategy receive higher activation. If, as connectionist psycholinguists argue, a similar process of competition for activation occurs in the human brain when processing long dependencies, simulating that competition could be a very insightful exercise.

2.2 Input

To further illustrate how the dual-path model uses this information learned during its training to produce its own sentences, we will elaborate an example below. The GitHub repository for this project contains all of the input which was fed into the network and all 1004 sentences which were generated by the network, at <https://github.com/4944-iib/dissertation>.

Below is an example sentence generated by the model, and the message used to generate it:

- sentence: "a tall sister gives a soft toy to the father."
- message: AGENT=undef,sister;AGENT-MOD=tall;ACTION-LINKING=give;PATIENT=undef,toy;PATIENT-MOD=soft;RECIPIENT=def,father;EVENT-SEM=SIMPLE,PRESENT,AGENT,AGENT-MOD,PATIENT,PATIENT-MOD,RECIPIENT

Note that, as mentioned in the previous section, the meaning system binds the semantic roles in the message to particular words of an applicable part of speech before commencing sentence generation, whereas the sequencing system produces these bindings during sentence generation. The ‘EVENT-SEM’ variable in the message tells the event-semantics units in the network which roles, and how many of each, to activate. The definite/indefinite article variable is set randomly upon encountering ‘det’ in the sentence structure for a given message. A specific parameter is set to True to account for languages in which adjectives precede nouns (like English) despite the ‘modifier’ message variable following the noun it modifies; see the ‘corpus generator’ module of the network on GitHub for more detail.

The above message is paired with two different sentence structures in the structures .csv file which was fed into the network. Those structures are the following:

1. det adj:animate noun:animate verb:double:present det adj:inanimate noun:inanimate prep:to det noun:animate (‘a tall sister gives a soft toy to the father’)

2. det adj:animate noun:animate verb:double:present det noun:animate det
 adj:inanimate noun:inanimate ('a tall sister gives the father a soft toy')

Once the model has been trained, it moves onto the testing phase, wherein a randomly-selected message from the 'structures' input file provides it with a plan of a target sentence to generate. It fills the semantic roles in the plan with randomly-selected items from the 'lexicon' input file which happen to be in the appropriate category to fill a certain slot in the structure. For instance, if a structure calls for `adj:animate` at a certain position, the network will randomly select an item from the lexicon file which is tagged as `adj`, `animate` ('kind', 'intelligent', 'happy' and so on).

The network must decide which of the possible sentence structures corresponding to a given message to 'translate' the message into. In this example, (1) is a non-DLM structure and (2) is a DLM structure; this is because the extra adjective in the direct object makes the verb/direct object dependency (marginally) longer than the verb/indirect object (or verb/oblique) dependency, and according to the DLM strategy, the longest dependency should be at the end of the sentence. Since the sentence which was eventually generated conforms to the structure of (1), we can say that DLM effects did not cause (2) to outcompete (1) in this instance. This would lend some support to the idea that relatively long direct object dependencies are unproblematic in sentence processing and can be placed in the middle of a sentence.

It is important to note that the sentences generated by the network are extremely heavily influenced by what it picked up in its training phase. While being trained, the aim of the model is to learn to generate sentences which are maximally similar to the target sentences in the training set (see the description of the sequencing system in section 2.1 for more detail). As a result, activation strengths are shaped by the training set. In our experiment, the training items were generated from the 'structures' and 'lexicon' files such that there was an even split across messages and possible structures (i.e. no particular structures were more prevalent in the set than any others). Therefore, in theory, if dependency length does not influence word order then every structure in the input file should occur equally often in the model's output, because there are no significant differences in activation strength between the DLM structure and the non-DLM alternative. Other syntactic processing studies using the dual-path architecture have made use of this principle (see [21]).

In order to model the corpus study carried out by Jing et al. [9] as closely as possible, we also used the Universal Dependencies data to select the dependency types to include in training (more information on these types is in section 1.1). Since the main findings of this study concerned obliques and nominal modifiers, it was necessary to train the model to produce these dependencies in order to verify whether they induce DLM effects when long. The oblique and nominal modifier constructions we used are the following:

- Recipient constructions involving 'to' (oblique; note that the structure without 'to', in example (2) above, is normally counted as an indirect object and not an oblique [10])
- Origin constructions involving 'from' (oblique)
- Possessive constructions involving 'of' (nominal modifier)
- The 'English genitive', -'s (nominal modifier)

We included direct objects as a control to compare the oblique and nominal modifier results with, since the corpus analysis found no significant dependency length effects for direct objects. In addition, direct objects are often the NPs involved in heavy-NP shift, so they are a particularly important category to include. Adverbial modifiers - specifically 'still', 'more often' and 'for many years', were included as well - but due to a bug in the model (discussed further in section 4.2) the results for adverbs are not very reliable. They have been included in the summary data anyway for completeness.

The other dependency types included in the corpus analysis are nominal subjects and adjectival modifiers. However, in a preliminary analysis of the UD English Web Treebank data (available at <https://github.com/4944-iib/dissertation/blob/main/lexiconewithlength.csv>) we found low variance in dependency length. Indeed, the 'standard' constructions for these dependency types - equivalent to recipient constructions for obliques - simply have a dependency length of 1, due to nominal subjects being placed adjacent to their verb and adjectival modifiers adjacent to their nouns. We chose to exclude these dependency types from the analysis, rather than selecting relatively rare constructions to lengthen them.

In addition to the dependency types, we also taught the model the following constructions, which served the sole purpose of lengthening the dependencies to create the conditions for a DLM strategy to be applied:

- Adjectives modifying any semantic role (represented in the message by the semantic roles `AGENT-MOD`, `PATIENT-MOD` etc.)
- Multiple adjectives joined by 'and' (represented by `AGENT-MOD-DOUBLE`, `PATIENT-MOD-DOUBLE` etc.)

- Multiple nouns joined by ‘and’ (represented by AGENT-DOUBLE, PATIENT-DOUBLE etc.)
- Relative copular clauses (the type ‘rel’ was added to the lexicon and its value was manually set to ‘that’ for every sentence with a relative clause)

Minimal semantic properties were added to the model, and so many of the sentences are rather nonsensical - however, we did add the property of animacy in order to disqualify certain nouns as agents, as well as adding the category of ‘relational nouns’ to yield more sensible possessive constructions (e.g. ‘the boy’s sister’ makes sense, but ‘the boy’s girl’ does not).

The generated sentences were subsequently run through a Python script which labelled them with their **sentence-final** dependency types - i.e. the type of dependency which was pushed to the very end of the generated sentence - as well as a binary value representing whether or not this sentence-final dependency was the longest dependency in the sentence. We used relative rather than absolute length to measure which dependencies qualify as ‘long’/‘heavy’, in keeping with [7]. Dependency lengths were calculated via spaCy’s builtin dependency parser, which was also used to generate the heavy-NP shift diagrams in section 1. These diagrams therefore also illustrate how dependency lengths were counted to produce the results tables below.

3 Results

Table 1: Summary statistics

Dependency type	Total sentence-final dependents	Long (‘heavy’) sentence-final dependents	Percentage (3sf, %)
direct object (obj)	210	63	30.0
oblique (obl)	506	392	77.5
nominal modifier (nmod)	269	204	75.8
adverbial modifier (advmod)	17	7	41.2

Table 2: χ^2 statistics

Dependency type	Long (Expected) [χ^2 statistic]	Non-long (Expected) [χ^2 statistic]
direct object (obj)	63 (140) [42.02]	147 (70.4) [83.4]
oblique (obl)	392 (336) [9.22]	114 (170) [18.3]
nominal modifier (nmod)	204 (179) [3.55]	65 (90.2) [7.04]
adverbial modifier (advmod)	7 (11.3) [1.64]	10 (5.70) [3.24]

χ^2 statistic = 168.2574, p -value < 0.0001

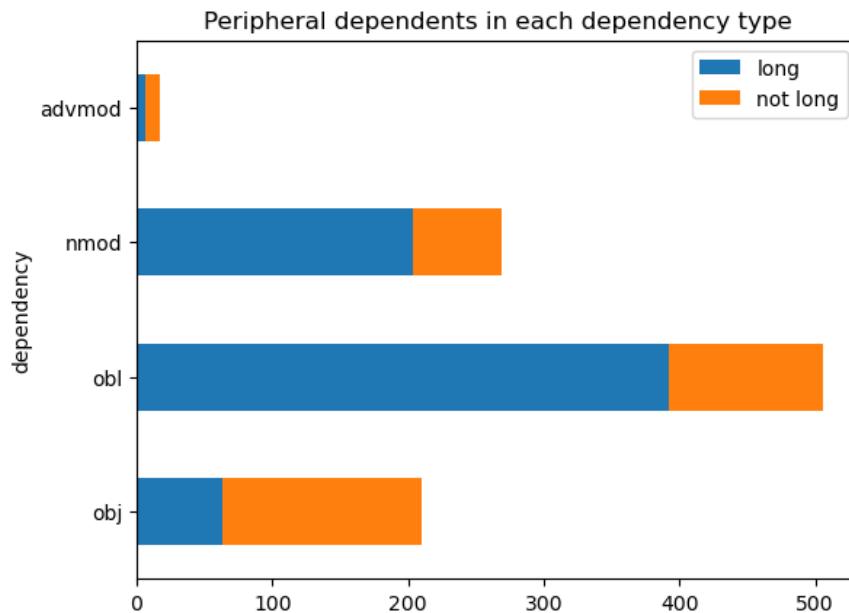


Figure 5: Graphic representation of summary statistics

Of the 1004 sentences generated by the model, Table 1 shows the proportion that had sentence-final dependencies of each type, as well as the proportion of sentence-final dependencies within each type which constitute the longest dependency in their sentence. The example sentence ‘a tall sister gives a soft toy to the father’, for instance, would be labelled as ‘obl, n’ because the final dependency in the sentence is an oblique (a recipient construction with ‘to’) and it is not as long as the direct object dependency which occurs earlier on in the sentence (so despite being lengthened by the adjective, the direct object dependency is not shifted to the end of the sentence). Therefore, it would not count towards the total of 392 heavy sentence-final obliques. The full, analysed dataset can be found in the ‘output’ folder of the GitHub repository.

If DLM effects are significant for a certain dependency type, we would expect to see long dependencies making up a high proportion of sentence-final dependencies of that type, because sentence-finality will mainly have been caused by DLM strategies. Conversely, if DLM effects are not significant, we may well still see sentence-final dependents in a certain category, but they will not be disproportionately long.

The results show that sentence-final oblique and nominal modifier dependencies tend to be the longest dependencies in their sentences, whereas sentence-final direct object dependencies do not (neither do adverbial modifiers, although they are underrepresented in the output data). The difference is rather significant: sentence-final oblique and nominal modifier dependencies are long/heavy more than twice as often as sentence-final direct object dependencies. This sizeable difference between dependency types in the proportion of sentence-final dependencies which are long is clearly visually represented in Figure 5.

The χ^2 test results further illustrate this conclusion: the final test statistic of 168 is significant at $p < 0.05$. Since $H_0 =$ no significant difference between the effects of the different dependency types on the incidence of long sentence-final dependencies, we accept H_1 , concluding that dependency type and incidence of long sentence-final dependencies are *not* independent. The expected values if H_0 were correct are also shown in Table 2. According to H_0 , it is expected that for each dependency type, there are roughly twice as many sentence-final dependencies which are the longest dependency in their sentence as there are sentence-final dependencies which are not. This is a reasonable prediction assuming that DLM is a significant strategy in language processing, and that it is not contingent on dependency type. By supporting H_1 , this particular statistical analysis appears to show that DLM effects exist - supporting [8] and other findings from human language production tasks - but only significantly affect obliques and nominal modifiers. Once again, this is in line with Jing et al.’s corpus study, which to our knowledge is the first empiricist study to make the distinction between dependency types. Since this paper is the first connectionist study to make this distinction, and has found the same result as the corpus analysis, this suggests that there may well be details of the relationship between dependency length and processing difficulty which are specific to certain dependency types. This opens the doors to many more research possibilities on this topic; see section 4.3 for further ideas.

We can take a closer look at the results by examining whether any particular constructions (either those generated to illustrate a certain dependency type or those generated to lengthen dependencies) had any individual effects on DLM, or any distributional particularities. Most notably, the results from our model appear to support the HSP, since relative clauses were the dependency-lengthening constituent which induced the most significant DLM effect. This is demonstrated by the two least common structures in our dataset (their messages are also provided):

1. AGENT=; ACTION-LINKING=; PATIENT=; RECIPIENT=; ACTION-LINKING=; ATTR=; EVENT-SEM=SIMPLE, PRESENT, AGENT, ACTION-LINKING, REL-ACTION-LINKING, REL-ATTR, ACTION-LINKING, PATIENT, REL, RECIPIENT, REL det noun:animate verb:double:present::singular det noun:animate rel:that verb:copular:present adj:animate det noun:inanimate
2. AGENT=; AGENT-MOD=; ACTION-LINKING=; PATIENT=; PATIENT-MOD=; RECIPIENT=; RECIPIENT-MOD=; POSSESSOR=; ACTION-LINKING=; ATTR=; EVENT-SEM=SIMPLE, PRESENT, AGENT, AGENT-MOD, PATIENT, PATIENT-MOD, RECIPIENT, RECIPIENT-MOD, POSSESSOR, ACTION-LINKING, ATTR, REL det adj:animate noun:animate verb:double:present det adj:animate noun:relational prep:of det noun:animate rel:that verb:copular:present adj:animate det adj:inanimate noun:inanimate

(Generic examples of these two structures would be (1) ‘the girl gave the boy that was intelligent the book’ and (2) ‘the kind girl gave the happy sister of the boy that was intelligent the long book.’)

The second of these constructions occurred just 4 times out of 45 ‘possessor’ (‘of’) constructions involving relative clauses. In both of these cases, it was much more standard for the model to shift the dependency containing the relative clause to the end of the sentence. For the first of these two structures, doing so would create a sentence-final oblique (with the longest dependency at the end); for the second structure, this would yield a heavy sentence-final nominal modifier dependency, which is much more common in the model output, comprising the majority of relative clause nmod constructions. Using the above example sentences again, we can see that this process of DLM-driven shifting yields sentences which sound much more natural to human ears: ‘the girl gave the book to the boy that was intelligent’ and ‘the kind girl gave the long book to the happy sister of the boy that was intelligent’ respectively. Thus, the model has produced a realistic simulation of how human speakers might deal with this particular case of DLM.

It is also notable that the preposition ‘to’ is overrepresented in the data, appearing 661 times out of 1004 sentences - despite the fact that every sentence containing a recipient was provided with a structure including ‘to’ and another without (since not every sentence has a recipient, we would therefore expect slightly fewer than half of the sentences to contain ‘to’ if no DLM effects were at play). This clearly reflects the preference of the model for shifting long indirect objects to the end of the sentence, thus creating sentence-final long obliques, since the preposition only appears in recipient constructions when the recipient (i.e. the oblique dependency) is sentence-final. It may also reflect the DLM-driving effect of relative clauses: even though sentence (2) above gets a heavy sentence-final nominal modifier when the dependency containing the relative clause is shifted, adding ‘to’ is still required to make this happen.

Aside from the general preference for sentence-final placement of long dependencies across all of the model’s output (relatively long dependencies are twice as likely as non-long dependencies to be sentence-final), adjectives which lengthen dependencies do not incur a significant additional preference for sentence-finality, unlike relative clauses. This is true regardless of whether the conjunction ‘and’ is also part of the lengthened dependency. In other words, dependencies which are the longest in their sentence and contain an adjective are no likelier than the average longest dependency in a sentence to be sentence-final.

Looking at the different constructions chosen to represent each dependency type, we see the more equal distribution of constructions in the output which we would expect to see without the influence of DLM. ‘Of’ and ‘-s’ nominal modifiers occur roughly equally as often in the output (276 vs. 252 occurrences), reflecting the fact that they were included in 12 and 16 of the messages in the structures.csv input file respectively (out of 75 total). Similarly, the oblique ‘from’ occurs in 10 out of 75 messages and 188 out of 1004 generated sentences. Since the messages were used to generate training sets in which every structure was equally represented, this illustrates that in most cases, the activation strengths which cause the model to select certain words over others are strongly informed by the prevalence of each word in the test sets - except, it seems, in the cases of the relative pronoun ‘that’ and the preposition ‘to.’

It is important to bear in mind the details of the calculations used in these summary statistics: rather than showing which dependency in the sentence was the longest and whether it was moved to the end of the sentence, they show which dependency was at the end of the sentence and whether it was the longest dependency. This was done for ease of implementation, and means that every DLM-motivated syntactic shift in the model’s output was captured by the statistical analysis (we can think of the DLM-driven shifts as the blue stacked bars in Figure 5). It also highlights cases where the sentence-finality of a certain dependency might result from factors other than DLM (this usually just reflects the activation strengths developed by the model when it was exposed to a certain structure with a non-heavy

sentence-final dependency during training). On the other hand, it means that for sentences marked ‘n’, we do not receive any information about the longest dependencies within the sentences for which DLM-motivated sentence structure was not observed. This issue is discussed further in section 4.2.

4 Discussion

4.1 Implications

Looking at the results of this study, it appears as if we have simply built a connectionist model which concurs in every regard with the corpus analysis carried out by Jing et al., and we can therefore support their conclusion that DLM effects are specific to dependency type. Indeed, connectionist studies often are used to confirm that the *exact* observations derived from a corpus study also hold from a ‘brain systems’ perspective. The original implementation of the dual-path model to simulate the neural mechanisms involved in heavy-NP shift reproduced the exact preferences for short-before-long and long-before-short ordering found in English and Japanese corpus studies respectively. This was then used as evidence that it is possible to account for cross-linguistic variability in the direction of heavy-NP shift, which was observed in corpus analyses, by appealing to how the brain processes long dependencies (and handles the lexical competition involved). [11]

With that in mind, is it possible to use the results from our study to account for DLM effects specific to dependency type, on the basis of neuropsychological processing? In some respects, this is broadly possible. Since long/heavy dependencies are considerably likelier to appear in sentence-final positions in our study, we have yet more evidence that some DLM effects can be accurately simulated by a dual-path architecture which makes use of the principle of lexical competition - though ample evidence has already been provided by [11] and [16]. More interestingly, we also have connectionist evidence of the HSP, since relative clauses in particular had such an outsize effect on whether a heavy dependency was moved to sentence-final position. In practice, this means that there was disproportionate delayed activation of the lexical unit introducing the relative clause dependency. This will have been triggered by some tendency to postpone the translation of the relative clause components of the message into the structure of the generated sentence.

The reason for this may lie with the event-semantic layer, which is told by the message that the relative clause adds two extra structural components to the sentence: REL-ATTR and REL-ACTION-LINKING. The inclusion of REL highlights the fact that these extra verb and attribute components ‘belong to’, or depend on, the relative pronoun. This means that potentially, from the perspective of the event-semantic layer, there are stronger bindings between the relative pronoun and other elements within/dependent on the relative clause, increasing the clause’s internal complexity according to the sequencing system, and causing the parsing of any element related to the relative clause to be delayed until all elements of the relative clause have been resolved. Practically, this means that the event-semantics layer holds off resolving the relative until it has parsed the rest of the event-semantic substring of the message (and in all messages corresponding to relative clause structures in the structures.csv input file, REL occurs at the end of the message). The effect of this event-semantics activation pattern must not be sufficiently inhibitory for the relative to be moved to the end 100% of the time, but must still be sufficiently significant for the model to overwhelmingly prefer sentence-final relative placement. This is merely a hypothetical explanation, as it is difficult to isolate individual processes within neural network productions. However, if it is the case, it very closely mirrors HSP-based processing hypotheses, which hold that rightward movement of relative clauses is due to their comparative heaviness - which could be formulated as their internal complexity.

The other disproportionately-occurring element in the output was the preposition ‘to’, which appears to have been more strongly associated with RECIPIENT constructions than the constructions without ‘to.’ This is partly accounted for by the preference of the model (or specifically the event-semantics layer) for structures that postpone the relative clause to sentence-final position, since these very often involve ‘to.’ It may also be because recipient constructions without ‘to’ were of course paired with the same message as those with ‘to’, but did not have any information replacing prep:to encoded in their structures. There is a possibility that, seeing prep:to exclusively associated with messages containing RECIPIENT, the model developed a strong preference for structures involving ‘to’ when translating RECIPIENT messages into structures. Due to this association, structures containing ‘to’ received higher activation and were therefore privileged more often over alternative structures in the lexical competition process. This would make the prevalence of ‘to’ constructions a learning-based outcome, i.e. purely a consequence of the associations learnt by the model during training. This may mean that, contrary to initial expectations, it has nothing to do with ease of processing at all and is merely a result of the preposition’s strong, exclusive association with RECIPIENT constructions making it easier for the model to ‘find’ structures containing ‘to’ when encoding a message with a RECIPIENT.

It is also important to consider whether the comparatively low proportion of long, sentence-final direct object dependencies actually implies that direct object dependencies have weaker DLM tendencies than do obliques or nominal modifiers. Jing et al. conclude in their corpus study that “While [the observed DLM effects for obliques and nominal

modifiers are] consistent with theories of DLM and theories that favor early placement of simple items, all these theories overgenerate, since they make the same predictions for other dependency types as well.” [9] While our data does not allow us to comment on these conclusions with respect to nominal subject, adjectival modifier, or adverbial modifier dependencies, our data concurred with Jing et al.’s study in that we found comparatively little evidence for the DLM-driven shifting of long direct object dependencies. As discussed above, the disproportionate activation of the lexical item ‘to’ may explain this lack of long sentence-final object dependencies, since the inclusion of ‘to’ directly rules out any structure with a sentence-final direct object. Since this is likely to be a learning-based association, it likely has little to do with ease of processing and more to do with ‘to’ being something of a default option for recipient constructions in general. This would make our findings concerning direct objects weaker evidence than they seem against the claim that long direct object dependencies pose less of a processing challenge.

Moreover, returning to Figure 3, we see that the corpus analysis found little DLM-driven *variation* in the order of verb and object. Noun/nominal modifier and verb/oblique dependencies started with head-final word order when dependency length was low and tended towards decreasing prevalence of head-final word order as dependency length increased. Direct object dependencies, meanwhile, had consistently low prevalence from the outset. This does not necessarily mean that DLM is not a concern when human speakers process long direct object dependencies. Rather, it could just mean that the ‘default’ ordering of direct object dependencies is favourable to DLM. This is reasonably consistent with our data, in that we found a reasonable number of sentence-final direct object dependencies in total, but sentence-finality is not specifically linked to longer dependencies in our output, whereas it is for obliques and nominal modifiers.

It may therefore be the case that the sentence-finality of direct object dependencies occurs when it is a ‘default’ option (in a similar way to constructions involving ‘to’ being the default option for sentences with recipients). This squares well with connectionist modelling outcomes, since the output of a connectionist model is so strongly informed by learned associations, and will therefore be strongly biased towards any structure which was learned due to a high frequency of analogical structures in the training input (further discussion on this in section 4.2). In practice, a connectionist model will develop an association of a certain structure with the ‘default’ way to parse a message due to the influence of other, similar/analogical messages (i.e. if the model is ‘translating’ a certain message into a sentence, and a sufficiently similar message had a sentence-final direct object, the activation from this message will have a non-negligible influence on the likelihood of a direct object to be selected for sentence-final position in the target message). All of this suggests that DLM effects may well apply just as strongly to direct object dependencies - but are simply less discernible for direct object dependencies because their default order looks more similar to how a DLM-motivated order would look, regardless of actual dependency length.

The conclusion that DLM effects still apply to direct object dependencies, but are simply less discernible because direct object dependencies are subject to less syntactic variation regardless, is compatible with experimental results on the processing of long direct object dependencies. Although Jing et al.’s study was - to our knowledge - the first corpus study of DLM to take dependency type into account, psycholinguistic studies have done so before. The production study which found that the relative length of a dependency had a stronger accessibility effect on NPs than absolute length, and led to stronger heavy-NP shift tendencies, *was* a study of direct object dependencies [7]. Another study, which used eye-tracking to determine which part of a sentence with a long dependency posed the most processing difficulty (causing the eyes to linger), also focused on direct object dependencies. It found that difficulty effects appear concentrated on the direct object itself or on the material intervening between the direct object and the verb (depending on the verb’s transitivity) [24]. These studies suggest that DLM-driven shifting of direct object dependencies *does* occur, and *is* motivated by processing difficulty. Therefore, although our connectionist model and Jing et al.’s corpus study appear to imply that there are weaker DLM effects for direct object dependencies, we should not write off the applicability of the DLM strategy to direct object dependencies given the experimental evidence to the contrary. This is particularly true given that the corpus data and the model output can partly be explained by the low positional variation of direct object dependencies, and the fact that they tend to appear in positions which enable DLM regardless.

4.2 Limitations

There were several limitations to the approach we used in this study, which should be taken into account. The first category of these is the technical limitations and bugs, which could probably have been fixed given more time. For instance, the model struggled to deal with adverbial modifiers correctly. Adverbs were encoded into the lexicon and structures using a similar mechanism to other parts of speech: they were included in the lexicon input file under the ‘adverb’ category, and this category was then included in a few different messages in the structures file, which were then paired with multiple sentence structures to account for different possible word orders. However, the adverbs ended up being parsed by the network as part of the ACTION-LINKING (verbal) component of the message, which threw off the structure parsing for the rest of the sentence, and led to the omission of adverbs entirely in many cases (hence the very low prevalence of adverbial modifiers in the output data). In addition, some sentences produced by the network

were ungrammatical, though there were not many of these and they were generally ungrammatical by omission (a particularly common error was the omission of determiners). There was also a bug which caused some sentence-final nominal modifiers to be incorrectly identified as shorter than others, though not enough to significantly impact the final summary data and the conclusions drawn.

Further issues concern the algorithm we used to label the model's output. If a sentence is labelled 'obl y', for example, this does tell us that the sentence has a heavy sentence-final oblique and therefore aligns with the hypothesis that DLM effects apply to heavy/long obliques. However, it does not tell us whether the sentence-finality of the oblique was caused by the length of the oblique or simply by the fact that it is an oblique. There are, after all, hundreds of sentence-final obliques in the data, and it is difficult to isolate the length of the oblique as a cause of sentence-finality with the information that we have, as opposed to postulating that obliques have a tendency to be placed sentence-finally in general. The label 'obl n', meanwhile, can be interpreted as supporting two opposing claims. On the one hand, it suggests that the final dependency in the sentence is a non-heavy oblique, and so it has been shifted to the end of the sentence despite not being the longest dependency. It therefore counts towards the 'non-long' total in Figure 5 and provides an example of a sentence-final oblique which is not due to DLM. However, if an oblique is in sentence-final position and is not the longest dependency in the sentence, there must be a longer dependency somewhere in the sentence which is not an oblique - which *supports* the claim that other dependency types incur weaker DLM effects. Therefore, the relevance of a non-long sentence-final dependency to the question of which dependency types exhibit the strongest DLM effects is not clear-cut. Non-long sentence-final obliques suggest that sentence-finality in obliques is not always due to DLM, but also suggests that the other, longer dependency in the sentence did not incur a DLM effect (because it did not move to sentence-final position), thus both supporting and opposing the claim that obliques provoke the strongest DLM tendencies.

The algorithm also missed certain details of dependency length: it computed relative length as a binary value (longest or not longest), ignoring any specific effects which might have emerged from the exact difference in length between the dependencies in a sentence. Moreover, it is sometimes difficult to separate out the effects of different dependency types, since sometimes dependencies of one type are embedded within dependencies of another type ("I gave the book to the boy's sister" contains a nominal modifier within an oblique, but would have been labelled as a sentence-final nominal modifier by the algorithm, ignoring any effect which the oblique may have had on word order).

Finally, it is worth commenting on the suitability of a machine learning approach for a research question like this. Our model receives a message to 'translate' into a natural language sentence and decides on which of two alternative structures to use, based on which structure corresponds to the lexical units which receive the highest activation at the point in the sentence where the two structures diverge. This is useful for simulating processes involved in sentence construction because such lexical competition has been postulated to take place in human language production. However, the output of a machine learning model is much more heavily based on the specificities of the data on which it was trained than any human speaker's sentence production. During the training of the model, it becomes optimised to produce sentences which are as close as possible to the training sentences by accurately predicting the next word in each sentence. This is why (as discussed in section 3) most words appeared in roughly the same proportions in the model's output as in the model's input, with the exception of mid-sentence relative clauses and 'to' (which is probably due to the associations developed by the event-semantics layer in particular, due to the ordering of certain elements in the message). As such, connectionist modelling is most suitable for simulating how humans derive information on how to construct a sentence from the input they are exposed to. This leaves us unable to use the data from this study to answer questions regarding the interaction of dependency type and DLM with semantic information, working memory constraints, and other factors which are separate from learning processes. In other words, connectionist models are necessarily 'learning-based accounts.'

These limitations do not preclude us from gaining useful information from this connectionist modelling exercise. We have learnt that DLM effects can be supported by a 'brain systems' approach, since sentence-final dependencies had a significant tendency to be long. We observe that particular ordering preferences developed by the model with respect to relative clauses and the preposition 'to' may be due to the development of stronger associations in the meaning system between the concept of a recipient and the preposition 'to' (and hence structures involving 'to'), and the way in which the event-semantics layer developed internal representations of the relationships between different components of the message. While the processing of relative clauses by the event-semantics layer is reasonably analogous to the human perception of relative clauses as more internally complex and rich (i.e. the HSP), the way the model deals with the preposition 'to' appears to simply be a learned association rather than a processing facilitation strategy like DLM. Therefore, we have achieved only partial success at using a connectionist model to simulate the neuropsychological processes involved in DLM. We also found that sentence-final direct object dependencies are less frequently the longest in their sentences than sentence-final obliques and nominal modifiers, but conclude that this is insufficient evidence to prove that DLM effects apply less strongly to direct objects, since it could simply be a result of direct object dependencies having fewer possibilities for positional variation (and therefore less propensity for shifting) than other

dependency types. This hypothesis concerning the low variance of direct object dependency placement is supported by Jing et al.'s corpus study.

4.3 Further research

With these conclusions in mind, we now turn to the opportunities for further research provided by this study. Since a connectionist model will inevitably produce outcomes that are more strongly influenced by learning constraints than processing capacity constraints, we suggest using more direct psycholinguistic methods, such as eye tracking, to better understand whether certain dependency types are inherently harder to process than others when in the middle of a sentence. Future connectionist studies could focus more specifically on how human speakers learn the DLM effects associated with certain dependency types from the input. This could be done from a child language acquisition perspective (as in [21]) or by controlling the distribution of structures in the training sets such that they conform to the distributions found in corpora, to see if the trends from our study which were specific to certain dependency types still hold when input is more realistic. This would make for a more similar connectionist approach to the landmark past tense morphology study, which showed that morphological irregularity patterns are learnable from input that a child could realistically have been exposed to [14]. If we can demonstrate via connectionist modelling that DLM patterns for each dependency type are similarly learnable from input, there may be reason to believe that DLM effects are somehow derived from input rather than from a processing optimisation strategy. We concur with Jing et al. that more research on the specific effects of dependency type on DLM - and psycholinguistic study of DLM in general - is necessary.

Acknowledgements

Many thanks to Professor John Williams for supervising this project, to Professor Franklin Chang for providing conceptual insights on the dual-path model, and to Dr Chara Tsoukala for the original Python implementation of the model (and for being extremely responsive over GitHub and email!)

References

- [1] Richard Futrell, Kyle Mahowald, and Edward Gibson. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341, 2015.
- [2] Edward Gibson. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76, 1998.
- [3] Daniel Grodner and Edward Gibson. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive science*, 29(2):261–290, 2005.
- [4] John Robert Ross. *Constraints on variables in syntax*. ERIC, 1967.
- [5] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.
- [6] John A Hawkins. A parsing theory of word order universals. *Linguistic inquiry*, 21(2):223–261, 1990.
- [7] Lynne M Stallings and Maryellen C MacDonald. It's not just the "heavy NP": relative phrase length modulates the production of heavy NP shift. *Journal of Psycholinguistic Research*, 40(3):177–187, 2011.
- [8] Lynne M Stallings, Maryellen C MacDonald, and Padraig G O'Seaghdha. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy NP shift. *Journal of Memory and Language*, 39(3):392–417, 1998.
- [9] Yingqi Jing, Paul Widmer, and Balthasar Bickel. Word order variation is partially constrained by syntactic complexity. *Cognitive science*, 45(11):e13056, 2021.
- [10] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. Universal dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*, 2020.
- [11] Franklin Chang. Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61(3):374–397, 2009.
- [12] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- [13] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [14] David E Rumelhart, James L McClelland, PDP Research Group, et al. *Parallel distributed processing*, volume 1. IEEE New York, 1988.

-
- [15] David C Plaut. A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, 23(4):543–568, 1999.
- [16] Franklin Chang and Hartmut Fitz. Computational models of sentence production: a dual-path approach. *The Oxford handbook of language production*, pages 70–87, 2014.
- [17] Anna Y Klintsova and William T Greenough. Synaptic plasticity in cortical systems. *Current opinion in neurobiology*, 9(2):203–208, 1999.
- [18] Ton Dijkstra and Koenraad De Smedt. *Computational psycholinguistics: AI and connectionist models of human language processing*. Taylor & Francis, 1996.
- [19] John A Hawkins. *Efficiency and complexity in grammars*. OUP Oxford, 2004.
- [20] Franklin Chang. Symbolically speaking: A connectionist model of sentence production. *Cognitive science*, 26(5):609–651, 2002.
- [21] Katherine E Twomey, Franklin Chang, and Ben Ambridge. Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, 73:41–71, 2014.
- [22] Jeffrey L Elman. Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1):71–99, 1993.
- [23] Jean K Gordon and Gary S Dell. Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, 27(1):1–40, 2003.
- [24] Adrian Staub, Charles Clifton Jr, and Lyn Frazier. Heavy NP shift is the parser’s last resort: Evidence from eye movements. *Journal of memory and language*, 54(3):389–406, 2006.