# Modeling the regular/irregular dissociation in non-fluent aphasia in a recurrent neural network

**Alexandra Krauska (akrauska@umd.edu)**
Department of Linguistics, 7814 Regents Dr
College Park, MD 20740 USA

**Naomi H. Feldman (nhf@umd.edu)**
Department of Linguistics and UMIACS, 7814 Regents Dr
College Park, MD 20740 USA

## Abstract

In the debate between single-route and dual-route models of verb inflection, the dissociation between regular and irregular verbs in the non-fluent variety of aphasia has been a key sticking point for the proponents of the single-route model. This paper adopts a state-of-the-art neural model which has previously been used to learn inflectional morphology, and shows that it can also be used to model data from non-fluent aphasia. This challenges the assumption that a dual-route model is necessary to capture apparent dissociations in aphasia data and encourages a reanalysis of the deficits involved in non-fluent aphasia.

**Keywords:** non-fluent aphasia; Encoder-Decoder network; past-tense debate; inflection

There has been heated debate over the past 40 years about how people represent regular and irregular inflectional morphology (Clahsen et al., 1992; Goebel & Indefrey, 2000; Hahn & Nakisa, 2000; Hare et al., 1995; Laaha et al., 2006; Lachter & Bever, 1988; MacWhinney & Leinbach, 1991; Marcus et al., 1995; McClelland & Patterson, 2002; Plunkett & Marchman, 1991; Plunkett & Juola, 1999; Pinker & Prince, 1988; Pinker, 1998; Pinker & Ullman, 2002; Rumelhart & McClelland, 1986). Proponents of the single-route model have argued that the phonological relationship between inflected and uninflected forms can be captured within a single pattern associator. In contrast, proponents of the dual-route model have argued that separate mechanisms are responsible for the regular and irregular inflections: regular inflection is performed by the grammar, while irregular inflection is stored in the lexicon.

Recently, state-of-the-art single-route neural models have been able to achieve high degrees of accuracy when learning the English past-tense (Kirov & Cotterell, 2018), addressing many of the initial issues that previous single-route models faced (with some limitations; see McCurdy et al., 2020). However, a key sticking point for the single-route approach has been the dissociation between regular and irregular inflections in people with different types of aphasia, an acquired language disorder that arises after a brain injury such as stroke (Ullman et al., 1997). In non-fluent aphasia, there seems to be a selective deficit for regular inflections, while in fluent aphasia, there seems to be a selective deficit for irregular inflections.

In this paper, we reassess the strength of that evidence. We apply a state-of-the-art single-route network which has previously been used to learn the past-tense inflection, and show that it can also model data from non-fluent aphasia in English. This model indicates that the dissociation does not require multiple mechanisms to capture this particular pattern of deficits. These findings contribute to the ongoing past-tense debate in cognitive science, providing an alternate single-route interpretation for one of the key pieces of evidence in support of the dual-route model.

Whereas previous work on the single-route model has cast it in opposition to the traditional idea of grammatical rules in linguistics, we pursue a different interpretation in our discussion of these results, based on recent advances in linguistic theory. We situate the single-route model as performing a particular function that any traditional model of grammar would need to carry out: converting the underlying hierarchical syntactic structure into a linear order and making calculations over phonological space to determine its form. We argue that non-fluent aphasia should be approached as a deficit not in syntactic operations, but in phonological operations.

We begin by reviewing the evidence from aphasia, discussing empirical data as well as previous models. We then give an overview of the model from Kirov & Cotterell (2018), which we adopt in our simulations. The next section describes simulation results showing that under certain training conditions, this model can simulate the dissociation in non-fluent aphasia. We conclude by discussing the relationship between the past tense debate and linguistic theory in light of substantial recent advances in both fields.

## Morphological Deficits in Aphasia

As an extension of contemporary models of memory circuits, Ullman (2004) developed the declarative/procedural model to connect language to more domain-general cognitive processes. This theory suggests that the grammar is a subdivision of procedural memory, like other motor and cognitive skills, while the lexicon is a subdivision of declarative memory, which stores arbitrary relations, facts, and events. Consequently, their model predicts that regular and irregular forms should be served by different neural substrates associated with the different types of memory - procedural memory corresponding to the basal ganglia and the frontal cortex, and declarative memory corresponding to the hippocampus and temporo-parietal regions. Data from language disorders such as aphasia can provide exactly the kind of evidence needed to test this hypothesis. Aphasia is an acquired language disorder that occurs as result of a brain injury - such as a stroke, tumor, or head injury - that impacts areas of the brain associated with language. There are many varieties of aphasia, but two key varieties are fluent and non-fluent aphasia. The non-fluent variety of aphasia is characterized by “telegraphic speech”, which tends to include few function words and morphology, so it has often been analyzed as a ‘grammatical’ or syntactic deficit. Meanwhile, the fluent variety of aphasia often involves “empty speech”, syntactically well-formed sentences that lack meaning or message, so it has often been analyzed as a ‘lexical’ deficit. Though a brain injury is not guaranteed to be isolated to one area of the brain, analysis of lesion patterns and behavioral deficits can provide convincing evidence for functional dissociations.

Ullman et al. (1997) identified two patients, FCL and JLU, with opposite lesion patterns, and tested their ability to produce the past tense form of regular and irregular verbs. Their results are shown in Figure 1. FCL, who had non-fluent aphasia, produced the regular past-tense with only 20% accuracy, while producing the irregular past-tense with 69% accuracy. In contrast, JLU, who had fluent aphasia after injury to the left temporo-parietal area, produced regular verbs with 90% accuracy, and produced irregular verbs with 63% accuracy.

The two patients, FCL and JLU, are the key evidence for the proposed double dissociation between regular and irregular inflections. They show opposite patterns of accuracy rates for regulars and irregulars, as would be expected under the dual-route model. Previous models have shown that dissociations can arise through a single learning mechanism (Penke & Westermann, 2006; Plaut, 1995), but even in those cases, the models developed functional specialization that was different for regulars and irregulars. Based on these data points alone, there does not seem to be an obvious explanation that would support the single-route account.

There are several reasons why we think these data should be reexamined. First, whereas the rhetoric surrounding the original past tense debate characterized dual-route models as more consistent with linguistic theory, this has changed due to advances in linguistics. The Pinker & Prince (1988) model was built in a ‘lexicalist’ framework, which assumes that syntactic operations do not extend below the word level, making morphological operations distinct from syntactic ones. This assumption has been challenged in the past 40 years, especially with the development of theoretical frameworks such as Distributed Morphology (Halle et al., 1993) and non-semiotic approaches (Preminger, 2021). These kinds of theories argue that words and phrases are built using the same syntactic processes, and functional morphemes such as the English past-tense do not have a ‘built-in’ phonological form (Embick, 2015). In Distributed Morphology, their form is instead supplied in the post-syntactic operation of Vocabulary Insertion; in Preminger’s theory, the form is provided by a mapping from sets of syntactic units to sets of phonological units. According to these views, from a syntactic perspective, the verbs “ran” and “walked” do not differ in terms of their hierarchical structure—both involve a verb and the past-tense morpheme combined in a particular way. The only difference is how that past-tense morpheme conditions the form of the output, and how complex and frequent those transformations may be.

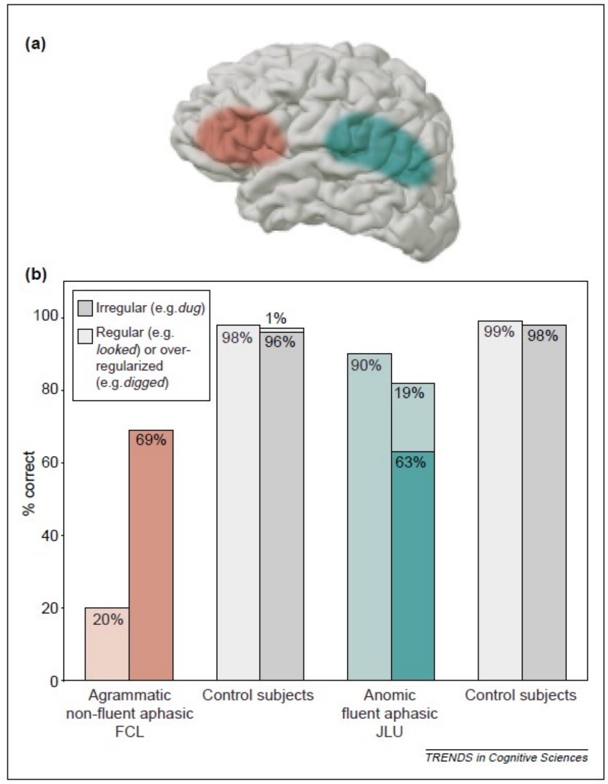


Figure 1: Patients FCL and JLU are used to demonstrate a neural dissociation; FCL, in red, exhibits left anterior perisylvian lesions and a greater deficit for regular inflections, while JLU, in blue, exhibits left temporo-parietal lesions and a greater deficit for irregular inflections (Ullman et al., 1997)

Second, studies of non-fluent aphasia cross-linguistically suggest that the pattern of deficits even in one variety of aphasia seems to be more nuanced, and not entirely consistent with the predictions of the dual-route model. Faroqi-Shah (2007)’s meta-analysis of non-fluent aphasia across seven different languages compared participants’ performance on the regular and irregular past-tense inflection. The analysis included 25 studies published between 1980 and 2006. The resulting dataset contains 66 different participants, and includes speakers of Catalan, Dutch, English, German, Greek, Italian, and Spanish. Figure 2 summarizes the participants’ performance.

According to the dual-route model, non-fluent aphasia should be an impairment in the grammar, and thus a selective deficit for regular verbs. If this were the case, Figure 2 would show many more participants - from all language groups - performing at or near ceiling for the irregular verbs, with lower accuracy on regular verbs. This would mean more points near the top of the plot or above the solid line, and few below the line. Instead, this shows a large number of German- and Dutch-speaking participants that have high accuracy on regular verbs but varying accuracy on irregular verbs, as well as a number of English-speaking participants with better accuracy on regular verbs than irregular verbs. The ‘dissociation’, consequently, is less straightforward than Pinker & Ullman (2002) suggested,

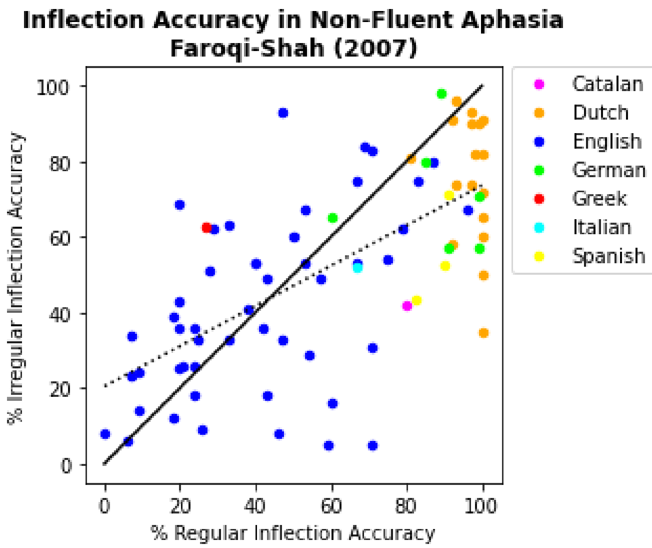


Figure 2: Plot summarizing the inflection accuracy of participants in the studies cited in the Faroqi-Shah (2007) meta-analysis of non-fluent aphasia (n=66). The solid line corresponds to equal accuracy on regular and irregular inflections. The dotted line corresponds to the line of best fit ($y = 0.53x + 20$, $R^2 = 0.429$).

especially cross-linguistically.

These data create a puzzle. It is not only challenging for the dual-route model, because it does not present an obvious selective deficit for regular inflections, but also for the associated analyses of non-fluent aphasia. Other theories of the functional organization of language, such as Matchin & Hickok (2020), suggest that the brain area implicated in non-fluent aphasia, the inferior frontal gyrus (IFG), is involved in linearization and phonological transformations rather than rule-based syntactic operations. This model suggests that non-fluent aphasia is not a syntactic deficit, but a phonological one.

If the deficit is phonological, rather than a strict dissociation, patterns of deficits in aphasia could be the result of an interaction between irregularity, the complexity of the phonological transformation, and frequency, or a number of other factors. For example, whereas regular verb inflection in English can make word-final consonant clusters more complex (as in walk $\sim$ walked) or add a syllable (load $\sim$ loaded), as discussed by Bird et al. (2003), irregular verbs can vary in the complexity of the transformation: “sing” only changes one vowel sound, whereas “be” and “go” undergo suppletion. Irregular verbs also tend to have higher frequencies than regular verbs, partly due to factors involved in language change (Bybee, 1995; Lieberman et al., 2007). Because every language involves different kinds of phonological transformations with varying complexity, and with varying frequency distributions, non-fluent aphasia may result in different patterns of deficits cross-linguistically based on these properties, even if the lesion pattern is similar across patients. This kind of explanation could account for the differences between language groups in Faroqi-Shah (2007), where the dual-route model cannot.

Recent work that has revisited the original past-tense debate using state-of-the-art neural models, such as the Encoder-Decoder (ED) network used by Kirov & Cotterell (2018), has addressed many of the initial critiques that were directed at the Rumelhart & McClelland (1986) model. Specifically, they showed that a model trained on English inflections exhibits high accuracy rates, produces human-like novel forms, and mirrors some child-like learning patterns. It can also achieve high accuracy rates when trained on multiple inflection classes such as the gerund, past participle, and third-person singular forms. The success of this neural model suggests that a single-route model is sufficient for both regular and irregular inflections, but it does not explain how the aphasia data discussed above might arise. Non-fluent aphasia has been modeled before by Penke & Westermann (2006), which simulated the pattern of deficits for German-speakers. However, their model was not fully homogeneous, and it is not clear if their model would see the same success for English data.

Given this background, if non-fluent aphasia is not a deficit in rule-based syntactic operations, but in the phonological transformations involved in inflection, then it should be possible to simulate the effects of non-fluent aphasia in a single-route neural model that performs that kind of phonological transformation. In this paper, we demonstrate that a state-of-the-art single-route model can be used to account for the pattern of deficits in non-fluent aphasia. In Simulation 1, as a proof of concept, we show that training an ED network on datasets with different frequency distributions leads to greater effects of lesioning in one set of verbs. In Simulation 2, we show that a model trained on an English-like dataset using perplexity-based sampling can lead to an English-like pattern of deficits. As predicted by the single-route approach, the dissociation can be captured without appealing to separate lexical and grammatical mechanisms.

## Model

The model we use in the present study is a homogeneous, single-route, single-mechanism model that - when lesioned - can simulate the pattern of deficits observed in non-fluent aphasia in English-speaking populations. In doing so, this model indicates that the dissociation that has been a key piece of evidence for the Dual-Route model does not, in principle, require multiple mechanisms to capture this particular pattern of deficits. The effect of fluent aphasia requires further research and will be investigated in future work.

Simulation 1 tests datasets with different proportions of regulars and irregulars. Simulation 2 looks at the effect of using perplexity—essentially, the model's degree of surprise at a particular data point—to constrain learning.

***Architecture.*** Our models use the Encoder-Decoder network architecture specified by Kirov & Cotterell (2018), which includes the architecture described by Bahdanau et al. (2014) and hyperparameters set by Kann & Schütze (2016). This network operates over strings of characters which represent phonemes. The encoder includes a bidirectional LSTM with

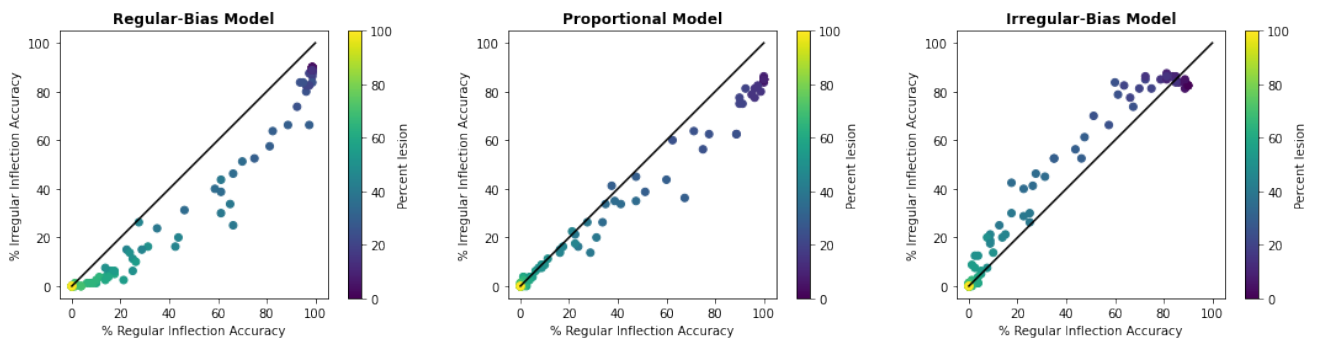


Figure 3: Results from Simulation 1. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where less than 5% of the dataset is made up of irregular verbs. **Center:** Plot summarizing the performance of the model trained on the Proportional dataset, where 26% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the Irregular-Bias model, where 95% of the dataset is made up of irregular verbs.

two layers. Each character has an embedding size of 300 units. The dropout value between layers is 0.3. The encoder and decoder have 100 hidden units each. The Adadelta training procedure (Zeiler, 2012) was used, with a learning rate of 1.0 and a minibatch size of 20. The model was trained for 100 epochs. The decoder uses a beam search (k=12). The model was implemented using OpenNMT-py 2.0 (Klein et al., 2020).

***Training data.*** The base dataset is also the same one used by Kirov & Cotterell (2018), which includes 4,039 English verbs from the CELEX database (Baayen et al., 1996). Of these verbs, 168 have irregular inflections, and the remaining 3,871 have regular inflections. 20% of this dataset was held out for validation, so the training dataset included only 3,232 verbs, 141 of which were irregular. The testing dataset included 80 regular verbs and 80 irregular verbs which were all seen in training, since we are more interested in the model's performance after lesioning than how accurately it can generalize to new forms.

From this base dataset, three datasets were developed. The **Regular-Bias** dataset has each verb in the corpus appear exactly once, so that only 4.6% of the verbs in that training dataset are irregular, with a total of 3,232 items. The **Proportional** dataset had each verb appear proportional to its COBUILD frequency, so the dataset was composed of 26% irregular verbs, with a total of 15,424 items. In the **Irregular-Bias** dataset, a subset of high-frequency regular verbs were selected from the base dataset, and these appear only once; the remaining 141 irregular verbs are repeated proportional to the inflected forms' COBUILD frequency per 1 million words. The resulting dataset is composed of 5% regular verbs and 95% irregular verbs, with a total of 15,645 items.

The models were trained for the same number of train steps, so that each one was exposed to the same amount of training data despite the differences in the sizes of each dataset.

***Lesioning.*** After training, each model was 'lesioned' by randomly resetting connection weights to 0.[1] The proportion of weights that were reset ranged from 5% to 100% by increments of 5%. After lesioning, the models' accuracy on the test dataset was measured. This was repeated 5 times for each model in order to observe the effect with different random seeds.

## Simulation 1

In Simulation 1, one model was trained for each of the three datasets, and its accuracy was measured before and after lesioning. We make several predictions. Firstly, if any of the single-route models we test can simulate a dissociation similar to what was observed for FCL (Ullman et al., 1997), this would provide a proof in principle that the dissociation can be captured without requiring multiple distinct mechanisms. Secondly, if models that observe different proportions of regulars and irregulars in their training data exhibit different behavior after lesioning, this would suggest that the frequency distribution of regulars and irregulars impacts how those transformations are encoded. This would also suggest that the cross-linguistic differences observed by Faroqi-Shah (2007) might depend on such factors.

Figure 3 shows the performance of each of the models. These plots compare accuracy on regular and irregular inflections. Each point represents a different instance of the model, with different proportions of connections lesioned, and with different random seeds. The black diagonal line represents equal performance on both sets of verbs, so points above the line represent models performing better on irregular verbs, while points below the line represent models performing better on regular verbs.

For the model that was trained on the Regular-Bias dataset, the performance was consistently better on regular inflections than irregular inflections. The model trained on the Proportional dataset shows better performance on irregular inflections than the Regular-Bias model, but most points still fall below

[1]During a stroke or brain injury, lesions are not randomly distributed across a brain area, but instead localized to a set of adjacent neurons. In the neural network, there is no sense of 'adjacency'; all of the neurons in one layer are connected to all of the neurons in the neighboring layers. Randomly lesioning the neural network provides a reasonable approximation of the effect of a brain injury, given the differences between neural architectures and neural networks.

the line of equivalence. For the model that was trained on the Irregular-Bias dataset, the model performs better at irregular verbs than regular verbs at nearly every point. Though the pattern in Figure 3 does not perfectly resemble the pattern found in Faroqi-Shah (2007), shown in Figure 2, this demonstrates that it is possible to achieve a range of outcomes with a model trained on regular and irregular verb inflections without appealing to a grammatical or lexical distinction. Because the effect was only achieved for the model that was trained on a dataset composed predominantly of irregular verbs, these frequency statistics must play a key role in how the two sets of verbs are encoded in the model.

This simulation shows that the frequency distribution of the training dataset does play a role in the pattern of deficits observed after lesioning the model, suggesting that this factor could be partially responsible for the cross-linguistic differences observed by Faroqi-Shah (2007). Although the lesioning did not quantitatively replicate the proportions of regular and irregular verbs produced correctly by FCL, it did show that a single-route model can produce qualitative asymmetries. The fact that the asymmetries in the model do not quantitatively resemble FCL may be because that the models do not generate much variety within the groups of lesioned models in terms of accuracy; the points on the plot are relatively closely grouped together. We return to this point in the Discussion.

While this provides a proof in principle that the relevant patterns can be captured in a single-route model, the model trained on the Proportional dataset, the one with the most realistic proportions of English verb inflections, was not able to capture the pattern of deficits observed for the patient FCL. This difference motivates Simulation 2.

## Simulation 2

In Simulation 2, a second set of models were created that used perplexity-based sampling (Fernandez & Downey, 2018). One model was trained on each of the three datasets.

Perplexity-based sampling involves calculating a 'perplexity score' for each item in the training dataset, which measures how well the model was able to predict the output form given its current state. The model then selects the items with the highest score for the next training step. This simulates the attention, surprisal, and reanalysis that a learner may experience when presented with a form that was not predicted (Clark, 2013). If models trained using perplexity-based sampling better capture the empirical data than models trained with random sampling, this would suggest that surprisal, attention, and reanalysis play a role in shaping neural representations.

The results for Simulation 2 are shown in Figure 4. The model trained on the Regular-Bias dataset with perplexity-based sampling exhibited similar performance to the Regular-Bias model in Simulation 1, where most of the points fell below the solid line. The model trained on the English-like dataset with perplexity-based sampling falls close to the equivalence line. The difference between this model and the equivalent one in Simulation 1 is slight, but closer inspection shows that the model trained with perplexity-based sampling has higher accuracy on irregular inflections than the model trained with random sampling when matched on the proportion of lesions. Though this would not be able to capture patient FCL, in Faroqi-Shah (2007), the English speaking group was broadly centered around the equivalence line, especially at higher rates of lesioning. Lastly, the model trained on the Irregular-Bias dataset with perplexity-based sampling was once again almost completely above the equivalence line, performing better at irregulars than regulars at every point. The pattern observed for the Proportional model in Simulation 2 suggests that perplexity-based sampling does shift it closer to the pattern observed for the Irregular-Bias model, to a degree. In addition, the difference in performance between these three models in Simulation 2 indicates that the frequency distribution of regulars and irregulars has an impact for these models even in the context of perplexity-based sampling.

Overall, both simulations show that a single-route model can capture dissociations between regulars and irregulars, and that this relationship depends on specific characteristics of the training data. Based on the complexity and the frequency of the transformation, a network may encode the two classes differently, and this encoding can then affect performance. For example, a more complex transformation may lead to lower accuracy rates in an non-lesioned network, but if it has a more distributed representation it may be more resilient to lesions. The model trained on a dataset biased toward irregulars illustrates this effect using a dataset based on English. If the model were trained on inflections in another language, with different frequency distributions and different degrees of complexity, we might observe a different pattern of results. Furthermore, the degree of 'perplexity' that the learner experiences as they encounter new forms may vary across languages, depending on the patterns of sub-regularity. In this way, patterns that vary across language groups can arise not because of representational differences, but because of differences in the complexity and frequency of the transformations.

## Discussion

This paper demonstrates that lesioning a fully homogeneous state-of-the-art neural model can simulate the effects of non-fluent aphasia. This result provides an alternate account for the apparent dissociation between regular and irregular inflections observed by Ullman et al. (1997), which was previously a sticking point for the single-route model in the past tense debate. A critical component of this finding is the difference in frequency distributions between regular and irregular inflections in English. Simulation 1 demonstrates that when the model is trained on a dataset that has a greater frequency of irregular verbs, it is able to simulate the pattern of deficits in aphasic populations. All languages should have a frequency distribution somewhere between these two extremes, and so their pattern of deficits may be predicted based on that. In Dutch, for example, the frequency distributions of regular verbs and irregular verbs overlap more than they do in En-

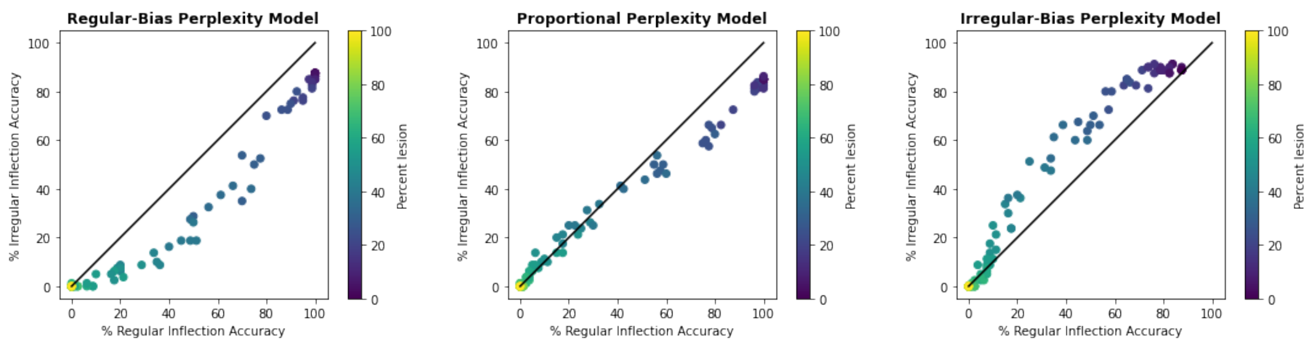


Figure 4: Results from Simulation 2, with perplexity-based sampling. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where less than 5% of the dataset is made up of irregular verbs. **Center:** Plot summarizing the performance of the model trained on the Proportional dataset, where 26% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the model trained on the Irregular-Bias dataset, where 95% of the dataset is made up of irregular verbs.

glish (Tabak et al., 2005), so a model trained on Dutch verbs may exhibit a greater deficit for irregular verbs when lesioned, closer to the Regular-Bias model in Simulation 1. This would align with the clinical data discussed by Faroqi-Shah (2007).

After training on an English-like dataset, the model is able to mimic the patterns that Faroqi-Shah (2007) found in English speakers, particularly when the network is trained using perplexity-based sampling. Thus, rather than assuming that the pattern of deficits in non-fluent aphasia arises because of syntactic or lexical features of the verbs, we can argue instead that the process involved in the IFG is a phonological transformation, conditioned by frequency and form predictability.

By placing Ullman et al. (1997)'s findings into a broader context of non-fluent aphasia across different languages, we demonstrate that FCL's performance may be an outlier when compared to other individuals with non-fluent aphasia, consistent with the argument in Plaut (1995). The data described by Faroqi-Shah (2007) do not clearly support the dual-route model as proposed by Pinker & Ullman (2002).

Faroqi-Shah's meta-analysis demonstrated that there is a significant amount of variation between individuals with aphasia, even when the injury impacts very similar brain areas. This could emerge due to individual differences in how each form is encoded neurally, or due to differences in which groups of neurons are impacted and whether they are able to be reorganized through neural and synaptic regeneration processes or by leveraging alternate pathways through the brain (Campbell et al., 2019). Other forms of variation could be introduced to these models by using a different lesioning method, such as adding Gaussian noise to all weights between layers, unit ablation—severing all outgoing connections from some units—or adding Gaussian noise to the activations of units. Each strategy can have a different effect on the performance of the model, as discussed in Guest et al. (2020).

Once again, these findings should not be interpreted as a challenge to the view that words and sentences have hierarchical structure that is relevant during on-line language comprehension and production. Rather, we interpret this model as performing a calculation over phonological space. It represents a post-syntactic operation such as 'Vocabulary Insertion', as characterized by Distributed Morphology, or 'mapping from sets of syntactic units to sets of phonological units', as described by Preminger (2021).

There are many differences between our model and the actual IFG. This means that it is not a perfect model of what might be happening at the neural level after a brain injury such as a stroke. Designing a model to more closely reflect the architecture of the IFG might yield different results or further insights in future work. However, even neural networks that are designed to have similar functions or mechanisms as the human brain can sometimes behave very differently (Rajalingham et al., 2018). Our approach is simpler, but is nevertheless sufficiently similar to the hypothesized the function of the IFG to allow us to test predictions of what should happen when the mechanism is damaged in a controlled way.

Future work should simulate the effects of fluent aphasia, or the accuracy data for the patient JLU from Ullman et al. (1997). Fluent aphasia is not as well-understood as non-fluent aphasia, and there are many other competing theories that may not be as easy to test with this kind of neural network. For example, because fluent aphasia often involves "empty speech" (syntactically well-formed sentences that lack meaning or message), it has sometimes been concluded that fluent aphasia involves difficulties with discourse coherence (Linnik, 2016), or a failure to properly inhibit incorrect lexical items that are retrieved (Prather et al., 1997). These may be difficult to implement as a neural model, though additional investigation in this area could be fruitful.

In the future, it will also be important to identify where different languages fall along the continuum between the Irregular-Bias dataset and the Regular-Bias dataset, as discussed above for Dutch. If different languages discussed in Faroqi-Shah (2007) fall at different points along this continuum, then networks trained on different languages may exhibit different effects of lesioning. This would allow us to observe the impact of different language inputs and frequency distributions, and to simulate the variation in non-fluent aphasia cross-linguistically.

## Acknowledgements

We thank Nika Jurov for her endless patience and help getting the code to run, and members of the UMD Linguistics Department for very constructive discussions.

## References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The celex lexical database (cd-rom).* University of Pennsylvania.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bird, H., Ralph, M. A. L., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language*, *48*(3), 502–526.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425-455.

Campbell, B. C., De Silva, D. A., Macleod, M. R., Coutts, S. B., Schwamm, L. H., Davis, S. M., & Donnan, G. A. (2019). Ischaemic stroke. *Nature Reviews Disease Primers*, *5*(1), 1–22.

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, *45*, 225-255.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181–204.

Embick, D. (2015). *The morpheme*. De Gruyter Mouton.

Faroqi-Shah, Y. (2007). Are regular and irregular verbs dissociated in non-fluent aphasia?: A meta-analysis. *Brain Research Bulletin*, *74*(1-3), 1–13.

Fernandez, J., & Downey, D. (2018). Sampling informative training data for rnn language models. In *Proceedings of acl 2018, student research workshop* (pp. 9–13).

Goebel, R., & Indefrey, P. (2000). A recurrent network with short-term memory capacity learning the German -s plural. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (p. 177-200). Oxford: Oxford University Press.

Guest, O., Caso, A., & Cooper, R. P. (2020). On simulating neural damage in connectionist networks. *Computational brain & behavior*, *3*(3), 289–321.

Hahn, U., & Nakisa, R. C. (2000). German inflection: Single route or dual route? *Cognitive Psychology*, *41*, 313-360.

Halle, M., Marantz, A., Hale, K., & Keyser, S. J. (1993). Distributed morphology and the pieces of inflection. *1993*, 111–176.

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes*, *10*, 601-630.

Kann, K., & Schütze, H. (2016). Med: The lmu system for the sigmorphon 2016 shared task on morphological reinflection. In *Proceedings of the 14th sigmorphon workshop on computational research in phonetics, phonology, and morphology* (pp. 62–70).

Kirov, C., & Cotterell, R. (2018). Recurrent neural networks in linguistic theory: Revisiting pinker and prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, *6*, 651–665.

Klein, G., Hernandez, F., Nguyen, V., & Senellart, J. (2020). The opennmt neural machine translation toolkit: 2020 edition. In *Proceedings of the 14th conference of the association for machine translation in the americas (amta 2020)* (pp. 102–109).

Laaha, S., Ravid, D., Korecky-Kröll, K., Laaha, G., & Dressler, W. U. (2006). Early noun plurals in German: regularity, productivity, or default? *Journal of Child Language*, *33*(2), 271-302.

Lachter, J., & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning - a constructive critique of some connectionist learning models. *Cognition*, *28*, 195-247.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, *449*(7163), 713–716.

Linnik, A. (2016). *Coherence and structure in aphasic and non-aphasic spoken discourse*. Unpublished doctoral dissertation, Universität Potsdam.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *49*, 235-290.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*(3), 189-256.

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*(3), 1481–1498.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*(11), 465-472. Retrieved from `http://www.cnbc.cmu.edu/ jlm/`

McCurdy, K., Goldwater, S., & Lopez, A. (2020). Inflecting when there's no majority: Limitations of encoder-decoder neural networks as cognitive models for german plurals. *arXiv preprint arXiv:2005.08826*.

Penke, M., & Westermann, G. (2006). Broca's area and inflectional morphology: Evidence from broca's aphasia and computer modeling. *Cortex*, *42*(4), 563–576.

Pinker, S. (1998). Words and rules. *Lingua*, *106*, 219-242.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73-193.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456-463.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291-321.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, *23*(4), 463-490.

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

Prather, P., Zurif, E., Love, T., & Brownell, H. (1997). Speed of lexical activation in nonfluent broca's aphasia and fluent wernicke's aphasia. *Brain and language*, *59*(3), 391–411.

Preminger, O. (2021). *Natural language without semiosis.* (University of Maryland)

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, *38*(33), 7255–7269.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart, & t. P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. volume 2: Psychological and biological models* (p. 216-271). Cambridge, MA: MIT Press.

Tabak, W., Schreuder, R., & Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in dutch. *Linguistic evidence—Empirical, theoretical, and computational perspectives*, 529–555.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–270.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of cognitive neuroscience*, *9*(2), 266–276.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.