

## Article

Sara Budts\*

# A connectionist approach to analogy. On the modal meaning of periphrastic DO in Early Modern English

<https://doi.org/10.1515/cllt-2019-0080>

Received December 9, 2019; accepted June 9, 2020; published online August 17, 2020

**Abstract:** This paper innovatively charts the analogical influence of the modal auxiliaries on the regulation of periphrastic DO in Early Modern English by means of Convolutional Neural Networks (CNNs), a flavour of connectionist models known for their applications in computer vision. CNNs can be harnessed to model the choice between competitors in a linguistic alternation by extracting not only the contexts a construction occurs in, but also the contexts it could have occurred in, but did not. Bearing on the idea that two forms are perceived as similar if they occur in similar contexts, the models provide us with pointers towards potential loci of analogical attraction that would be hard to retrieve otherwise. Our analysis reveals clear functional overlap between DO and all modals, indicating not only that analogical pressure was highly likely, but even that affirmative declarative DO functioned as a modal auxiliary itself throughout the late 16th century.

**Keywords:** neural networks, connectionism, analogy, modality, periphrastic do

## 1 Introduction

One of the long-standing puzzles in the history of English is the rise of the auxiliary DO. Today, the use of DO as auxiliary, also known as periphrastic DO, has become the default option in questions (1a) and negatives (1b).

---

\*Corresponding author: Sara Budts, Universiteit Antwerpen, Antwerpen, Belgium,  
E-mail: sara.budts@uantwerpen.be

- (1a) Do you like ice cream? (\*Like you ice cream?)
- (1b) I do not like ice cream. (\*I like not ice cream.)

Periphrastic DO entered the language in Middle English, but only became obligatory in the Early Modern period. Various hypotheses have been put forward as to why the construction acquired its present day distribution. One of these revolves around analogy with the modal auxiliaries. The core modals – CAN, MAY, MUST, SHALL and WILL – had become significantly distinct from main verbs around 1550, shortly before periphrastic DO settled in its eventual distribution. This closeness in timing has raised questions about their interrelatedness: perhaps the newly established modals served as an analogical model for DO and collectively helped it to acquire full auxiliary status in the subsequent century (Warner 1993). The present article seeks to test this hypothesis empirically in a large-scale, EEBO-based corpus, by zooming in on the relation between periphrastic DO and the modal auxiliaries by the turn of the 16th century. A detailed manual analysis of the functional equivalence between DO and the modals reveals that analogical influence was likely but restricted in time. Analogy with the modals probably triggered the use of DO as an emphatic marker of truthfulness in affirmative declarative contexts in the middle of the 16th century, but it fails to account for the eventual regulation of the construction in the 17th century.

To chart the functional similarity between periphrastic DO and the modals, I make use of a connectionist model that identifies similarities in a fully data-driven, bottom-up way. Connectionist models, currently more widely known as Artificial Neural Networks, have rapidly gained fame as data-analysis tools in a wide range of applications. The method developed in this paper adopts a Convolutional Neural Network, one specific member of this family, which, as I show, can be successfully recruited for the study of linguistic alternations. When trained on a large body of corpus data, the network identifies which uses are prototypical of each competitor and, more importantly, where the functional profiles of several competitors overlap. This enables us to identify analogical influences within and across linguistic paradigms.

The paper is structured as follows. Section 2 contains a concise summary of previous research on the rise of periphrastic DO, focussing in particular on its relation to the modal auxiliaries. Section 3 will outline the methodology. Section 4 presents the results of the case study. It starts off with a qualitative analysis of prototypical periphrastic DO in the late 16th century and goes on to describe more peripheral uses of periphrastic DO and, in particular, how they relate to the modal auxiliaries. Section 5 discusses the results as well as the theoretical relevance and limitations of this methodology for tracing analogical influence holistically.

## 2 Periphrastic DO and the modal auxiliaries

### 2.1 The rise of periphrastic DO?

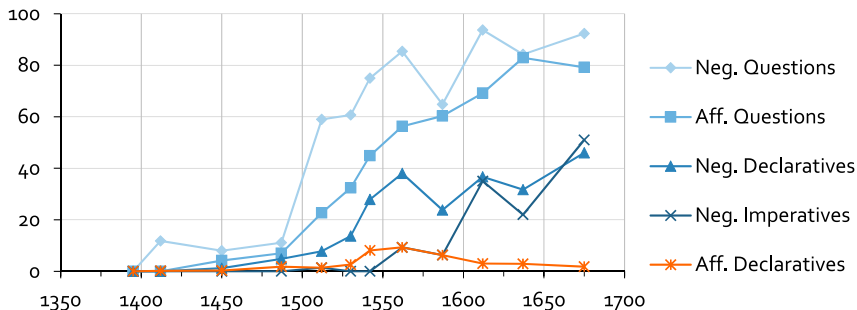
Today, when no auxiliary is present, periphrastic DO functions as a syntactically required “operator” in the so-called NICE-environments (Huddleston 1976), exemplified in (2) below:

- |   |   |
|---|---|
| (2a) Negation                             | I do/will/shall/must not love you.              |
| (2b) Inversion                            | Do/will/shall/must you love me?                 |
| (2c) Coding previously mentioned material | So you ate all the cookies, did you?            |
| (2d) Emphasis                             | I dó/will/sháll/múst love you!<br>(*I lóve you) |

DO has not always been obligatory in these environments. Its current distribution is a 16th century innovation. Until the end of the Middle English period, DO typically occurred in clusters where the verb still straightforwardly contributed to the meaning of the clause (Denison 1993). From the 16th century onwards, however, more and more uses of DO no longer seemed to contribute to the overall meaning anymore (Denison 1993; Ellegård 1953; Filppula et al. 2008; Garrett 1998; Van der Auwera and Genee 2002). The period between 1600 and 1800, traditionally termed the regulation stage, witnessed the systematic diffusion of the semantically empty construction in the NICE-environments.

Periphrastic DO spread across these different contexts at different rates. The spread was charted for the first time in a seminal study by Ellegård (1953), whose results are reproduced in Figure 1. Even though the study dates from the 1950s, the graph has been cited frequently and continues to inform a lot of contemporary research on the rise of the construction (e.g. Kroch 1989: 221; Stein 1990: 15; Nurmi 1996: 151, 2011: 343; Vulcanovic 2005: 2; Kauhanen and Walkden 2017: 2). According to Ellegård’s dataset, DO was present in about 80% of all negative questions around 1550, but only in 55% of affirmative questions (2b), and a mere 30% of negative declaratives (2a).

Interestingly, until 1550 the rise of DO systematically affected all contexts, including affirmative declaratives like “I do love you”, without emphasis on “do”. Nevertheless, the use of DO in affirmative declaratives never exceeded 10% of all declarative sentences without other auxiliaries and eventually waned. Judging by

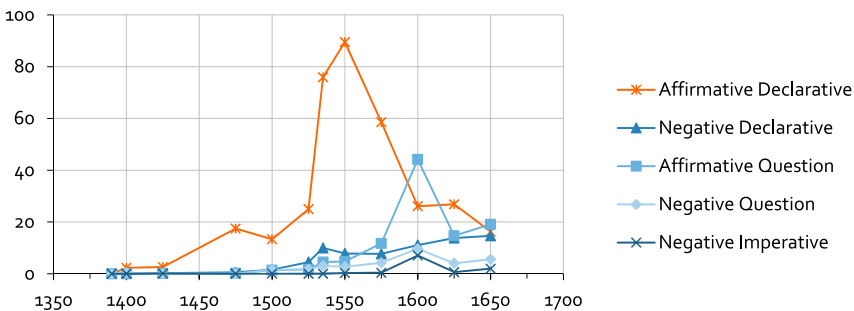


**Figure 1:** Relative frequency of DO. Adapted from Ellegård 1953: 162.

its relative frequency, the construction pattern seems only a footnote in DO's journey towards auxiliarihood.

If we take into account the frequency distribution of the contexts, however, a completely different picture emerges. The true pervasiveness of affirmative declarative DO is revealed by Figure 2, which shows the normalized frequencies of the data in Figure 1. Even if DO only accounted for at most 10% of all affirmative declaratives, the sheer frequency of its host clause made that affirmative declarative DO outnumbered the other construction patterns by a large margin. Despite its absolute frequency, however, the flourishing of affirmative DO was only short-lived. Within the timeframe of a mere century, the construction first rose to almost seven times the frequency of the other construction types combined, only to plummet dramatically in the next 50 years.

Moreover, the canonical graph in Figure 1 is a poor visualization of DO's overall frequency. Judging by its relative frequency across sentence types, periphrastic DO seems to have been on the rise significantly and consistently ever since the 16th century. As with the frequency of affirmative declaratives, however, this



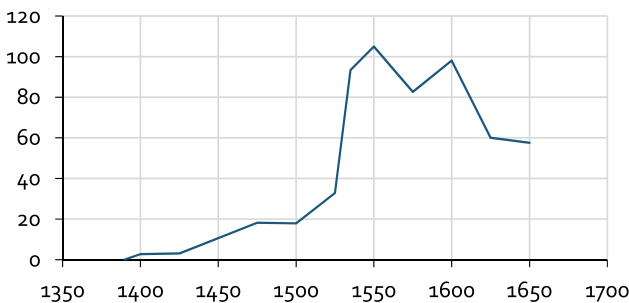
**Figure 2:** Normalized frequency of DO per thousand finite clauses. Based on counts in Ellegård 1953: 161–162.

appearance proves false when we take into account the frequency of the sentence types in Figure 1. After a sharp rise in the early 16th century, the frequency of periphrastic DO actually declined (cf. Figure 3), most notably so in the early 17th century, an era known ironically as that of “the rise of DO”. Because previous research has mainly been concerned with the spread of DO to the NICE environments, its most frequent construction type has usually remained below the radar.

## 2.2 Analogy with the modals

While there have been studies on the semantic, pragmatic and sociolinguistic properties of affirmative declarative DO (e.g. Nevalainen 1991; Rissanen 1991; Nurmi 2018), the importance of this construction for the eventual regulation of DO as a whole has rarely been explored. Rather, the regulation of DO has been explained in various other ways, from a variety of linguistic backgrounds. It has been related to the loss of V-to-I raising (Kroch 1989; Warner 1993), to functional pressures promoting ease of communication (Hudson 1997), to influence from Celtic (Van der Auwera and Genee 2002) and to the phonotactics of the DO INF combination (Stein 1990).

Warner (1993) suggests that a major cognitive factor was similarity with the modal auxiliaries. Just like DO, the modals became full-fledged auxiliaries in Early Modern English, around 1550 (cf. also Lightfoot 1979; Plank 1984). The close timing between the cluster of changes that marked the birth of the modal auxiliaries and the rise of periphrastic DO in questions led to the belief that the rise of periphrastic DO was influenced by the modals: “The coincidence of date here strongly suggests that the development in modals and in DO are interconnected. Any linguistic history must give some account of this interconnection if it is to be convincing”



**Figure 3:** Total normalized frequency of DO per thousand finite clauses, regardless of sentence type. Based on counts in Ellegård 1953: 161–162.

(Warner 1993: 221). In a self-proclaimed “unashamedly speculative section” (1993: 219), Warner describes this interconnection as follows: “It looks at the first sight as if the connection between the category change of modals and the rise of DO might be that it involved a reanalysis of DO as a unitary item expressing tense/mood in the late 15th century, when this characteristic in modals became central to a basic-level word class; or if this had already happened, some favouring of DO because of the word class’s new status” (Warner 1993: 223). He not only recognizes an influence of the modal auxiliaries on periphrastic DO, he hypothesizes the reverse relation as well (221–222). If the auxiliaries underwent influence from periphrastic DO in Early Modern English, that very influence sharpened the divide between modals and full verbs even more.

The hypothesis that the regulation of periphrastic DO was influenced by the modals ties in with research on the importance of analogy as a driving force in both language processing and language change (e.g. Fischer 2007; De Smet 2009). While many studies on analogy have focused on the regularization of outliers in (morphosyntactic) paradigms, e.g. the shift from “kine” to “cows” (Fertig 2013: 8), the mechanism is by no means restricted to these environments. In the present paper, analogy is defined in its broadest sense as “structural similarity” between (groups of) linguistic elements (cf. Itkonen 2005: 1), either because of their form, their function or their distribution. In addition, I adopt Traugott and Trousdale’s (2013) distinction between analogical thinking and analogization (35–39). Analogical thinking refers to a language user’s ability to perceive two constructions as related because of similarities in form or function. Analogization refers to a constructional change motivated by analogical thinking. While analogical thinking crucially enables analogization, not all instances of analogical thinking materialize into change. If we find any evidence that the regulation of periphrastic DO was influenced by similarity with the modal auxiliaries, the rise of DO might well be considered an instance of analogization.

### 2.3 Hypotheses

The present article seeks to investigate the analogical impact of the modal auxiliaries on DO’s journey towards auxiliarihood by zooming in on their interplay in the late 16th century. More specifically, it aims to look into the functional similarity between the modal auxiliaries and periphrastic DO in its use in affirmative declarative sentences. If analogy with the modals was of any importance, it might well have caused the sudden burst in affirmative DO.

A detailed analysis of the functional similarity between the modals and DO is the first step to be taken in order to test Warner’s claims about their interrelatedness. If

the brief rise of affirmative DO was the result of analogical attraction by the modal auxiliaries, we should find an elevated level of similarity between the modals and affirmative DO at the peak of its development. If functional similarity is attested, the analogy hypothesis has to be taken more seriously. Alternatively, if no traces of similarity can be found, analogy can be discarded as a regulating principle.

## 3 Data and methodology

### 3.1 Corpus and dataset

The relation between the modal auxiliaries and periphrastic DO will be pinned down on the basis of the Antigoon corpus. Antigoon is an 800-million-word selection from EEBO-TCP (Early English Books Online – Text Creation Partnership), an exhaustive collection of all English texts printed between 1477 and 1700 that survived the period, which in total covers more than 1.5 billion words of British English (cf. <http://eebo.chadwyck.com/>). The selection in Antigoon covers the period from 1580 to 1700 and is divided into six 20-year periods. The first three periods are an exhaustive collection of the corresponding data in EEBO. The last three periods consist of a random sample of EEBO texts, cut off at about 160 million words per period. The starting point of 1580 was chosen pragmatically; while earlier texts would have been useful, there were not enough of them in comparison to what is available for the later periods. To enhance (automated) querying and processing, the original EEBO-TCP-transcriptions were lowercased and have been enriched with tokenization<sup>1</sup> and spelling normalization (VARD2, Baron and Rayson 2008).

Because Ellegård's data indicate that the rapid surge of affirmative declarative DO took place in the second half of the 16th century, for this paper the analysis will be restricted to data drawn from Antigoon's first 20-year slice, which covers 1580–1600 and contains slightly under 80 million words. I queried the corpus for all attestations of forms of DO and the canonical modals (cf. Table 1). The form *does* was excluded from the analysis because there were only 511 attestations in the corpus (0.1% of the total number of attested DO forms), most of which were plurals of the animal *doe*.

I did not query for forms with deviant spellings, since most of them were covered by Antigoon's native spelling normalization. I applied specific post-processing in two other cases: (1) all instances of *cannot* and *shallbe* that were tokenized as one word have been split up into two separate tokens to ensure a uniform treatment of all negated modals and modals followed by *be*; (2) nominal

---

<sup>1</sup> The tokenization of Antigoon was carried out in parallel with the tokenization of the EMMA corpus (Petré Anthonissen et al. 2019).

**Table 1:** Overview of token count per form involved.

Present tense modals				Past tense modals				Forms of do	
Form	# Hits	Form	# Hits	Form	# Hits	Form	# Hits	Form	# Hits
<b>can</b>	107,950	<b>canst</b>	2973	<b>could</b>	48,752	<b>couldst</b>	480	<b>do</b>	117,398
<b>may</b>	116,496	<b>mayst</b>	4700	<b>might</b>	68,722	<b>mightst</b>	579	<b>dost</b>	3691
<b>shall</b>	188,153	<b>shalt</b>	13,121	<b>should</b>	143,002	<b>shouldst</b>	651	<b>doth</b>	109,579
<b>will</b>	151,921	<b>wilt</b>	6489	<b>would</b>	110,437	<b>wouldst</b>	2528	<b>did</b>	116,154
<b>must</b>	84,935								
Total present tense modals: 726,738				Total past tense modals: 374,974				Total do: 346,822	
Total dataset size: 1,348,534									

uses of *may* (i.e. the month) and *will* have been semi-automatically removed from the dataset<sup>2</sup>. Lexical uses of *will* as a transitive verb have not been removed and were taken up in the analysis. After these corrections I ended up with a dataset of 1,348,534 attestations in total. The token count for each form is provided in Table 1.

## 3.2 Convolutional neural networks

The turn of the 1970s witnessed the emergence of a new framework of cognitive science. Known as “connectionism”, or “parallel distributed processing”, the framework aspired to be a neurologically more plausible model of the human mind than the models that had been developed so far (Hinton and Anderson 1981; Rumelhart and McClelland 1986). Innovatively, connectionist research revolves around a mathematical model trained to map input patterns to output patterns. These models are the predecessors of Artificial Neural Networks, including the one used in the present case study. In the following sections, I will briefly outline my methodology and its assets for linguistic research. Sections 3.2.1, 3.2.2 and 3.2.3 provide a conceptual overview of my approach. Details on the technical implementation of the models are given in Appendix I.

<sup>2</sup> I compiled an annotated dataset of 500 nominal and 1500 verbal uses of both forms and used this set as training and test data for a CNN like the one described in Section 3.2.4. Once the models had learned the nominal/verbal distinction – they both achieved an accuracy of more than 99% on their respective test sets – I had them predict a label for all other instances of *will* and *may*. I accepted all its predictions when it was more than 90% confident of its decision and manually annotated the remaining cases.



### 3.2.1 A sentence as a picture

My analysis of the interplay between the modals and *DO* is centred around Convolutional Neural Networks (CNNs), a family of connectionist algorithms whose original application area is image recognition (LeCun et al. 1998). In image recognition, the learning algorithm is trained to label images based on the objects they depict. To the algorithm, an image is merely a grid of pixels. During the training process, it learns to map input pictures to output labels by sliding over the picture and keeping track of the groups of neighbouring pixels it has seen.

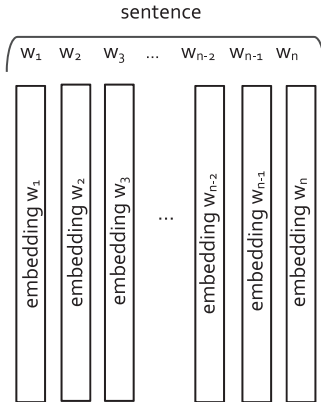
Recently, Convolutional Neural Networks have been applied to linguistic input as well, by treating a sentence as a picture (Collobert and Weston 2008; Dauphin et al. 2017; Vanni et al. 2018, but not yet to historical linguistics). A sentence can be thought of as a list of words, and a word as a list of word features. Such feature lists take the shape of real-valued vectors called “word embeddings”. Word embeddings are derived fully automatically from the distribution of words in a large corpus and implicitly encode semantic properties (like near-synonymy) as well as syntactic properties (like part of speech) (e.g. Mikolov et al. 2013; Hamilton et al. 2016; Budts and Petr  2020) for neural embeddings; Hilpert and Perek 2015 for non-neural embeddings; Dubossarsky et al. 2017; Tahmasebi et al. 2018 for a comparison between various types of embeddings). If the embeddings of all the words in a sentence are placed right next to each other, the sentence representation forms a 2D-grid not unlike a picture. This idea is visualized in Figure 4.

Importantly, algorithms like these can be harnessed to automate (parts of) a study on linguistic alternations. The idea is as follows: we take two (or more) alternating constructions (like periphrastic *DO* and the modals); we collect their attestations in a large corpus with a fixed context window<sup>3</sup>; we mask the words that express the alternating variant (i.e. the form of *DO* or the modal) in every context; we feed the masked input sentences to a convolutional network and train it to reconstruct which alternating competitor the sentence originally contained. By masking the words whose behaviour we want to study, we force the model to retrieve the contextual elements that optimally distinguish one competitor from the others. This, in turn, gives us insight in the factors that govern the alternation.

This process is exemplified in Table 2, where the context of a sentence originally containing *do* and one originally containing *may* are shown. In each sentence, the competitor (i.e. *do* and *may*) is masked. The sentences are fed to the

---

<sup>3</sup> I opted for a symmetric context window of 50 words in total. This choice was made pragmatically: a 50-word context window is long enough to be interpretable by a human annotator, but short enough to be computationally tractable.



**Figure 4:** A sentence as a picture.

algorithm as we teach it to learn which (groups of) words in the context are most predictive of, respectively, *do* and *may*. If we repeat this procedure for all sentences containing either a modal or a form of DO, the algorithm will be able to generalize over the individual attestations and pick up on more abstract contextual features that maximally discriminate DO from each of the modals as well as the modals among themselves. This is a valuable source of information from a theoretical point of view, because it provides insight into the factors that determine the choice between competitors in a completely bottom-up fashion.

### 3.2.2 Discriminatory features: Flexible N-Gram templates

When figuring out how to classify the contexts of competing constructions, the model gradually grows sensitive to N-grams that are predictive of one competitor,

**Table 2:** Example of input sentences with masked target along with their corresponding output labels. The beginning and end of the context window is marked with ///.

Input sentence	Output label
/// feast, when the poor are ready to famish? Was there never but one Dives, and one Lazarus upon the earth? Or <TARGET> we want wit, or will, or grace to apply a parable? Here I may well cry out and say to the ///	do
/// and of the Britons, and did incorporate that duchy to his realm and crown of France, as in the history of France it <TARGET> appear at large. In July this year was a prest levied for the king in the city of London , of four thousand ///	may

**Table 3:** Two groups of N-grams predictive of *may*.

5-grams predictive of <i>may</i>	
Desire verb + pronoun + <i>that</i>	Comparatives
desired marcellus that they <TARGET>	<TARGET> you the more easily
wish, that some occasion	<TARGET> i the better believe
beseech him that he <TARGET>	we <TARGET> the better attain
desiring him that he <TARGET>	<TARGET> you so rather be
prayed her that she <TARGET>	i <TARGET> with more facility
instantly craving that he <TARGET>	thou <TARGET> the better comprehend

but not the others. Importantly, these N-grams do not need to be the same literal string every time.

As an illustration, Table 3 lists two groups of 5-grams that the model identified as predictive of *may*<sup>4</sup>. The group on the left instantiates a template with a verb of desire, optionally followed by a pronoun in the objective, followed by *that*. The group on the right instantiates a template sensitive to evaluative comparatives, whether expressed as one word (“the better”), periphrastically (“the more easily”) or embedded in a prepositional phrase (“with more facility”). From both groups it is obvious that there is no need for the templates to be lexically specified. The patterns *can* grow sensitive to a single word, but they might just as well grow sensitive to a sub-word feature, regardless of the lexical items that instantiate it.

The two examples presented above are intuitive patterns that are easily interpretable by humans. It should be noted though that not all patterns picked up by the model reflect intuitive categories. Still, the patterns reflect genuine statistical tendencies and should not be disregarded.

Importantly, the model’s flexibility stems from the representation of the input data. As mentioned before, the model’s input consists of vector embeddings of input words rather than the input words themselves. Because the vectors capture aspects of the words’ semantic and syntactic behaviours in a large corpus, the model can infer how collocates relate to one another. This allows it to grow sensitive to (combinations of) sub-word features, such as verb semantics or syntactic category. In other words, the model grows sensitive to abstract N-gram templates that are tailored for the alternation at hand, in that they encode exactly the patterns

<sup>4</sup> The N-grams have been retrieved by convolving the learned feature templates with a 10K random sample of input sentences one N-gram at a time, collecting for each template the 50 N-grams that resulted in the highest convolution score.

that allow the model to discriminate best between the competitors in the alternation.

This naturally solves a couple of issues that collocational approaches of the slot-and-filler type still struggle with (e.g. Distinctive Collexeme Analysis [Gries and Stefanowitsch 2004], multidimensional scaling based on collocates [Hilpert 2016])). First, collocate synonymy is automatically dealt with. CNNs know how collocates relate to one another and pick up on abstract tendencies expressed collectively by a large diversity of lexical items.

Second, CNNs automatically perform ambiguity resolution on polysemous collocates. If a word has two senses, the word's embedding vector will capture features of both, but these features are stored at different positions in the vector. Some values in the embedding of *may*, for example, will reflect its meaning as a month, whereas others will betray its meaning as a modal verb. A template sensitive to temporal nouns will only pay attention to the positions where temporal semantics and syntactic nominality are stored, and will be underspecified for all other positions. As such, the template will effectively recognize *may* as a month when it is looking for a month, ignoring the modal semantics of its homonym. Similarly, when *may* is matched with a template sensitive to modal verbs, the template will ignore the vector positions expressing temporal semantics, only paying attention to the positions salient for modal verbs.

### 3.2.3 Probability distribution

The features that the model has grown sensitive to are used as input for the eventual classification, where the model computes for each context a probability distribution over the competing variants. For every context, the algorithm assigns a score to each competitor, as an indicator of the degree to which the competitor fits in that particular context. To illustrate this idea, Table 4 contains three examples from the dataset along with their probability distribution over the uninflected present tense forms. The first example is a context that originally contained *may*. The model has figured that out, since it assigned a significantly higher probability to *may* than to all other forms. The second example originally contained *do*. While the model recognized *do*'s suitability in this contexts, it puts forward *can* as a likely candidate too. The third sentence originally contained *do* as well, but the model failed to recognize this and suggests *may*, *must*, *shall* and *will* as likely targets instead.

These examples illustrate how we can harness CNNs to stratify the dataset in terms of how prototypical the attestations are for each competitor in the alternation. A sentence like the first one, where the model easily identified the correct competitor, probably instantiates a prototypical usage pattern of that competitor.

**Table 4:** Probability distribution over competitors.

Target Sentence	can	may	must	shall	will	do	other
<i>may</i> /// unto his will in all things. Amen. O Lord increase my faith. O Lord open thou my lips, that my mouth <TARGET> extol thee with praise, and be thankful unto thee for my benefits, & grant that I speak nothing but that which may ///	0%	<b>98%</b>	0%	0%	0%	0%	2%
<i>do</i> ///amongst which the first is our sins, not only those that be mortal, but also venial sins, because these, albeit they <TARGET> not extinguish charity in vs, yet do they slack and make cold the fervor of charity, which is as it were devotion ///	<b>45%</b>	2%	0%	0%	3%	<b>46%</b>	4%
<i>can</i> ///been used in divers and sundry fashions. The first thing that is required of him that shall take this powder is, that he <TARGET> prepare himself with good diet, & good order, keeping himself from all things that may offend health, and use these ///	1%	<b>32%</b>	<b>25%</b>	<b>18%</b>	<b>17%</b>	1%	6%

By inspecting a large sample of such easily classifiable sentences, we can reconstruct the prototypical use(s) of that competitor in opposition to the other forms in the alternation.

The main assets of CNNs as a tool to uncover analogy, however, resides in its treatment of contexts like the second and the third example, where several competitors are put forward as suitable fillers. CNNs allow us to extract not only the contexts a construction occurs in, but also the contexts the construction could have occurred in, but did not. Bearing on the idea that two forms are perceived as similar if they occur in similar contexts, the algorithm provides us with pointers towards potential loci of analogical attraction that would be very hard to retrieve otherwise.

## 4 Results

The remainder of the article will focus on the prototypical uses of periphrastic DO at the end of the 16th century, as well as the sites of functional overlap with the modal auxiliaries. For reasons of scope, I will restrict my analysis to the present tense forms not inflected for 2nd person singular. Section 4.1 sketches the prototypical use of periphrastic DO in the late 16th century, while Section 4.2

discusses the functional overlap between DO and the modal auxiliaries. The model treated *do* and *doth* as separate forms and assigned them both a separate score. In the analysis below, the scores for the two forms have been added to eliminate the competition between them and bring them more in line with the modal auxiliaries.

#### 4.1 Prototypical use: The modal meaning of periphrastic DO

To investigate the prototypical use of *do* and *doth* at the end of the 16th century, I randomly selected 500 easily classifiable<sup>5</sup> attestations of both forms and subjected them to a traditional corpus analysis. The prototypical samples of *doth* and *do* consist of affirmative declaratives for more than 70 and 80% respectively. This confirms the tendencies observed in Ellegård's dataset and clearly indicates that affirmative declarative DO was a salient construction pattern in the late 16th century.

Semantically, affirmative declarative DO often occurs in sentences with a strong sense of universality or habituality. With plural subjects, prototypical affirmative declaratives with DO express either typical behaviour or habits (3a) or generic traits<sup>6</sup> of a group of people (3b). Singular subjects, as in (3c), are typically inanimate and occur in propositions framed as a plain fact. In total about 63% of prototypical *do* and 42% of prototypical *doth* occurs in generic statements like the ones in (3).

- (3a) They **do** maintain themselves with fish which they **do** take on the coast, and of wild beasts which they **do** kill in the mountains. [92.5% **do**]
- (3b) Like as lovers **do** always behold somewhat in their mistress whereby they think her to excel all others. [91.8% **do**]
- (3c) for idleness breedeth lothsomnes, but labour engenders hunger: which hunger **doth** make diverse hard meats marvelous savoury, that lothsomnes made unsavoury. [88.8% **doth**]

---

<sup>5</sup> An attestation was considered as easily classifiable when the model identified its target correctly with more than 80% confidence.

<sup>6</sup> In most cases, the genericity is expressed at the level of the entire sentence, rather than at phrasal or discourse level (cf. the distinction made in Behrens 2005: 279).

Syntactically, about 6% of the universal/habitual attestations mentioned above occur in an *as*-clause, where they serve as a metaphorical means of comparison to clarify a claim (4). As metaphors are by definition easily accessible generic truths, *DO*'s connotation of universality in affirmative declaratives is well suited for this purpose.

- (4a) the church takes all her light and brightness of Jesus Christ, the true sun of righteousness, **as the moon doth take of the sun all her light**, which thing the papists do not. [95.0 % **doth**]
- (4b) **as sailors do apply themselves to the changes of the winds**: so do wise men to the affections of the mind. [96.1% **do**]

In addition to the broad universal/habitual usage pattern, the prototypical samples reveal two semantically more coherent extensions. The first spin-off comprises scientific texts (e.g. grammar in [5a] and medicine in [5b]). Scientific uses account for 10% of the prototypical sample of affirmative *do* and 7% of prototypical *doth*. The use of *DO* in scientific writing aligns with its universal/habitual uses both in terms of topic and in terms of writer commitment. Scientific writing tends to deal with general truths or regularities rather than one-off situations, and the authors typically show a high degree of commitment to the truth/validity of their statements.

- (5a) in [...] “virgilius legitur me”, the verb passive “legitur” **doth** govern the ablatiue case of the doer “me” [97.3% **doth**]
- (5b) hereof the discreet physician may plainly see the causes why these waters **do** manifestly cure wounds and ulcers [97.4% **do**]

The second spin-off comprises reportative or argumentative uses of *DO*, where the speaker mentions a claim made by a reliable source (e.g. *the epistle of Saint Iude* [6a], *three of our evangelists* [6b]) in order to back-up his own narrative. This use is in line with Stein's observations that *DO* tends to occur at key points in the discourse, and that “talking to or about God, Saints, and whoever may constitute an authority, as a rule a classical figure, triggers *do*” (Stein 1990: 64). It also ties in with Ellegård's (1953: 167) findings that the verbs “write”, “translate” and “add” are among the strongest collocates of the periphrastic construction.

- (6a) likewise **the epistle of Saint Iude doth show well**, with what vehement zeal and ardent affection, he exhorted men in his time unto repentance [93.1% doth]
- (6b) **Three of our evangelists do report**, that Iesus did blank divers of the learnedest pharises with alleging only these words of Daud; [92.1 %]

The reportative/argumentative use is reasonably frequent with prototypical *do* (17%), but is especially dominant with *doth*, where it accounts for about 38% of all affirmative declaratives. In these clauses, periphrastic DO imposes a sense of truthfulness onto the proposition, even if the proposition itself is not necessarily a priori true. This contrasts with universal truths and scientific facts, because they are true regardless of how they are framed. When statements are undeniably true by themselves, affirmative declarative DO merely emphasizes their inherent truth. In reportative/argumentative cases, by contrast, affirmative declarative DO imposes truthfulness rather than emphasize it, and therefore serves as a rhetorical device to frame an interpretation as a well-established fact.

While the universal/habitual use and its two spin-offs account for the vast majority of all attestations of prototypical affirmative DO, there are a couple of cases that lack universal semantics. With *do*, there are a few instances with first person subjects and a performative verb (e.g. *beseech* in [7a]), which – by definition – expresses an imminent action rather than a universal truth. A second group of uses lacking universal semantics occurs in poetry, where affirmative DO is used merely for metric purposes (7b)<sup>7</sup>.

- (7a) **I do beseech** your graces all to pardon me: I am bound by oath, I may not do it. [90.0 % do]
- (7b) then Cynthia he desires to show her face, // and bids her nightly chariot upward slide, // then **doth** he pray the clouds for to disgrace // the darkened night, and with their vailes to hide [86.2 % doth]

Overall, the prototypical samples point out that by the turn of the 16th century, affirmative DO was by no means semantically empty. Instead, it functioned as a marker of truthfulness, either by emphasizing the inherent truth of an easily accessible fact, or by framing a potentially subjective interpretation as an undeniable fact for rhetorical purposes (cf. “emphatic do” in Ellegård 1953: 147). Rather

---

<sup>7</sup> The line breaks (“//”) in the poetic example have been inserted manually for purposes of clarity. The model did not have access to this kind of information when classifying the sentence.



than contributing to the proposition itself, affirmative DO reflects the speaker's commitment towards the truth of the proposition. This is in line with Traugott's (1972) observation that when the subjunctive system broke down in Middle English, "the function of the indicative as assertion of truth was in part taken over by *do + Tense + Verb*, while the expression of uncertainty or noncommitment was left to the simple *verb + Tense*." (139)<sup>8</sup>. This explains the frequent occurrence of affirmative DO in passages where the hearer might doubt the correctness of the speaker's perception or claim (Traugott 1972: 139).

In contexts where DO emphasizes or imposes truthfulness, it essentially functions as an epistemic modal marker. The modal meaning of DO is explicitly argued for in Stein (1990: 88–89, 267): "if we assume a cline between the values of 1 and 0 as the epistemic area, modality is concerned with a comment from the speaker's side as to the truth or not – or shades in between like "possible", "likely", etc. – of the proposition. (88–89)." Periphrastic DO, then, emphatically imposes (more than) the maximal likelihood onto a proposition and has to be "located beyond the epistemic scale" (89).

While the idea of periphrastic DO as an epistemic marker might not be new as such, it does become of special significance when the rise of DO is discussed in relation to the modal auxiliaries. The 50-year gap in between the modals' collective coming-of-age and DO's syntactic regulation witnessed the drastic rise-and-fall of a construction that not only has the syntax of a (modal) auxiliary, but also displays modal semantics. In that particular time frame, affirmative DO might have been just as modal as the modal auxiliaries. This is a first indication that the 16th century explosion in affirmative DO was triggered by analogy with the modal auxiliaries. The next section addresses this issue explicitly by zooming in on the functional similarity between affirmative DO and the modals in various contexts.

## 4.2 Functional overlap with the modal auxiliaries

To chart the overlap between DO and the modals, I collected and annotated the sentences where functional equivalence was most likely<sup>9</sup>. The size of

---

<sup>8</sup> As one reviewer pointed out, the Present Day English translations of virtually all examples provided would have a bare verb form in the simple present (as opposed to the progressive). This observation, in combination with Traugott's observation that periphrastic *do* was related to the loss of the subjunctive, indicates that the semantics adopted by affirmative *do* might well have been aspectual as well as modal.

<sup>9</sup> I annotated all sentences where the product of the highest scoring form of DO and the highest scoring modal exceeded the threshold of 0.1875. This ensures that both probabilities lie between 0.25 and 0.75 and it favours an even distribution over a skewed one.

**Table 5:** Overview of overlap dataset sizes.

	Overlap with <i>do</i>	Overlap with <i>doth</i>
<b>can</b>	292	397
<b>may</b>	180	244
<b>must</b>	78	43
<b>shall</b>	249	294
<b>will</b>	422	314

the annotated dataset is shown in Table 5. The distribution of the dataset reflects the relative sizes of the overlap with the different forms, but it should by no means be regarded as an exhaustive collection of all contexts where overlap is plausible. For a more fine-grained estimate of the relative magnitude of the overlap sets, the reader is referred to the tables in Appendix II.

In the following sections, I discuss the overlap with CAN, MAY, SHALL and WILL. The overlap with MUST, by far the smallest dataset, will not be discussed for reasons of scope.

#### 4.2.1 DO and CAN

CAN differs from the other modals in its predisposition for negated clauses. The algorithm picked up on this tendency and assigned a relatively high probability score to *can* in all negated sentences. While the rest of the article is concerned with affirmative DO, the overlap with CAN focuses on negated attestations instead, as negation accounts for 90% of the overlap with both *do* and *doth*.

Both *doth* and *do* overlap with *can* in clauses with third person subjects expressing universal impossibility. Indeed, in sentences like (8), the two forms are naturally close in meaning. CAN expresses the impossibility of a situation to occur or the inability of a person to perform an action. In terms of truth value, this is equivalent to epistemic DO expressing emphatically that a situation does not hold or that someone did not perform an action. Even though CAN and DO themselves have different semantics, they by and large occur in similar environments.

- (8a) Our bodies and souls **do** not make vs members of Christ, but our faith and obedience. [41.1% **do**; 48.7% **can**]
- (8b) The good tree **doth** not bear ill fruit. [37.7% **doth**; 51.1% **can**]
- (8c) They which **can** not valiantly expose themselves to dangers, become slaves to those which assail them. [40.2% **can**; 52.4% **do**]

In addition, DO and CAN overlap with first person singular subjects and negated verbs of perception and cognition, as in (9) below. Again, while DO and CAN might not be identical in these contexts, they are equivalent from a functional point of view: the inability (*can not*) to perceive something inevitably leads to the perception not taking place (*do not*).

- (9a) I have many secret sins in me, which I **do** not see because I do not perfectly understand the law of God. [45.5% **do**; 44.2% **can**]
- (9b) I **can** not see who else may be so well antichrist, and a seducer of people. [37.4% **can**; 52.9% **do**]

#### 4.2.2 DO and MAY

A vast amount of the overlap between *doth* and MAY consists of clauses providing external evidence for a claim made elsewhere in an argumentative text. The subject is usually *it*, the infinitive is often *appear* (10), but verbs like *signify* and *follow* occur as well (cf. Ellegård 1953: 167)

- (10a) Concerning the members, **it may sufficiently appear** by this that is said already, who they are that do appertain to that account [53.1% **may**; 40.8% **doth**]
- (10b) but most true it is, that this was the cause of the emperors deposing, as **it doth appear** by the assises of pickering and Lancaster. [45.3% **doth**; 44.7% **may**]

The overlap with *do* features a functionally similar construction in clauses with first person subjects and verbs of cognitive perception (11). Just like in (10), *do* and MAY explicitly invite the reader to draw a conclusion based on the evidence provided. MAY can be paraphrased like “it is safe to assume that ...”, while *do* imposes a reading like “I invite you to interpret it as such”. This essentially evidential use is reminiscent of prototypical DO’s reportative/argumentative function discussed above.

- (11a) From which definition **we may clearly gather**, that the cause and fountain of contingency is the free will of man [71.6% **may**; 26.9% **do**]
- (11b) **Hereby do we see** that the errors of the mind do enforce the hatred of the heart. [35.3% **do**; 54.7% **may**]

Another context where DO and MAY overlap is centred around the pattern “*it may/doth come*” and is mainly found in scientific treatises, more specifically in lists with likely causes of a particular condition (12). This links up with the use of prototypical DO in scientific writing, as well as with its habitual readings: in the normal course of events, it can be expected that the condition is caused by the situation described.

- (12) This impediment **doth** come of corrupt gross flume, certain times it **doth** come of caterva<sup>10</sup>, some times of a pleurisy, it **may** come of superabundance of other gross humours. [61.1% doth; 35.7% may]

Interestingly, the habitual readings of DO turn out to overlap with deontic uses of MAY as well as with epistemic ones. For example, ships will only anchor somewhere on a regular basis (13a) if they are allowed to do so (13b). As the permission to perform an action is a prerequisite for the action’s regular performance, contexts with a strong sense of habituality are likely to feature both deontic MAY and habitual DO.

- (13a) Thence to the mouth of Thawan are three miles, wherunto **ships do come at will** [42.5% do; 48% may]

- (13b) About three miles from the town of Bardes, lies a place where **the portingals ships may anchor**. [64.4% may; 34.1% do]

#### 4.2.3 DO and SHALL

As a marker of futurity, the temporal semantics of SHALL differ from those of DO. As such, the two forms overlap mainly in contexts where the temporal mismatch is backgrounded. This is the case, for example, in temporal subclauses (14a–b) where DO has future reference. Alternatively, the mismatch fades in sentences with indefinite future reference (14c–d), where the temporal span denoted by SHALL stretches from now until eternity, which aligns well with the universal/habitual senses of DO.

- (14a) The iii. Swanimote shall be kept in the beginning of xv. days before the feast of S. John Baptist, **when that our gistakers do mete** to hunt our dere. [34.1% do; 63.1% shall]

---

<sup>10</sup> During spelling normalization, “catarue” in the original text was erroneously changed into “caterva” rather than “catarrh”.

- (14b) **until that day shall come**, & these shadows fly away, I will go into the mountain of myrrh, & into the hill of incense. [48.5% shall; 43.1% do]
- (14c) [...] also set before vs the reward of the good, which is, that glory, and that everlasting life, which **the blessed saints do enjoy in the kingdom of heaven**. [44.9% do; 46.7% shall]
- (14d) for Christ saith; **their angels shall always behold** the face of my heavenly father [32.0% shall; 60.8% do]

In addition, DO and SHALL engage in a genre-specific overlap in legal texts (17). In general, SHALL seems to predict some future state of affairs rather than to convey the intentions of the speaker. The use of DO in these contexts adds a layer of truthfulness and objectivity to the prediction, which is rhetorically desirable in the legal genre.

- (17a) And if the said T. N. his heirs, executors or administrators, **do** fail or make default, and do not well and truly acquit, discharge, or save harmless the said T. S. G. F [...] [52.2% do; 44.7% shall]
- (17b) Only excepted, which the said deane and canons and their successors **shall** bear and pay, during the said term. [66.4% shall; 30.8% do]

#### 4.2.4 DO and WILL

Of all modal auxiliaries, WILL shows the most extensive overlap with DO. WILL and DO overlap in first person performatives, often with a discourse-structuring function (15). Even if first person performatives only accounted for a small minority of the prototypical sample of *do*, the pattern is rather salient in the overlap with WILL. Because performatives by definition express an imminent action, they cancel out the temporal mismatch between present tense DO and future WILL.

- (15a) Which here **I do omit** for brevity sake. [33.5% do; 61.0% will]
- (15b) But there be two things, **which I will propound** unto thee to be declared. [51.0% will; 38.9% do]

A second and more extensive source of overlap between WILL and DO consists of universally true statements and (persistent) habits. Just like DO, WILL implies that a situation occurs regularly (and therefore is expected to continue to occur under similar conditions in the future) or that a statement is universally true. The examples in 16a–b, for instance, both express a typical trait of a generic, undefined

group of people. Interestingly, a large share of the overlap between WILL and DO features (usually violent) action taken by almighty subjects (16c–d). This makes sense: the intentions and decisions of almighty subjects typically result in situations that last until eternity, which naturally aligns with the universal semantics of DO. Moreover, almighty subjects have the power to actualize their intentions whenever they want, which decreases the functional difference between WILL, as a marker of intention, and DO, as a marker of actual occurrence.

- (16a) Desires **do** kill the slothful man [46.7% **do**; 42.2% **will**]
- (16b) Malicious wicked persons **will** seldom or never regard good counsel [37.8% **will**; 51.7% **do**]
- (16c) **Almighty God doth severely punish** the wicked, who have persecuted his church or any member thereof [47.8% **doth**; 46.2% **will**]
- (16d) It is to be feared, and that greatly, that **God will very severely punish** vs for our presumption [27.0% **will**; 70.4% **doth**]

Finally, WILL also occurs in scientific writing (19), a context which I described before as typical of affirmative DO.

- (19a) If cold fleume do putrefy in the small veins and arteries which are about the flesh, it **doth** cause a long and languishing ague, which they call a cotidian. [48.5% **doth**; 40.8% **will**]
- (19b) [...] the juice whereof **will** cause the skin to blister: some call it the travellers joy. [32.8% **will**; 63.2% **doth**]

## 5 Discussion and conclusion

### 5.1 Periphrastic DO as a modal auxiliary

The prototypical samples of *do* and *doth* have shown that affirmative declarative DO in the late 16th century functioned as an evidential marker of truthfulness. More importantly, the epistemic uses of affirmative declarative DO display functional similarities with all modal auxiliaries. The universal/habitual use of affirmative declarative DO aligns with CAN when it expresses generic negation, with MAY signalling permission, with SHALL in legal texts and above all with WILL in universal truths and complaints about persistent habits. Its spin-off use in scientific writing is reminiscent of the use of WILL in the same genre, as well as the use of MAY in lists

of causes of scientific phenomena. Its second spin-off, the reportative/argumentative use, links the construction with evidential *MAY*, while its occurrence with first person subjects and verbs of communication echoes the use of first person *WILL* with performatives.

Overall, the results indicate that shortly after the modal auxiliaries had become markedly different from main verbs, *DO*'s slumbering uses as a syntactic auxiliary took off rapidly, especially in contexts that license an epistemic reading of the verb. Although a synchronic snapshot of functional similarity between the modals and affirmative *DO* does not suffice as evidence to claim that analogy has in fact taken place, the suddenness of the developments and the epistemic nature of the eventual construction suggest that its brief 16th century take-off was the result of modal influence.

This does not mean, however, that analogy with the modals was also at play in *DO*'s eventual syntactic regulation. After all, epistemic *DO* did not push through. It had the right syntax, the right semantics and the right overall frequency to fit in with the modals, yet it failed to do so. In fact, the loss of affirmative *DO* seems to suggest that the story of periphrastic *DO* and the modals in the 17th century is one of divergence rather than attraction. Perhaps affirmative *DO* had grown so similar to the modals, especially to *WILL*, that the two forms entered in a competition that pushed *DO* to the fringes of the paradigm, allowing it to preserve its auxiliary syntax but forcing it to let go of its modal semantics. Obviously, this hypothesis remains speculative until a separate study of the interaction between *DO* and the modals throughout the 17th century has been carried out. This study is underway.

## 5.2 Connectionism and analogy

An important objective of my choice of methodology is to showcase the power of connectionist algorithms to model language as a complex, dynamic system. The models do not just measure the immediate influence of individual static collocates on a dynamic construction, they also take into account semantic and syntactic properties of the collocates, enabling a more flexible, robust and holistic approach than, for example, collocational analyses of the slot-and-filler type. The analysis of periphrastic *DO* shows how flexible connectionist models are in terms of feature extraction and semantic modelling by applying them to research on modality, a semantic (mine)field that is notoriously challenging even to human annotators because of the complexity and intricacy of the nuances involved.

One area of linguistic research where connectionist models are particularly promising, both on practical and theoretical grounds, is that of analogy. Practically, connectionist models match well with analogy because of their ability to

pinpoint contexts where several forms could be used in a functionally similar fashion. This allows us to investigate not only the contexts in which a form frequently occurs, but also the contexts in which it could have occurred, but did not. The ability to (re)construct alternative realities stems, again, from the model's flexibility. Tracing functional equivalence beyond the superficial level is only possible with a model capable of generalizing over individual lexemes and picking up on semantic or syntactic tendencies that are expressed, implicitly or explicitly, by a large variety of lexical items and constructions. For applications in historical linguistics, the approach has the additional advantage to be fully bottom-up, which makes the analyses less prone to modern reinterpretations of datasets where native speaker intuitions are a long lost dream.

Importantly, however, connectionist models not only *uncover* analogy, they also radically *implement* it. In her discussion of a neural language model, Fischer (2007: 142–143) argues that the pervasiveness of analogy in linguistic processing ties in nicely with connectionist approaches to language not unlike the model described above, where new input is forced to be analysed in terms of features that have been identified as important in the input the model has received so far. Crucially to such a connectionist approach, the model does not store a new instance by itself, but immediately interprets it in function of how it relates to previous instances. This kind of behaviour makes a connectionist model a powerful operationalization of the analogy-driven language model described in Fischer (2007: 135–145) for two reasons.

First, a connectionist network is a straightforward implementation of how repeated exposure to innovative constructions can lead to structural change. Upon its first encounter with a novel construction, the network will classify the construction in terms of what it has seen before, but it will fail to find a good match. With repeated exposure, the network will recognize the importance of the novel construction in its own right and adjust its weights accordingly, effectively incorporating a prototype of the novel construction in its network configuration.

Second, it naturally implements the idea that changes in the wider grammar can induce constructional change. When the frequency of a certain construction increases, the model will attach more weight to that construction's prototypical features in its language model as a whole, effectively increasing the power of that construction as an analogical example for other input. When the model is presented with an instance of another construction, it will be biased to analyse it more as an instance of the recently-grown-salient construction than as an instance of the construction that originally licensed it. As such, it explains how a (probabilistic) change elsewhere in the grammar may cause a construction to change analogical allegiance and eventually trigger reanalysis (cf. also De Smet 2009; Traugott and Trousdale 2013: 57–59).



### 5.3 Limitations of the approach

While CNNs provide us with a handle to tackle a concept as fleeting as analogy, the approach comes with a number of drawbacks. In the ideal scenario, for instance, the analysis would also have included sentences without an auxiliary. Simple form sentences would have formed a solid baseline for the rise-and-fall of affirmative *do*, but their inclusion proved to be practically infeasible.

As for the limitations of the models more generally, there are two drawbacks to note. First, there is no evidence yet that the scores assigned by the model reflect human intuitions. That kind of evidence would require a comparison of the model's guesses with native speakers judgements in an experimental setting. This is obviously impossible for the present case study, as it deals with historical data. Despite the lack of evidence on the model's cognitive plausibility, however, it is promising that the overlapping uses it retrieved are not merely a collection of isolated cases where it cannot decide on a suitable target form. Instead, the sentences retrieved by the model occur in clusters. Instantiations of the same underlying structure are assigned similar scores. The consistency of the model's judgements is an argument in favour of its reliability as well as of its robustness in tracing cognitive associations. Two competitors are not necessarily cognitively associated if the network has trouble distinguishing between them in an isolated context, but when such confusion happens repeatedly and consistently across many sentences that turn out, upon human inspection, to be structurally or semantically highly similar, it is more likely that the competitors are cognitively entangled.

Second, even if the model's predictions turn out to reflect human intuitions, it remains unclear which degree of overlap two constructions should exhibit in order to become associated in speakers' minds. As is apparent from the tables in Appendix II, CNNs model functional equivalence as a gradient phenomenon rather than a binary choice and it feels arbitrary even to estimate what degree of overlap is sufficient for analogy to kick in. Nevertheless, I do not consider the inability to provide a cut-off point a methodological flaw of this particular model. Instead, I regard the issue as a profoundly linguistic problem that this model draws special attention to, simply because it goes beyond existing models in terms of trying to quantify analogical attraction. Perhaps the question of when similarity becomes meaningful can be addressed by comparing the model's estimates with experimental data, but even then it is not obvious how cognitive entanglement in participants' minds would need to be measured.

All in all, I think the predictive and associative power of connectionist models offer a profoundly different way of looking at data in comparison to descriptive

statistics. With this study, I have tried to illustrate the potential of this methodology for theoretical linguistics, particularly in terms of flexibility and scope. As with any new technique, several aspects of the approach need further scrutiny in order to fully realize its potential. In this particular case, I believe comparison of the model's performance with native speaker judgements (either of individual speakers or a representative sample of speakers collectively) is the avenue for future research that will prove most fruitful.

**Acknowledgements:** I am grateful to the three anonymous referees for their helpful comments and to the Vlaams Supercomputer Centrum for the use of their HPC infrastructure.

## Appendix A: Supplementary Material

Supplementary material to this article can be found online at <https://doi.org/10.1515/cllt-2019-0080>.

## References

- Budts, Sara & Peter Petré. 2020. Putting connections centre stage in diachronic Construction Grammar. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and networks in diachronic construction grammar*, 317–352. Amsterdam: John Benjanmins.
- Baron, Alistair & Paul Rayson. 2008. VARD2: A tool for dealing with spelling variation in historical corpora. In *Paper presented at the Postgraduate Conference in Corpus Linguistics*. Birmingham, UK: Aston University.
- Behrens, Leila. 2005. Genericity from a cross-linguistic perspective. *Linguistics* 43(2). 275–344.
- Collobert, Ronan & Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *ICML25, Helsinki, Finland*. 160–167.
- Dauphin, Yann N, Angela Fan, Michael Auli & David Grangier. 2017. Language modeling with gated convolutional networks. In *ICML34, Sydney, Australia*. 933–941.
- De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119. 1728–1755.
- Denison, David. 1993. *English historical syntax: Verbal constructions*. London & New York: Longman.
- Dubossarsky, Haim, Daphna Weinshall & Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark*. 1136–1145.
- EEBO = *Early English Books Online*. eebo.chadwyck.com.
- Ellegård, Alvar. 1953. *The auxiliary do. The establishment and regulation of its use in English*. Stockholm: Almqvist & Wiksell.
- Fertig, David L. 2013. *Analogy and morphological change*. Edinburgh: Edinburgh University Press.

- Filppula, Markku, Juhani Klemola & Heli Paulasto. 2008. *English and Celtic in contact*. New York: Routledge.
- Fischer, Olga. 2007. *Morphosyntactic change. Functional and formal perspectives*. Oxford: Oxford University Press.
- Garrett, Andrew. 1998. On the origin of auxiliary *do*. *English Language and Linguistics* 2. 283–330.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspectives on ‘alternations’. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Hamilton, William L, Jure Leskovec & Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *AMACL45, Berlin, Germany*. 1489–1501.
- Hilpert, Martin. 2016. Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames* 8(1). 66–85.
- Hilpert, Martin & Florent Perek. 2015. Meaning change in a petri dish: Constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard* 1(1). 339–350.
- Hinton, Geoffrey E. & James A. Anderson. 1981. *Parallel models of associative memory*. Hillsdale: Erlbaum.
- Huddleston, Rodney. 1976. Some theoretical issues in the description of the English verb. *Lingua* 40. 331–383.
- Hudson, Richard. 1997. The rise of auxiliary *do*: Verb-non-raising or category-strengthening?. *Transactions of the Philological Society* 95(1). 41–72.
- Itkonen, Esa. 2005. *Analogy as structure and process: Approaches in linguistics, cognitive psychology and philosophy of science*. Amsterdam: John Benjamins.
- Kauhanen, Henri & George Walkden. 2017. Deriving the constant rate effect. *Natural Language & Linguistic Theory* 36(2). 483–521.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1. 199–244.
- LeCun, Yann, Léon Bottou, Yoshua Bengio & Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *IEEE* 86(11). 2278–2324.
- Lightfoot, David. 1979. *Cambridge studies in linguistics 23: Principles of diachronic syntax*. New York: Cambridge University Press.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. *NIPS* 26. 3111–3119.
- Nevalainen, Terttu. 1991. Motivated archaism: The affirmative *DO* in Early Modern liturgical prose. In Dieter Kastovsky (Ed.), *Historical English syntax*, 303–320. Berlin and New York: Mouton de Gruyter.
- Nurmi, Arja. 1996. Periphrastic *do* and *be+ing*: Interconnected developments?. *Language and Computers* 15. 151–166.
- Nurmi, Arja. 2011. The rise and fall of periphrastic *DO* in Early Modern English, or “Howe the Scotts will declare themselv’s”. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg & C. B. McCully (Eds.), *Generative theory and corpus studies: A dialogue from 10 ICEHL 31*, 373–394.
- Nurmi, Arja. 2018. Periphrastic *do* in eighteenth-century correspondence. Emphasis on no social variation. In Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (Eds.) *Patterns of change in 18th-century English. A sociolinguistic approach*. New York: John Benjamins. 117–135.
- Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas Arévalo, Emma-Louise Silva, William Standing & Odile Aurora Oscar Strik. 2019. Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals’ languages. *ICAME Journal* 43(1). 83–122.

- Plank, Frans. 1984. The modals story retold. *Studies in Language* 8. 305–364.
- Rehurek, Radim & Petr Sojka. 2010. Software framework for topic modelling with large corpora. *LREC 2010*.
- Rissanen, Matti. 1991. Spoken language and the history of do-periphrasis. In Dieter Kastovsky (Ed.) *Historical English syntax*. Berlin and New York: Mouton de Gruyter. 321–342.
- Rumelhart, David & James C. McClelland. 1986a. Parallel distributed processing. Explorations in the microstructure of cognition. In *Foundations*, vol. I, Cambridge: MIT Press.
- Stein, Dieter. 1990. *The semantics of syntactic change. Aspects of the evolution of do in English*. Berlin: Mouton de Gruyter.
- Tahmasebi, Nina, Lars Borin & Adam Jatowt. (2018). Survey of computational approaches to diachronic conceptual change detection. Preprint at arXiv. <https://arxiv.org/abs/1811.06278>.
- Traugott, Elizabeth C. 1972. *A history of English syntax*. New York: Holt, Rinehart & Winston.
- Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Van der Auwera, Johan & Inge Genee. 2002. English do: On the convergence of languages and linguists. *English Language and Linguistics* 6(2). 283–307.
- Vanni, Laurent, Mélanie Ducoffe, Damon Mayaffre, Frédéric Precioso, Dominique Longrée, Veeresh Elango, Nazly Santos Buitrago, Juan Gonzales Huesca, Luis Galdo & Carlos Aguilar. 2018. Text Deconvolution Saliency (TDS): A deep tool box for linguistic analysis. In *AMACL56, Melbourne, Australia*.
- Vulanovic, Relja. 2005. The rise and fall of periphrastic do in affirmative declaratives: A grammar efficiency model. *Journal of Quantitative Linguistics* 12(1). 1–31.
- Warner, Anthony R. 1993. *English auxiliaries. Structure and history*. Cambridge: Cambridge University Press.