

Simulating phonological and semantic impairment of English tense inflection with Linear Discriminative Learning

Maria Heitmeier and R. Harald Baayen
University of Tübingen

January 10, 2020

Abstract

This study applies the computational theory of the ‘discriminative lexicon’ (Baayen et al., 2019) to the modeling of the production of regular and irregular English verbs in aphasic speech. Under impairment, speakers with memory loss have been reported to have greater difficulties with irregular verbs, whereas speakers with phonological impairment are described as having greater problems with regulars. Joanisse and Seidenberg (1999) were able to model this dissociation, but only by selectively adding noise to the semantic units of their model. We report two simulation studies in which topographically coherent regions of phonological and semantic networks were selectively damaged. Our model replicated the main findings, including the high variability in the consequences of brain lesions for speech production. Importantly, our model generated these results without having to lesion the semantic system more than the phonological system. The model’s success turns out to hinge on the use of a corpus-based distributional vector space for representing verbs’ meanings. Joanisse and Seidenberg (1999) used one-hot encoding for their semantic representation, under the assumption that semantically regular and irregular verbs do not differ in ways relevant to impairment in aphasia. However, irregular verbs have denser semantic neighborhoods than do regular verbs (Baayen and Moscoso del Prado Martín, 2005), and we show that in our model this greater density renders irregular verbs more fragile under semantic impairment. These results provide further support for the central idea underlying the discriminative lexicon: that behavioral patterns can, to a considerable extent, be understood as emerging from the distributional properties of a language and basic principles of human learning.

1 Introduction

In English, the past tense of regular verbs is produced by attaching *-ed* to the stem. The way the past tense of irregular verbs is formed differs from this rule to varying degrees. An extensive and unresolved debate has addressed the question of whether the past-tense system of English is best modelled by a classical division of labor between representations

and rules, as hypothesized by the so-called dual mechanism model (Pinker, 1991), or by a single statistical system as first explored by Rumelhart and McClelland (1986).

The dual route model as laid out by Pinker (1991, 1999) comprises a computational component which concatenates *-ed* to the stem, and an associative memory in which irregular verb forms are stored. The two subsystems are supposedly located in different parts of the brain. Following up on this dual-mechanism approach, several other models have subsequently been proposed (see Marusch et al., 2017, for an overview). Typically, these models work with classical constructs such as stems, affixes, and allomorphs. The Stem-Assembly model (Cholin et al., 2010), for instance, assumes that a lexicon stores all possible stems for a given verb together with indicators stating which affixes can be combined with which stems (see also Smolka et al. (2014) for a similar approach for comprehension).

In contrast, the computational model proposed by Rumelhart and McClelland (1986) and subsequent studies (e.g., MacWhinney and Leinbach, 1991, Joanisse and Seidenberg, 1999) sought to show that past tense forms, irrespective of whether they are regular or irregular, can be produced by a single connectionist network that obviates the need for separation of rules and memory.

The dual-mechanism model makes some clear predictions. Since two different systems are hypothesized to be involved in inflecting regular and irregular verbs, it should be possible to identify two different brain areas that subservise these two systems. This prediction has been tested using brain imaging (e.g. Meteyard et al., 2013) and priming paradigms (Marslen-Wilson and Tyler, 1997) and has also been investigated through developmental surveys (Pinker, 1991).

In this line of research, dissociations between regular and irregular verbs in aphasic patients have played an important role. Ullman et al. (1997) claimed that since the dual-route model distinguishes between a rule-like process and storage in memory, agrammatic patients should have problems with inflecting regulars while patients with memory loss should have problems with irregulars. They therefore ran experiments on the one hand with patients with Alzheimer’s disease and Posterior Aphasia, both associated with memory loss, and on the other hand with patients with Anterior Aphasia and Parkinson’s disease, who typically show deficits producing grammatically correct language. The performance of these patients in a simple production task was roughly in line with Ullman et al. (1997)’s predictions. However, the dissociations were not as clear cut as predicted by the theory. While patients typically had more problems with one specific group of verbs, they were also impaired on the other one. Other studies, such as Shapiro and Caramazza (2003), reported similar results.

However, Joanisse and Seidenberg (1999) noted that the patients with memory problems actually had semantic deficits and that the agrammatic patients suffered from phonological difficulties. They therefore proposed a connectionist model with banks of units for semantics, auditory input, and speech output, as well as multiple hidden layers. Depending on whether they damaged the production layer or the semantic layer, the model selectively had more problems with regular verbs or with irregular verbs. Importantly, difficulties were not restricted to either regulars or irregulars, but rather there was a tendency for one type of past tense formation to be more impaired than the other. Joanisse and Seidenberg (1999) argued that damage to semantics disables the ability to identify specific words. As a consequence, patients with memory loss cannot find idiosyncratic

past tense forms. Damage to the phonology, on the other hand, was claimed to interfere with the ability to produce forms by analogy to other forms.

Multiple follow-up studies were conducted. Bird et al. (2003) reported a group of ten patients with phonological deficits who performed worse on regulars than irregulars. Patterson et al. (2001) tested a group of eleven patients with semantic dementia. As predicted, these patients performed worse on irregular as compared to regular verbs. However, not all results reported are this consistent. Miozzo (2003) reported the case of a patient with phonological deficits who none the less had more problems with inflecting irregular verbs. And three out of four patients with semantic dementia studied by Tyler et al. (2004) did not have significantly more problems with irregular than with regular verbs.

Unfortunately, these studies have various methodological shortcomings. Inherent to studies of neuropsychological disorders is the small number of subjects included. Statistically, these studies are vastly underpowered (see Westfall et al., 2014). Some studies failed to control for crucial variables such as frequency (an overview is given in Faroqi-Shah, 2007). Bird et al. (2003) and Nickels and Howard (2004) reported that selective difficulties with regular verbs disappeared once words were controlled for their phonological complexity: inflected forms of regular verbs tend to be longer and more phonologically complex than those of irregular ones (see also Burzio, 2002). As pointed out by Faroqi-Shah (2007), most studies conducted subsequently nevertheless failed to control for phonological complexity. A further complication is that tasks varied across studies (e.g. free production, multiple choice, . . .), and that dissociation patterns may actually vary within patients depending on the task (Faroqi-Shah, 2007). Their meta-analysis came to the conclusion that there is no compelling evidence for reliable dissociations.

In the light of this uncertainty about the empirical facts, computational modeling studies are of special interest, as they make it possible to rigorously work out the consequences of lesions for lexical processing given specific constellations of explicit assumptions about the architecture of the mental lexicon. In what follows, we first review classical computational models for aphasic speech production. We then introduce the computational framework that we use in the present study to address the vexed question of double dissociations and their relation to types of impairment. Finally, we present two simulation studies addressing the effect on the production of past tense verbs of selectively lesioning the form system or the semantic system.

2 Previous computational modeling studies

Several computational modeling studies have specifically addressed the issue of whether the different effects for regular and irregular verbs reported across types of impairment can be predicted from artificial neural networks. Juola (2000) constructed a network that produced inflected forms from stems. He then randomly lesioned connections, and observed that among a large number of lesioned networks, a small number of networks were selectively impaired for regular verbs, and that likewise a small number of other networks was selectively impaired for irregulars. He argued that selective impairment is, in all likelihood, quite rare among aphasic patients, and that there is a bias in the literature for reporting the rare cases in which either regulars or irregulars are (relatively) spared.

Such cases can, apparently, arise also in randomly lesioned artificial neural networks, albeit infrequently.

Westermann and Ruh (2009) proposed an artificial neural network that self-organizes in such a way that over time, part of the network becomes specialized for regular verbs, and another part takes care of irregular verbs. In such a network, it is possible to impose localized lesioning such that either regular or irregular verbs are predominantly impaired.

Because the models of Juola (2000) and Westermann and Ruh (2009) do not incorporate semantics, they cannot account for the — contested — double dissociation of type of aphasia and morphological regularity. The parallel distributed processing (PDP) model proposed by Joanisse and Seidenberg (1999) differs in this respect. These authors developed a network for speech production that can be driven either from semantic input or from auditory input. The model contains banks of so-called cleanup units for both the semantic layer and for the production layer. For the semantic layer, localist representations were used, with one specific unit on for a given verb and all other units off, and with in addition one unit that was on or off depending on tense. For the modeling of semantic impairment, the connections between the semantic units and the corresponding cleanup units were lesioned. In addition, they investigated the consequences of adding Gaussian noise to the semantic units. For the modeling of phonological impairment, they severed the connections from the production units and the corresponding cleanup units. With this setup, the authors were able to model the selective impairment of irregulars for patients with memory impairment, and the selective impairment of regulars for agrammatic patients suffering from phonological difficulties.

Although this model makes the correct predictions — granted the assumption that double dissociations of type of aphasia and regularity are indeed empirically well-attested — aspects of its architecture are less convincing. First, Joanisse and Seidenberg (1999) claimed that a localist representation of verb meanings is justified because semantic similarity, even though it is “crucial for other phenomena . . . , it is not important for the past tense.” (p. 7593). However, it has since then been established that semantics can co-determine whether a verb is inflected regularly or irregularly (Ramscar, 2002), and that distributionally there are consistent semantic differences between regular and irregular verbs (Baayen and Moscoso del Prado Martín, 2005) in several Germanic languages. Second, severing of only connections is biologically implausible. It is much more likely that clusters of neurons, together with their synapses, are destroyed, rather than that an individual synapse is taken out of commission while at the same time the neuron and its other synapses keep functioning as before the lesion. Furthermore, connections and neurons that are lesioned should be close together in the network, as brain lesions typically involve blood flow in coherent areas of the cortex. Third, it is not clear why Gaussian noise is added to the semantic layer, but not to the production layer. Finally, the slot coding used for the production layer is less than optimal, as it is unclear how prefixed verbs such as *undo* and their simple counterparts such as *do* can be modeled without a certain amount of hand-engineering that is foreign to the spirit of the parallel distributed processing enterprise (see, e.g., Pinker and Ullman, 2002). Furthermore, having multiple slots where the same phone can appear possibly is at odds with evidence suggesting that cortical areas may specialize for specific phone identities (Blakely et al., 2008).

The model introduced in the next section sidesteps these problems, while making use of an architecture that is much simpler mathematically. The specific goals to be pursued

in the remainder of this study are, first, to clarify whether the dissociations reported in the literature are straightforwardly predicted by our model, without requiring selective additional mechanisms such as the addition of Gaussian noise only to semantic units as in the model of Joanisse and Seidenberg (1999), second, to clarify whether the greater semantic density that characterizes irregular verbs helps explain the vulnerability of irregular verbs to semantic impairment, and third, to provide, by means of our simulations, improved insight into the prevalence of the specific dissociations reported in the literature (cf. Juola, 2000). The next section provides an introduction to the theory underlying our simulation studies.

3 The discriminative lexicon

The framework of linear discriminative learning (LDL) provides a computational theory of the mental lexicon in which processes for visual and auditory comprehension, as well as processes for speech production, are brought together and integrated (see Baayen et al., 2019, for detailed discussion). Linear discriminative learning, and its conceptual twin Naive Discriminative Learning (Baayen et al., 2011), have been found useful for explaining a wide range of phenomena in the experimental literature on morphological processing, including masked priming (Milin et al., 2017), overt priming (Baayen and Smolka, 2019), auditory comprehension (Arnold et al., 2017, Shafaei-Bajestan and Baayen, 2018) and speech production (Tomaschek et al., 2019). Given that the LDL framework has broad empirical coverage, it is a promising tool for studying the consequences of lesions for the processing of regular and irregular verbs.

For the modeling of comprehension and production, we need (minimally) two tables, one defining words’ forms, and one defining words’ meanings. These tables can be represented by matrices. The $n \times k$ matrix for words’ forms, \mathbf{C} , specifies, for each of n words, which of k letter or phone trigrams is realized in that word’s form. Presence is indicated by 1, absence by 0.

The $n \times m$ matrix \mathbf{S} specifies, for each of the n words, a semantic vector. In computational linguistics, such vectors are known as word embeddings (see, e.g., Mikolov et al., 2013). The m values (typically small real numbers) in a word’s semantic vector as used in LDL represent the collocational strengths of that word with each of the top m best represented lexemes. Lexemes are the elementary semantic units of the model. They comprise both onomasiological units such as MOOSE and WRITE, and inflectional and derivational functions such as PLURAL (as in *mooses*) and AGENT (as in *writer*). Baayen et al. (2019) provide details on how LDL estimates these collocational strengths. In what follows, we make use of the semantic vectors developed in their study, which are derived from the TASA corpus (Ivens and Koslin, 1991).

Above, we mentioned that for English, irregular and regular verbs have been found to differ not only with respect to their past tense forms, but also with respect to their semantics. Baayen and Moscoso del Prado Martín (2005) observed that irregular verbs have denser semantic neighborhoods compared to regular verbs. One of the measures that they considered is the similarity between the semantic vectors of regular and irregular verbs. The semantic vectors that we use in the present study provide further support for irregulars having denser semantic similarity neighborhoods than is the case for regu-

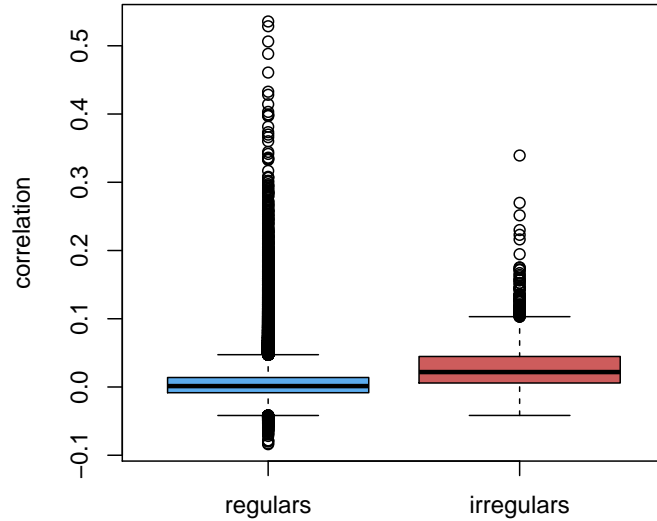


Figure 1: Box-and-whisker plots for the correlations between the semantic vectors of regulars (left, $n = 304590$ comparisons) and the correlations between the semantic vectors of irregulars (right, $n = 2850$ comparisons).

lars. This greater density emerges clearly when the similarities between semantic vectors (evaluated using the Pearson correlation) are compared within the set of regular verbs, and separately within the set of irregular verbs. Using the dataset described in Baayen et al. (2019), the distribution of similarities between 781 regular verbs (left boxplot in Figure 1, median 0.0013) is shifted downwards ($p < 0.0001$, Wilcoxon test) compared to the distribution of similarities between 78 irregular verbs (right boxplot, median 0.0218). Because linear discriminative learning works with corpus-based semantic vectors, rather than with localist semantic units, the different semantic densities of regular and irregular verbs can play out naturally in the simulations reported below.

Of further relevance to the present study is how the semantic vectors for inflected words are constructed. Let \overrightarrow{walk} denote the semantic vector for the regular verb ‘to walk’, \overrightarrow{run} the semantic vector for the irregular verb ‘to run’, \overrightarrow{past} the semantic vector for the past tense, and $\overrightarrow{3sg}$ the semantic vector for the third person singular. The semantic vectors for inflected forms are obtained by summing the semantic vectors of the pertinent lexemes:

$$\begin{aligned}
 walk: & \quad \overrightarrow{walk} \\
 walks: & \quad \overrightarrow{walk} + \overrightarrow{3sg} \\
 walked: & \quad \overrightarrow{walk} + \overrightarrow{past} \\
 run: & \quad \overrightarrow{run} \\
 runs: & \quad \overrightarrow{run} + \overrightarrow{3sg} \\
 ran: & \quad \overrightarrow{run} + \overrightarrow{past}
 \end{aligned} \tag{1}$$

Note that tense is treated as a privative opposition (Lyons, 1968): the present tense is semantically unmarked, whereas the past tense is semantically marked. It is only for this tense that an inflectional semantic vector, the vector specific to the past tense, is added to the semantic vector of the verb.

Given the matrices \mathbf{C} and \mathbf{S} specifying words' forms and meanings, the simplest set-up for comprehension is to posit a direct mapping \mathbf{F} that transforms words' form vectors (the row vectors of \mathbf{C}) into words' semantic vectors (the row vectors of \mathbf{S}):

$$\mathbf{CF} = \mathbf{S}$$

(For detailed discussion of indirect mappings passing through more abstract phonological representations, see Baayen et al., 2019). The mapping \mathbf{F} , which is itself a matrix of dimension k by m , can be obtained using standard linear algebra as used in multiple regression. For production, we are interested in the $m \times k$ mapping \mathbf{G} that transforms a word's semantic vector into its form vector:

$$\mathbf{SG} = \mathbf{C}.$$

Matrix \mathbf{G} is estimated in the same way as matrix \mathbf{F} . Since the mappings (or linear transformation) \mathbf{F} and \mathbf{G} are estimated from the data, the predicted form and meaning vectors will be close to, but not identical to, the row vectors of \mathbf{S} and \mathbf{C} . In what follows, we denote the matrices with predicted meaning and form vectors, following standard notation in statistics, by $\hat{\mathbf{S}}$ for comprehension and $\hat{\mathbf{C}}$ for production. The following paragraph provides a technical description of this process and can be skipped by readers not interested in the underlying mathematics.

The mathematical engine that formalizes and quantifies the computational processes in the discriminative lexicon is that of multivariate multiple regression. In standard multiple regression, a response vector \mathbf{y} of length n , with n the number of observations, is predicted from p predictors (including an intercept term), which are brought together as the column vectors of an $n \times p$ matrix \mathbf{X} . The predictors are weighted by a vector of coefficients $\boldsymbol{\beta}$ of length p :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta},$$

see Baayen et al. (2018) for a non-technical introduction to this notation and its geometric interpretation. Multivariate multiple regression takes a matrix of k response variables, brought together as the column vectors of an $n \times k$ matrix \mathbf{Y} , and also models it as a function of the predictors \mathbf{X} . For each response vector in \mathbf{Y} , however, there now is a corresponding column vector of coefficients in an $n \times k$ matrix \mathbf{B} :

$$\mathbf{Y} = \mathbf{XB}.$$

In the discriminative lexicon, the vector $\hat{\mathbf{c}}_i$ predicted for the form of word i , a row vector of $\hat{\mathbf{C}}$, quantifies the amount of support the different triphones receive from this word's semantic vector. Typically, the triphones that actually occur in the word will receive substantial support, whereas those that do not occur in the word will have support close to zero. However, the order of the triphones in the predicted form vector is arbitrary, and hence, the predicted triphone vector provides incomplete information about a word's pronunciation. Fortunately, the triphones contain partial information about their order.

For instance, the triphones `abc` and `bcd` can form the string `abcd`, whereas the triphones `abc` and `xyz` do not overlap and hence cannot be joined to form a legal longer phone string. This partial information in the triphones can be exploited to calculate all legal sequences of triphones. In LDL, this is accomplished by means of algorithms from graph theory. Triphones are defined to be the vertices of a graph, and a directed edge between two vertices is posited whenever two triphones can be joined to form a longer phone string. Thus, there is an edge from `abc` to `bcd`, but there is no edge between `abc` and `xyz`. Legal pronunciations are now represented by paths in the triphone graph. Such paths start with a word-initial triphone (e.g., `#ab`) and end with a word-final triphone (e.g., `yz#`). Technically, paths are reconstructed by means of the `all_simple_paths` function from the `igraph` package (Csardi and Nepusz, 2006). In order to reduce computational load, prior to searching for all simple paths, the graph is thinned by removing all vertices and corresponding edges that are not well supported by the semantics. In other words, triphones with a value below a threshold θ in a predicted form vector \hat{c} are not taken into account when constructing the graph. As shown by Baayen et al. (2018) and Chuang et al. (2019), this algorithm succeeds in predicting words’ forms with high accuracy.

Of specific importance for the present study is that the triphone graph can be projected onto a cartesian plane. This is usually done for purposes of visualization. When a particular visualization algorithm, `graphopt`, is used, triphones can sometimes show some clustering by inflectional function (Baayen et al., 2018). The `graphopt` algorithm, implemented in the `igraph` package (Csardi and Nepusz, 2006), is designed to create optimal layouts for large graphs by using basic principles of self-organization in physics. A node is given both mass and an electric charge, while edges are modelled as springs. In this way, the graph is modelled as a system in which there are attracting and repelling forces between vertices. The dynamics of this system are simulated until an equilibrium is reached. Through this process, a natural layout of all nodes is created, which serves as a proxy of the layout of neurons in the brain. For related work addressing the topological organization of phone sequences in a two-dimensional space, but using temporal self-organizing maps, see Ferro et al. (2011), Chersi et al. (2014). For localization of phones in the cortex, see Blakely et al. (2008).

4 Lesioning a discriminative lexicon

In this section, we specify how we combine the mathematically simple yet computational powerful engine of multivariate multiple regression with topological self-organization to study the consequences of locally cohesive lesioning of form and meaning.

4.1 Topological maps for form and meaning

In order to create a two-dimensional topological map, the first step is to build an adjacency matrix. An adjacency matrix is a square matrix with vertices on its rows and columns. The values in the matrix are zero, except for those cells for which the row vertex has an edge to the column vertex. Here, the cell value is set to 1. An adjacency matrix fully defines a graph. Given the graph, the `graphopt` self-organization algorithm generates the coordinates of the vertices in the cartesian plane.

The adjacency matrix for the form graph is straightforwardly obtained from the word forms, by extracting, for each word, its successive pairs of triphones and entering a 1 in the pertinent cell of the adjacency matrix. This adjacency matrix defines the triphone graph \mathcal{C} . Since in LDL the input for articulation is a path in this triphone graph, and since the `graphopt` algorithm lays out the graph onto a 2-D plane, a word form is represented by a path in this phonological topological map. As a consequence, lesioning an area of this map will affect only those words for which the triphone path passes, at least in part, through this area.

The adjacency matrix for the semantic graph is obtained from the n (words) \times m (lexomes) semantic matrix \mathbf{S} . Above, following Baayen et al. (2016, 2019), we used the row vectors of \mathbf{S} as semantic vectors, and showed that these vectors support the difference in semantic density between regulars and irregulars previously reported by Baayen and Moscoso del Prado Martín (2005). However, the column vectors of \mathbf{S} also capture semantic similarity, albeit to a somewhat reduced extent. The difference in semantic density between regulars and irregulars is also reflected in the column vectors (median correlation regulars: 0.0012, median irregulars: 0.0184, $p < 0.0001$, Wilcoxon test). In the discriminative lexicon, the lexomes are the most basic points in semantic space, they are the points (or, equivalently, vectors) from which the points (semantic vectors) of inflected words are created dynamically. Importantly, they also figure in the mapping \mathbf{G} , which is mathematically equivalent to a two-layer network with m input units for the lexomes, and k output units for the triphones. Therefore, the column vectors of \mathbf{S} are used for constructing the semantic topological map.

As a first step, we calculated the correlation matrix for the column vectors of the semantic matrix \mathbf{S} . We then converted this correlation matrix into an adjacency matrix by setting to 1 all cells with an absolute value exceeding a threshold correlation, and setting all other cells to zero. The threshold correlation was set such that the graph density (the ratio of the number of edges to the number of possible edges) of the semantic graph \mathcal{S} matches that of the form graph \mathcal{C} . Finally, we again used the `graphopt` algorithm to lay out this graph in a 2-D plane. In this way, the n word meanings, originally represented as n points in an m -dimensional space, are projected onto a 2-D semantic topological map, such that semantic similarity in the original space is largely preserved.

Interestingly, because of the greater semantic density of irregular verbs, these verbs are somewhat more likely to have edges to other irregular verbs in the semantic graph \mathcal{S} . As a consequence, in its two-dimensional graph layout, the semantic topological map, they will be more likely to be close together. It follows that under lesioning of the map, irregular verbs will be somewhat more likely to be affected jointly.

The specifics of the layout produced by `graphopt` depend on a random seed. By choosing different seeds, we can create different graph layouts, which can be viewed as models of the same lexicon but implemented for different brains. Figure 2 presents examples of maps generated for triphones and lexomes.

4.2 Lesioning topological maps

In order to damage a locally cohesive area in a phonological or semantic map, we defined rectangular areas of a specified size (percentage of the total area of the map) with varying height and width. Vertices and edges in these areas are then deleted from the map. Note

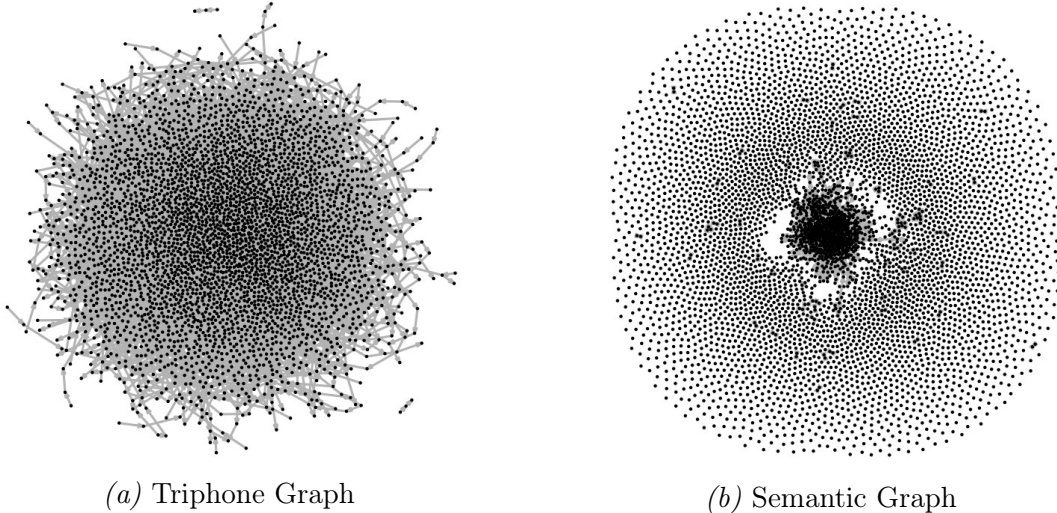


Figure 2: Plots of a triphone and a semantic graph for English verbs. Labels are omitted for clarity. The proportion of edges to the overall number of possible edges is the same in both graphs.

that depending on where exactly the model is damaged, the extent of the damage can vary significantly. Since at the fringes of the graph nodes tend to be much less dense (see Figure 2), if the damaged area happens to be there, the impairment will not be too strong. If on the other hand it is somewhere closer to the centre, the impairment can be quite strong even if the area is comparably small.

Mathematically, lesioning the j -th triphone in the phonological map amounts to setting the j -th's column in \mathbf{C} to zero. This renders the triphone inert so that it can no longer propagate activation to the semantics. The weights on the efferent connections from this triphone, which are given by the j -th row of the network (transformation matrix) \mathbf{F} can be set to zero as well, to reflect that synapses become inert when the neuron is damaged. Mathematically, however, this is not strictly necessary once the triphone has been disabled in the \mathbf{C} matrix.

Similarly, for the semantic topological map, the lesioning of a lexome is equivalent to setting all values in the corresponding column in \mathbf{S} to zero. To explicitly take out of commission the lexome's efferent connections, the corresponding row in the network (transformation matrix) \mathbf{G} can be set to zero as well. Figure 3a illustrates the lesioning of the \mathbf{G} network. The input units, the lexomes, are shown on the right in orange. The output nodes, the triphones, are shown on the left in blue. The orange nodes are organized topologically in the semantic map. In this map, there are undirected edges between the lexomes S1 and S2, and between S2 and S3. The blue nodes are organized topologically in the phonological map, with directed edges from #ab to abc and from abc to bc#. The directed edges (in black) from the orange nodes to the blue nodes represent the connections in the \mathbf{G} network. Each of these connections is associated with a regression weight as specified in the corresponding matrix. The left panel of Figure 3 illustrates semantic impairment, and the right panel phonological impairment. In this figure, a single unit and its connections are removed. For the lesions that we implemented, multiple units located in a rectangular area of the topological map are taken out, together

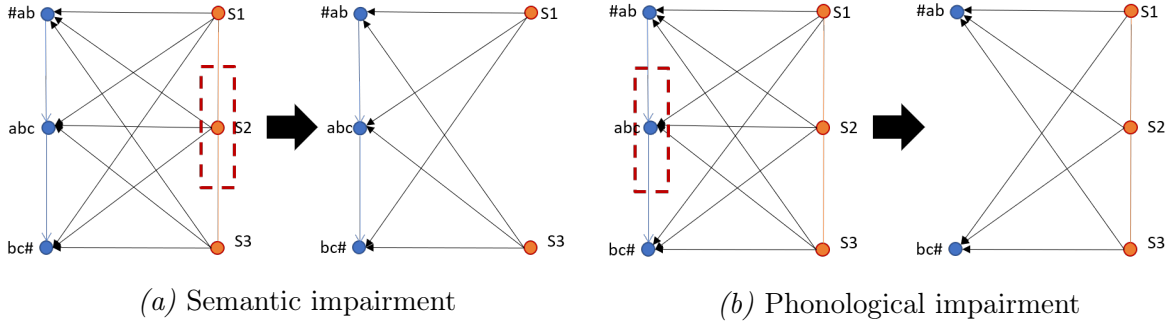


Figure 3: Illustration of a trigram graph (blue), a semantic graph (orange) and the connections between them (black). When a triphone (right panel) or a lexome (left panel) is deleted, all the pertinent connections are deleted. In the simulations, rectangular areas covering multiple triphones or lexomes are used to implement lesions in the network.

with their connections.

5 Simulations

In what follows, we report two simulation experiments. The first experiment investigates the consequences of lesioning either the phonological topological map or the semantic topological map. This simulation experiment follows up on the findings reported for semantic dementia by Patterson et al. (2001) and those reported by Bird et al. (2003) for patients with phonological deficits, as well as on the simulation study of Joanisse and Seidenberg (1999). A central question is whether, when (1) cells and all their connections are taken out of commission in localized areas, (2) the semantic representation for regular and irregular verbs does justice to the difference in semantic density between regulars and irregulars, and (3) the model is trained with more comprehensive coverage of English verbs, the double dissociations modeled by Joanisse and Seidenberg (1999) can still be observed.

The second simulation experiment, following Nickels and Howard (2004) and Bird et al. (2003), studies the consequences of matching stimuli for phonological complexity. If the model replicates earlier findings, the selective impairment of regular verbs in phonologically impaired patients should vanish.

Across these simulation experiments, the variability in the consequences of lesioning the discriminative lexicon, as observed across many simulation runs, is of special interest, given that the findings reported in the literature are often conflicting and inconclusive (Faroqi-Shah, 2007). In the light of the simulation results of Juola (2000), who observed occasional double dissociations in a large series of simulation runs with the same PDP network, it is important to ascertain whether theoretical claims are supported by a majority of our simulation runs, or arise only occasionally.

For setting up the model, we used the English data set discussed in Baayen et al. (2019), from which we extracted 4906 verb forms of in total 2084 different verbs. This data set was derived from the TASA corpus (Ivens and Koslin, 1991, Landauer et al., 1998), which was tagged for part of speech and inflectional function, and then used to

	regular	irregular
number of verbs	22	22
mean frequency of past tense form	218.09 (281.24)	233.56 (221.48)
mean length of past	4.5 (0.86)	3.68 (0.72)
percentage of stem changes	0%	82%

Table 1: Properties of the stimuli for Experiment 1. Standard deviations in parentheses.

derive semantic vectors for content and inflectional lexemes (see Baayen et al., 2019, for detailed discussion and validation). In this way, we obtained a semantic matrix \mathbf{S} of size 4906 verb forms \times 5487 lexemes. The semantic vectors of inflected words were obtained by summing the semantic vectors of their lexemes as explained above in (1). For inflectional homophones such as *walked*, which can be either a past tense or a past participle, only one form was retained.

Words’ triphones were encoded using the DISC keyboard phonetic alphabet, which represents phones with single characters, as given by the transcriptions in the CELEX database (Baayen et al., 1995). Words’ triphones were subsequently used to construct the triphone matrix \mathbf{C} .

After setting up the \mathbf{C} and \mathbf{S} matrices, the transformation matrices (networks) \mathbf{F} and \mathbf{G} were derived, thereby completing the model definition. Overall production and comprehension accuracy of the model was 92.7%.

To keep the carbon footprint of the simulations down, the model was set up for verbs only; for a more comprehensive model also including nouns and adjectives, see Baayen et al. (2019). Because the simulations discussed below are conditional on a lexicon with only verbs, we checked whether for the semantic space defined by \mathbf{S} it still holds that irregular verbs have higher semantic density compared to regular verbs. As expected, this was indeed the case (median regulars: 0.0004; median irregulars: 0.0502; $p < 0.0001$, Wilcoxon test).

5.1 Experiment 1: Stimuli controlled for frequency

Materials and Procedure Experiment 1 addresses the effects on production accuracy of lesioning either triphones or lexemes. Lesioning of triphones is expected to give rise to more errors for regulars, and lesioning of lexemes to more errors for irregulars, according to the theory of Joanisse and Seidenberg (1999). To test these predictions for our model, we took the design of Bird et al. (2003) as point of departure. Twenty-two regular and irregular verbs respectively were chosen such that they were matched in the mean for past-tense frequency. Of these verbs, 13 are identical to those used by Bird et al. (2003). The full list of verbs can be found in Appendix A of the present study. Table 1 provides an overview of mean length and mean frequency broken down by regularity.

We selected six seeds (315 to 320) to be used with the `graphopt` algorithm, for constructing different topological maps for triphones and lexemes. In this way, we simulated six different brains. For each of these “brains”, we defined rectangles of varying sizes, from 0% to 100%. Triphones or lexemes located within these rectangles are targets for deletion. For each of 21 rectangle sizes, we implemented 10 replications, thus allowing the location of the rectangle to vary. Length and width of the rectangles were chosen

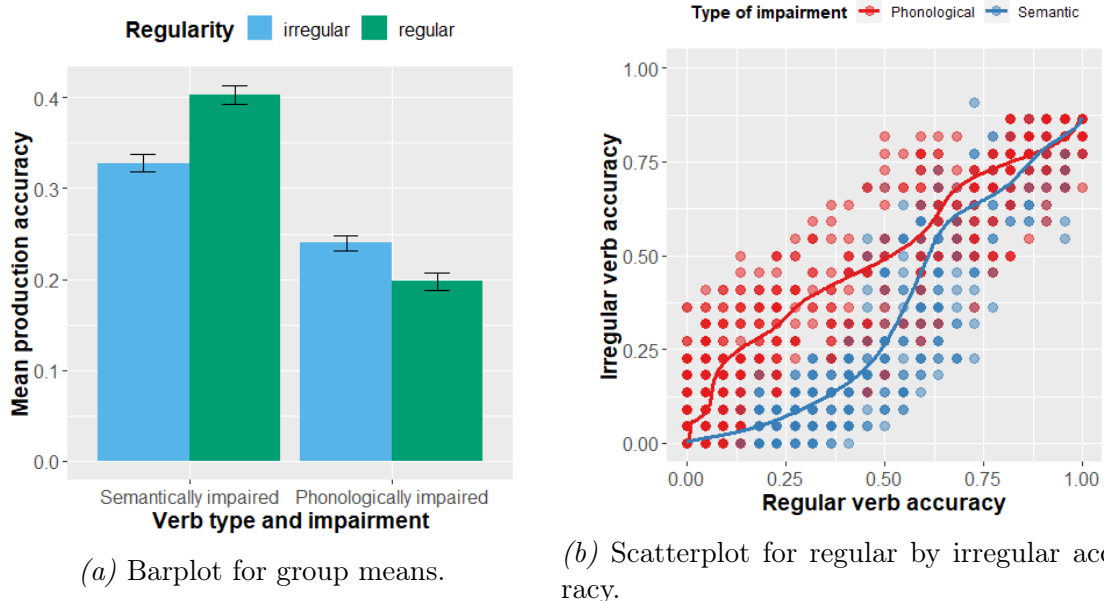


Figure 4: Experiment 1: Interaction of Regularity by Type of Impairment. Left panel: barplot for group means, scores are calibrated with respect to the baseline performance of the undamaged model. Right panel: scatterplot for the simulation runs. The less transparent points are, the more points overlap. Overlapping red and blue points are shown in purple. The curves connect accuracies as predicted from proportion of deleted nodes using GAMs for each of the four subsets defined by Regularity \times Impairment.

randomly subject to the constraint on their size. In this way, we obtained a total of 6 brains \times 21 damage sizes \times 10 replications = 1260 simulation runs for semantic lesioning, and another 1260 runs for phonological lesioning. For each run, the selected nodes were deleted from the \mathbf{S} or \mathbf{C} matrices. For each simulation run, the lesioned networks were used to produce the past tense form of every verb in the stimulus set, and the form produced and production accuracy were recorded. Simulations in which the size of the damage was zero were used as a baseline for unimpaired performance. Furthermore, simulation runs for which accuracy on both regular and irregular verbs was zero were excluded, as patients unable to get a single verb correct are unlikely to have been included in neuropsychological studies.

	Estimate	Std. Error	z value	$\Pr(> z)$
Intercept (Reg=irreg, Imp=Phonol)	1.9402	0.0288	67.3709	< 0.0001
Reg=reg	-0.2302	0.0325	-7.0768	< 0.0001
Imp=Semantic	0.7257	0.0307	23.6154	< 0.0001
Prop. deleted nodes	-6.2721	0.0472	-133.0166	< 0.0001
Reg=reg by Imp=Semantic	1.2361	0.0429	28.7854	< 0.0001

Table 2: Generalized Linear Model fitted to production accuracy in Experiment 1 (Reg: Regularity, Imp: Impairment, treatment dummy coding).

Results A generalized linear model was fitted to the accuracy data with as predictors the factors Regularity (regular, irregular), Impairment (phonological, semantic), and the proportion of deleted nodes as control covariate. The model coefficients and associated statistics are given in Table 2. As expected, accuracy decreased substantially as the proportion of deleted nodes increased. Regularity and Impairment entered into an interaction, visualized in the left panel of Figure 4. When the semantic map was damaged, irregular verbs suffered more than regular verbs. Conversely, when the phonological map was lesioned, accuracy showed a severe drop, and now irregular verbs emerged with somewhat higher scores compared to irregulars. In general, semantically impaired models perform better than phonologically impaired ones.

The scatterplot shown in the right-hand panel of Figure 4 presents an overview of results for the simulation runs. The horizontal axis represents accuracy for regular verbs, and the vertical axis accuracy for irregular verbs. Data points are arranged on a grid. The transparency of the color of a datapoint codes how many simulation runs fall in a small square around that datapoint. Simulation runs for which the phonological map was damaged are shown in red, the blue dots represent simulation runs with lesions to the semantic map. Purple dots represent runs for which both types of lesioning are present in the counts. The red and blue curves represent the mean accuracy for regulars (x-coordinate) and irregulars (y-coordinate) as predicted by logistic generalized additive models fitted to the accuracy data for the simulation runs with phonological (red) and semantic (blue) lesions.

In conformance with the factorial results, the simulation runs with semantic impairment (blue dots) show higher accuracy for regulars than for irregulars. The reverse holds for simulation runs with phonological impairment. This pattern is most pronounced for lower accuracies. Of interest is the wide scatter, with occasional simulation runs with semantic lesioning siding with the runs with phonological lesioning, and vice versa, among accuracies greater than 0.5, exactly opposite to the general pattern. Here, our results mirror those of Juola (2000). Given that the findings reported in the literature are highly variable (as can be seen by comparing Bird et al. (2003) and Patterson et al. (2001) with Miozzo (2003), Tyler et al. (2004), Lambon Ralph et al. (2005), see Table 3 for further details), it is interesting that our simulations indicate that the pattern of results reported by Patterson et al. (2001) (eight patients with semantic dementia) and Bird et al. (2003) (ten patients with phonological deficits), which was modeled by Joanisse and Seidenberg (1999), is the most common pattern, and that diverging patterns are certainly possible, but less frequent.

Error types were analysed for all simulation runs where the topological magnitude of damage was 10% of the total area (overall accuracy: mean 0.75, standard deviation 0.21). Table 4 presents statistics for the kind of errors made, broken down into the classes of Ullman et al. (1997, 2005) supplemented with additional classes required for the additional error types produced by our model, specifically under phonological impairment. From this overview, it is clear that our semantically and phonologically impaired models produce almost all of the errors reported by Ullman et al. (2005), as well as those given by Miozzo (2003), Bird et al. (2003), Patterson et al. (2001), Lambon Ralph et al. (2005).

The highest number of error types beside “no response” was “unmarked”. In both phonologically and semantically impaired models this number is much higher for irregulars than for regulars. The reason for this is that the undamaged model does not perform

Study	#Subjects	Frequency-controlled?	Impairment	Theory	Regular accuracy	Irregular accuracy
Patterson et al. (2001) (excluding ultra-high-frequency)	8	yes	semantic	single system	97%	62%
Bird et al. (2003) (screening)	10	yes	phonological (non-fluent aphasics)	single system	20%	37%
Lambon Ralph et al. (2005)	8	no	phonological	single system	63%	59.9%
Miozzo (2003)	1	yes	phonological	dual-route	98.5%	71.5%
Tyler et al. (2004) (Ullmann-test)	4	no	semantic	dual-route	90%	85%
Ullman et al. (1997)	10	no	semantic	dual-route	74%	84.5 %
Ullman et al. (1997)	12	no	phonological	dual-route	62.25%	76.67%
Ullman et al. (2005)	2	no	phonological (non-fluent aphasics)	dual-route	20%	47%
Ullman et al. (2005)	6	no	semantic (fluent aphasics)	dual-route	85.83%	71.33%

Table 3: Overview of various studies testing aphasic patients with past tense generation tasks. In order to make results comparable, only tests are shown which are similar to the original past tense production task by Ullman et al. (1997). Only results of tasks in which data sets were not controlled for phonological complexity are shown.

	Semantically impaired		Phonologically impaired	
	irregular	regular	irregular	regular
Distortion	1(<.01)	0(.00)	8(.01)	1(<.01)
en-suffixed	3(.01)	0(.00)	15(.02)	0(.00)
ing-suffixed	0(.00)	2(<.01)	-	-
Multiple suffixed	0(.00)	1(<.01)	-	-
No response	0(.00)	2(<.01)	118(.13)	326(.37)
Other errors	1(<.01)	0(.00)	-	-
Overregularisation	10(.02)	0(.00)	-	-
s-suffixed	12(.03)	14(.03)	6(.01)	19(.02)
Unmarked	241(.57)	77(.18)	239(.27)	51(.06)
Word intrusion	16(.04)	33(.08)	-	-
Word substitution	2(<.01)	5(.01)	9(.01)	13(.01)
en-suffixed onset	-	-	8(.01)	0(.00)
Onset	-	-	8(.01)	14(.02)
Onset intrusion	-	-	16(.02)	9(.01)
Onset substitution	-	-	5(.01)	12(.01)
Past perfect	-	-	3(<.01)	0(.00)
Past perfect onset	-	-	2(<.01)	0(.00)
Unmarked onset	-	-	0(.00)	2(<.01)

Table 4: Classification of errors in 10 semantically and 10 phonologically impaired models (at 10% damage size, accuracy 75% on average), with counts and proportions relative to the overall number of errors.

perfectly on the irregular test set and as a consequence these errors naturally reoccur in damaged models.

With respect to the high number of “no response” trials in our simulations, we note the following. First, in actual studies, patients producing large numbers of no responses are likely to have been excluded in pre-tests, as it does not make sense to test patients on a task they can no longer perform. Second, real patients are able to try again, and in order to say something, simply produce something that happens to come to mind as an alternative. Third, the production algorithm thins the triphone graph before calculating words’ paths in the graph, which considerably speeds up computation time. By lowering this threshold (which was left at its default value), the number of no responses is expected to become somewhat lower.

Turning to the other error types, “Onset” errors happen when a word is produced wrongly because the first (couple of) phonemes are wrong, usually they are omitted. These errors can be further divided into simple onset errors (*dance* → *nced*), en-suffixed onset (*speak* → *ken*), “onset intrusion” when the onset error leads to the production of a word that actually exists (*play* → *laid*), accordingly an “onset substitution” when the onset of a produced form is wrong and the word not inflected, but the produced word still exists (*agree* → *read*) and “Overirregularised past perfect” (*shrink* → *runk*). Onset errors most likely arise when a critical onset triphone has been deleted from the \mathbf{C} matrix. For example, for the word *played* (**p11d**) the triphone **#p1** might have been deleted, but still leaving the model with good support for **11d**. When the triphone **#11** is spared, the model can still predict *laid* (**11d**) as best approximation. These kind of errors are not described in the literature on aphasia, but it seems likely that they do actually occur. For instance, in the Buckeye corpus (Pitt et al., 2005) of conversational Ohio English, there are realisations of *topic*, *careless*, *telling*, *time*, *can*, *computer*, and *turned* without the initial consonant. The past-tense form *agreed* is realized once without the initial vowel. Such reduced forms are abundant in spontaneous speech (Johnson, 2004), and classifying such forms as errors actually makes sense only when the written canonical form is taken as the yardstick.

5.2 Experiment 2: Controlling for phonological complexity

Materials and Procedure According to Nickels and Howard (2004), the number of phonemes is a significant predictor for production accuracy in aphasics. Moreover, Bird et al. (2003) reported that when controlling the test set for number of phonemes, the selective impairment of regular verbs in phonologically impaired patients disappeared. The stimuli of Experiment 2 were therefore constructed such that 25 irregulars and 25 regulars were matched pairwise with respect to their CV structure (10 CVC, 5 CVCC, 6 CCVCC, 2 CVCVC, 2 CCCV), following Bird et al. (2003). The full test set can be found in Appendix A. As can be seen in Table 5, matching for length was possible at the cost of no longer matching for frequency of occurrence. The procedure for the simulations for this second data set was identical to that used for Experiment 1.

Results Table 6 presents estimates and associated statistics for a logistic regression model fitted to the accuracy data. As for Experiment 1, a greater proportion of deleted nodes predicted a substantial decrease in accuracy, unsurprisingly. The main effect of regularity was not significant, and the contrasts for semantic impairment for both the

	regular	irregular
number of verbs	25	25
mean frequency of past tense form	295.04 (520.35)	6300.56 (19452.49)
mean length of past	4.16 (1.21)	4.16 (1.21)
percentage of stem changes	0%	92%

Table 5: Properties of the stimuli for Experiment 2. Standard deviations in parentheses.

reference level of regularity (irregular) and for regular verbs were very similar (0.5675 and 0.4862). Further subanalyses confirmed that regularity is only a significant predictor for semantically impaired models ($p < 0.001$), but not for phonologically impaired ones ($p = 0.19$). The left panel of Figure 5b summarizes this result graphically.

	Estimate	Std. Error	z value	Pr(> z)
Intercept (Reg=irreg, Imp=Phonol)	1.8640	0.0271	68.7322	< 0.0001
Reg=reg	0.0396	0.0289	1.3697	0.1708
Imp=Semantic	0.5675	0.0278	20.3928	< 0.0001
Prop. deleted nodes	-5.9917	0.0448	-133.6870	< 0.0001
Reg=reg by Imp=Semantic	0.4862	0.0388	12.5441	< 0.0001

Table 6: Generalized Linear Model fitted to production accuracy in Experiment 2 (Reg: Regularity, Imp: Impairment, treatment dummy coding).

The right panel of Figure 5b visualizes the effects of regularity and impairment for the simulation runs. The red line, which is close to the line $y = x$, highlights that there is no difference in accuracy for regular and irregular verbs under phonological impairment. Under semantic impairment, regulars show higher accuracy than irregulars, and as in Experiment 1, this effect is most pronounced for lower accuracies. Again, there is substantial scatter, with outlier simulation runs with atypical accuracy patterns, especially in the mid accuracy ranges.

Experiment 2 replicates exactly the results of Bird et al. (2003). To our knowledge, this is the first simulation study showing that under phonological impairment, regular and irregular verbs pattern in the same way. An important advantage of our approach is that we can not only simply state that equally phonologically complex words should yield the same results in production, but moreover why exactly this happens. The model is simple enough that its behaviour is straightforward to understand. The longer or more phonologically complex a word is, the more likely it will be affected by damage to the phonology, as it is more probable that one of its triphones will no longer be available.

5.3 Discussion

What Experiments 1 and 2 have in common is that under semantic impairment, regular verbs suffer less than irregular verbs. At the same time, semantic impairment allows higher levels of accuracy to be maintained compared to phonological impairment. This finding follows straightforwardly from the fact that phonological lesions take triphones out of commission. As a consequence, words with such triphones can no longer be produced.

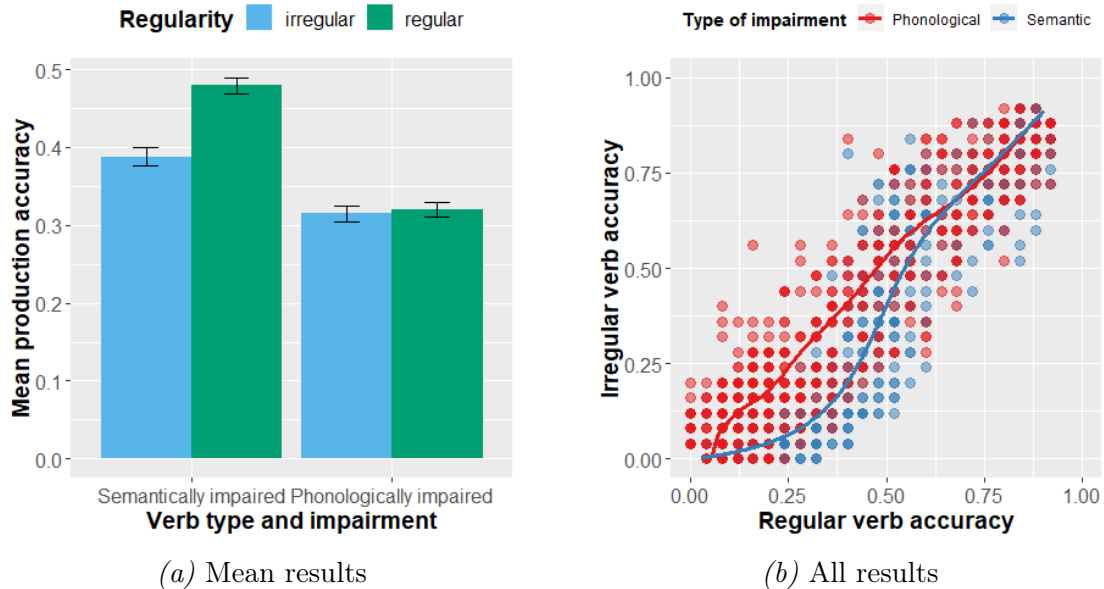


Figure 5: Experiment 2: Interaction of Regularity by Type of Impairment. Left panel: barplot for group means, scores are calibrated with respect to the baseline performance of the undamaged model. Right panel: scatterplot for the simulation runs. The less transparent points are, the more points overlap. Overlapping red and blue points are shown in purple. The curves connect accuracies as predicted from proportion of deleted nodes using GAMs for each of the four subsets defined by Regularity \times Impairment.

By contrast, semantic impairment has as its consequence that the mapping onto triphones will be less accurate, but words’ paths in the triphone graph remain available in principle, albeit with decreased and less accurate support from the semantics.

The question then remains why it is that in the simulations, across both experiments, regular verbs retain higher levels of accuracy compared to irregular verbs when the semantic topological map is damaged. After all, the discriminative lexicon model is not informed about whether a verb is regular or irregular. There is a single mapping from semantics to phonological form that has been optimized for regulars and irregulars simultaneously.

One possible reason could be that there is an asymmetry in the ‘convergence of codes’ for regulars and irregulars that is independent of their semantics. However, this possibility can be ruled out given the simulation study of Joanisse and Seidenberg (1999). These authors took for granted that semantic similarity is irrelevant for the past tense in English. This made it possible for them to represent verbs’ meanings using 1-hot encoding, effectively making all verb meanings fully orthogonal and completely unrelated. An interesting property of their model is that, in addition to lesioning connections between the semantic clean-up units, and independently between the phonological clean-up units, Gaussian noise was added to the semantic units, but not to the phonological units. If lesioning of connections between clean-up units and semantic units had been sufficient to obtain this increased vulnerability of irregular verbs, it would not have been necessary to add further noise to the semantic units. From this we conclude that the statistical properties of the relation between form and meaning, under the assumption that verbs’

meanings are to all practical purposes unrelated, do not support an asymmetry in the convergence of codes for regulars and irregulars.

However, above we documented that irregular verbs have greater semantic density compared to regular verbs. We showed that this greater semantic density is actually visible in the semantic space that we are using in the present study. Interestingly, this difference in density re-emerges in the materials of Experiments 1 and 2. For each of the experimental words, we calculated the number of semantic neighbors with a correlation exceeding 0.1. For regulars, there were on average 23 fewer such neighbors. This difference was numerically smaller for Experiment 2, but, as can be seen in Table 7, the effect of Experiment did not reach significance in a model regressing semantic density on regularity and experiment. (In a model including an interaction of regularity and experiment, this interaction was not significant, $p = 0.4543$.) Although the stimuli of Experiment 1 were matched for frequency, and those of Experiment 2 for phonological make-up, neither experiment matched for semantic density. Therefore, in both experiments, regularity and semantic density are confounded.

	Estimate	Std. Error	t value	Pr(> t)
Intercept (exp1, irregular)	72.5063	7.1911	10.0828	< 0.0001
Regularity=regular	-22.9890	8.1472	-2.8217	0.0059
Experiment=exp2	14.0082	8.1699	1.7146	0.0899

Table 7: Predicted number of semantic neighbors with a correlation exceeding 0.1, predicted from regularity and experiment.

In our model, the greater semantic density of irregular verbs renders them more vulnerable to semantic lesioning. Specifically, in the semantic graph, which has lexomes as vertices and an edge between lexomes if they are correlated above a certain threshold, irregular verbs tend to be connected to more hubs than is the case for regular verbs. Thus, if a hub lexome is lesioned, more irregular verbs are affected compared to regular verbs.

To see this, we proceeded as follows. We first inspected more closely the matrix \mathbf{S} , the row vectors of which specify the collocation strengths of verbs (rows) with lexomes (columns). From this matrix, we derived the correlation matrix for its column vectors. The 75-th percentile of the resulting set of correlation values was 0.056. We chose this as a threshold for defining an adjacency matrix for a graph with lexomes as vertices and (undirected) edges between any pair of lexomes if and only if the correlation between their column vectors in \mathbf{S} exceeds 0.056. From this adjacency matrix, we calculated for each vertex its degree, and defined the 500 lexomes with the highest degrees to be the ‘hubs’ of the graph (0.132% of the total number of lexomes). For these hub lexomes, degrees ranged from 1513 (for ALLOWANCE) to 1762 (for WEIGH).

Having defined the hubs, we next calculated, for each verb, the number of strong collocational strengths with hub lexomes. We counted a collocational strength as strong if it exceeded 0.001; this is the case for 1.9% of the collocation strengths (median: -0.00004). For the regular verbs, the median number of strong connections to hubs was 8, whereas for irregular verbs, the median was 14, a significant difference according to a Wilcoxon test ($p < 0.0001$; a t-test for log-transformed counts likewise supported a difference in strong connectivity to hubs $t_{(111.52)} = 6.90, p < 0.0001$, mean log count irregulars: 2.57,

mean log count regulars: 1.9). Since irregular verbs have more strong connections to hubs than is the case for regular verbs, it follows that if a hub is lesioned, production is likely to be impaired for more irregular verbs than is the case for regular verbs.

	Estimate	Std. Error	t value	Pr(> t)
Intercept (regtype = both, hubtype = hub)	1.4183	0.0167	84.91	< 0.0001
hubtype = nonhub	0.1557	0.0179	8.69	< 0.0001
regtype = irregular	-1.1072	0.0236	-46.87	< 0.0001
regtype = regular	-0.5001	0.0236	-21.17	< 0.0001
hubtype = nonhub × regtype = irregular	-0.1676	0.0254	-6.61	< 0.0001
hubtype = nonhub × regtype = regular	-0.0612	0.0254	-2.41	0.0158

Table 8: Linear model fitted to mean absolute collocation strengths from lexomes to triphones, broken down by whether the lexome is a hub (**hubtype** hub/nonhub) and by subsets of triphones that are unique to regulars, unique to irregulars, or that occur in both kinds of verbs (**regtype**: regular, irregular, both).

Irregular verbs are also characterized by more frail connection strengths between lexomes and triphones — a direct consequence of their irregularity. The connection strengths at issue are those defined by the matrix \mathbf{G} , which transforms a word’s semantic vector into its form vector. Figure 6 plots on the vertical axis words’ mean absolute connection strengths from lexomes to triphones, broken down by vertex type (hub, nonhub, on the horizontal axis) and triphone type (unique to regulars, unique to irregulars, or used by both regulars and irregulars, color-coded with blue, red, and black respectively). The triphones unique to irregular verbs have the smallest connection strengths, the triphones unique to regular verbs have stronger connection strengths, and the triphones that occur in both types of verbs have the strongest connection strengths. We think that it is this frailty of the connection strengths for specifically triphones unique to irregulars that makes it possible for the addition of random noise to the semantic nodes of the model of Joanisse and Seidenberg (1999) to affect irregular verbs more strongly than regular verbs.

Interestingly, for the triphones used (exclusively, or non-exclusively) by regular verbs, connection strengths are reduced for hubs as compared to nonhubs. By contrast, for triphones unique to irregular verbs, there is a trend in the opposite direction ($t_{(644.64)} = 2.7995, p = 0.0053; W = 876230, p = 0.0215$). Thus, when a hub lexome is lesioned, it is not only the case that more irregular verbs are affected, but in addition, the triphones that connect up to these hubs have connection weights that for the irregular verbs are stronger (and hence somewhat more important) compared to those to nonhubs. As a consequence, when hubs are lesioned, irregular verbs are hit twice, once because hubs are shared by more irregulars, and a second time because for irregular verbs it is the connections between their triphones and hubs that are the strongest.

6 Summary and conclusions

Studies of the production of English verbs in aphasia have reported dissociations between regular and irregular verbs (Ullman et al., 1997). It has been claimed that agrammatic patients with anterior aphasia or Parkinson’s disease have problems with inflecting regulars while patients with memory loss (Alzheimer’s disease or posterior aphasia) have

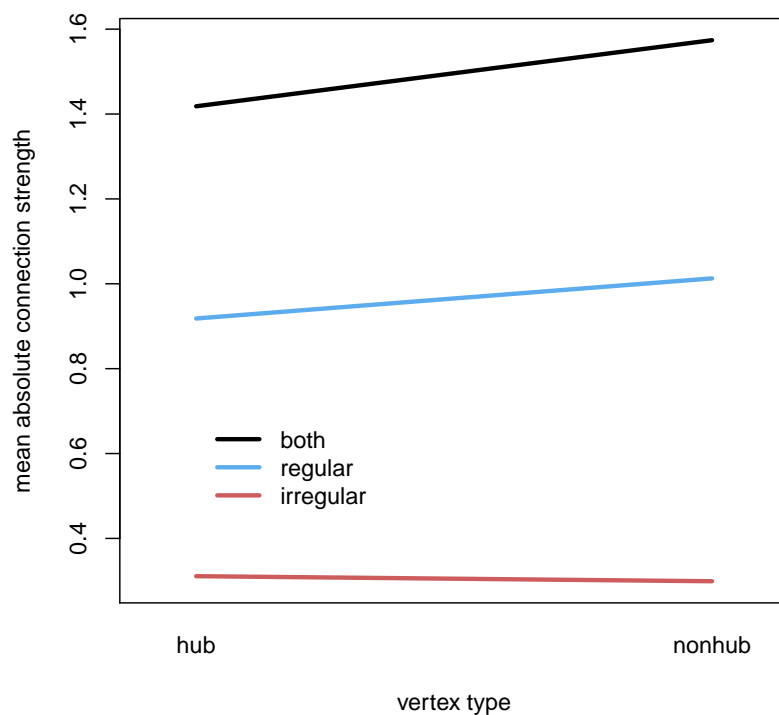


Figure 6: Mean absolute collocation strengths from lexomes to triphones, broken down by whether the lexome is a hub (hub type hub/nonhub) and by subsets of triphones that are unique to regulars, unique to irregulars, or that occur in both kinds of verbs (regularity type: regular, irregular, both).

problems with irregulars. However, Joanisse and Seidenberg (1999) noted that patients with memory problems actually had semantic deficits and that the agrammatic patients suffered from phonological difficulties. They therefore proposed a connectionist model with banks of units for semantics, auditory input, and speech output, as well as multiple hidden layers. Depending on whether they damaged the production layer or the semantic layer, the model selectively had more problems with regular verbs or with irregular verbs. But in order to get their model to produce the desired interaction of type of impairment by regularity, Joanisse and Seidenberg had to add Gaussian noise to the semantic representations, without in parallel adding noise to the phonological representations. In other words, in this study, the desired interaction was hand-crafted into the model architecture.

Joanisse and Seidenberg assumed that semantics is irrelevant for past-tense forms in English. However, Baayen (2005) showed that English irregular verbs have higher semantic densities than regular verbs. In the present study, we replicated their finding that irregular verbs cluster more closely together in semantic space compared to regular verbs, using a different method for constructing semantic vectors (a.k.a. word embeddings). By using empirical semantic representations, instead of one-hot encoding of word meanings, we were able to show for English that when topographically coherent areas of phonological and semantic networks are lesioned, the interaction of regularity by type of impairment follows straightforwardly. The reason that specifically irregular verbs are vulnerable to semantic lesioning follows from discrimination learning (Baayen et al., 2011, 2019, Baayen and Smolka, 2019) in interaction with the distributional properties of English. Because of their greater semantic density, irregular verbs are more likely to be connected to other irregular verbs in the semantic network. As a consequence, in the two-dimensional network layout of the semantic topological map, irregular verbs will be more likely to be close together. Hence, when the map is lesioned, irregular verbs are more likely to be affected jointly.

Returning to the model of Joanisse and Seidenberg (1999), we note that it is likely that if its one-hot encoded semantic units were to be replaced by more realistic semantic vectors such as can be derived with methods from distributional semantics, it would be able to capture the interaction of regularity by type of impairment without requiring additional noise superimposed on the semantic units. Whereas computational models such as Ferro et al. (2011) share some of the principles of self-organisation and discriminative learning with the present theory of the discriminative lexicon, they lack the incorporation of semantics. Joanisse and Seidenberg (1999) on the other hand did include semantics, but used one-hot encodings which fail to capture semantic similarity. Because their model does in principle integrate both phonology and semantics, we expect it to yield similar results to ours if real semantic vectors are used instead of one-hot encodings. This would make adding Gaussian noise to the semantic representations unnecessary, resulting in a model where the effect is not crafted into the architecture.

The present study is the first to provide a biologically more plausible way of implementing lesioning in artificial neural networks. In patients with a stroke it is unlikely that random connections across the entire network are damaged, given that bloodflow is typically constrained to some small subarea. This locality of brain damage is taken into account in our model, which constructs phonological and semantic maps. While we used the `graphopt` algorithm (Csardi and Nepusz, 2006) to obtain two-dimensional layouts of our model’s networks, there are other viable alternatives, such as t-SNE (Maaten and

Hinton, 2008) or the self-ordering maps used by Ferro et al. (2011).

We do not by any means claim that our model is, as it stands, anywhere close to biological truth. A first step forward to make it more realistic is to set up the model with spiking neurons (Maass, 1997, Bellec et al., 2019) instead of perceptron-like neurons. Since mathematically our mappings are extremely simple, such a conversion is feasible. A challenge for this line of research is to develop principled ways in which topographical maps of spiking neurons can be set up. The model of the discriminative lexicon (Baayen et al., 2019), that we have extended in the present study to help understand phonological and semantic impairment in aphasia, is a functional linguistic computational model that, by using the mathematics of multivariate multiple regression and network science, provides a framework for obtaining a more precise understanding of how the distributional properties of language and basic principles of human learning interact.

A Test sets

1. Verbs controlled for frequency (Baayen et al., 1995) as used in Experiment 1. The first 13 verbs are taken from Bird et al. (2003).

Regulars: clear, dust, fork, graze, lace, march, play, rake, roar, rock, step, tap, test, agree, change, dance, die, drill, glaze, hook, hug, warn

Irregulars: eat, fall, freeze, lend, sit, slit, speak, steal, stand, stride, swim, weep, write, buy, break, build, cast, drive, grow, hide, forbid, shrink

2. Verbs where past tense is matched for phonological complexity as used in Experiment 2.

Regular infinitive	Regular past tense	Irregular infinitive	Irregular past tense	CV-Structure
row	r6d	come	k1m	CVC
weigh	w1d	do	dId	CVC
cheer	J7d	give	g1v	CVC
bow	b6d	get	gQt	CVC
die	d2d	make	m1d	CVC
share	S8d	say	sEd	CVC
tie	t2d	think	T(t	CVC
dare	d8d	take	tUk	CVC
peer	p7d	be	wQz	CVC
vow	v6d	be	w3R	CVC
call	k(1d	keep	kEpt	CVCC
like	l2kt	leave	lEft	CVCC
look	lUkt	mean	mEnt	CVCC
move	muvd	tell	t51d	CVCC
seem	simd	go	wEnt	CVCC
twitch	twIJt	creep	krEpt	CCVCC
twirl	tw31d	drink	dr{Nk	CCVCC
trudge	trV=d	shrink	Sr{Nk	CCVCC
storm	st(md	slink	s1VNk	CCVCC
place	pl1st	spend	spEnt	CCVCC
drag	dr{gd	stink	st{Nk	CCVCC
dart	d,tId	forbid	f@b{d	CVCVC
hate	h1tId	forget	f@gQt	CVCVC
screw	skrud	spring	spr{N	CCCV
stray	str1d	stride	str5d	CCCV

References

- Arnold, D., Tomaschek, F., Lopez, F., Sering, T., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLOS ONE*, 12(4):e0174623.
- Baayen, R. H. (2005). Data mining at the intersection of psychology and linguistics. In Cutler, A., editor, *Twenty-first century psycholinguistics: Four cornerstones*, pages 69–83. Erlbaum, Hillsdale, New Jersey.
- Baayen, R. H., Chuang, Y., and Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, 13(2):232–270.
- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*.
- Baayen, R. H., Milin, P., Filipović Durdević, D., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118:438–482.
- Baayen, R. H. and Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, 81:666–698.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition, and Neuroscience*, 31(1):106–128.
- Baayen, R. H. and Smolka, E. (2019). Modelling morphological priming in German with naive discriminative learning.
- Bellec, G., Scherr, F., Hajek, E., Salaj, D., Legenstein, R., and Maass, W. (2019). Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. *arXiv preprint arXiv:1901.09049*.
- Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., and Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language*, 48(3):502–526.
- Blakely, T., Miller, K. J., Rao, R. P., Holmes, M. D., and Ojemann, J. G. (2008). Localization and classification of phonemes using high spatial resolution electrocorticography (ecog) grids. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4964–4967. IEEE.

- Burzio, L. (2002). Missing players: Phonology and the past-tense debate. *Lingua*, 112:157–199.
- Chersi, F., Ferro, M., Pezzulo, G., and Pirrelli, V. (2014). Topological self-organization and prediction learning support both action and lexical chains in the brain. *Topics in cognitive science*, 6(3):476–491.
- Cholin, J., Rapp, B., and Miozzo, M. (2010). When do combinatorial mechanisms apply in the production of inflected words? *Cognitive Neuropsychology*, 27(4):334–359.
- Chuang, Y.-Y., Loo, K., Blevins, J. P., and Baayen, R. H. (2019). Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning. *PsyArXiv*, pages 1–19.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Faroqi-Shah, Y. (2007). Are regular and irregular verbs dissociated in non-fluent aphasia? *Brain Research Bulletin*, 74(1-3):1–13.
- Ferro, M., Marzi, C., and Pirrelli, V. (2011). A self-organizing model of word storage and processing: implications for morphology learning. *Lingue e linguaggio*, 10(2):209–226.
- Ivens, S. H. and Koslin, B. L. (1991). *Demands for Reading Literacy Require New Accountability Methods*. Touchstone Applied Science Associates.
- Joanisse, M. F. and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: a connectionist model. *Proceedings of the National Academy of Sciences*, 96:7592–7597.
- Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.
- Juola, P. (2000). Double dissociations and neurophysiological expectations. *Brain and cognition*, 43:206–324.
- Lambon Ralph, M. A., Braber, N., McClelland, J. L., and Patterson, K. (2005). What underlies the neuropsychological pattern of irregular regular past-tense verb production? *Brain and Language*, 93(1):106–119.
- Landauer, T., Foltz, P., and Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25:259–284.
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge.
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671.

- Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.
- MacWhinney, B. and Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition*, 40:121–157.
- Marslen-Wilson, W. D. and Tyler, L. K. (1997). Dissociating types of mental computation. *Nature*, 387(6633):592–594.
- Marusch, T., Jäger, L. A., Burchert, F., and Nickels, L. (2017). Verb morphology in speakers with agrammatic aphasia. *The Mental Lexicon*, 12(3):373–403.
- Meteyard, L., Price, C. J., Woollams, A. M., and Aydelott, J. (2013). Lesions impairing regular versus irregular past tense production. *NeuroImage: Clinical*, 3:438–449.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., and Baayen, R. H. (2017). Discrimination in lexical decision. *PLOS-one*, 12(2):e0171935.
- Miozzo, M. (2003). On the processing of regular and irregular forms of verbs and nouns: evidence from neuropsychology. *Cognition*, 87(2):101–127.
- Nickels, L. and Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology*, 21(1):57–78.
- Patterson, K., Ralph, M. A. L., Hodges, J. R., and McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, 39(7):709–724.
- Pinker, S. (1991). Rules of language. *Science*, 253(5019):530–535.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. Basic Books, New York, NY, US.
- Pinker, S. and Ullman, M. (2002). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences*, 6:472–474.
- Pitt, M., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, 45(1):89–95.
- Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45(1):45–94.
- Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, volume Volume 2: Psychological and biological models, chapter On Learning the Past Tenses of English Verbs, pages 216–271. MIT Press, Cambridge, MA.

- Shafaei-Bajestan, E. and Baayen, R. H. (2018). Wide learning for auditory comprehension. In Yegnanarayana, B., editor, *Proceedings of Interspeech 2018*, pages 966–970, Hyderabad, India: International Speech Communication Association (ISCA).
- Shapiro, K. and Caramazza, A. (2003). Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia*, 41(9):1189–1198.
- Smolka, E., Preller, K. H., and Eulitz, C. (2014). ‘verstehen’(‘understand’) primes ‘stehen’(‘stand’): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, 72:16–36.
- Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Modeling the duration of word-final s in english with naive discriminative learning. *Journal of Linguistics*. <https://psyarxiv.com/4bmwg>, doi = 10.31234/osf.io/4bmwg.
- Tyler, L. K., Stamatakis, E. A., Jones, R. W., Bright, P., Acres, K., and Marslen-Wilson, W. D. (2004). Deficits for semantics and the irregular past tense: A causal relationship? *Journal of Cognitive Neuroscience*, 16(7):1159–1172.
- Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., and 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.
- Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., and Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, 93(2):185–238.
- Westermann, G. and Ruh, N. (2009). Synthetic brain imaging of english past tense inflection. In *Proceedings of the Cognitive Science Society*, volume 31.
- Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5):2020–2045.