

What makes the past perfect and the future progressive?

Romain, Laurence; Ez-zizi, Adnane; Milin, Petar; Divjak, Dagmar

DOI:

[10.1515/cog-2021-0006](https://doi.org/10.1515/cog-2021-0006)

License:

None: All rights reserved

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Romain, L, Ez-zizi, A, Milin, P & Divjak, D 2022, 'What makes the past perfect and the future progressive? Experiential coordinates for a learnable, context-based model of tense and aspect', *Cognitive Linguistics*, vol. 33, no. 2, pp. 251-289. <https://doi.org/10.1515/cog-2021-0006>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

The final publication is available at www.degruyter.com. <https://doi.org/10.1515/cog-2021-0006>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Laurence Romain*, Adnane Ez-zizi*, Petar Milin and
Dagmar Divjak

What makes the past perfect and the future progressive? Experiential coordinates for a learnable, context-based model of tense and aspect

<https://doi.org/10.1515/cog-2021-0006>

Received January 11, 2021; accepted November 3, 2021; published online January 31, 2022

Abstract: We examined how language supports the expression of temporality within sentence boundaries in English, which has a rich inventory of grammatical means to express temporality. Using a computational model that mimics how humans learn from exposure we explored what the use of different tense and aspect (TA) combinations reveals about the interaction between our experience of time and the cognitive demands that talking about time puts on the language user. Our model was trained on *n*-grams extracted from the BNC to select the TA combination that fits the context best. It revealed the existence of two different sub-systems within the set of TA combinations, a “simplex” one that is supported lexically and is easy to learn, and a “complex” one that is supported contextually and is hard to learn. The finding that some TA combinations are essentially lexical in nature necessitates a rethink of tense and aspect as grammatical categories that form the axes of the temporal system. We argue that the system of temporal reference may be more fruitfully thought of as

Laurence Romain and Adnane Ez-zizi: Joint first authors.

***Corresponding authors: Laurence Romain**, Department of Modern Languages, University of Birmingham, Birmingham, UK, E-mail: oominds@oominds.org; and **Adnane Ez-zizi**, Department of Modern Languages, University of Birmingham, Birmingham, UK; and School of Engineering, Arts, Science and Technology, University of Suffolk, Ipswich, UK, E-mail: oominds@oominds.org

Petar Milin, Department of Modern Languages, University of Birmingham, Birmingham, UK.
<https://orcid.org/0000-0001-9708-7031>

Dagmar Divjak, Department of Modern Languages, University of Birmingham, Birmingham, UK; and Department of English Language & Linguistics, University of Birmingham, Birmingham, UK.
<https://orcid.org/0000-0001-6825-8508>

the result of learning a system that is steeped in experience and organised along a number of functional principles.

Keywords: cognitive grammar; computational modelling; naïve discriminative learning; tense and aspect; usage-based approaches to language

1 Introduction

In a usage-based approach to language, languages and our knowledge of them are considered a product of our interaction with the world. Three dimensions that play a crucial role in this process are the commonality of the human experience, our shared cognitive abilities and our social nature. Languages are designed to facilitate social interaction and are learned in social interaction: language knowledge emerges from exposure to usage during such interactions. The cognitive capacities that enable us to gradually build up knowledge of the communicative code are the same as those that support other, non-linguistic, areas of cognition. They are also those that constrain the forms languages can take: there is surprising unity in the diversity of formal expressions exhibited by the languages of the world (WALS Online 2013). This is in no small part due to the commonality of the human experience, which is processed with the same cognitive system.

In this paper, we look into those categories that are used to classify the temporal dimensions of our experience, as expressed in language. Expressions of temporality are among the most challenging to describe and even closely related languages such as Germanic and Slavic display striking differences in how temporality is expressed linguistically (Binnick 1991; Comrie 1985; Dahl 1985). Using English as our basis, we set out to explore what the use of different tense and aspect combinations reveals about the interaction between our experience of time, and the cognitive demands that talking about time puts on the language user. Rather than taking a behavioural or neurological approach, we model computationally how language supports the expression of temporal relations. Focusing on the raw linguistic input language users receive, we aim to establish how the expression of the English tense/aspect (TA) combinations differs in the linguistic support they each require, and what these differences tell us about the experiential basis and the cognitive constraints on the linguistic expression of time.

In Section 1, we will summarily review the mechanisms that cognitive grammarians have proposed for conceptualising tense and aspect. We will dwell on

those that are likely to pose a challenge for machine learning as they assign a key role to the conceptualiser of the situation. In Section 2, we will explain our psychologically plausible computational learning algorithm which includes n -grams as cues and is thus representative of how actual language learning is thought to unfold on a usage-based account. The results, including information on what guides the choice of TA combinations, are presented in Section 3 and compared to the findings from a short survey designed to compare our model's results with speakers' choices. In Section 4, we put our findings in a broader context and explore those dimensions of the interaction between language, experience and cognition that play a key role in maintaining the English TA system as we know it.

1.1 Tense, aspect and perspective

It is generally assumed that tense is used to locate a situation in time, relative to the time of speaking (Comrie 1985). And prototypically, in default or canonical viewing arrangements (Langacker 2001), the past tense would be used to locate events before the moment of speaking, the present at the time of speaking, and the future after the time of speaking. There is, however, discussion as to whether the present actually does locate an event at the time of speaking (Langacker 2008, 2011), as we will discuss in Section 1.2.

Also, not everyone accepts that the future is a tense in English because, in English, the future is not morphologically marked in the same way as the past and present are. While past and present attach inflections directly to the verb base, the future requires the use of the (modal) auxiliary *will*. The future is thus a “periphrastic construction”, and Michaelis (2006: 224), for instance, distinguishes these from “actual tenses” where the (main) verb bears the actual tense marking. Furthermore, beyond the aforementioned formal differences, and more central to the *cognitive plane*, another crucial distinction between the past and present on the one hand and the future on the other is their relation to reality. As illustrated by Langacker's (1991: 244) Elaborated Epistemic Model, the present is closely associated with the Immediate Reality of the speaker, the past with their Known Reality and the future with a possible Non-Reality (cf. Brisard 2013). Therefore, while past and present are associated with reality, the future is not because it has not been experienced.

This Elaborated Epistemic Model leads us into a discussion of aspect and of how it interacts with tense to form combinations whose semantics, Brisard (2013:

220) argues, indicate the reality status of a situation. Most researchers agree that tense and aspect are conceptually separable (Comrie 1985; Michaelis 2006, *inter alia*), but it is also clear that they are closely related and intertwined (Hornstein 1991: 9). Despite the ability of aspect to occur on its own (e.g., *I saw the kids playing in the park*), this option hasn't received much coverage in the literature.

Aspect is traditionally defined as “different ways of viewing the internal temporal constituency of a situation” (Comrie 1976: 3); that is, how the event relates to time and how that relation is described. The two prototypical markers of aspect in English, the perfect and the progressive, each have their characteristics with the progressive denoting an event in progress and the perfect an event that is completed. However, it has been argued that the choice of one aspectual form over another is not necessarily fully given by the situation that is being rendered. The choice of aspect can be considered as an instance of construal (see Croft [2012] for the theoretical elaboration and Janda and Reynolds [2019] for an application), which “refers to our manifest ability to conceive and portray the same situation in alternate ways” (Langacker 2008: 43). Furthermore, it is useful to hypothesise an underlying *contradistinction* between the freedom of linguistic expressiveness and situational (including perceptual) constraints: whenever the linguistic packaging of an event is not strongly constrained by the event itself, the linguistic choices we make reflect our freedom to conceptualise the event we communicate. In this regard, for example, Croft (2012: 40) compares the different construals of *She coughed (once)* where the use of the past simple leads to the construal of the event as a cyclic achievement and *She coughed for five minutes/She was coughing* where the use of either the past simple with the duration adverbial or the past progressive leads to the construal of the event as an activity.

Also relevant to construal is the notion of scope, paraphrased by Langacker (2001: 253) as the ‘onstage region’. Maximal scope is roughly the entire situation, and the immediate scope is the part of the situation that is relevant for a particular purpose. Thus, a speaker will contextualise or view a situation relative to the immediate scope (Boogaart and Janssen 2007: 253; Kermer 2016; Langacker 2001; Niemeier and Reif 2008), i.e., the general locus of viewing attention. That is, a speaker's choice of a TA combination indicates the angle they choose for describing the event. Take the cab-driver example, borrowed from Michaelis (2006: 221): *I took a cab back to the hotel. The cab driver was Latvian*. Here, we know that the cab driver did not stop being Latvian after the speaker got out of the cab or that the driver is now dead, which would be the usual interpretation of a past simple. Rather, the past tense is used because the situation/event is no longer part of the

speaker's immediate scope. This sentence is an example of a non-default use of the past simple, or a non-canonical viewing arrangement (Langacker 2001), and is thus more cognitively complex, which explains why it requires conceptual back-up. Note that, here, the required contextual back-up is provided outside of utterance boundaries. Without the appropriate (wider) context, a speaker would have most likely chosen the present simple to describe the situation.

The past simple is not the only TA combination that can be used in a non-canonical viewing arrangement: the present forms are particularly prone to this. According to Brisard (2013), the present simple has a "strictly perfective character" (Brisard 2013: 227). In fact, Brisard (2013: 227) uses this 'strictly perfective' feature of the present simple and the fact that it is used to describe a "structural aspect of reality" to explain the uses of the present simple in non-canonical viewing arrangements. As is well-known, a present simple can be used to denote a situation that is expected to occur in the future (e.g., *Our train leaves at 4*). Brisard explains this futurate use of the present simple (with an action verb) through the fact that the event is accepted as part of the speaker's reality (Brisard 2013). However, as mentioned by Bergs (2010: 218), such uses of the present require a temporal adverbial to be interpreted as a future, in our case here, the prepositional phrase *at 4*. Given the proper contextual background, which can be linguistic (and possibly fall outside the sentence) or extralinguistic, as Langacker (2001) points out, the present simple can be used for non-default situations. The need for contextual elaboration also holds for non-default uses of the present progressive, as in *I am having lunch with Lina tomorrow* where the adverb *tomorrow* points towards a future interpretation.

English TA combinations are, in other words, used to refer to a wider range of temporal situations than their name might lead one to expect: while it might seem efficient to reserve specific TA combinations for specific event types, such a crisp division is not actually implemented in usage. The nuances in the various uses of each TA combination enable speakers to present situations in the way they conceptualise them, with relatively few formal options; at the same time, this flexibility makes it very hard for a machine to pick up the intended meaning. These conceptualisations can be addressed with reference to conceptual operations such as viewing arrangements and construal. It should be noted, however, that the flexibility in the use of TAs to convey the conceptualisation of a situation relies on the felicitous use of context: without adequate contextual support, a non-default arrangement risks being misinterpreted. Based on this observation, we will argue that the more complex the TA arrangement, and the more intricate its conceptual relationship to the situation it renders, the rarer it will be and the more contextual support it will require. And contextual elaboration is something a machine can pick up.

1.2 Experience: complexity, distance and uncertainty

A major functional principle of linguistic organisation is complexity, which comes in many forms. Rohdenburg (1996: 151) presented the complexity principle, based on Hawkins's (1992) analysis of phrasal verbs: "In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favoured in more complex environments." Rohdenburg finds that the more complex the sentence (formally and/or cognitively), the more explicit it is, i.e., the more syntactic elements are present to guide the interpretation. This is in line with research from psycholinguistics. Studies that have looked at the processing of subordinate clauses, for example, have found that the marker *that* is more likely to be used as the proposition becomes more complex (Hale 2001, 2003; Jaeger 2010; Levy 2008).

We expect this relation to hold true for TA combinations as well: the different TA combinations embody differences in the conceptual complexity of the situation described, with simpler forms reserved for describing cognitively simpler situations. However, we expect that differences in formal complexity extend beyond the verb, into the context, and affect the amount of contextual support needed to achieve an accurate representation of the intended conceptualisation of the event. This relates to our argument about the language-situation *contradistinction*, in which a language's expressive power for conceptualising *any* given situation will show flexibility to ensure accommodation, not only of the specific linguistic complexity but of the broad complexity of the actual (e.g., physical) situation.

Michaelis (2006: 224) points out a difference between simpler, more direct forms and complex, more indirect TA combinations. Compare here three different expressions of pastness, the past simple *skated*, the past perfect *had skated* and the past perfect progressive *had been skating*. Arguably, the past simple *skated* is simpler both in terms of the form of the verb and in terms of the relation it expresses between the time of the situation (Event Time) that is being described and the time of speaking (Speech Time), which are distinct (Reichenbach 1947). In the past simple, Reference Time (the time which corresponds to the viewpoint chosen by the speaker) coincides with Event Time and is thus also situated in the past and hence detached from the present. The past perfect *Tonya had skated* represents an intermediate level of formal and conceptual complexity, as it requires the auxiliary *had* and is used when Event Time is anterior to Reference time which is anterior to Speech Time. The past perfect progressive *had been skating*, represents the most complex TA combination of the three. It combines complexity in the form itself with complexity in the conceptualisation of the situation. We see how the various elements that *had been skating* is made up of guide the hearer through the conceptualisation offered by the speaker, with *had* indicating that the event is over by

Speech Time and Reference Time and the progressive *-ing* indicating an event that extended over a period of time.

In determining cognitive complexity, epistemic distance, i.e., the location of an event relative to a speaker's sphere of knowledge (cf. Langacker 1987), plays an important role, not in the least because epistemic distance affects the certainty or uncertainty with which the situation can be rendered (for a discussion of epistemic distance with regards to tense and aspect see, for example, Stanojević 2011 for Croatian). Epistemic distance has been mainly applied to tense. Langacker (2011), for example, describes the schematic characterisation of the present in English in terms of epistemic immediacy: the present is epistemically immediate whereas the past is non-immediate. Chilton (2013) argues that things or situations that are spatially close are also epistemically close (see also Bender and Beller 2014 for a discussion of the mapping of spatial frames onto time). We apply this notion of epistemic distance/immediacy together with temporal distance to TA combinations and explain how epistemic distance increases cognitive complexity and hence the need for context. As for the categories we have described as aspect—namely simple, progressive and, perfect—the notion of immediacy applies in a slightly different way. The simple could be considered as neutral in terms of epistemic distance as it is taken to be the default and does not add much to event description, which will be dependent on the tense and type of verb (action vs. state). With a progressive aspect, the event can be considered more immediate or epistemically close in that it is in the speaker's immediate scope (Langacker 2011: 57) and thus seen from an internal point of view. De Wit and Brisard (2014) consider the present progressive to be *epistemically contingent*, i.e., it is used in non-structural situations “whose actualisation at the time of speaking could not have been predicted” (De Wit and Brisard 2014: 88), as opposed to the simple form which they consider to be part of the speaker's structural reality. As such, the progressive is somewhere between the simple and the perfect in terms of immediacy: it is part of the speaker's immediate scope but is unbounded. In the perfect, the event is over at the time of Reference (except with stative verbs such as *live*) and is anterior to Speech Time. Radden and Dirven (2007), who consider the present perfect as a complex tense, argue that it is used for an anterior event or an anterior phase of an event with a backward-looking stance from the present (Radden and Dirven 2007: 212). As such, the perfect is also somewhat hybrid in its immediacy as it combines the anteriority of an event and a present or more immediate viewpoint.

Overall, we can say that only what has been experienced is certain, and that only that which is occurring at the time of speech is close. As such, the past is distant but certain, the present is close and can be certain (with a simple form) or somewhat uncertain (with the progressive, cf. De Wit and Brisard 2014) and the future is both distant and uncertain (cf. also Janda 2015 for a similar statement). We

should note here that it is the combination of a tense and an aspect that amounts to various levels of complexity, not just tense or aspect on their own. The present and past simple are simpler by virtue of their denoting a structural reality, implying that the event described is considered in its entirety (cf. De Wit and Brisard 2014) and can stand on its own. A progressive, on the other hand, denotes an event that is viewed internally, i.e., the event is seen as ongoing by the speaker, which may thus require contextual backup that anchors the relationship between speaker and event. With the perfect, the event denoted is anterior to Speech Time but considered with continuity to Reference Time; we expect this particular (rather complex) viewing arrangement to leave a trace in the context as well. As to the future, it is also epistemically more distant, in the sense that while the past and present are relatively fixed, and have been or are being experienced by the speaker, the future is open and thus less certain (cf. de Brabanter et al. 2014).

In sum, we would expect morphologically and conceptually simpler forms to require less supporting information whereas morphologically and conceptually more complex forms are likely to require more contextual support, as illustrated in Figure 1. Note that these two scales can conflict or coincide. A combination of future and simple may well require more contextual support than that of a present and a

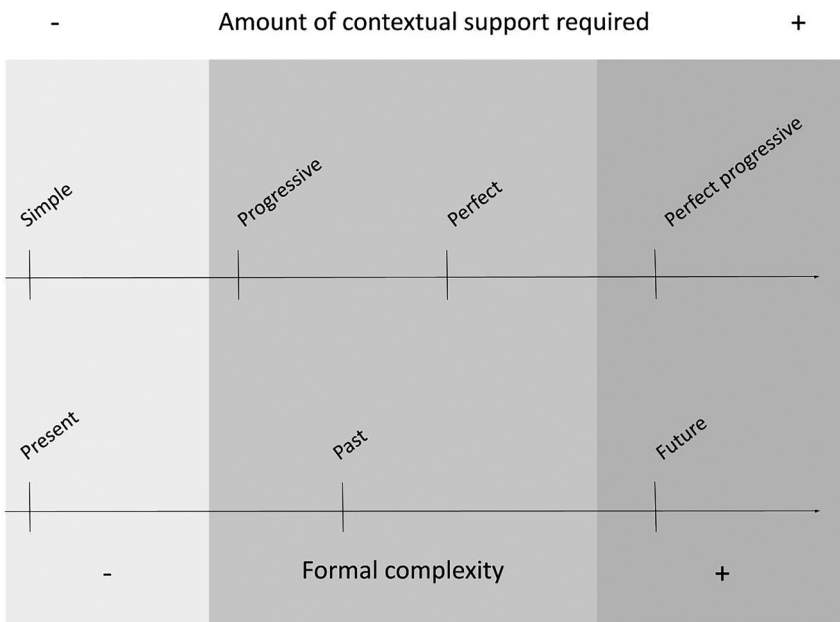


Figure 1: A projection of the increasing amount of contextual support needed for various TA combinations (little contextual support on the left, more contextual support on the right).

perfect, for example. We use three blocks of different shades of grey to mark the three levels of morphological complexity for grammatical aspect and tense marking.

In the next section, we will delve deeper into principles of learning from context as they apply to learning TA categories from exposure to language.

1.3 Support: the role of frequency and context in the learnability of tense and aspect

Recently, the concern has been raised that the categories with which (cognitive) linguists operate may not be those that drive the language cognition of the average user (Dąbrowska 2016: 221; Divjak 2015; Divjak et al. 2016b). Although descriptively accurate and often didactically relevant, many of the core concepts and categories on which cognitive linguists rely require exquisite sensitivity to minute differences in expression that may not be shared by the average, untrained language user. To take an example from the description of TA combinations in English, concepts such as (un)boundedness or viewing arrangements and intricate distinctions between achievements and accomplishments could not be expected to be described by naïve speakers, let alone in such terms. Research in this area remains sparse, with most theoretical and descriptive categories lacking empirical verification of their cognitive relevance (but see Rissman and Majid 2019 for a review of the empirical evidence regarding the thematic roles Agent, Patient, Goal, Recipient and Instrument, for example).

Our ‘litmus-test’ of cognitive reality, or a category’s potential for cognitive reality, is learnability: if language knowledge emerges from exposure to usage, any knowledge we assume exists in the minds of language users should, at the very least, be learnable from input (Divjak et al. 2015; Milin et al. 2016, 2017b). Whereas earlier work explored whether abstract labels can be learned and whether they map onto distributional patterns in the input (cf., Divjak et al. 2015), more recent work has brought these strands together by modelling directly how selected abstract labels would be learned from input using a cognitively realistic learning algorithm (cf. Divjak and Milin 2020; Milin et al. 2017b). Following these foundational assumptions, we proceed to explore to what extent the use of the different English TA categories is learnable from (sentential) context (see Bielak and Pawlak 2011; Kravchenko 2012 for similar attempts to teach TA combinations from a cognitive perspective) and which (sentential) cues would trigger the use of a specific TA combination.

To achieve this, we rely on insights from learning. Cognitive linguistics is predicated upon the premise that languages are dynamic systems shaped by usage in a process that is mediated by general cognitive abilities and functional considerations. The general cognitive abilities that have received most attention to

date are classification, abstraction and imagination (metaphor, metonymy). Processes or functions that would enable ‘growing’ a system from use have, however, been conspicuously absent from usage-based accounts. In fact, learning constitutes our very own “elephant in the room”. Learning was exiled vigorously from the linguistic landscape by Chomsky’s (1959) criticism of Skinner (1957) and although recent years have seen a resurgence in the interest in learning (Baayen et al. 2011; Ellis 2006a, 2006b; Ramsar and Yarlett 2007; Ramsar et al. 2010), learning is still to make a full comeback onto the linguistic scene (Divjak 2019: 6).

However, the distributional patterns in the input are abundant and “growing” a system or structure from use requires focusing on the relevant subset, i.e., filtering out the important information. To achieve this, we rely on the basic principle of error-correction learning. Error-correction learning is biologically realistic (cf., Chen et al. 2007; Trimmer et al. 2012) and has previously been used in conjunction with a cognitive linguistic framework (Milin et al. 2016). We rely on the Rescorla-Wagner (1972) learning rule, and its implementation as Naïve Discriminative Learning (NDL; cf., Baayen et al. 2011; Milin et al. 2017a etc.; see Section 2.3. for details). The Rescorla-Wagner rule defines the change in strength of association (weight) between a cue (for example, a word or a phrase) and an outcome (which can be a word, or a linguistic abstraction, an experiential notion, and so on). The change in association can be driven by positive evidence, when both cue and outcome are present, or by error-correction, when a cue is present, but the outcome is absent (cf., Pearce and Hall 1980). Thus, the model presents the cues and outcomes as discrete units, that can be either present or absent; the change in association weights also happens in discrete time steps. This is conveniently intuitive from a usage-based perspective, as language corpora ‘naturally’ offer such discrete cue or outcome units (letter pairs or triplets, words and phrases, annotations etc.), in discrete time steps (as we crawl through consecutive word sequences, phrases, clauses, sentences).

Using this approach, we aim to explore how much of what language users know about tense and aspect can be explained by cue-outcome associations between information that is available in textual input. However, we do not mean to imply that detecting associations between formal cues and outcomes is all there is to language learning. Language learners have access to other information than the formal information available in the speech signal, e.g., information gleaned from the wider experiential context. The rich, conceptual structure that language users are assumed to have access to is unlikely to be reducible to learning associations between formal cues and outcomes in textual input alone, but such associations may make a bigger contribution than is currently acknowledged. At the same time, higher-level knowledge, too, may be learned via associative mechanisms: recent behavioural research demonstrates that associative learning is a major force guiding the development of complex human behaviour, including decision

making and high-level cognitive functions (for an overview see Heyes 2012, 2016). The associative process is not as simple or limited as it is often presented to be (cf., Rescorla 1988; also see Tolman 1932 for similar points made earlier, and Hanus 2016 for more recent discussions).

Of particular importance for this study are the characteristics of the dynamics of learning by association. First, the strengthening or weakening of the cue-outcome association does not happen in isolation: as the number of cues increases, positive evidence will become less important, while error-correction ‘steals the show’ (hence the name ‘error-correction learning’). Therefore, any outcome can be strongly associated with (or well predicted by) no cue at all, one or a handful of cues, or a large(*r*) number of cues. If an outcome’s *contingency* on one particular cue is such that the two frequently and exclusively occur together, then their association weight will likely become strong and positive. In the case of many cues that co-occur more or less systematically with a given outcome, the possible associations can become rather complicated. This is typical of natural languages, which makes them an exciting challenge for learning theory.

For one, whether these *sets of cues* are presented to the learning system as combinations of *elements* or as *configurations* is still an important point of debate within learning theory (see, for example, Ghirlanda 2015; Kokkola et al. 2019; Pearce 1994; Rescorla 1997; Wagner 2003). More broadly, however, multiple cues (or their configurations) give rise to a multitude of *imperfect contingencies*, which are recognised as crucial for learning and behaviour (Tolman and Brunswick 1935; Wasserman et al. 1993). The point here is that cues and outcomes often do not or cannot connect in “simple one-one, univocal (*eindeutig*) fashion [...] and [any] one type of [cue] is found to be causally connected with differing frequencies with more than one kind of [outcome] and vice-versa” (Tolman and Brunswick 1935: 44). This facilitates the rich dynamics of learning, by the learning system and for the sake of its successful interaction with the environment (see Rescorla 1988).

Thus, because TAs differ both in their frequency of occurrence and in their grammatical complexity, this will define the complexity of the cue-outcome connections – never perfect contingencies. And from this we can draw the hypotheses that these two factors will affect learning: (i) the less frequent the outcome, the more unlearning it will induce, which will be reflected in weak or even negative associations with relevant cues; (ii) the more complex the contingencies that the outcome has, the less important (or informative) the individual cue will be.

2 Methods

To establish to what extent TA categories relate to experience as captured in usage, we trained a learning algorithm on a large dataset. After an overview of the English

Table 1: The 12 tense-aspect combinations of English investigated in this paper.

	Past	Present	Future
Simple	<i>Tonya skated.</i>	<i>Tonya skates.</i>	<i>Tonya will skate.</i>
Perfect	<i>Tonya had skated.</i>	<i>Tonya has skated.</i>	<i>Tonya will have skated.</i>
Progressive	<i>Tonya was skating.</i>	<i>Tonya is skating.</i>	<i>Tonya will be skating.</i>
Perfect progressive	<i>Tonya had been skating.</i>	<i>Tonya has been skating.</i>	<i>Tonya will have been skating.</i>

TA system and the various forms it comprises (Section 2.1), we present our data and the method used for the automatic annotation of TA combinations (Section 2.2). Finally, we present our model, NDL, and discuss how we extended it to include n -grams as cues (Section 2.3).

2.1 The expression of tense and aspect in English

For the purposes of this study, we consider that there are 12 possible TA combinations in English, as illustrated in Table 1 (despite the caveats mentioned in Section 1.1).

Note that, for the purposes of computationally modelling unannotated data, we are restricted to formally expressed properties. For this reason, we use the term *aspect* to refer to grammatical aspect, as marked by *have* + past participle and *be* -*ing*, a combination of both or the simple forms, to the exclusion of lexical aspect (Bach 1986; Croft 2012; Dowty 1979; Rothstein 2004; Verkuyl 1993). We remain theoretically agnostic as to whether the future should be considered a tense of English and consider the futurity marker *will* as an indication of futurity.

2.2 Data

For training, we relied on data from the British National Corpus (BNC), which includes approximately 100 million words carefully sampled across functional styles, including both spoken and written materials (Leech 1992). The choice of the BNC was motivated by the fact that it is currently the most exhaustive and balanced corpus of British English available. The BNC is made up of 90% written data and 10% spoken data, drawn from a total of 4,049 different texts, in 40,000 word increments. The written texts are drawn from a wide variety of sources, while the spoken data are transcribed speech recorded in formal and informal settings (<http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=numbers>).

The BNC is annotated for part of speech only, and thus tenses are not identified. Given the limited number of verb tags in the BNC, we developed additional

heuristics (55 in total) to improve the identification of verb forms in a sentence. Our approach was highly accurate (96.2%) at detecting TA categories as shown by a comparison with manual tense annotation (by the first-listed author) on a random sample of 1,000 sentences that contain a single verb. We rely on the `ndltenses` Python package by Kwakpovwe (2021) to annotate the data for tense and aspect.

After extracting all sentences from the BNC that contain at least one verb tag, we automatically labelled all verb forms in each sentence. We removed all sentences that contained fewer than three or more than 60 words (i.e., sentences below the first percentile or above the 99th percentile), or had a main verb in the imperative or modified by a modal (other than *will*, as mentioned above); this allowed us to focus on the 12 TA combinations. We automatically identified the lemmas of the verb forms using Python's Natural Language Toolkit (Bird et al. 2009). We manually checked that verbs with a frequency higher than 10 were correctly lemmatised and corrected them wherever necessary, as these lemmas were used later to train our computational model. Words in American English spelling were automatically converted to British English spelling. All numbers, punctuation symbols and words that contain non-alphabetical characters were removed from the sentences.

A sample of the final data set is shown in Table 2. In the table, some of the sentences are duplicated because we created as many instances of a sentence as there are verbs in the sentence as our goal is to predict a verb's TA combination from its surrounding context.

The distribution of tenses in our final data set is clearly Zipfian, as shown in Table 3. The bulk of the data is made up of the present simple (46.09%) and the past

Table 2: Illustration of the format of the final data set. Sentences were extracted from the BNC corpus along with variables that characterise the verbs within those sentences.

Sentence	Tense	Verb Form	Main Verb	Lemma
the difficulties which they met were formidable and as much political as scientific	past simple	met	met	meet
the difficulties which they met were formidable and as much political as scientific	past simple	were	were	be
no but ian will not have said she will frighten me	future perfect	will have said	said	say
no but ian will not have said she will frighten me	future simple	will frighten	frighten	frighten
she gave a tremulous smile	past simple	gave	gave	give

Table 3: Frequency of tense-aspect labels in our sample.

TA combination	Frequency	Percentage (%)
Present simple	3,229,514	46.09
Past simple	2,636,030	37.62
Present perfect	338,791	4.84
Future simple	271,345	3.87
Past perfect	253,110	3.61
Present progressive	139,878	2
Past progressive	108,682	1.55
Present perfect progressive	11,278	0.16
Future progressive	8,021	0.11
Past perfect progressive	7,032	0.10
Future perfect	2,955	0.04
Future perfect progressive	30	<0.01
Total	7,006,666	100

simple (37.62%), with the remaining 10 TA combinations making up between 4.84% and less than 0.01% of the sample each. Note that we had to exclude all instances of the future perfect progressive for training as there were only 30 cases in total.

Interestingly, the proportions we report closely resemble those reported in other sources. Kramsky (1969) took different samples of three different styles of English—novels, plays (as proxies for spoken language), and specialized (academic and technical) texts. He analysed 20,000-word samples from each text and found that the simple present accounts for more than half of the verbs used in English speech; that the simple tenses are the top three verb tenses; and that the five most commonly used verb tenses (in his data, the three simple ones plus the present perfect and the present progressive) total over 95% of usage. The genres differ according to the tense they prefer: spoken language and specialised texts rely overwhelmingly on the present tense, but fiction prefers the past tense in over half of all cases. The same proportions are reported by Biber et al. (1999: 456) on the basis of the 40-million word Longman Spoken and Written English (LSWE) corpus which contains five million word samples of the four core registers (conversation, fiction, news, academic prose): overall, present tense is somewhat more common than past tense but the distribution of past and present differs considerably across registers with conversations and academic prose strongly preferring the present tense, fiction preferring the past tense and news using both tenses to a similar extent.

It is generally believed that genres also differ according to aspect, and the progressive aspect is typically thought to be more common than the simple aspect, especially in spoken language (Biber 2006). However, Kramsky (1969) reports that

the simple aspect makes up over 85% of usage, across all three genres, followed by the perfect, the progressive and finally the perfect progressive; our data shows the exact same trend. Biber et al. (1999: 461) report that the simple aspect is “overwhelmingly the most common” across all four registers, accounting for about 90% of all verbs; the perfect makes up between 5 and 10% in all registers, while the progressive is slightly less common than the perfect; the perfect progressive was omitted from their study because it occurred in fewer than 0.5% of all cases. Again, this is very similar to what we see in our data. Römer (2005: 32–33) reviews existing literature that puts the frequency of the progressive at around 5%; in our sample it is just under 4%. Biber et al. (1999: 462–463) do report that the progressive is much more common in American English conversation than in British English conversation, occurring at a ratio of 4:3. However, it would be wrong to think that the progressive is more common than the simple. Biber (2006) shows, against a sample of real data, that the progressive is not the unmarked form in conversation: the simple is the most frequent form in conversation. However, it is indeed in conversation that Biber (2006) finds the progressive to be the most frequent as opposed to other registers.

2.3 Computational modelling

2.3.1 Naïve discriminative learning

Naïve Discriminative Learning (NDL; Baayen et al. 2011) is an adaptation of the Rescorla–Wagner learning rule (Rescorla and Wagner 1972) for modelling language learning and language comprehension in particular. NDL has been used in many studies on language (for details see, e.g., Baayen et al. 2011; Chuang and Harald Baayen 2021; Milin et al. 2017a; Pirrelli et al. 2020; for integration in usage-based approaches see Milin et al. 2017b; Divjak et al. 2021).

An NDL network basically divides the language inputs into cues and outcomes. Outcomes are those stimuli whose occurrence needs to be learned from exposure to input. In language learning, outcomes are typically form units. Different from previous studies, here, outcomes are the verb TA labels, e.g., present perfect, future simple, etc. (see Table 3). Cues are stimuli that are used to predict the occurrence of the outcomes and they can be form units or abstract units (cf. Milin et al. 2017b). The cues used in this study are word n -grams rather than individual words or sub-lexical units (letter pairs and triplets). For example, the 2-g from the sentence ‘John is currently learning Russian’ are ‘john#is’, ‘is#currently’, ‘currently#learning’ and ‘learning#russian’, while the 3-g for this sentence are ‘john#is#currently’, ‘is#currently#learning’ and ‘currently#learning#russian’.

Given that we are interested in learning which TA combination is appropriate in a particular sentential context, we lemmatised all verb forms to force the prediction of TA combination. We consider all n -gram cues with n ranging from 1 to 4 to be able to identify the various adverbials that may be distinctive for each TA combination. Therefore, the sentence *John is currently learning Russian* has as cues ‘john’, ‘currently’, ‘russian’, ‘john#currently’, ‘currently#russian’, ‘john#currently#russian’, ‘LEARN’ (note here that there are no 4-g cues since there are only three words in the context, excluding forms belonging to the verb) and as outcome ‘present progressive’. As such, our method combines the lexical focus of a collocation analysis (Stefanowitsch and Gries 2003) with the contextual scope of traditional corpus linguistic approaches (Anthony 2013), making it possible – for the first time – that both foci are combined and applied on a large scale.

Each instance in which cues and outcomes co-occur constitutes a learning event. In our case, a learning event is one sentence, and there are as many learning events as there are verbs in a sentence. Note that the input is realistic at the sentence level, but sentences are considered individually in the input. That is, sentences are presented to the model one by one, and did not connect into meaningful sequences. Furthermore, as per the design of the BNC, no more than 40,000 words were extracted from any one source (Leech 1992) and hence the texts that make up the corpus do not form a coherent narrative. This approach, although standard in computational modelling, of course, impoverishes the input the model receives, and hence what can be learned from that input which may affect different TA combinations differently; we will return to this point in the Discussion.

The fourth component of the model is the set of association weights between cues and outcomes. An association weight is a measure that encodes the tendency of an outcome (here a TA label) to be triggered by the presence of a particular cue. The collection of these weights represents the network of linguistic knowledge specific to the task at hand, as is represented in Figure 2, where ‘ $w(c_i, o_j)$ ’ is the measure of association between a cue (c) and an outcome (o). To choose an outcome (i.e., a TA label) in a given learning event, the model combines the association weights of all cues that appear in the event and calculates the activation of each possible outcome—a function that measures how likely each outcome is to be chosen by the learner given a certain set of cues. The most activated outcome in the network is selected by the model.

Note that, even though the input NDL received is not what a child would receive, this is not necessarily a problem for our account. Overall, the proportions of TA combinations in language are stable and hence not severely affected by variations in the type of input the learner receives. Additionally, we learn throughout our lifespan and hence keep on recalibrating our association weights. Therefore, our model would be representative of any adult with experience of a reasonable variety of genres.

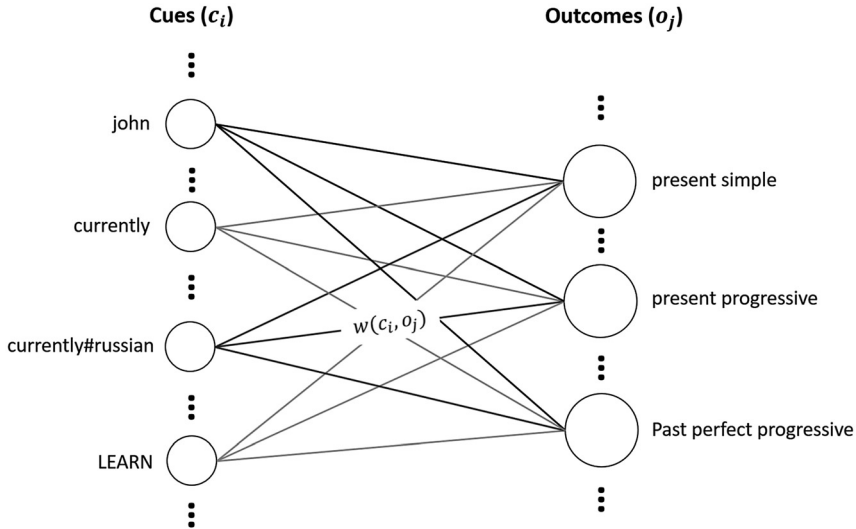


Figure 2: Illustration of the NDL network. The linguistic knowledge is encoded in the set of association weights between the cues and outcomes.

2.3.2 Model implementation

We trained NDL utilising the `ndltenses` Python package (Kwakpovwe 2021), which inherits the functionality of the `pyndl` library, the Python 3.x implementation (Sering et al. 2017) of the Naive Discrimination Learning computational framework (Baayen et al. 2011). More specifically, we used the 10,000 most frequent n -grams from each n level, i.e., one, two, three and four word strings (40,000 out of 99,152,531 possible n -grams). For the verb lemmas, we retained those with a frequency higher than 10 (4,938 out of 14,723). We therefore had 44,938 cues and 11 outcomes (recall that the future perfect progressive was excluded due to its low frequency of occurrence in the corpus). Prior to running simulations, we divided the data into training (90%), validation (5%) and test (5%) sets. The training set, as the name suggests, was used to train NDL and estimate the weights. The validation set was used to select the parameter set that maximises accuracy. The parameters that we tuned were the learning rate (we considered seven values: 0.00001, 0.00005, 0.0001, 0.0005, 0.001, 0.005 and 0.01) and the number of epochs, i.e., repetitions through the training data set (we considered integer values between 1 and 10), so we simulated the model with 70 combinations of parameters in total. The test set was used to assess the accuracy of the model on data that was never seen by the model.

2.3.3 Activation support

Besides activation, another useful measure that can be extracted from NDL and which can be used to rank the cues/events by how informative they are for the different outcomes is activation support. This measure can be used to quantify the evidence derived from NDL supporting an outcome for a given event (or a set of cues). Activation support is defined as the difference between the activation of the outcome of interest and the activation of the other outcome, for a binary classification task:

$$\begin{aligned} & \textit{activation support}(\textit{outcome}|\textit{set of cues}) \\ & = \textit{activ}(\textit{outcome}|\textit{set of cues}) - \textit{activ}(\textit{other outcome}|\textit{set of cues}) \end{aligned} \quad 1$$

To allow the use of more than two outcomes, as is the case in our task, we adjust the formulas as follows:

$$\begin{aligned} & \textit{activation support}(\textit{outcome}|\textit{set of cues}) \\ & = \min_{\textit{all other outcomes}} \textit{activ}(\textit{outcome}|\textit{set of cues}) - \textit{activ}(\textit{other outcome}|\textit{set of cues}) \end{aligned} \quad 2$$

In other words, the activation support for an outcome is the (minimal) difference between the activation of the outcome of interest and the next most strongly supported outcome. The higher this value, the stronger the activation of the outcome of interest relative to the other outcomes, and thus, the stronger the support provided by the present set of cues for the outcome of interest. We also used the activation support measure to rank the cues by how informative (or discriminative) they are for each TA; for this, in Eq. (2), ‘set of cues’ is replaced by the cue to be considered. In essence, if raw weights are an *absolute* measure of association between the cue and outcome, association support is a measure of *relative* association strength, i.e., relative to other possible outcomes.

3 Results and discussion

In this section we will first discuss the model’s prediction accuracy, overall and in relation to specific TA combinations (Section 3.1). As we will see, the model displays the expected frequency effect and is better at predicting the most frequent TA combinations in the dataset. Next, we will zoom in on the most predictive cues for a selection of TA combinations and what they tell us about each TA combination (Section 3.2). Looking at these cues in more detail reveals crucial differences between the elements that support the learning of various TA combinations. In Section 3.3., we compare NDL’s performance with data from an

online gap-filling task with native speakers of English who undertook the same performance task as NDL did.

3.1 Categorising usage instances: NDL's prediction accuracy

After selecting the best parameters using the grid search procedure described above (the best parameter values found were 0.0001 for the learning rate and 10 for the number of epochs), we assessed our optimised NDL model by calculating its overall prediction accuracy on our test set (containing unseen data) along with its accuracy for each TA combination. The model performed strongly overall, reaching a test accuracy of 68.0%, which is well above the reference accuracy thresholds of 9.1% if predictions for each of the 11 categories are made randomly and of 46.1% if the most frequent TA (present simple) is always predicted. The accurate predictions, however, mainly pertain to the present simple (86.2%) and past simple (73.9%), remaining low for the other TA combinations, as can be seen in Figure 3. For comparison, we ran the model on a balanced dataset (with equal numbers of instances for each TA), the details are provided in Appendix.

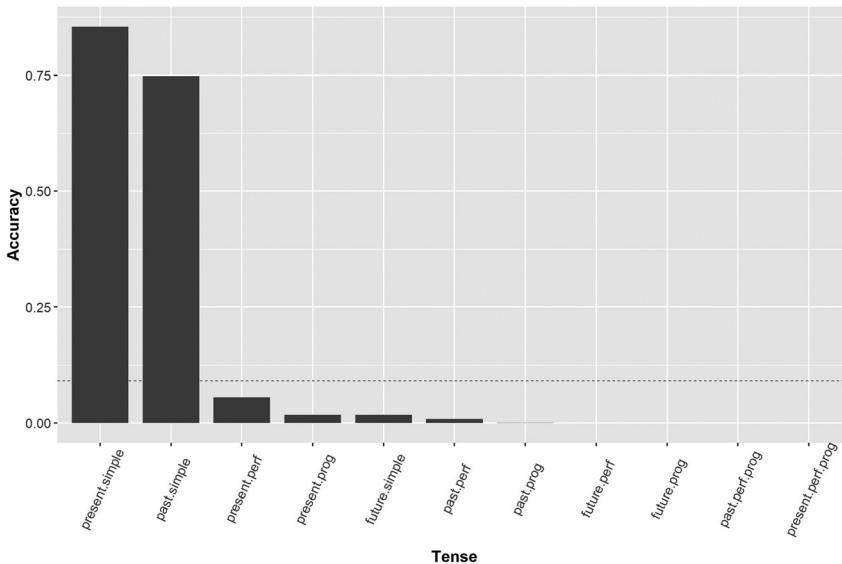


Figure 3: Prediction accuracies of NDL per TA combination. The dashed horizontal line represents the accuracy level that would be achieved in case of an 11-way choice if predictions were made randomly.

The question now is whether the model has learned some useful information about the complex TA combinations despite its (beneficial) bias towards the two simple ones (it is beneficial in the sense that the bias captures the real distribution of TAs in the data as it is experienced by language users). Since the model struggles to predict complex TA combinations when only given one option, we assessed the test accuracy of the three most activated TA combinations for each event (i.e., the three TAs whose activations are the highest) instead of only the most activated one. In other words, we checked how frequently the ‘true’ TA is among the top three TAs that are predicted by the model based on activations. The accuracy of the top three predictions reached 93.0%, substantially higher than the reference accuracy threshold of 27.3% (= 3/11) if the three predictions are made randomly, and also higher than the accuracy threshold of 88.3% if the three most frequent TA combinations (present simple, past simple and present perfect) are always predicted. Crucially, the accuracy was well above chance level across complex TA combinations, markedly for the past perfect (77.6%), present perfect (74.8%), and future simple (67.1%), as shown in Figure 4.

Another possible explanation for the model’s observed low accuracy on complex TA combinations is that many sentences allow the use of more than one

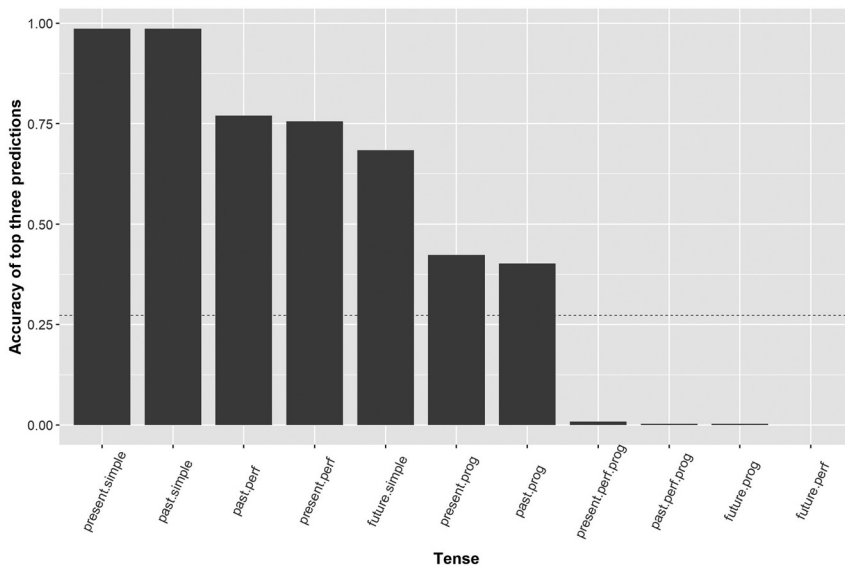


Figure 4: Prediction accuracy for each tense-aspect combination, based on the three most activated outcomes. The dashed horizontal line represents the accuracy level that would be achieved if predictions are made randomly.

TA. Short sentences are particularly prone to this since they have more limited contexts, they are less likely to constrain the type of TA that can be used, but they are also less likely to support those TAs that need contextual support. As examples, consider these two sentences extracted from our data set: “*So, how long have you been waiting?*” and “*The civil aviation authority has launched an inquiry into the incident.*” At least three TA combinations could fit the context within each of the two sentences. It is very difficult for our computational model to learn from this type of unconstrained sentences as they do not contain any cues that could help discriminate between the TA combinations, and hence they will end up activating multiple TAs to a similar extent.

3.2 Hierarchies of cues: types and strengths of cues

Now that we have established that our NDL model is able to determine from mere exposure to usage (and in the proportions encountered in the BNC) what sort of TA construal is likely, at least for some TA categories, we can examine which cues, and ideally which kinds of cues, guide the model’s choice. To extract the most informative cues for a given outcome, we ranked the cues based on their activation support for that outcome. The higher the activation support for an outcome given a certain cue, the more informative (or predictive) the cue is for that outcome. The top 20 positive cues for the two most frequent TAs, present simple and past simple, are displayed in Table 4 together with the top 20 positive cues for the future simple (the other simple form) and the present perfect, for comparison with the past simple. Note that these same cues may well occur with other TAs, e.g., the lexemes listed as cuing present or past simple will be used in other TAs as well. They are, however, not among the strongest cues for those TAs.

Overall, we notice that the most informative cues for each TA combination vary in two ways: (i) their activation support and (ii) their form. For example, Table 4 shows that the activation support for the top 20 most predictive cues is between 0.5 and 0.75 for the present simple but only 0.17 and 0.41 for the future simple. Such a difference is expected on the learning approach outlined above: cue strength (and similarly activation support) will increase as the frequency and exclusivity of the cue-outcome relationship increases. When many cues co-occur more or less systematically with a given outcome, the associations are less decisive. This shows in the distribution of types of cues between the simpler forms (present simple and past simple) which rely on local cues and the more complex TA combinations that are supported by more complex sets of cues (recall configurations in Section 1.3). This is a general trend for these TAs among the top 500 strongest positive cues. It is

Table 4: Top 20 most predictive positive cues for the present simple, past simple, future simple and present perfect.

Present simple	Activation support	Past simple	Activation support	Present perfect	Activation support	Future simple	Activation support
THANK	0.713973	REPLY	0.676626	we#already	0.260298	you#able#to	0.224399
SUPPOSE	0.693332	MURMUR	0.523947	i#already	0.231633	we#you	0.167946
HOPE	0.678446	WHISPER	0.523767	since#then	0.225969	i#you#a	0.15813
REMEMBER	0.584896	NOD	0.516041	CHANGE	0.219764	it#is#hoped	0.142354
MIND	0.563391	SMILE	0.45984	i#ever	0.194032	i#you	0.128989
MEAN	0.548977	MUTTER	0.435074	DEVELOP	0.187542	BENEFIT	0.121223
LET	0.541533	SAY	0.389772	PROVE	0.186498	hopefully	0.119438
HATE	0.539788	SHOUT	0.385171	so#far	0.171343	if#he#a#statement	0.114184
BELIEVE	0.536496	PAUSE	0.383107	ACHIEVE	0.168491	in#a#minute	0.113367
LIKE	0.535944	COMMENT	0.370932	EVOLVE	0.168171	i#you#what	0.110949
RECKON	0.5273	SIGH	0.367944	IDENTIFY	0.165994	will#not#it	0.109195
KNOW	0.513459	GRIN	0.354293	INCREASE	0.160831	i#back	0.099544
WISH	0.499022	GLANCE	0.346232	CRITICISE	0.156794	will#not#you	0.099031
WANT	0.492674	SHAKE	0.337477	traditionally	0.151746	and#i#you	0.098357
NEED	0.489823	SNAP	0.337016	i#never	0.142617	we#hope	0.092754
REPRESENT	0.476806	SHRUG	0.33552	recently	0.139817	predict	0.092629
CONTAIN	0.465639	ANSWER	0.329707	over#the#last	0.138407	tomorrow	0.092123
DEPEND	0.465618	ANNOUNCE	0.313471	you#ever	0.136708	hope#that	0.08779
SOUND	0.464448	STAR	0.309236	over#the#years	0.130634	you#not#able#to	0.082998
REQUIRE	0.461634	DECLARE	0.298843	FINISH	0.120898	they#you	0.07786

worth noting, however, that some cues are shared by two or more TA combinations but with different activation support.

Cues also differ in their form and different TA combinations are characterised by different *types* of strong predictive cues. In Table 4, capital letters signal lexical cues, i.e., the verb lemmas that occur frequently in these TA combinations. The other cues are the various *n*-grams that we identified automatically. A quick look at Table 4 suffices to notice that the present simple and past simple are mostly lexically supported whereas the other TA combinations are more contextually supported. Lexical cues are verb lemmas such as *thank*, *reply* and *murmur* while contextual cues can be temporal adverbials such as *since then* or phrases such as *it is hoped*. While we will not linger on the semantics associated with each TA combination, a few remarks on the most informative cues for these TAs are in order. First of all, the BNC contains mostly written language, which explains the *type* of lemmas found in the past simple: that is, verbs for reported speech, e.g., *reply*, *say*, *comment* or denoting reactions, e.g., *nod*, *smile*, *shrug* (a similar observation was made by Biber et al. 1999: 459). These verbs are very commonly found in fiction, which Biber et al. (1999: 456) report has the most predominant use of the past. Interestingly, however, fiction makes up only 25% of the BNC's written section, while the remaining 75% are composed of informative writings. These verbs of reporting speech and reaction are also commonly found in news outlets, where journalists report on different people's positions. As to the verbs that are most informative for the present simple, we confirm Biber et al.'s observation that the verbs that are the most representative of the present (simple) are stative verbs such as *hope*, *hate* or *mean* (Biber et al. 1999: 459). As mentioned, both the progressive and perfect take more contextual cues in the form of *n*-grams. For example, as shown in Table 4, cues for the perfect are either of a temporal nature: *since then*, *over the last*, or they seem to express some form of telicity, as with the adverb *already* found in multiple *n*-grams (cf. Wulff et al. 2009 for further discussion of telicity across the progressive and the perfect). Despite their focus on verbs, Biber et al. (1999: 469–470) do note that the perfect is mostly (70% of the time) found with time adverbials and/or in dependent clauses. This observation aligns with our observation about these TAs requiring more contextual support.

The proportion of *n*-grams and lexical cues (i.e., verb lemmas) within the top 100 cues (based on activation support) for each TA combination is given in Figure 5 (a plot based on cue weights is available in Appendix together with the corresponding version of Table 4).

Figure 5 shows that the simpler the form of the TA, the higher the proportion of lexical cues and the more complex the TA, the higher the proportion of contextual cues. This appears to confirm our hypothesis that situations that are conceptually more complex require more contextual support. We also note that frequency, and

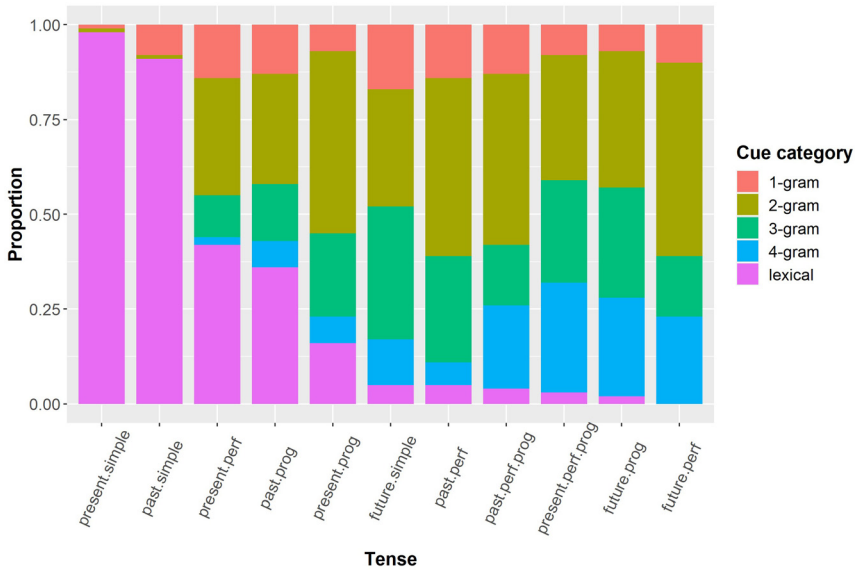


Figure 5: Distribution of each cue type (contextual versus lexical) in the top 100 cues for each tense-aspect combination.

in particular the Zipfian distribution of the TAs in the input, plays a crucial role in the cue-outcome dynamics. It becomes clear that TAs that occur less often than there are verbs in the input are cued by n -grams, whereas TAs that occur overwhelmingly more often than there are verbs in the input are cued by (their own) lexemes. This assumption, that verbs must be repeated more often in the most frequent TAs, is borne out by a tally of types and tokens. Table 5 shows that frequently used TA combinations such as Present Simple and Past Simple are instantiated by 3,363 and 3,676 different verb types, respectively, and that those verb types are repeated many times, yielding 162,421 instances of the Present Simple and 131,899 instances of the Past Simple.

Another observation that jumps out from Table 4 is that the cues for simple forms have a much higher activation support than those of more complex TA combinations. As previously mentioned, more complex TAs have more contextually distributed cues (configurations), but a larger number of cues makes it harder for any cue to develop a strong relationship with the outcome. As is visible from Table 6, which presents the combined weights and activation support for the top 100 most predictive positive cues for each TA combination, we find that combinations in the simple form, which are morphologically less complex, tend to

Table 5: Tally of types and tokens per TA combination as attested in our training sample.

	TYPE	TOKEN	Token/type	Type/token
Present simple	3,393	161,794	47.68	0.02
Past simple	3,653	131,831	36.09	0.03
Present perfect	1,550	16,732	10.79	0.09
Future simple	1,275	13,826	10.84	0.09
Past perfect	1,596	12,484	7.82	0.13
Present progressive	1,019	7,056	6.92	0.14
Past progressive	1,024	5,227	5.10	0.20
Present perfect progressive	243	576	2.37	0.42
Future progressive	181	409	2.26	0.44
Past perfect progressive	161	342	2.12	0.47
Future perfect	90	143	1.59	0.63

be more strongly activated. Note that the future simple is in the same range as its morphologically complex equivalents, the presents and pasts perfect and progressive. At the bottom of the table, there are the most complex combinations: the futures other than simple and the perfect progressives. It seems that there is a correlation between the strength of the cues and the type of experience: TAs which describe situations/events that are known or certain (such as present and past), or that express a simple temporal relation (such as the simple) are more strongly supported.

Table 6: The sum of activation support and weights for the top 100 most predictive positive cues for each TA combination.

TA	Sum of activation support	Sum of weights
Present simple	43.22092417	40.70124996
Past simple	33.90935375	32.33607789
Present perfect	16.34280621	11.3738899
Future simple	15.76685668	8.834190179
Past perfect	14.86026528	7.963369709
Present progressive	14.2889809	6.972837081
Past progressive	13.11763362	5.585983481
Present perfect progressive	12.64396006	1.270793048
Future perfect	12.6311508	0.373777585
Future progressive	12.62350075	0.65108672
Past perfect progressive	12.58568224	0.794178674

3.3 The human yard stick: is NDL's performance human-like?

On the assumption that categories can only be cognitively realistic if they are learnable from exposure to usage, we set out to establish how well each TA can be learned from exposure to usage alone. Therefore, in the final step, we probe the similarity between NDL's predictions and characteristics of what was actually learned, on the one hand, with native (and naive) speakers' performance, on the other hand. We ran an online survey using data from our testing dataset so that we could test the model and the language users on the exact same sentences (for attempts similar in spirit, see Divjak et al. 2016a; Milin et al. 2016). Note that this comparison does not assume that language users develop their TA knowledge based on isolated sentences; instead, the comparison tests which TA combination language users would select based on the information provided in an isolated sentence, and how that choice compares to what NDL would predict would be chosen.

A total of 65 participants took our survey. They were all native speakers of English and represented a total of six varieties of English distributed as follows: 29 speakers of British English, 19 speakers of Irish English, 12 speakers of American English, three speakers of Australian English, 1 speaker of Canadian English and 1 speaker of New Zealand English. Participants were shown twenty-five sentences and were asked to choose the best TA combination from a total of five. First, we chose sentences of moderate length. These sentences were divided into four categories based on NDL's predictions and how they matched the expected outcome: five sentences where NDL's first prediction was a match, five where its second prediction was a match, five where the third prediction was a match. Finally, the last 10 sentences were chosen from instances where NDL did not predict the expected outcome. For all sentences where NDL predicted the expected outcome, we gave participants NDL's top three predictions and two fillers. For sentences for which NDL did not predict the expected outcome, we replaced one of the fillers with the actual outcome.

Overall, a slight majority of respondents preferred the actual TA label as attested in the corpus (60%). The results of this survey are presented in Table 7, where columns correspond to the respondents' choices and rows to NDL's prediction accuracy, for example 35.4% of respondents chose NDL's first choice when NDL's accurate prediction was its second (first column, second row). Table 7 shows that most participants (81.2%) retrieved the actual TA label when NDL did too, while participants were more likely to prefer another TA label when NDL

Table 7: A comparison of NDL's predictions and native speakers' choices. Cases where speakers' choices match those attested in the corpus are highlighted in bold.

NDL's prediction	Respondents' choice				
	NDL's first prediction	NDL's second prediction	NDL's third prediction	Fourth option	Fifth option
First prediction correct	81.2%	12%	4%	1.5%	1%
Second prediction correct	35.4%	49.6%	4%	6.2%	3.1%
Third prediction correct	15.7%	13.9%	45.9%	21.2%	0.6%
No correct prediction	12.8%	9.1%	11.1%	57.2%	6.2%

made a different first prediction, as illustrated in the bottom three rows of Table 7. Participants' responses were thus more varied when NDL's first prediction was not the original TA combination. Disagreement among speakers as to which TA combination is the most appropriate when the sentence context is ambiguous may well illustrate the speakers' choice of scope: given the limited context, they will pick whichever immediate and maximal scope fits best. Our algorithm may be biologically and psychologically plausible, it is, however, not human and as such, it can only take into account what it has access to. In this case, the algorithm was provided with sentences. As we noted in Section 1.1, human minds have a variety of construal operations at their disposal that allow them to choose how they present a situation or event. Such choices are typically supported by the context, and very often this context exceeds the sentence boundaries. For now, this is not information our algorithm has access to.

To explore the link between participants' choices and our NDL model further, we probed whether a measure like activation support can directly explain (some of the) variability in participants' proportions of choices. We hypothesise that larger values of activation support for an outcome will be associated with a higher likelihood of users selecting that outcome (i.e., a higher level of agreement between participants for selecting that outcome). Therefore, we ran a linear regression model in R (R-Core-Team 2013) that predicts proportions of choices of TAs (logit-transformed) based on their activation support in the examples that participants encountered in the survey. As expected, the higher the activation support for a TA, the higher the proportion of participants that selected that TA ($\beta = 1.68$, $p = 0.002$, 95% CI = [0.62, 2.74]). This suggests that the strongest cues as identified by our model match those used by speakers.

4 General discussion

Overall, training a biologically plausible learning algorithm has shown that TA categories can indeed be learned from exposure: our model's success is an attestation of this, with a 68% prediction accuracy even on to the most stringent assessment. This success mirrors the distribution of TA categories across our dataset. We observe a typical Zipfian distribution in the input: the bulk of the data is made up of instances of the present simple and past simple (about 84% combined). In what follows, we will discuss how these results emerge from a system that is steeped in experience and organised along a number of cognitive and functional principles, and what these findings entail for the cognitive theories of tense and aspect.

The “simple” TA combinations, simple present and simple past, are arguably simpler than the other TA combinations in two ways. They are not only formally simpler, but also conceptually simpler: they usually encode situations where the relation between speech time, reference time and event time is rather straightforward (but see Hirtle [1988] for a discussion of the semantics of the simple form). Besides being formally and conceptually simpler, the present simple and past simple are also experientially central: they occupy a central position in Langacker's elaborated epistemic model. They are used to depict events that are part of speakers' immediate or known reality—or that are construed as such. They are thus more directly experienced, which makes them cognitively more readily available to speakers, and hence more frequently used. The other TA combinations are more complex, both formally and conceptually and this affects their frequency of use. From a cognitive perspective, this is not surprising: the working memory load that the sequencing of temporal components imposes is significant (Oakhill 2020). In addition, the future has not yet been experienced and is part of a possible non-reality. It is thus experientially inaccessible and epistemically distant.

This discrepancy as to whether the event is construed as being part of a speaker's immediate or known reality or not, and its related conceptual simplicity, also shows in the cues that are the most predictive for these TA combinations: the present simple and past simple, which refer to events that have been experienced, rely nearly exclusively on lexical support. To the contrary, the reliance of the future simple on contextual support indicates the lack of immediate experience of the event and the need to place the event in a potentially dependent and/or uncertain context. Qualitatively, we also find that these contextual cues are indicative of uncertainty: among the most predictive cues for the future are *n*-grams such as *it is hoped*, *hopefully* and *if*-phrases. As for

complex TA combinations, we see that the kind of experience they denote also plays a role as to the types of cues that are the most predictive for these TA combinations. This difference shows qualitatively: while the *n*-gram cues for the perfect are temporal adverbs (cf. Table 4), the *n*-grams that are the most predictive for the progressive appear to be more situational: e.g., *we#about, are#not#you, understand#what* and *know#what#you*.

The formal and conceptual simplicity of the present simple and past simple explains their frequency: because they are more immediately experienced, these temporal arrangements are easier to conceptualise and talk about. The cognitive load that complex TA combinations bring out can be illustrated with an example taken from our dataset. Compare the following two sequences, where (2) is the original sentence and (1) has been created to express the same events with simple forms only:

- (1) *The officers **tried** to stem the flow of blood. The paramedics **arrived** and **took over**.*
- (2) *Paramedics were at the scene in 4 min of the emergency call and took over from officers who **had been trying** to stem the flow of blood.*

The first sequence (1) is straightforward and depicts events chronologically and simply. The second sequence (2), on the other hand, organizes the events as they are construed by the speaker, with a focus on the paramedics, which places the officers in the background and requires the use of a complex TA combination: the past perfect progressive.

Recall that if an outcome's *contingency* on one particular cue is such that the two frequently and exclusively occur together, then their association weight will likely become strong and positive. In the case of many cues that differ with respect to how systematically they co-occur with a given outcome, we end up with *imperfect contingencies*. The reason why lexical cues can come out strong for simple TAs lies in the difference in the frequency of use between simple and complex TAs: because simple TA combinations are used more often, the verbs they are used with have to be repeated (recall here the tallies presented in Table 5).

Given that our account depends on the frequency with which the TAs occur in the input to which it is exposed, the question might arise to what extent our overall findings would change with different input frequencies. Children, for example, learn for many years from conversational language, while L2 learners may long be predominantly exposed to textbook language before being exposed

to other genres. However, in a usage-based approach to language, language users continue to adapt their linguistic knowledge to the input they receive throughout their lifespan. By the time they reach adulthood, many will have accumulated an experience that resembles what we found in the corpus as different studies have found that the proportions of TA combinations appear surprisingly stable across large samples of data and across registers (as reported above in Section 2.2).

However, learning is typically gradual and evolves over time and the fact that a given linguistic abstraction, such as a TA combination, is stable in the long-run and cross-sectionally, says very little about the dynamics that are triggered by the order of exposure. For example, a given TA combination might be highly probable in a particular context in the long run, but if it is introduced later than a competitor TA combination, the later used combination will likely suffer from a *blocking effect* – from being associated with that given context by the earlier used combination, as the general learning principle teaches us (cf., Kamin 1969; Rescorla and Wagner 1972). This might affect L1s who are not exposed to the range of TAs through reading, and L2s who do not venture beyond their textbooks. Furthermore, if we take the reasonable assumption that the more complex the TA combination, the more imperfect contingencies might incur (viz. Tolman and Brunswick 1935), the learning dynamics will become increasingly more complex (Beckner et al. 2009; Ellis 2016). This is where our *learning-from-usage approach* becomes particularly informative.

All in all, everything conspires for us to learn these simple TAs really easily: they are conceptually simple, which makes them frequent. This, in turn, brings out the lexical cues as distinctive for these TA combinations: lexical cues are localised and hence easy to spot (cf. Table 4). In line with previous research, we thus find that the input is distributed in such a way that it makes certain things easier to learn (Wulff et al. 2009; Boyd and Goldberg 2009, inter alia). But we also find that this same distribution makes other things harder to learn. Other TA combinations such as the future simple and the perfects and progressives are used to describe conceptually more complex temporal arrangements. These arrangements are cognitively more demanding to conceptualise and require more complex linguistic expressions. These complexities make them less frequent. Because of the limited repetition, they do not exhaust the lexical cues, hence *n*-grams remain the more stable and more often repeated elements. Yet, different from lexical cues, *n*-grams are diffuse, and therefore more difficult to detect. Together, these factors explain why complex TAs are more difficult to learn. In other words, the observed learnability effects are the work of more than

frequency alone: conceptual complexity drives frequency of occurrence, which in turn affects the types of cues that support a specific TA combination. It would not be surprising, then, to see the simple aspect take on an even larger share over time since forces appear to be conspiring to push complexity out, especially if there is a simpler form to rely on. Biber et al. (1999: 463), for example, report that American English appears as the most “advanced” variety in their corpus, as the trend-setter, so to speak. They noted that American English (already?) uses the simple past where British English relies on the present perfect.

5 Conclusions

Sinclair famously declared that language looks very different when you look at a lot of it at once (Sinclair 1991: 100). The English tense-aspect system is no exception. Using English as our case study, we examined how language supports the expression of temporality within sentence boundaries. We trained our psychologically plausible learning algorithm that mimics how humans learn from exposure on *n*-grams extracted from the BNC to select the TA combination that fits the context best. Our findings are by and large in line with what corpus grammarians have proposed: the same lexical and contextual preferences were observed. Yet, the much larger scale of our study, which ran on the entire BNC (minus a very small number of sentences, cf. Section 2.2), revealed the existence of two different sub-systems within the set of TA combinations: a “simplex” one that is supported lexically, and a “complex” one that is supported contextually. Taking a cognitive linguistic, experiential view on language, and considering languages and our knowledge of them a product of our interaction with the world, allowed us to explore what the use of different tense and aspect combinations reveals about the interaction between our experience of time, and the cognitive demands that talking about time puts on the language user.

We have argued that the observed bifurcation between a simplex and a complex system follows from the cognitive complexity of the conceptualisation of the associated temporal event, which affects their frequency of use. Our data shows that the need for contextual support increases when the temporal events expressed are temporally distant, epistemically uncertain or cognitively complex to construe. The more straightforward the relation between time of speech and time of the event (and thus the simpler the construal of the event), the simpler the form and its support: past and present simple are simple forms that are lexically supported. Temporal immediacy and epistemic certainty resonate in the verb form and the need for contextual backup. This relation is not direct, but mediated by frequency: complexity and frequency, jointly, affect learnability and the types

of cues that emerge as informative. The existence of a system, steeped in human experience and sculpted by the blade of cognition, provides an elegant account of how the English TA system is learned without reverting to item-by-item memorisation and storage of lexical and contextual preferences (cf. Divjak et al. 2021).

The finding that the simplex TA combinations, which are the most frequent TA combinations, are essentially lexical in nature and that the more complex TA combinations typically require contextual support necessitates a rethink of tense and aspect as grammatical categories. Instead of a separation of tense and aspect as such, it appears that the distinction lies in a simplex versus complex paradigm that emerges from the interaction of language use and language cognition, learning in particular. This insight paves the way for new approaches to tense and aspect, across the range of areas that have shown an interest in tense and aspect, and which run the gamut from the multimodal expression of TA categories (cf. Parrill et al. 2013) to the teaching of TA categories to foreign language learners.

6 Data availability statement

The scripts and data used in this paper can be found through the following link: <https://github.com/oominds/ndltenses> and a dummy version of the survey we used can be found here: https://birminghamcoaal.eu.qualtrics.com/jfe/form/SV_9S2qhPQIJ2fZkUK.

Acknowledgments: We would like to thank Dagmar Hanzlíková for her work on developing the annotation heuristics for tense and Oghenetekevwe Kwakpovwe for creating a user-friendly Python package with which the results reported in this manuscript were obtained. Special thanks go to Thomas Herbst for reading a draft of this paper and providing relevant feedback. We would also like to thank the audiences to which we presented this research (members of the STL lab at the Université de Lille as well as UK-CLC and EDLL attendants) and the three anonymous journal reviewers for their very helpful comments. We are particularly grateful to Reviewer #2 whose extremely detailed feedback helped improve this paper considerably.

Research funding: The work reported on in this manuscript was funded by Leverhulme Trust Leadership Grant RL-016-001 to Dagmar Divjak which funded all authors. Dagmar Hanzlíková and Oghenetekevwe Kwakpovwe were also funded by RL-2016-001 to Dagmar Divjak.

Conflict of interest: John Newman was appointed as Editor-In-Chief for the review purposes of this article. The submission was blinded in the submission system ScholarOne and the standard double-blind peer-reviewed procedure was adhered to.

Appendix

Appendix A: Simulations on a balanced dataset

Since our dataset was skewed towards the present simple and past simple, we decided to run our model on a balanced dataset. This dataset was reduced to TA combinations that occur at least 100,000 times in the BNC. The simulations run on the balanced dataset (which contains equal numbers of examples for each TA combination) confirm that the skew in the distribution of TA combinations has most likely caused the low accuracy for TAs other than the present simple and past simple that we describe in Section 3.1 of the paper. That is, the TA combinations that we referred to as complex TAs in Section 1.3 are better predicted in a balanced dataset, as illustrated in Figure A1 below.

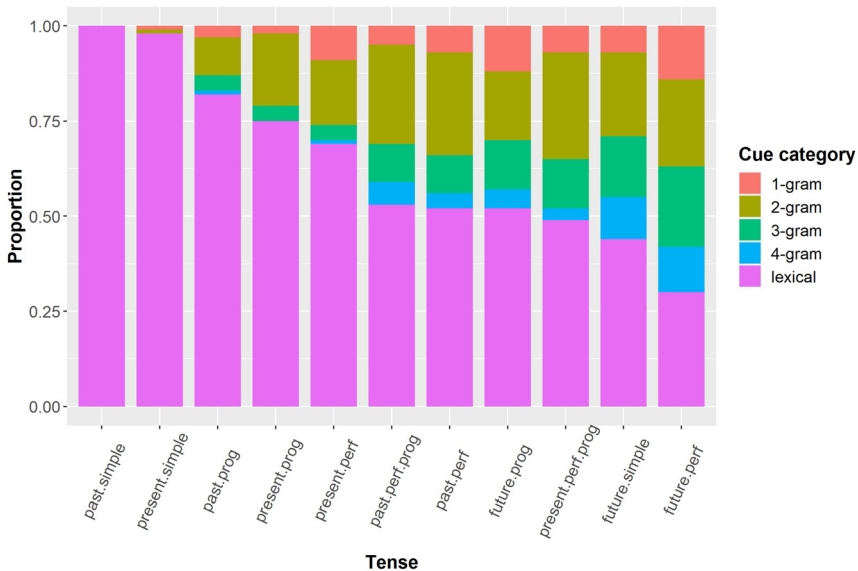


Figure A1: Prediction accuracies of NDL per TA combination on a balanced dataset.

Table A1: Top 20 most predictive positive cues for the past simple, present simple, present perfect and future simple.

past.simple	past.simple. weights	present.simple	present.simple. weights	present.perf	present.perf. weights	future.simple	future.simple. weights
REPLY	0.676532	THANK	0.713724	we#already	0.260918	you#able#to	0.227417
NOD	0.57591	SUPPOSE	0.693115	since#then	0.234537	BENEFIT	0.19586
SMILE	0.551272	HOPE	0.681297	i#already	0.231499	we#you	0.17322
WHISPER	0.526316	REMEMBER	0.584738	CHANGE	0.222712	HELP	0.15879
MURMUR	0.523826	MIND	0.563529	i#ever	0.196533	i#you#a	0.15861
SAY	0.52059	MEAN	0.54862	DEVELOP	0.190194	i#is#hoped	0.145141
MUTTER	0.441012	LET	0.543124	PROVE	0.186529	CONTINUE	0.13275
SHOUT	0.43565	HATE	0.539626	so#far	0.175844	hopefully	0.12902
GLANCE	0.430791	BELIEVE	0.536128	EVOLVE	0.172319	i#you	0.128942
ANSWER	0.429004	LIKE	0.535603	ACHIEVE	0.171793	ENSURE	0.126743
PAUSE	0.428964	RECKON	0.526818	IDENTIFY	0.166688	in#a#minute	0.121689
SHAKE	0.415516	KNOW	0.512973	INCREASE	0.164056	AFFECT	0.120868
SIGH	0.408754	WISH	0.498653	CRITICISE	0.160095	RESULT	0.118277
ANNOUNCE	0.408571	WANT	0.492362	traditionally	0.151686	i#he#a#statement	0.114097
LAUGH	0.404451	NEED	0.489425	LAUNCH	0.149933	will#not#it	0.113167
LAY	0.40047	REPRESENT	0.477873	recently	0.148805	tomorrow	0.112821
COMMENT	0.389857	DEPEND	0.465561	over#the#last	0.148533	i#you#what	0.110918
STAR	0.383384	CONTAIN	0.465307	FINISH	0.148013	COST	0.11066
GRIN	0.382588	SOUND	0.464387	SIGN	0.145532	will#not#you	0.108579
ADD	0.381551	REQUIRE	0.461266	APPOINT	0.144684	DEPEND	0.107034

Figure A1 shows that complex TA combinations are better predicted in a balanced dataset. However, this balanced dataset does not do justice to the type of exposure speakers get. Speakers usually encounter TA combinations in the frequencies we, and others have described where the simple forms are the most frequent regardless of genre (cf. Section 2.2).

Appendix B: Cue weights versus activation support

Figure A2 and Table A1 represent prediction accuracies in terms of association weights. In the main paper, predictions are discussed in terms of activation support (cf. Section 2.3.3), our preferred measure of the relative strength, which magnifies the separation between two types of cues: lexical versus contextual. While this separation is still present (i.e., the amount of lexical support for the present simple and past simple, on the one hand, and the other TAs on the other hand), with the raw association weights, the absolute measure, the separation is more graded and forms a cline.

References

- Anthony, Laurence. 2013. A critical look at software tools in corpus linguistics. *Linguistic Research* 30(2). 141–161.
- Baayen, R. Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–481.
- Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16.
- Beckner, Clay, Nick C. Ellis, Richard Blythe, John Holland, Joan Bybee, Jinyun Ke, Morten H. Christiansen, Diane Larsen-Freeman, William Croft & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59. 1–26.
- Bender, Andrea & Sieghard Beller. 2014. Mapping spatial frames of reference onto time. A review of theoretical accounts and empirical findings. *Cognition* 132. 342–382.
- Bergs, Alexander. 2010. Expressions of futurity in contemporary English: A Construction Grammar perspective. *English Language and Linguistics* 14(2). 217–238.
- Biber, Douglas. 2006. Parsing and Grammar Description, Corpus-Based. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 2 edn., 197–205. Boston: Elsevier.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bielak, Jakub & Mirosław Pawlak. 2011. Teaching English tense and aspect with the help of cognitive grammar: An empirical study. *Studies in Second Language Learning and Teaching* 1(3). 365–400.
- Binnick, Robert I. 1991. *Time and the verb: A guide to Tense and Aspect*. New York and Oxford: Oxford University Press.

- Bird, Steven, Ewan Klein & Edward Loper. 2009. *Natural Language Processing with Python*. Sebastopol, CA, USA: O'Reilly Media Inc.
- Boogaart, Ronny & Theo Janssen. 2007. Tense and aspect. In *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press.
- Boyd, Jeremy K. & Adele E. Goldberg. 2009. Input effects within a constructionist framework. *The Modern Language Journal* 93(3). 418–429.
- de Brabanter, Philippe, Mikhail Kissine & Saghie Sharifzadeh. 2014. *Future times, future tenses*. Oxford: Oxford University Press.
- Brisard, Frank. 2013. An account of English tense and aspect in Cognitive Grammar. In Kasia M. Jaszczolt & Louis de Saussure (eds.), *Time*, 1st edn. Oxford: Oxford University Press.
- Chen, Zhe, Simon S. Haykin, Jos J. Eggermont & Suzanna Becker. 2007. *Correlative learning: A basis for brain and adaptative systems*. Hoboken, NJ: Wiley.
- Chilton, Paul. 2013. Frames of reference and the linguistic conceptualization of time: Present and future. In Kasia M. Jaszczolt & Louis de Saussure (eds.), *Time*, 236–258. Oxford: Oxford University Press.
- Chomsky, Noam. 1959. Review of Skinner's Verbal Behavior. *Language* 35. 26–58.
- Chuang, Yu-Ying & R. Harald Baayen. 2021. Discriminative learning and the lexicon: NDL and LDL. *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
- Croft, William. 2012. *Verbs. Aspect and causal structure*. Oxford: Oxford University Press.
- Dąbrowska, Ewa. 2016. Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27(4). 479.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- De Wit, Astrid & Frank Brisard. 2014. A Cognitive Grammar account of the semantics of the English present progressive. *Journal of Linguistics* 50(1). 49–90.
- Divjak, Dagmar. 2015. Four challenges for usage-based linguistics. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Cuyckens Hubert (eds.), *Change of Paradigms: New Paradoxes*, 297–309. Berlin: De Gruyter.
- Divjak, Dagmar. 2019. *Frequency in Language. Memory, Attention and Learning*. Cambridge: Cambridge University Press.
- Divjak, Dagmar & Petar Milin. 2020. Exploring and exploiting uncertainty: Statistical learning ability affects how we learn to process language along multiple dimensions of experience. *Cognitive Science* 44(5). e12835.
- Divjak, Dagmar, Ewa Dąbrowska & Antti Arppe. 2016a. Machine meets man: Evaluating the psychological reality of corpus-based probabilistic models. *Cognitive Linguistics* 27(1). 1–33.
- Divjak, Dagmar, Natalia Levshina & Jane Klavan. 2016b. Cognitive Linguistics: Looking back, looking forward. *Cognitive Linguistics* 27(4). 447.
- Divjak, Dagmar, Nina Szymor & Anna Socha-Michalik. 2015. Less is more: Possibility and necessity as centres of gravity in a usage-based classification of core modals in Polish. *Russian Linguistics* 39(3). 327–349.
- Divjak, Dagmar, Petar Milin, Adnane Ez-Zizi, Jarosław Józefowski & Christian Adam. 2021. What is learned from exposure: An error-driven approach to productivity in language. *Language, Cognition and Neuroscience* 36(1). 60–83.
- Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Amsterdam: Springer.
- Ellis, Nick C. 2006a. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.

- Ellis, Nick C. 2006b. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking and perceptual learning. *Applied Linguistics* 27(2). 164–194.
- Ellis, Nick C. 2016. Salience, Cognition, Language Complexity, and Complex Adaptive Systems. *Studies in Second Language Acquisition* 38(2). 341–351.
- Ghirlanda, Stefano. 2015. On elemental and configural models of associative learning. *Journal of Mathematical Psychology* 64. 8–16.
- Hale, John. 2001. A probabilistic early parses as a psycholinguistic model. *Paper presented at the NAACL 01: Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, Pittsburgh, PA.
- Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32. 101–123.
- Hanus, Daniel. 2016. Causal reasoning versus associative learning: A useful dichotomy or a Strawman battle in comparative psychology? *Journal of Comparative Psychology* 130(3). 241–248.
- Hawkins, John. 1992. Syntactic weight versus Information structure in word order Variation. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, Vol. 4, 186–219. Opladen: Westdeutscher Verlag.
- Heyes, Cecilia. 2012. What's social about learning? *Journal of Comparative Psychology* 126(2). 193–202.
- Heyes, Cecilia. 2016. Homo imitans? Seven reasons why imitation couldn't possibly be associative. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1686). <https://doi.org/10.1098/rstb.2015.0069>.
- Hirtle, Walter H. 1988. Events, Time and the Simple Form. *Revue québécoise de linguistique* 17(1). 85–105.
- Hornstein, Norbert. 1991. *As time goes by*. Cambridge, MA: MIT Press.
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61. 23–62.
- Janda, Laura A. 2015. Tense aspect mood. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 616–634. Berlin/Boston: De Gruyter-Mouton.
- Janda, Laura A. & Robert J. Reynolds. 2019. Construal vs. redundancy: Russian aspect in context. *Cognitive Linguistics* 30(3). 467–497.
- Kamin, Leon J. 1969. Predictability, surprise, attention and conditioning. In Byron A. Campbell & Russell M. Church (eds.), *Punishment and Aversive Behaviour*. New York: Appleton-Century-Crofts.
- Kermer, Franka. 2016. *A Cognitive Grammar Approach to Teaching Tense and Aspect in the L2 Context*. Newcastle-Upon-Tyne: Cambridge Scholars Publishing.
- Kokkola, Niklas H., Esther Mondragón & Eduardo Alonso. 2019. A double error dynamic asymptote model of associative learning. *Psychological Review* 126(4). 505–549.
- Kramsky, Jiri. 1969. Verb-Form Frequency in English. *Brno Studies in English* 8. 111–120.
- Kravchenko, Alexander V. 2012. Cognitive Grammar and EFL Methodology: The Case for Tenses. *Актуальные проблемы филологии и педагогической лингвистики* 14. 343–358.
- Kwakpovwe, O. 2021. ndltenses: A package for training an NDL model to learn tenses using pyndl (Version 1.0) [Computer software]. <https://doi.org/10.5281/zenodo.5680781>.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*, vol. 1. Stanford: Stanford University Press.

- Langacker, Ronald W. 1991. *Foundations of cognitive grammar: Descriptive application*, vol. 2. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 2001. The English present tense. *English Language and Linguistics* 5(2). 251–272.
- Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Langacker, Ronald. 2011. The English Present. Temporal coincidence vs. epistemic immediacy. In A. Patard & F. Brisard (eds.), *Cognitive Approaches to Tense, Aspect and Epistemic Modality*, 45–86. Amsterdam/Philadelphia: John Benjamins.
- Leech, Geoffrey. 1992. 100 Million Words of English: The British National Corpus (BNC). *Second Language Research* 28. 1–13.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.
- Michaelis, Laura. 2006. Tense in English. In Bas Aarts & April McMahon (eds.), *The Handbook of English Linguistics*, 220–243. Oxford: Blackwell Publishing Ltd.
- Milin, Petar, Dagmar Divjak, Strahinja Dimitrijević & R. Harald Baayen. 2016. Towards cognitively plausible data science in language research. *Cognitive Linguistics* 27(4). 507.
- Milin, Petar, Dagmar Divjak & R. Harald Baayen. 2017a. A learning perspective on individual differences in skilled reading: Exploring and exploiting orthographic and semantic discrimination cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(11). 1730–1751.
- Milin, Petar, Laurie Beth Feldman, Michael Ramscar, Petar Hendrix & R. Harald Baayen. 2017b. Discrimination in lexical decision. *PLoS one* 12(2). e0171935.
- Niemeier, Susanne & Monika Reif. 2008. Making progress simpler? Applying cognitive grammar to tense-aspect teaching in the German EFL classroom. In Sabine De Knop & Teun De Rycker (eds.), *Cognitive Approaches to Pedagogical Grammar*, 325–356. Berlin, New York: Mouton de Gruyter.
- Oakhill, Jane. 2020. Four decades of research into children’s reading comprehension: A personal review. *Discourse Processes* 57(5–6). 402–419.
- Parrill, Fey, Benjamin K. Bergen & Patricia V. Lichtenstein. 2013. Grammatical aspect, gesture, and conceptualization: Using co-speech gesture to reveal event representation. *Cognitive Linguistics* 24(1). 135–158.
- Pearce, John M. 1994. Similarity and discrimination: A selective review and a connectionist model. *Psychological Review* 101(4). 587–607.
- Pearce, John M. & Geoffrey Hall. 1980. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* 87(6). 532–552.
- Pirrelli, Vito Marzi, Ferro Claudia, Marcello, Franco Alberto Cardillo, R. Harald Baayen & Petar Milin. 2020. Psycho-computational modelling of the mental lexicon. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word usage*, 23–82. Berlin: De Gruyter Mouton.
- Radden, Günter & René Dirven. 2007. *Cognitive English Grammar*. Amsterdam/Philadelphia: John Benjamins.
- Ramscar, Michael & Daniel Yarlett. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science* 31(6). 927–960.

- Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957.
- R-Core-Team. 2013. *R: A language and environment for statistical computing*.
- Reichenbach, Hans. 1947. *Elements of symbolic logic*. New York: Macmillan.
- Rescorla, Robert A. 1988. Pavlovian Conditioning. It's not what you think it is. *American Psychologist* 43(3). 151–160.
- Rescorla, Robert A. 1997. Summation: Assessment of a configural theory. *Animal Learning & Behavior* 25(2). 200–209.
- Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton Century Crofts.
- Rissman, Lilia & Asifa Majid. 2019. Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review* 26(6). 1850–1869.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.
- Römer, Ute. 2005. *Progressives, Patterns, Pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam/Philadelphia: John Benjamins.
- Rothstein, Susan. 2004. *Structuring Events: A Study in the Semantics of Aspect*. Oxford: Blackwell Publishing Ltd.
- Sering, Konstantin, Marc Weitz, David Elias Künstle & Lennart Schneider. 2017. Pyndl: Naive Discriminative learning in Python.
- Sinclair, John. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Skinner, Burrhus Frederic. 1957. *Verbal Behavior*. New York: Appleton_Century-Crofts.
- Stanojević, Mateusz-Milan. 2011. Modeling epistemic distance: The Croatia I-participle. *Suvremena lingvistika* 37(72). 219–239.
- Stefanowitsch, Anatol & Stefan T Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Tolman, Edward C. 1932. *Purposive behavior in animals and men*. New York: Century.
- Tolman, Edward C. & Egon. Brunswick. 1935. The organism and the causal texture of the environment. *Psychological Review* 42(1). 43–77.
- Trimmer, Pete C., John M. McNamara, Alasdair I. Houston & James A. Marshall. 2012. Does natural selection favour the Rescorla–Wagner rule? *Journal of Theoretical Biology* 302. 39–52.
- Verkuyl, Henk J. 1993. *A Theory of Aspectuality: The Interaction between Temporal and Atemporal Structure*. Cambridge: Cambridge University Press.
- Wagner, Allan R. 2003. Context-sensitive elemental theory. *The Quarterly Journal of Experimental Psychology: Section B* 56(1). 7–29.
- WALS Online. 2013. Retrieved from <https://wals.info/>.
- Wasserman, Edward A., Susan M. Elek, Diane Lynn Chatlosh & Andrew G. Baker. 1993. Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(1). 174–188.
- Wulff, Stefanie, Nick C. Ellis, Ute Romer, Kathleen Bardovi-Harlig & Chelsea J. Leblanc. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal* 93(3). 354–369.