

Article

Dagmar Divjak*, Petar Milin and Maciej Borowski

On the learnability of aspectual usage

<https://doi.org/10.1515/cllt-2023-0117>

Received December 10, 2023; accepted March 14, 2025; published online June 2, 2025

Abstract: We examine the learnability of grammatical aspect in Polish through the application of an error-correction learning algorithm to corpus data. We explore whether mastery of aspectual usage can be attained solely through exposure to co-occurrence patterns, or if an understanding of the abstract semantic distinctions discussed in the aspectual literature are a prerequisite. To validate the model, we compare the corpus-based findings with data on aspectual usage from a survey of Polish L1 users. Our findings show that aspectual preferences are best accounted for by a model that relies on co-occurrence information to predict a particular lemma, without reference to aspectual information. The models incorporating aspectual information, either by predicting aspect from co-occurrence information or from abstract semantic labels, performed adequately but were only able to capture usage preferences relating to one of the two aspects. This study contributes insights into the role of usage in the acquisition of linguistic categories and underscores the importance of integrating diverse evidence sources and methodologies for the development and validation of linguistic theories.

Keywords: Slavonic aspect; abstraction; error-correction learning; computational simulation; usage; behavioural profiling

1 Introduction

Traditionally, aspect is considered a grammatical category that renders how a state, event, or action expressed by a verb extends over time. In many Slavonic languages,

Dagmar Divjak and Petar Milin share first authorship on this paper.

***Corresponding author: Dagmar Divjak**, Department of Linguistics & Communication and the Department of Modern Languages, University of Birmingham, Edgbaston, Ashley Building, B15 2TT, Birmingham, UK, E-mail: d.divjak@bham.ac.uk. <https://orcid.org/0000-0001-6825-8508>

Petar Milin and Maciej Borowski, Department of Modern Languages, University of Birmingham, Ashley Building, B15 2TT, Birmingham, UK. <https://orcid.org/0000-0001-9708-7031> (P. Milin)

the aspectual system comprises two variants – perfective and imperfective. Over the past decades, a variety of concepts has been proposed to capture the temporal semantics of imperfective and perfective, challenging both the linguists who try to describe aspectual usage and the second language learners who try to master it.

Grounding our research in usage-based linguistic theory, we explore the interrelated questions of whether abstractions such as imperfective and perfective, and the many semantic concepts that have been proposed to explain their usage, are prerequisites for proficient use and, hence, whether it is reasonable to assume that these linguistic concepts arise naturally from the data, meaning they are learnable from exposure (Divjak 2015). Staying true to the cognitive commitment (Lakoff 1990, 1991), we rely on a computational model of learning based on the Rescorla–Wagner model (Rescorla and Wagner 1972; Wagner and Rescorla 1972) known as the Naive Discriminative Learner (Baayen 2011; Milin et al. 2017a, 2017b) to determine whether the use of aspect can be explained by word-level contingencies in usage or whether abstractions over these contingencies are indispensable. Finally, we validate our findings by testing both our models and a sample of L1 Polish users on the same selection of items.

1.1 The traditional dimensions of aspect

To honour the tradition, we must start with the caveat that what follows is by no means a comprehensive account of the vast literature on aspect (see Sasse 2002 for a more comprehensive overview). As Łaziński (2020: 309) put it: there is an ‘abundance and variety of existing theories of aspect, which have produced a multitude of often inconsistent approaches and terms’. Because our central concern is with the driving force behind aspectual usage and the necessity and learnability of abstractions over that usage, we will limit ourselves to outlining the two main approaches – grammatical and lexical – very briefly below.

The prototypical Slavonic verbal paradigm distinguishes, across two aspects, at least three tenses (past, present, future), three moods (indicative, imperative, conditional), two voices (active and passive) and two numbers (singular and plural) with three persons each (first, second and third); aspect is additionally marked on non-finite forms such as the infinitive, participle and gerund. Inflectionally, there are notable aspectual gaps that apply without exception: perfective verbs lack a present tense and have a simple future tense while imperfective verbs have a present tense but require an auxiliary to express the future. Similarly, only imperfectives have active present participles and present gerunds, while passive past participles and past gerunds are exclusive to perfectives. Different from most languages, Slavonic languages do not use consistent inflectional aspectual markers: Slavonic aspectual

morphology is highly irregular, and clear inflectional categories cannot easily be distinguished. Imperfective and perfective variants relate to each other in one of three ways: through prefixation, infixation or stem changes and suppletion.

Firmly embedded since Jakobson (1932 [1984]), the grammatical approach assumes a binary division into perfective and imperfective. In a subset of languages, such as Polish, the perfective and imperfective are morphologically marked on nearly every verb form (Bańko 2002; Grzegorzczkowska et al. 1999; Wróbel 2001); Łaziński (2020: 310) estimates this to be the case for 90 % of all Polish verbs, the only exceptions being simplex (im)perfectives and biaspectuals. Because, in some cases, either aspect is possible and the choice for one over the other is said to express the perspective that users take on the event, grammatical aspect is also known as ‘viewpoint aspect’. Interestingly, while 90 % of verbs exist in both perfective and imperfective, only about 10 % were found to be used equally frequently in both aspects a large-scale corpus analysis (Divjak et al. 2024).

Much work on grammatical aspect has been dedicated to establishing the invariant meaning of one or both of the aspectual classes (see Janda 2004 for an assessment). While intuitively there appears to be a meaning difference between the imperfective and perfective form of a verb, pinning down the invariant meaning of each aspect has proved challenging, not in the least because the invariant meaning is supposed to be a ‘one size fits all’ label for a phenomenon that has many dimensions. It is, therefore, not surprising that different accounts have been put forward.

One central concept is that of *completion* (the Polish terms for imperfective and perfective verb are *czasownik niedokonany* or ‘incompleted verb’ and *czasownik dokonany* or ‘completed verb’). Even though sentences (1) and (2) describe a book-reading situation that happened in the past, (2) informs us that the reading of the book was completed, while (1) presents the situation as ongoing at some point in the past.

- (1) Wczoraj czytałem książkę.
Yesterday read.1SG.PAST.IMPF book
I was reading a book yesterday.
- (2) Wczoraj przeczytałem książkę.
Yesterday read.1SG.PAST.PF book
I read a book yesterday.

The difference between the perfective and imperfective can also be explained with reference to *totality*. Comrie (1976: 16) proposes that the perfective presents the situation ‘as a whole’ while ‘the imperfective pays essential attention to the internal structure of the situation’. Likewise, Forsyth (1970: 8) sees totality as ‘the action as a total event summed up with reference to a single specific juncture’. In other words,

and glossing over many intricate differences between these two accounts, (2) presents reading as a singular situation, without concern for the individual steps in the process while in (1) it is precisely the process of reading itself that is in focus.

Others have appealed to the notion of *boundedness* (e.g. Smith 1986). The basic aspectual distinction would be one between unbounded and bounded situations: situations may be conceived of as including their starting points or endpoints or both, or may be conceived of as persistent situations with no boundaries implied. For our examples, this would mean that (2) includes both the beginning and the end (or the initial and final boundary) of the action, whereas (1) does not include these boundaries.

Klein (1995) offers a rather different account of aspect and present a *time-relational* analysis. Klein argues that aspects are temporal relations between the time when a situation is in effect and the time about which something is said in the utterance describing the situation. It does not rely on metaphorical concepts but operates with notions that are needed outside the study of aspect too, i.e. time intervals, temporal relations between these intervals and the notion of assertion. Simply put, the imperfective is used when the assertion time overlaps with the action it describes, i.e. the source state, leaving the target state out of consideration. The perfective is used when assertion time encompasses the entire lexical content, i.e. includes the target state. However, the choice for one aspect over another seems to be much more limited than expected on this account, as Divjak et al. (2024) showed: knowing which lexical time is used to report on a situation and knowing when the situation happened with respect to the moment of speaking predicts aspectual use correctly in more than 92 % of all cases in the imperfective and in 86% of all cases in the perfective.

In addition to the grammatical and lexical approaches, some accounts centre on the discourse functions that aspect fulfils (Hopper 1982; Thelin 1990); these approaches consider the function of aspect to be primarily one of discourse organisation, and only secondarily as a property of individual propositions. In functional discourse approaches, aspect is usually analysed in terms of the notions of 'foreground' and 'background'. Foregrounded properties are typically attributed to or equated with the story line of a narrative discourse, which depicts events in a sequence. The scenery, in which the story line is embedded, constitutes the background. Furthermore, in the successive predications that constitute part of a cohesive text, situations may not only be presented in a *sequence*, as typical for the story line, but also as *simultaneous* or as intersecting each other, known as *incidence*.

As we can see, apart from Klein's account, these concepts are rather abstract if not metaphorical and at the same time, they appear related to each other, if not

synonymous (Janda 2004). Crucially, it must be borne in mind that the concepts that have been proposed are merely ‘points of orientation’ that can be fully instantiated, further supplemented or suspended in specific instances of usage (Łaziński 2020: 120). While Russian has a prolific tradition of describing ‘specific’ aspectual meanings, such as ‘general factual’ for the imperfective (where emphasis is shifted away from the result, making an imperfective possible), this approach has remained marginal for Polish (Łaziński 2020: 121), and Dickey (2000) reminds us that accounts developed for one Slavonic language should not be unquestioningly transposed to another.

What has generated much research in Polish aspectology is the observation that grammatical aspectual distinctions do not affect all verbs in the same way. These lexical peculiarities seem to be naturally accommodated by classifications of situation types, a.k.a. lexical approaches to aspect, for which the meanings of the verbs themselves are the point of departure. However, the exact number of categories that can – and should – be distinguished differs between authors. One of the foundational views is concerned only with the notion of an endpoint, and much like the grammatical approach, distinguishes only two classes: telic and atelic verbs (Garey 1957). However, Vendler (1957), whose classification of events is probably most influential in the studies of lexical aspect, proposed four classes – states, activities, achievements and accomplishments –, based not only on the notion of endpoint but also on the way in which this endpoint is achieved. Events that lack an endpoint can be static (states) or dynamic (activities), while events that do have an endpoint can reach that endpoint instantly (achievements) or gradually (accomplishments). Focusing on the dynamicity, telicity and duration of events, Comrie (1976) and later Smith (1997) also distinguish *semelfactives*, i.e. events that do not have endpoints, are dynamic and punctual, i.e. occur instantly.

The lexical approach has also been used to classify events in Polish where it has been found to crosscut the aspectual categories of the grammatical approach. The leading approach, described in Laskowski (1999), distinguishes two main classes – states and dynamic actions, further divided into eight categories based on the combination of four semantic features: dynamicity, change of state, telicity and control (of the subject). Laskowski also shows that only verbs belonging to the two groups that capture telic verbs, that is actions (which have all of the described features) and processes (which are not controlled), can be said to participate in the coveted ‘neutral’ aspectual opposition (where the semantic difference between imperfective and perfective forms is restricted to the meaning difference triggered by the aspectual difference) that is key to grammatical approaches. In other cases aspectual pairs can be formed, but the meaning changes between the aspects. Such lexical-aspectual relations, unlike aspectual relations of telic verbs, are not

regular and depend on the semantics of the verbs. This observation represents the long-established insight that Slavonic-style aspect cannot be fully understood without a clear understanding of grammatical *and* lexical aspect, and how these two dimensions interact.

Despite the difficulties in pinning down aspectual meaning, in Slavonic, there are compelling morphological and distributional reasons for distinguishing a binary category. Recall that Polish verbs necessarily express grammatical aspect and there are clear restrictions on use. At the same time, it is well-known that the concepts that have been proposed are merely ‘points of orientation’. Hence, the questions of learnability and cognitive plausibility of grammatical aspect categories are particularly important for Slavonic linguistics. Next, we will review behavioural studies exploring whether grammatical categories are cognitively plausible (Section 1.3) and whether there is enough structure in the usage of perfective and imperfective for users to learn the categories from input alone (Section 1.4).

1.2 Grammatical aspect and behavioural studies

Existing experimental evidence for English suggests that perfective and imperfective affect conceptualisation of events differently. Madden and Zwaan (2003) showed that participants who read English sentences in the past simple tense chose pictures showing a complete event more often and faster. Morrow (1985) showed that people were more likely to place a character along a path rather than at a goal when they heard a progressive description of a situation. Similarly, Anderson et al. (2008) reported that, having heard progressive sentences, participants were more likely to place the human character at the centre or the beginning of the path. Having heard past simple sentences, however, they were more likely to place the character towards the end of the path.

Other studies have shown that grammatical aspect influences what mental models users form of the events and have demonstrated that these mental models affect the accessibility of other components of the event. For example, Ferretti et al. (2007) demonstrated for English that participants were faster to read the names of locations (e.g. ‘arena’) when they were related to the progressive rather than the simple past verb phrase primes they saw before the name (e.g. ‘was skiing’). The results suggest that other elements of the event, such as locations, are more activated when the situation is presented as ongoing and described using an imperfective aspect. Other studies have found similar activation effects for participants and instruments (Carreiras et al. 1997; Madden and Zwaan 2003; Magliano and Schleich

2000) and Golshaie and Incera (2020) reported, for Persian, that participants were more likely to erroneously indicate that an instrument was mentioned in the sentence if the sentence contained an imperfective verb.

Unfortunately, despite the vast amount of descriptive literature on Slavonic aspect, the study of how the highly grammaticalised version of aspect that typifies the Slavonic languages is processed and represented mentally is in its infancy (Sekerina 2006, 2017). Exceptions here are studies using self-paced reading on Polish data (Klimek-Jankowska et al. 2018), eye-tracking on Russian (Bott and Gattnar 2015; Klimek-Jankowska et al. 2018) and EEG on Polish (Błaszczak and Klimek-Jankowska 2016; Błaszczak et al. 2014; Klimek-Jankowska and Błaszczak 2020). Yet the findings of these studies have remained equivocal. For example, Klimek-Jankowska and Błaszczak's (2020) findings support the view that the domain of aspectual interpretation in Polish is the verb phrase, not just the verb; that the verb phrase would play a role is expected on a lexical aspectual approach. Yet, earlier, Bott and Gattnar (2015), using eye-tracking, had reached the opposite conclusion for Russian where the domain of aspectual interpretation appears to be just the verb.

Taken together these studies suggest that grammatical aspect is not merely a descriptive category proposed by linguists but also corresponds to a cognitively plausible distinction made by L1 users. However, the studies discussed above are mainly concerned with differences in conceptualisation and comprehension, not with learnability. That is, they already assume categories of aspect, without asking how these categories come to be. In this study, we approach the problem differently, assuming a usage-based learning position (Divjak et al. 2021). Therefore, we first ask what categories would be learned based on usage, and only then validate our models experimentally. The next section briefly outlines usage-based theory and discusses usage patterns of aspect.

1.3 Usage-based linguistics and aspect

Usage-based linguistic theory posits that language knowledge emerges from domain-general cognitive mechanisms (such as abstraction, categorisation and imagination) working on input. This allows a (language) system to be built from experience. This system is sensitive to the properties of experience, and hence probabilistic in nature. The distributional properties of input are exploited in the process of language development (e.g. Saffran et al. 1996) and the frequencies with which elements co-occur have been shown to influence learning, comprehension and production (for a review see Divjak 2019). To put it simply, things that go together are perceived to be

belonging together and users start to expect and use one whenever they see the other.

Patterns, biases and preferences can also be found in aspectual usage. In Section 1.2, we discussed one example of a distributional constraint – the fact that in Polish perfective verbs cannot be used in the future using a compound tense. However, research shows that there is more to the distributional story. First of all, a number of studies show statistical biases of aspectual use with certain elements of context. As many descriptive grammarians have noted, aspectual choice is aided by the use of temporal adverbials. Koranova and Bermel (2008) used the Czech National corpus to investigate the aspectual preference of tense-aspect indicators such as conjunctions, adverbs and temporal adverbial phrases. They showed that a number of these indicators reveal an aspectual preference and may serve as indicators of the choice reliably enough to be used as ‘helpers’ in teaching aspect to L2 learners. Similarly, Reynolds (2016) showed that temporal adverbial phrases tend to be reliable predictors of aspectual choice in Russian. He notes, however, that despite their reliability, those cues tend to appear very infrequently – a finding that Koranova and Bermel could not arrive at, since they only considered how often each marker appears with a given aspect, but not how often aspects appear without any tense-aspect indicator. The availability of adverbial cues is indeed very low, and Reynolds (2016) estimates it to occur in only 2 % of all sentences. Nonetheless, it is assumed that these elements can be used by learners in the process of acquiring the perfective and imperfective distinction: learners might pick up on the fact that some verbs are used with one type of adverbial, but not the other, and therefore form one group. Divjak et al. (2024) provide a learning-based justification for the low incidence of these adverbial cues, showing that they only occur in conjunction with verbs that occur approximately equally frequently in both aspects and hence do not have an aspectual preference themselves.

However, a perfect correlation is not necessary to form strong associations between elements in the environment. Clear biases in usage – reliable but not perfect – can also facilitate the process of categorisation, including categorising forms as perfective or imperfective. Corpus studies on aspect show several strong tendencies in aspectual use. Rice and Newman (2004) showed that in English, certain prepositions, such as *around*, *over* and *about*, show aspectual preference: they tend to appear with either progressive or stative forms. Jurkiewicz-Rohrbacher (2019), who conducted a corpus analysis of Polish aspect and its Finnish correlates, showed that the type of reflexive pronoun, the semantic role of the subject and object and even the person and number of the subject may matter for aspectual choice in Polish, as the statistical analyses show biases towards one or the other aspect. The fact that these statistical biases exist means that there is potentially enough information for learners to acquire categories of grammatical aspect. Learning what in the context

serves as a good cue and what verb it predicts might help them notice which verbs share similar cues and, therefore, facilitate the formulation of more general categories, such as grammatical aspect.

1.4 Learning from usage and the Naive Discriminative Learner

The idea that co-occurrences and their frequencies play a key role in L1 language learning is not new. Decades of research have proposed and tested various frequency-based measures (for a review see Pecina 2010; Divjak 2019) that would link frequency distributions to usage preferences. However, the goal of cognitively oriented research should not merely be to identify a high-performing modelling method, but to develop one that mirrors the processes of a typical language user. In other words, modelling algorithms should be grounded in what is known about cognition, and here we more specifically pursue their grounding in the biology and psychology of learning (Divjak and Milin 2023).

To model what could be learned from exposure to language, and devise a cognitively realistic way to link the frequency distributions that L1 users are exposed to with their usage preferences, we employ the well-known Rescorla–Wagner model (Rescorla and Wagner 1972; see also Widrow and Hoff 1960), as implemented in the computational modelling framework known as Naive Discriminative Learning (NDL: Baayen et al. 2011; Milin et al. 2017a, 2017b; for an overview see Divjak and Milin 2023). We begin with a brief presentation of the Rescorla–Wagner (RW) model, focusing on its relation to the domain-general learning mechanisms postulated in usage-based approaches; technical details are provided in the Methods section.

In essence, the RW model learns gradually through iterations of newly encountered information about relationships between events in the environment (Rescorla 1988). These events, referred to as *cues* and *outcomes*, may exhibit more or less systematic co-occurrences, which are encoded as continuously evolving cue–outcome association weights. If, at the first *learning event*, a cue A appears with an outcome X, the weight between them will increase. But if the next time the cue A is encountered with another outcome Y, this learning event will have two important consequences: first, the association weight between A and Y will increase; at the same time, the weight between A and X will decrease. These adjustments capture the relationships exhibited by cue A with the two outcomes. Naturalistic learning often involves many cues and outcomes, representing multiple co-occurring events in the environment.

There are three additional characteristics of learning according to the RW rule. First, the outcomes are treated as strictly independent, which is what is ‘naïve’ about NDL. Second, the cues compete to become associatively relevant, making the process

‘discriminative’. Third, as the number of cues increases, positive adjustments become smaller while negative adjustments grow larger. This dynamic is likely why this type of learning mechanism is referred to as ‘error-correction’. Nevertheless, individual adjustments are typically rather small which in a way pays respect to previously accrued experience, known as the principle of minimal disturbance in machine learning (Haykin 1999). Over time, associations become more stable and reliable and the learner becomes an experienced and fluent operator in a given environment, able to use knowledge about what can or cannot happen if a given cue is present.

There is plenty of evidence that this deceptively simple learning mechanism accurately models animal and human behaviour, from learning to associate a tone with shock (Rescorla 1988) to category learning (e.g. Gluck and Bower 1988; Wasserman et al. 2015). More importantly, it has also been successfully applied to conceptualise language learning (Ellis 2006) as well as to model it computationally (Baayen et al. 2011; Divjak et al. 2021; Milin et al. 2017a, 2017b, 2023). As noted above, usage-based linguistics presumes that language is learned from experience, relying on domain-general mechanisms that are sensitive to frequency. The learning process described by the Rescorla–Wagner model is a strong candidate for such a mechanism: the contingencies between cues and outcomes across various cognitive domains can be learned based on the same general principle, where properties of experience, such as frequency, are essential. The ‘knowledge’ is experiential and, thus, continuously changes with new evidence.

This is how the RW model can help us understand how language knowledge emerges from the flux of experience, and why it is difficult to extract from experience a parsimonious set of general and crisp rules. The model also shows why simple counts do not offer the complete picture as the strength of cue–outcome associations is influenced not only by presence of co-occurrence but also by absence of co-occurrence, i.e. the likelihood of the outcome when the cue is absent (cf. Ellis 2006; Milin et al. 2017a, 2017b). Despite such complexity and dynamics, experience allows for structured and systematic relationships to emerge (cf. Milin et al. 2023). Learning models resonate well with what is assumed in cognitive, usage-based linguistics (e.g. Boyd and Goldberg 2011): input alone provides enough evidence to learn what is and what is not possible in language.

Slavonic aspect is notoriously elusive for most second-language learners, which makes it an ideal test case for the usage-based emergence of learnable relationships from mere input. Divjak et al. (2024) report that the aspectual preference at the lexeme level is so strong that aspectual usage can be effectively explained using lexical information, combined with information about the relationship between the time of the event and the time of reporting on it. It is thus possible that L1 users of Polish use perfective and imperfective verbs accurately, not because they have

identified an invariant meaning of these aspects and consciously choose how to encode events but simply because they have learned which lexical and contextual cues are associated with each aspectual form.

In what follows, we present an empirical study testing this idea. First, we ask whether language users need to know the grammatical category of aspect at all, and we test the performance of a model that relies solely on the co-occurrences of verb forms and contextual cues. Next, we investigate how well aspectual distinctions can be learned based on the conceptual definitions of invariant meanings proposed in the literature. Finally, we test the performance of a model trained to predict aspect from contextual cues. We then validate our models against behavioural data collected through an online survey in which participants were asked to fill gaps in discourse chunks by selecting one of the two provided aspectual variants.

2 Corpus-based modelling

2.1 Data extraction and annotation

We relied on Polish Wordnet¹ to obtain an extensive list of aspectual pairs, as it is the largest database containing both information on the aspectual links between the verbs as well as annotation of the type of their aspectual relation (i.e. ‘pure’ or ‘secondary’). Based on the annotations of the aspectual relationship included in Wordnet, we retained only ‘pure’ pairs. That is, if an imperfective verb had multiple perfective counterparts, we retained the one that did not introduce any additional shades of meaning, so that the difference between the imperfective and perfective variant was only in the semantics of aspect. For instance, we considered *pisać* (write.IMPF) – *napisać* (write.PERF) to be a ‘pure’ aspectual pair, whereas pairs such as *pisać* (write.IMPF) – *podpisać* (sign.PERF) were excluded, because adding the perfective prefix *pod-* also modified the meaning of the verb. Next, we also excluded pairs for which we could not reliably count the occurrence frequencies by looking at the word form due to polysemy; this was achieved by inspecting and dropping duplicates. For instance, *stać* can either be an imperfective meaning ‘stand’ or a perfective meaning ‘become’. Finally, to ensure that our sample was representative, for each pair of verb lemmas on the filtered list, we calculated the so-called perfective bias, which we defined as the proportion of the perfective lemma in the sum of occurrence frequencies of the aspectual pair; if a lemma has a perfective bias of 0.38,

¹ The list of aspectual pairs along with the additional information that allowed for filtering was extracted from the Polish version of Wordnet (Słowniec) by mgr.inż. Tomasz Naskręt and dr Agnieszka Dziob.

it occurs in the perfective in 38 % of all cases and in the imperfective in the remaining 62 %. Because the lemmas in the tags of the annotated version of the Araneum Corpus of Polish (Benko 2014) are not always correct, we first listed, then counted the occurrences of all the inflected forms of the verb and summed them to obtain lemma frequencies. The final verb list contained 3,324 pairs of verbs; of these, 456 pairs had a very strong bias towards the perfective, occurring in the perfective in at least 90 % of all cases. We sampled along the continuum of perfective bias to control for the effect of relative frequency and selected 9 aspectual pairs (18 verbs, 9 imperfective and 9 perfective) using the `bandsample` function from the `pyndl` python package (Sering et al. 2022), similar to the application in Divjak et al. (2021). The selected verbs, their frequencies and perfective bias are given in Table 1. Note that all of these pairs are suffixation pairs, which may differ in behaviour from the so-called ‘true’ pairs with simplex imperfectives in terms of ‘perfective bias’.

For each of the 18 verbs, we sampled 100 sentences that contained the verbs of interest and up to six preceding sentences (see `SupMat_A` for further details). The exact number of sentences depended on the number of preceding sentences available: if the target sentence in the chunk happened to be the first sentence in the document, there was no preceding context, and the chunk consists of only the target sentence. The size of the chunk was decided through reference to memory studies: Light and Anderson (1985) suggested that the proportion of correct recall of a pronoun referent in the text decreases with distance and is only slightly above chance level for distances of 6–7 sentences.

With the help of information in the preceding context, the sample of 1,800 target sentences was annotated for 30 variables to create distributional Behavioural Profiles (BPs) of the verbs (Divjak and Gries 2006). In addition, they were annotated for

Table 1: The final sample of 18 verbs.

Lemma	Translation	Lemma frequency	Aspect	Competitor	Competitor lemma frequency	Competitor aspect	Perfective bias
rozliczyć	square/appraise	9,467	perfective	rozliczać	15,500	imperfective	0.38
urzec	charm	3,364	perfective	urzezać	3,647	imperfective	0.48
nabrać	gain/kid	25,432	perfective	nabierać	21,423	imperfective	0.54
rozczesać	comb	624	perfective	rozczesywać	421	imperfective	0.60
wywrócić	topple	2,504	perfective	wywracać	1,382	imperfective	0.64
zastąpić	replace	50,153	perfective	zastępować	22,092	imperfective	0.69
podmienić	switch	1,782	perfective	podmieniac	593	imperfective	0.75
opracować	develop	69,870	perfective	opracowywać	14,542	imperfective	0.83
odszukać	find	44,391	perfective	odszukiwać	580	imperfective	0.99

the seven abstract semantic distinctions most commonly used in the aspectual literature. We list the variables below and refer to SupMat_B.1 for the annotation manual.

2.1.1 Concrete, usage-related variables

The set of 30 concrete, usage-related variables can be divided into three groups that capture the discourse in which the verb form is used by describing the verb, its immediate context and important elements further back in the chunk, using a Behavioural Profiling approach (Divjak and Gries 2006). The first group captures the properties of the verb, as used in the target sentence. We annotated for tense, mood, presence of auxiliary verbs and reflexive pronouns. The second group describes the features of the clause the target verb is used in. We marked whether the verb is negated or not, and whether the clause is main or dependent. We also annotated for the properties of agents, patients and recipients, if present. We decided to use these labels instead of the syntactic counterparts subject, direct object and indirect object because the variables were intended to capture features that are distinguishable for L1 users who lack linguistic training. The third group of contextual variables covers elements in the wider chunk rather than in the immediate context of the verb. We annotated for adverbials, if they were related to the verb of interest, and aspectual triggers – expressions that are known to be used often or exclusively with only one aspect. The list of triggers was compiled on the basis of information extracted from two grammar books for learners of Polish: Swan (2002) and Sadowska (2012) and is included as SupMat_B.2.

2.1.2 Abstract aspectual-semantic variables

The set of abstract variables captures semantic dimensions presented in the literature as possible candidates for the invariant meaning of aspect. Given the abundance of distinctions, we decided to focus on the most prominent ones, following Janda (2004). From the grammatical approaches we adopted *boundedness*, which describes whether any boundary (i.e. beginning or end) of the action is signalled (e.g. Smith 1986), with unbounded events extending indefinitely in time on both ends; *totality* indicates whether the stages and development of the event are important or whether the event is treated as one indivisible thing (e.g. Comrie 1976); *perspective* goes back to Comrie's (1976) remark on imperfective aspect showing an action from 'within' as opposed to perfective aspect that presents an action from a different temporal perspective; *specificity* captures whether the actions are distinct, performed by identifiable individuals and can be located on the timeline (Divjak 2009). From the lexical tradition, we include *resultativeness*, which describes whether the action

resulted in a change of state (e.g. Laskowski 1999). Other variables capture discourse-related consideration, i.e. *foregrounding* evaluates whether the action is presented as the main event on the timeline or rather serves as a backdrop against which other actions move the plot forward (e.g. Hopper 1982; Thelin 1990); *sequentiality/simultaneity* captures whether the action described happens alongside other actions or whether the actions form a sequence of events (e.g. Hopper 1982; Thelin 1990).

The annotation for abstract variables entailed an increase in the degree of subjectivity that might affect the replicability of our findings. The inter-annotator agreement for a random sample of 100 chunks was rather low, with the best scores in the low ranges of ‘fair’ (see also Divjak et al. 2015); for more details, see SupMat_B.3. However, that disagreement is not necessarily negative. Different users can have different abstractions that are useful to them in different ways and – even when given an operationalization – they can still diverge in how they interpret each case. When discussing the annotation, the two annotators could easily come up with several interpretations on which both could agree.

2.2 Modelling

The data were modelled with the Naïve Discriminative Learner (NDL; Baayen et al. 2011), which implements the Rescorla–Wagner rule at scale: for each cue (*i*) and outcome (*j*), the association weight (*w*) at the next time step (*t* + 1) is a result of association at the current time step (*t*) and the change – an increase or decrease (Δw):

$$w_{ij}^{t+1} = w_{ij}^t + \Delta w$$

The change is critical for learning, and for this particular rule, it depends on whether a cue and an outcome co-occur or not, as captured by the equations used for calculating the changes in association weights presented in Table 2.

When a cue is absent – which effectively means that there is no new information – there is no learning, and no change. If cue and outcome co-occur, their association is (re)affirmed and strengthened. Otherwise, if a cue is present but the

Table 2: Equations used in NDL for calculating the changes in association weights (Baayen et al. 2011; adapted from Divjak and Milin 2023).

the cue is absent	no change	$\Delta wt = 0$
the cue is present; the outcome is present	positive evidence; association strengthened	$\Delta wt = \gamma(1 - \sum(w \cdot j))$
the cue is present; the outcome is absent	negative evidence; association weakened	$\Delta wt = \gamma(0 - \sum(w \cdot j))$

outcome is not, such negative evidence weakens the association. The change in association strength is proportional to the difference between that outcome's presence (1) or absence (0) and the sum of weights for all relevant cues to that outcome ($\sum(w_j)$). In effect, that sum of weight represents all the support a given outcome receives from the cues in the (linguistic) environment. The final component of this learning rule is a free parameter, the *learning rate* (γ), that controls how gradual or abrupt, smooth or fidgety the learning should be. If the principle of minimal disturbance is respected, then the learning rate is suitably small (Haykin 1999).

In our dataset, a learning event was defined as a row in which one outcome was present (i.e. a perfective or imperfective verb form). As available cues, the values of annotated linguistic categories were used (e.g. AgentNumber:Plural). The complete dataset, i.e. the annotated corpus, was divided into a training, test and validation set. The training and test sets were randomly split in a 90:10 % ratio, and the validation set was a 10 % random sample from the training set. The same test set was used to determine the accuracy of all the models. Each model was trained with the learning rate set to $\gamma = 0.0001$, for 500 repeated runs, randomizing the order of the learning events each time; i.e. after going through all the learning events in the training set, the model went back to the beginning and continued learning on the same dataset presented in a different order, continuing to update the association weights. This allows to counter any order effect and yields a stable model, i.e. a convergence of weight strength.

Given the types of annotated variables – the cues, concrete or abstract – and lemma- or aspect-based outcomes, we trained three different models.² Table 3 details what combinations of cues and outcomes were used for each model.

Table 3: Cues and outcomes used for each of the models trained to predict aspectual labels or lemmas.

Model	Cues	Outcomes
outcome-cue		
Lemma-concrete	Values of the contextual variables for each chunk (e.g. AgentNumber.plural)	All 18 lemmata, listed in Table 1
Aspect-concrete	Values of the contextual variables for each chunk (e.g. AgentNumber.plural)	Both aspects (perfective, imperfective)
Aspect-abstract	Values of the abstract variables for each chunk (e.g. Boundedness.bounded)	Both aspects (perfective, imperfective)

² We trained further three models to evaluate whether it is possible to acquire the abstract semantic labels, proposed in the literature, from contextual information. For details, see SupMat_C.

After training, the three models were probed against the test set. For the aspect-concrete and aspect-abstract models, accuracy was calculated in the same way. First, we obtained the activations of outcomes by summing the weights of present cues for each outcome (perfective and imperfective). The outcome with the highest activation was taken as the model’s preferred outcome. Then, we calculated the percentage of cases in which the model’s preference matched the aspect used in the annotated chunk. The lemma-concrete model had to be evaluated in a slightly different way: the model had to choose from 18 outcomes (lemmas), instead of 2, as in the case of the other models. To make the task comparably difficult for all three models, we limited the set of choices and calculated the accuracy of the lemma-concrete model by comparing only the activations of the original lemma and its counterpart. This way, the model was ‘choosing’ only between the perfective lemma and its imperfective counterpart, without any information of their relation or the aspect of their form.

2.3 Modelling results and discussion

Table 4 presents the accuracies of the main models trained to learn either lemmas or aspectual labels. Overall, all three models perform quite well: the aspect-abstract model (87.22 %) performs slightly better than the context-based model of aspect (aspect-concrete; 85 %), while the model using lemmas as outcomes was the least accurate of the three (lemma-concrete; 79.44 %). To determine whether any model’s performance is superior, we conducted Bayesian A/B tests comparing success rates, using priors set to 0.01 for no difference and 0.99 in favour of the model with the higher success rate (Hoffmann et al. 2022). Only the difference between the aspect-abstract and lemma-concrete models suggested a slight edge for the former, better-performing model, with the Bayes Factor indicating that the aspect-abstract model is

Table 4: Summary of the performance of the three models trained to learn either lemmas or aspectual labels.

Model	Overall accuracy	Accuracy per aspect	Significance of difference
Lemma-concrete	79.44 %	Perfective: 84.44 % imperfective: 74.44 %	$BF_{10} = 2.158$
Aspect-concrete	85 %	Perfective: 88.89 % imperfective: 81.11 %	$BF_{10} = 1.671$
Aspect-abstract	87.22 %	Perfective: 76.67 % imperfective: 97.78 %	$BF_{10} = 1748.194$

more than three times more likely to outperform the lemma-concrete model ($BF_{10} = 3.252$). The remaining comparisons – aspect-abstract versus aspect-concrete and aspect-concrete versus lemma-concrete – did not show a clear preference for the more successful model ($BF_{10} = 0.494$ and $BF_{10} = 1.168$, respectively).

Interestingly, both models based on concrete variables seem to be better in using the perfective, while the model trained on semantic abstractions performs well for the imperfective (almost 98 %) and underperforms on perfectives (76.67 %). The same Bayesian A/B test indicated that the performance difference per aspect was significant only for the aspect-abstract model, which performed notably better on the imperfective ($BF_{10} = 1748.194$). The other two models showed no significant difference in success rates across the two aspects ($BF_{10} = 2.158$ and $BF_{10} = 1.671$, for lemma-concrete and aspect-concrete models, respectively).

It seems then that imperfective is easier to learn when we use semantic distinctions as cues. This finding is particularly interesting in the light of the issues discussed in the first sections of this paper where we saw that the debate regarding the invariant meaning focused on the perfective aspect: all the proposed distinctions were said to capture the semantics of the perfective, as a marked form. However, the uneven performance of the aspect-abstract model suggests that it is the meaning of the imperfective which can be more reliably described using semantic labels, at least in usage. While we cannot pinpoint the cause for this discrepancy with the data we have, it is possible that this effect is the consequence of the more restricted applicability of the perfective, which in turn leads to a prioritised focus on distinguishing the unmarked and more broadly applicable imperfective form—an outcome consistent with a learning-based approach.

More generally, the models trained on our corpus data paint a complex picture by supporting both usage-based and abstract label-based accounts. Specifically, our data show that it is possible to achieve high performance in determining the appropriate aspect of a verb in a given context using either concrete cues that describe various dimensions of usage or abstract semantic labels that capture the stable definitions of aspect. However, corpus-based studies can only reveal whether patterns are present in the data, which is essential for hypothesis formation. Experimental confirmation is required as the next step to determine whether language users are sensitive to these patterns, learn them, and apply them in use.

3 Behavioural validation

To verify empirically which of the three corpus-based models best captures the actual behaviour of language users, we conducted a two-alternative forced-choice (2AFC) gap-filling study, where participants selected either the imperfective or

perfective form of a verb given a context. In this task, participants were presented with appropriately adapted paragraphs taken from our annotated corpus sample and were asked to choose one of the aspectual variants of the verb. Their choices were also predicted with each of the three models – lemma-concrete, aspect-concrete, and aspect-abstract – to evaluate the predictive power of the three learning models, each based on different sets of input cues. The study design allowed us to identify patterns in users' aspectual choices without needing to ask them directly, as L1 users often find it difficult to articulate the reasoning behind their linguistic decisions. By observing their natural choices, we could gain insights without relying on potentially unreliable introspective self-reports.

The task was inspired by Janda and Reynold's study (2019), where participants evaluated aspectual choices in context. However, we introduced two important changes. First, since the accuracy for the models was calculated by selecting only one of the options, we wanted to limit participants' choices in a similar way in order to make comparisons between the models and language users more direct. Secondly, we were interested in which of the verbs fit the context best. If both the perfective and the imperfective are equally acceptable, this should be visible from the proportion of answers to each question – the more balanced the answers, the stronger the indication that both verbs can be used in this context. However, the question that seemed more interesting to us was which of the verbs is preferred, and why. After all, during production we do not have the luxury of using both aspects.

3.1 Stimuli selection and preparation

To minimise participant fatigue and reduce the likelihood of incomplete surveys, a subset of 10 lemmas (5 pairs) was selected along the perfective bias continuum. The potential pool of stimuli for these lemmas was restricted to chunks from the annotated corpus used in the past or future tense, as the present tense allows only imperfective forms. Additionally, only chunks containing morphological forms with both imperfective and perfective options were included. For example, chunks where the target was a perfect or present adverbial participle were excluded.

The selection of items was guided by the lemma/aspect activations from the three learning models. First, network activations were calculated for each lemma/aspect and its competitor. Second, these activations were transformed into probabilities using the softmax function. Third, the odds ratios for lemma/aspect versus competitor were calculated. Finally, one randomly chosen chunk per lemma was selected for low (between the 0.125 and 0.375 quantiles) and high (between the 0.625 and 0.875 quantiles) odds ratios. This selection method ensured that the sample

included at least one instance of each lemma where each model strongly supported the original verb and at least one instance where it strongly supported its alternative.

The final sample consisted of 60 chunks. These chunks were manually prepared for the survey by replacing the target verbs with a dotted line of equal length in each sentence. The number of preceding sentences was limited, as 1–2 sentences of context generally provided sufficient information to annotate the sentences for our set of variables (mean = 1.78, mode = 1). For an example, see SupMat_D.

3.2 Participants

The study was conducted using Qualtrics software (Qualtrics 2024), and L1 participants residing in Poland were recruited via social media. All participants completed the full set of 60 gap-filling items, with both the order of questions and the order of response options randomised. A total of 77 complete responses were collected ($n = 57$ female, $n = 2$ undisclosed gender). Educational backgrounds varied, with 49 participants holding higher education degrees, 26 having completed high school, and two having vocational education. Participants ranged in age from 18 to 64 years, with a mean age of 33.61.

3.3 Model evaluation and discussion

To evaluate how well each of the three models predicted participants' choices, we fitted three Generalized Additive Mixed Models (GAMMs) using a binomial (logistic) link function. GAMMs handle continuous predictors more flexibly than traditional linear or generalised linear models by allowing for nonlinear relationships using smooth functions, while their mixed effects aspect allows them to capture variations across individuals or groups, repeated measurements and other hierarchical or nested designs (Wood 2017). The analysis utilised the **mgcv** (version 1.9-1: Wood 2011, 2017) and **itsadug** (version 2.4: van Rij et al. 2022) packages in **R** environment for statistical computing (version 4.4.2: R Core Team 2024). All stimuli, raw responses and model predictions are presented in Materials; 'correct' indicates that the originally used aspect was retrieved by a participant.

The target predictors were the network's activation and diversity. Activation quantifies the support an outcome receives from the cues that are present, while diversity reflects the competition among these cues for all possible outcomes (Divjak et al. 2021; Milin et al. 2017a, 2017b). The primary control variable was the originally used aspectual form, included both as a main effect and in interaction with the

activation and diversity covariates. Random effects of item and participant were also tested and retained when statistically significant.

Activation and diversity were rank-to-normal transformed to adjust their spiky distributions to a more symmetric, Gaussian-like shape, facilitating statistical modelling and interpretation of the results (for details on spiky distributions of NDL-network weights, see Milin et al. 2017a, 2017b; for rank-to-normal transformation, see Aulchenko et al. 2007; Johnson 1949). The full dataset consisted of 4,620 data points, of which 539 (11.7 %) were visually identified and removed as extremes, following the guidelines in Baayen and Milin (2010). This resulted in a final dataset of 4,081 data points (88 % of the original data).

Finally, for both categorical variables – accuracy in retrieving the original aspect (the dependent variable) and the original aspect itself (a fixed factor) – sum-to-zero (sigma) coding was applied to facilitate result interpretation (Fox and Weisberg 2018). For all three learning models, the nonlinear interaction of activation and diversity was strongly supported; hence, the template model specification was as follows:

```
gam ( OriginalAspectRetrieved~
  OriginalAspect+
  te (Activation[Lemma/Concrete/Abstract], Diversity[Lemma/Concrete/Abstract],
      by = OriginalAspect, k = c(3, 3))+
  s (Item, bs='re')+
  s (Participant, bs='re'),
  family = binomial,
  method='ML',
  ...)
```

with a tensor product of the two predictors of support (activation) and competition (diversity) by original aspect, allowing a moderate nonlinearity by specifying the respective smooth term with three knots for each component ($k = c(3, 3)$). As none of the three tested models – each using activations and diversities from their respective learning networks (lemma-concrete, aspect-concrete, and aspect-abstract) – showed a significant effect of by-participant adjustments, this random-effect term was removed, and the models were refitted. Conversely, by-item adjustments were justified, albeit only weakly for aspect-concrete predictors ($p = 0.0213$) compared to the other two statistical models ($p < 0.0001$).

The models are visualised in Figure 1 (lemma-concrete), Figure 2 (aspect-concrete) and Figure 3 (aspect-abstract). Comparatively, among the three tested models, the best overall fit was obtained with the lemma-concrete learning measures,

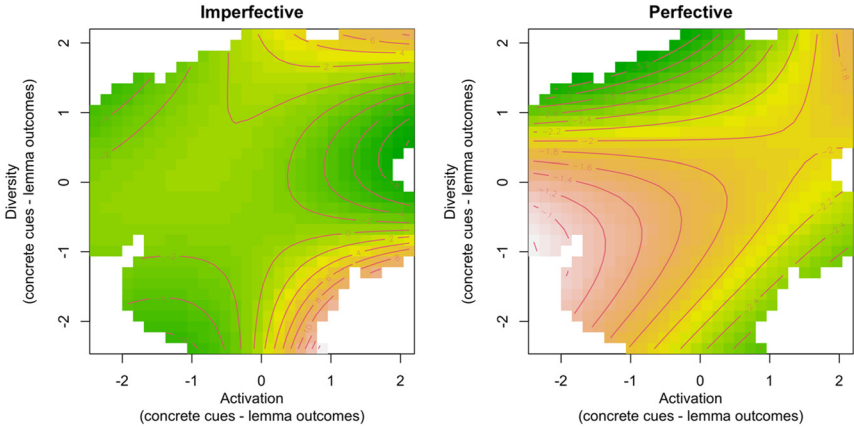


Figure 1: GAMM smooth terms for lemma-concrete activation and diversity on gap-filling task accuracy in retrieving the original verbal aspect. The left panel depicts effects for trials where imperfective was the original choice, while the right panel shows effects for trials where perfective was the original choice. Lower accuracy is presented in green, and higher accuracy in yellow-to-ochre colours. White patches indicate no activation/diversity data.

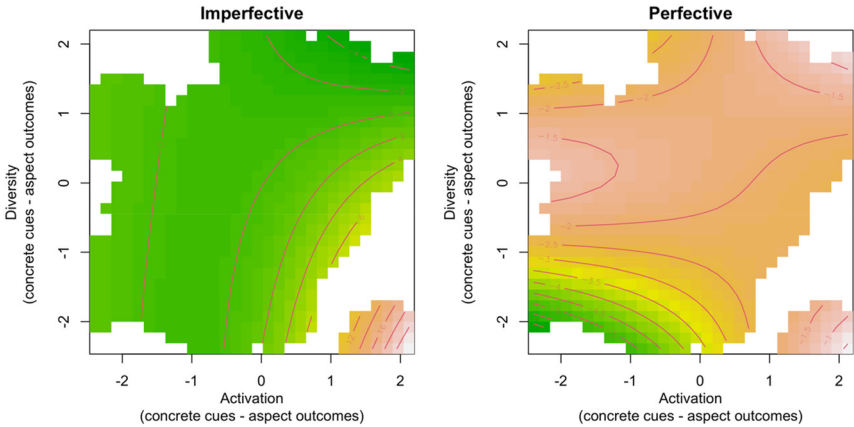


Figure 2: GAMM smooth terms for aspect-concrete activation and diversity on gap-filling task accuracy in retrieving the original verbal aspect. The left panel depicts effects for trials where imperfective was the original choice, while the right panel shows effects for trials where perfective was the original choice. Lower accuracy is presented in green, and higher accuracy in yellow-to-ochre colours. White patches indicate no activation/diversity data.

followed by the aspect-abstract and aspect-concrete measures of support (activation) and competition (diversity). The respective Akaike Information Criterion (AIC) values were 3,891.882, 3,954.420, and 4,036.741. These values suggest an Evidence

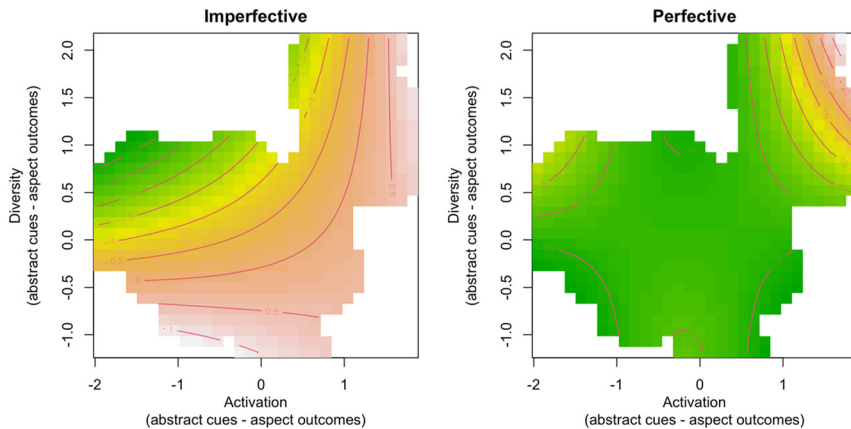


Figure 3: GAMM smooth terms for aspect-abstract activation and diversity on gap-filling task accuracy in retrieving the original verbal aspect. The left panel depicts effects for trials where imperfective was the original choice, while the right panel shows effects for trials where perfective was the original choice. Lower accuracy is presented in green, and higher accuracy in yellow-to-ochre colours. White patches indicate no activation/diversity data.

Ratio ($ER = \exp((|AICA - AICB|)/2)$) strongly favouring the lemma-concrete model over the aspect-abstract model ($ER = 3.80e+13$), as well as the aspect-abstract model over the aspect-concrete model ($ER = 7.51e+17$).

The three models are similar in showing a stronger effect of activation by diversity (i.e. tensor product) on retrieving the originally used aspect if that was an imperfective (aspect-abstract: Chi-square = 165.30, $p < 0.0001$; aspect-concrete: Chi-square = 154.14, $p < 0.0001$; lemma-concrete: Chi-square = 210.17, $p < 0.0001$) rather than a perfective (aspect-abstract: Chi-square = 55.79, $p < 0.0001$; aspect-concrete: Chi-square = 38.65, $p < 0.0001$; lemma-concrete: Chi-square = 18.52, $p = 0.0016$). The exact patterns of these interaction effects, that is, where and how they reveal the change in participants' retrieval accuracy, differ between the three models.

The best-performing model, utilizing measures of support (activation) and competition (diversity) from the lemma-concrete learning network, demonstrates that participants perform better when both support from cues and competition between cues are high (Figure 1, upper-right corner of both panels). Notably, within our sample of stimuli, there are no instances where one or both predictors exhibit very low values, limiting conclusions about performance in such scenarios (see the white patches in Figure 1, upper-left, lower-left and lower-right corners of both panels). For predicting imperfectives (Figure 1, left panel), the effect of competition gradually develops as a U-shaped pattern as support strengthens (visible on the right-hand side of the left panel, for X -axis values above approximately 0.0). This indicates

that participants are better at retrieving an originally used imperfective aspect when cue competition is either very weak or very strong, provided that the support they get from the cues remains high. In contrast, for predicting how good participants will be in retrieving an originally used perfective aspect (Figure 1, right panel), the model suggests a pattern that resembles an inverse U-shaped effect: when support from cues is weak, participants perform better in retrieving an originally used perfective aspect when cue competition is moderate but struggle as cue competition intensifies (visible on the left-hand side of the right panel, for *X*-axis values below approximately -0.5 and lower).

The model that uses the aspect-concrete learning measures of activation and diversity show, as we noted, a different relationship between these two predictors and participants' accuracy in retrieving the originally used aspect. The participants' accuracy appears constant when retrieving imperfectives (Figure 2, left panel, large green area without isolines). A small cluster of data appears in the region of low cue competition (diversity) and relatively strong cue support (activation); however, this patch is surrounded by an area with no data, making it unreliable (see white region in the lower-right corner of the left panel). The overall trend suggests a mild decline in performance for higher cue support as cue competition increases (for example, from the activation value around 1.0, going upward along the vertical axis, we cross a few isolines). For perfectives (Figure 2, right panel), participants perform worse when both cue support and cue competition are low. Performance improves as both predictors increase (the trend roughly goes along the lower-left to upper-right diagonal).

In contrast to the aspect-concrete model that shows little to no change in accuracy for imperfectives, the aspect-abstract model reveals a similar constant accuracy for perfectives (Figure 3, right panel) compared to imperfectives (Figure 3, left panel). Broadly speaking, when cue competition (diversity) is high, greater cue support (activation) leads to better performance. This trend is evident in the upper-right corner of the right panel in Figure 3. A similar pattern emerges for imperfectives (Figure 3, left panel). However, in this case, when cue support is weaker, increased cue competition tends to hinder performance. This suggests a nuanced interaction between cue support and cue competition, where strong cue support mitigates the negative effects of high cue competition for imperfectives.

Examining some of the discrepancies by looking at the stimuli (provided in Materials), we observe the following. In some cases, the majority of participants did not select the form that was originally used in the corpus. An example here is Stimulus 1, presented below as (3) with the target verb in *italics*, where only 4 respondents selected the originally used imperfective form (*nabierać*). The use of a perfective for the first infinitive (*zszczupłeć*) may have primed respondents to select a perfective also for the second infinitive (*nabrać*). This was predicted correctly by the AspectConcrete model, but not by the two other models.

- (3) Zmiany masy ciała mogą dwójako wpływać na budowę twojej sylwetki. Z jednej strony możesz zeszczupłeć, z drugiej zaś strony możesz nabierać masy ciała (mięśnie).

Changes in body weight can affect the structure of your figure in two ways. On the one hand, you can lose weight, and on the other hand, you can gain weight (muscles).

Overall, there are 8 stimuli (1, 53, 49, 56, 59, 51, 16, 5, in order of strength of preference for the competing form) where the originally used aspectual form was not preferred by the participants, and a further 5 (17, 6, 57, 43, 41) where about 40 % of respondents preferred the competing aspect over the originally used aspect. The remaining 47 stimuli did evoke a majority preference, defined as over 60 %, for the originally used aspect, but there was only one stimulus (47), presented below as (4) with the target *odszukać* in italics, where all participants made the same choice, and retrieved the originally used perfective form. All three models calculated positive activation for this form, i.e. higher than average activation, while the AspectConcrete and the AspectAbstract models also showed negative diversity, i.e. lower than average competition between the forms.

- (4) czy przyjedziesz może do mnie, podasz mi rękę , a później nagrzasz coś śmiesznego i będziesz miał ubaw? Bardzo źle zrobiłeś , bardzo mnie to dotknęło. [...] Pytasz co teraz robić? Prawdziwy mężczyzna w takiej sytuacji powinien za wszelką cenę doprowadzić do zniknięcia tego głupiego nagrania z internetu. A później powinienes tę organistkę *odszukać*, udać się do niej z kwiatami i przeprosić.

Maybe you will come to me, give me your hand, and then record something funny and you will have a blast? You did a very bad thing, it affected me very much. [...] You ask what to do now? A real man, in such a situation, should make this stupid video disappear from the Internet at all cost. And then you should find the organist, go to her with flowers and apologize.

At the same time, there were seven instances (12, 21, 30, 54, 45, 58, 41 in decreasing order of strength of participant agreement) where our models would have supported a different form than the majority of our participants preferred. Here, all three models computed a negative activation for the originally used form, which the respondents preferred, but the models differed in the diversity scores they returned (with low diversity from all models for 12 and 21; low diversity from ConcreteAspect only for 30, 54 and 45; low diversity from LemmaConcrete only for 58; and high diversity from AbstractAspect only for 41). Stimulus 12, rendered here as (5) with the target verb *podmienić* in italics, evoked little doubt in our respondents (with only 2 out of 77 preferring the non-original imperfective aspect over the original perfective

form); here too, respondents may have been primed by the perfective aspect of the preceding verb, which is not a cue our models take into account.

- (5) Do przenoszenia warstw służą przyciski ze strzałkami do góry i do dołu. Przeniesienie warstwy przenosi nie tylko samą teksturę, ale także informację o jej rozłożeniu. Przeniesienie samych informacji o rozłożeniu można zatem wykonać w dwóch krokach, najpierw przenieść całą warstwę, a na koniec *podmienić* tylko teksturę.
- Use the up and down arrow buttons to move layers. Moving a layer transfers not only the texture itself, but also information about its distribution.
- Transferring only the distribution information can therefore be done in two steps, first transfer the entire layer, and finally replace only the textures.

4 General discussion

Since structuralism, the dominant view on Polish aspect has been that all verbs come in two flavours: perfective and imperfective. Much of the research conducted over the past 75 years has focused on finding the underlying meaning of these two aspectual classes. Even though many abstract semantic distinctions have been put forward, the debate on what aspects really express is far from settled. A more granular, lexical approach proposes that a clear aspectual opposition can only be found in a subclass of verbs, thereby limiting the scope of an overarching aspectual meaning.

However, as we have seen, corpus-based research shows that in addition to being reliably used with certain temporal expressions, the two aspects show usage preferences along several other contextual dimensions. Even though those preferences do not determine use, they are not trivial. As posited by usage-based theory, those statistical biases are not a curiosity but a necessity: learners can acquire a language without explicit instructions, precisely because language is structured while bias, in fact, effectively reduces the uncertainty and makes the language system both learnable and processable. Yet, the presence of a statistical structure in language does not explain whether and how it is utilized by language users. The details of a frequency sensitive, domain-general learning mechanism that helps language learners detect and distil structure from usage remain to be fleshed out. This is where research in psychology and linguistics must come together.

In this paper, we followed Divjak and Milin (2023) and adopted an error-correction learning framework, based on the well-established Rescorla–Wagner model of learning. We showed that this learning framework provides computational

models that can be used to test the usage-based assumptions on language and helps answer the empirical questions that we asked.

First, we wanted to know how well the abstract semantic labels explain aspectual usage. Our aspect-abstract model, trained to use aspect based on the abstract semantic labels discussed in literature, is highly accurate. However, performance is significantly better for the imperfective than for the perfective. Next, we asked to what extent linguistic patterns guide aspectual use, and we pursued this alternative on two levels: we trained an aspect-concrete model to select the appropriate aspectual form of a verb based on the concrete variables that summarize the context alongside a lemma-concrete model, a fully concrete model that learns to use verb lemmas based on these same contextual variables. This latter model did not assume that the verbs are in any way connected by aspectual relation and was used to determine whether users need to form an abstract, difficult to define category of aspect to use verb forms correctly. The performance of the aspect-concrete model was indistinguishable from the performance of the aspect-abstract model, while at the same time performing equally well on imperfective and perfective cases. The lemma-concrete model, although it performed slightly worse than the other two models, still achieved high accuracy and did well on both imperfective and perfective forms.

All three models were then compared to the data from an online gap-filling experiment to determine which of the models fits the actual behaviour of the language users better. While all three models were found to fit the participants' responses well, the best overall fit was obtained with learning measures of support (activation) and competition (diversity) from the lemma-concrete network, followed by the measures derived from the aspect-abstract and aspect-concrete learning networks. As before, the lemma-concrete model performed well for both imperfective and perfective aspect. Second best was the aspect-abstract model, which, however, appears unpredictable for accuracy on perfectives. The same holds for the weakest performing model, the aspect-concrete model, for which accuracy on imperfectives turned out to be poorly predicted.

Taken together, the learning simulations and the gap-filling task suggest that contextual information is more important for selecting an appropriate verb form than abstract categories that generalise over usage. This, of course, is not to say that perfective and imperfective verbs do not carry any shared meaning or that they do not affect the conceptualisation of events in a similar way. On the contrary, the good performance of the aspect-abstract model on the corpus test-set shows that semantic notions can be associated with perfective and imperfective verbs in a way that would allow to tell them apart. However, our findings suggest that, even though aspectual categories can be used to highlight different nuances of meaning, they may well be surplus to requirement in everyday use (cf. Divjak et al. 2024). The choices users

make are more reliably explained by patterns of use; that is, users tend to choose one lemma over the other because it frequently occurs in a given context rather than because it carries a specific abstract aspectual meaning. This is not unexpected since generalisation requires detail to be sacrificed. In fact, an interesting question arising from this finding is whether deviations from patterns are more meaningful since the semantic contrasts might become particularly important in cases where users consciously choose a less preferred and less frequently occurring aspect, i.e. when a usage pattern suggests otherwise.

Our findings also suggest that an error-correction learning model is a good candidate for a domain-general mechanism posited by the usage-based theory of language. As we mentioned in the introduction, we selected an error-correction learning model specifically because of its theoretical grounding in the psychology of learning. There are certainly more powerful learning algorithms available, many of which now outperform humans on a range of linguistic tasks (Madabushi et al. 2022; Wang et al. 2018). However, our goal was not to achieve the best performance, but rather to see what performance can be achieved when we apply basic principles of human learning. Following Rescorla (1988), we want to stress that even the good performance of our models does not mean that forming associations is all that there is to learning. The results indicate, however, that it is an important part of it, certainly worth exploring further.

Finally, we want to highlight a few methodological implications that stem from the current study. Crucially, as indicated by our results, the evidence we use to evaluate our linguistic accounts should not come from only one source: corpus-based learning simulations support all three models, one of which does not even assume the existence of aspect as a category. Corroborating the models with behavioural data was a crucial step in establishing which model is more likely to model the processes driving user behaviour. We hope that our study further amplifies the call for an interdisciplinary approach to the study of language (Dąbrowska 2016; Divjak 2015; Ellis 2006; Milin et al. 2017a, 2017b; Stefanowitsch 2011, to mention just a few). On a theoretical level, such an approach should strive to integrate insights from linguistics and psychology. On a methodological level, it should incorporate formal modelling of authentic data and experiments to allow for formulation and falsification of hypotheses. This approach, as implemented in our study, points us towards a usage-based explanation of aspectual choice, which – instead of focusing on the elusive abstract – is based on the available concrete.

References

- Anderson, Sarah, Teenie Matlock, Caitlin M. Fausey & Michael J. Spivey. 2008. On the path to understanding the on-line processing of grammatical aspect. In Bradley C. Love, Ken McRae & Vladimir M. Sloutsky (eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society*. Cognitive Science Society.
- Aulchenko, Yurii S., Stephan Ripke, Aaron Isaacs & Cornelia M. van Duijn. 2007. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* 23(10). 1294–1296.
- Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada* 11(2). 295–328.
- Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28.
- 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.
- Bańko, Mirosław. 2002. *Wykłady z polskiej fleksji*. Warszawa: Wydawnictwo Naukowe PWN.
- Benko, Vladimir. 2014. Aranea yet another family of (comparable) web corpora. In *Proceedings of 17th international conference text, speech and dialogue*, 257–264.
- Błaszczak, Joanna, Patrycja Jabłońska & Dorota Klimek-Jankowska. 2014. An ERP study on aspectual mismatches in con-verbial contexts in Polish. In Linda Escobar & Vicenç Torrens (eds.), *The processing of lexicon and morphosyntax*, 89–126. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Błaszczak, Joanna & Dorota Klimek-Jankowska. 2016. Aspectual coercion versus blocking: Experimental evidence from an ERP study on Polish converbs. In Joanna Błaszczak, Anastasia Giannakidou, Dorota Klimek-Jankowska & Krzysztof Migdański (eds.), *Mood, aspect, modality revisited. New answers to old questions*, 381–436. Chicago: University of Chicago Press.
- Bott, Oliver & Anja Gattnar. 2015. The cross-linguistic processing of aspect – An eyetracking study on the time course of aspectual interpretation in Russian and German. *Language, Cognition and Neuroscience* 30(7). 877–898.
- Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87(1). 55–83.
- Carreiras, Manuel, Núria Carriedo, María Angeles Alonso & Angel Fernández. 1997. The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition* 25(4). 438–446.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Dąbrowska, Ewa. 2016. Cognitive linguistics' seven deadly sins. *Cognitive Linguistics* 27(4). 479–491.
- Dickey, Stephen McCartney. 2000. *Parameters of Slavic aspect: A cognitive approach*. Dissertations in Linguistics. Stanford: CSLI Publications.
- Divjak, Dagmar. 2009. Mapping between domains. The aspect-modality interaction in Russian. *Russian Linguistics* 33(3). 249–269.
- Divjak, Dagmar. 2015. Four challenges for usage-based linguistics. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms – New paradoxes: Recontextualizing language and linguistics*, 297–309. Berlin: De Gruyter.
- Divjak, Dagmar. 2019. *Frequency in language: Memory, attention and learning*. Cambridge: Cambridge University Press.
- Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60.

- Divjak, Dagmar & Petar Milin. 2023. *Ten lectures on language as cognition*. Leiden: Brill [Distinguished Lectures in Cognitive Linguistics, Vol. 28].
- 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.
- 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, Irene Testini & Petar Milin. 2024. On the nature and organisation of morphological categories: Verbal aspect through the lens of associative learning. *Morphology* 34. 243–280.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.
- Ferretti, Todd R., Marta Kutas & Ken McRae. 2007. Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(1). 182–196.
- Forsyth, James. 1970. *A grammar of aspect: Usage and meaning in the Russian verb*. Cambridge: Cambridge University Press.
- Fox, John & Sanford Weisberg. 2018. *An R companion to applied regression*. Toronto: Sage.
- Garey, Howard B. 1957. Verbal aspect in French. *Language* 33(2). 91–110.
- Gluck, Mark A. & Gordon H. Bower. 1988. From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117(3). 227–247.
- Golshaie, Ramin & Sara Incera. 2020. Grammatical aspect and mental activation of implied instruments: A mouse-tracking study in Persian. *Journal of Psycholinguistic Research* 50(4). 737–755.
- Grzegorzczkova, Renata, Roman Laskowski & Henryk Wróbel. 1999. *Gramatyka współczesnego języka polskiego: Morfologia*. Warszawa: Wydawnictwo Naukowe PWN.
- Haykin, Simon S. 1999. *Neural networks: A comprehensive foundation*. Saddle River (NJ): Prentice Hall.
- Hoffmann, Tabea, Abe Hofman & Eric-Jan Wagenmakers. 2022. Bayesian tests of two proportions: A tutorial with R and JASP. *Methodology* 18(4). 239–277.
- Hopper, Paul. 1982. Aspect between discourse and grammar: An introductory essay for the volume. In Paul Hopper (ed.), *Tense-aspect: Between semantics and pragmatics*, 3–18. Charlotte, NC, US: John Benjamins.
- Jakobson, Roman. 1932 [1984]. The structure of the Russian verb. In Linda R. Waugh & Morris Malle (eds.), *Russian and Slavic grammar studies, 1931–1981*. Berlin: Mouton [Janua Linguarum. Series Maior, Vol. 6].
- Janda, Laura A. 2004. A metaphor in search of a source domain: The categories of Slavic aspect. *Cognitive Linguistics* 15(4). 471–527.
- Janda, Laura A. & Robert J. Reynolds. 2019. Construal vs. redundancy: Russian aspect in context. *Cognitive Linguistics* 30(3). 467–497.
- Johnson, Norman L. 1949. Bivariate distributions based on simple translation systems. *Biometrika* 36(3/4). 297–304.
- Jurkiewicz-Rohrbacher, Edyta. 2019. *Polish verbal aspect and its Finnish statistical correlates in the light of a parallel corpus*. Helsinki: Helsingin yliopisto.
- Klein, Wolfgang. 1995. A time-relational analysis of Russian aspect. *Language* 71(4). 669–695.
- Klimek-Jankowska, Dorota & Joanna Błaszczak. 2020. How incremental is the processing of perfective and imperfective aspect in Polish? An exploratory event-related potential study. *Journal of Slavic Linguistics* 28. 23–69.
- Klimek-Jankowska, Dorota, Anna Czypionka, Wojciech Witkowski & Joanna Błaszczak. 2018. The time course of processing perfective and imperfective aspect in Polish: Evidence from self-paced reading and eye-tracking experiments. *Acta Linguistica Academica* 65. 293–351.

- Kořánová, Ilona & Neil Bermel. 2008. From adverb to verb: Aspectual choice in the teaching of Czech as a foreign language. In Craig Cravens, Masako U. Fidler & Susan C. Kresin (eds.), *Between texts, languages and cultures: A Festschrift for Michael Henry Heim*, 53–69. Bloomington (IN): Slavica.
- Lakoff, George. 1990. The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1. 39–74.
- Lakoff, George. 1991. Cognitive versus generative linguistics: How commitments influence results. *Language and Communication* 11(1–2). 53–62.
- Laskowski, Roman. 1999. Kategorie morfologiczne języka polskiego – charakterystyka funkcjonalna. In Renata Grzegorzczkova, Roman Laskowski & Henryk Wróbel (eds.), *Gramatyka współczesnego języka polskiego*, 150–224. Warszawa: Państwowe Wydawnictwo Naukowe.
- Łaziński, Marek. 2020. *Wykłady o aspekcie polskiego czasownika*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Light, Leah L. & Patricia A. Anderson. 1985. Working-memory capacity, age and memory for discourse. *Journal of Gerontology* 40(6). 737–747.
- Madabushi, Harish Tayyar, Dagmar Divjak & Petar Milin. 2022. Abstraction not Memory: BERT and the English article system. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 924–931.
- Madden, Carol J. & Rolf A. Zwaan. 2003. How does verb aspect constrain event representations? *Memory & Cognition* 31(5). 663–672.
- Magliano, Joseph P. & Michelle C. Schleich. 2000. Verb aspect and situation models. *Discourse Processes* 29(2). 83–112.
- 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, Peter Hendrix & R. Harald Baayen. 2017b. Discrimination in lexical decision. *PLoS One* 12(2). e0171935.
- Milin, Petar, Benjamin V. Tucker & Dagmar Divjak. 2023. A learning perspective on the emergence of abstractions: The curious case of phone(me)s. *Language and Cognition* 15(4). 740–762.
- Morrow, Daniel G. 1985. Prepositions and verb aspect in narrative understanding. *Journal of Memory and Language* 24(4). 390–404.
- Pecina, Pavel. 2010. Lexical association measures and collocation extraction. *Language Resources and Evaluation* 44(1/2). 137–158.
- Qualtrics. 2024. Qualtrics [Computer software]. *Qualtrics*. Available at: <https://www.qualtrics.com>.
- R Core Team. 2024. *R: A language and environment for statistical computing. (Version 4.4.2) [Computer software]*. R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Rescorla, Robert A. 1988. Pavlovian conditioning. It's not what you think it is. *The American Psychologist* 43(3). 151–160.
- Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II Current Research and Theory* 21(6). 64–99.
- Reynolds, Robert J. 2016. *Russian natural language processing for computer-assisted language learning*. UiT The Arctic University of Norway Doctoral Dissertation.
- Rice, Sally & John Newman. 2004. *Aspect in the making: A corpus analysis of English aspect-marking prepositions*. *Language, culture and mind*, 313–327. Stanford: CSLI Publications.
- Sadowska, Iwona. 2012. *Polish: A comprehensive grammar*. Routledge [electronic resource].

- Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294). 1926–1928.
- Sasse, Hans-Jurgen. 2002. Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology* 6(2). 199–271.
- Sekerina, Irina A. 2006. Building bridges: Slavic linguistics going cognitive. *Glossos – E-Journal of the Slavic and East European Language Research Center* 8.
- Sekerina, Irina A. 2017. Slavic psycholinguistics in the 21st century. *Journal of Slavic Linguistics* 25(2). 463–488.
- Sering, Konstantin, Marc Weitz, Elnaz Shafaei-Bajestan & David-Elias Küntle. 2022. Pyndl: Naive discriminative learning in python. *Journal of Open Source Software* 7(80). 4515–4518.
- Smith, Carlota. 1986. A speaker-based approach to aspect. *Linguistics and Philosophy* 9(1). 97–115.
- Smith, Carlota. 1997. *The parameter of aspect*. Dordrecht: Springer.
- Stefanowitsch, Anatol. 2011. Cognitive linguistics as a cognitive science. *Bidirectional Perspective in the Cognitive Sciences*. 296–309.
- Swan, Oscar E. 2002. *A grammar of contemporary Polish*. Bloomington (IN): Slavica.
- Thelin, Nils B. 1990. *Verbal aspect in discourse*. Amsterdam: John Benjamins.
- van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Dirk van Rijn. 2022. itsadug: Interpreting time series and autocorrelated data using GAMMs (version 2.4).
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160.
- Wagner, Allan R. & Robert A. Rescorla. 1972. Inhibition in pavlovian conditioning: Application of a theory. In Robert A. Boakes & M. Sebastian Halliday (eds.), *Inhibition and learning*. London: Academic Press.
- Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy & Samuel R. Bowman. 2018. GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.
- Wasserman, Edward A., Daniel I. Brooks & Bob McMurray. 2015. Pigeons acquire multiple categories in parallel via associative learning: A parallel to human word learning? *Cognition* 136. 99–122.
- Widrow, Bernard & Marcian E. Hoff. 1960. Adaptive switching circuits. In *IRE WESCON convention record*, vol. 4(1), 96–104.
- Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73(1). 3–36.
- Wood, Simon N. 2017. *Generalized additive models: An Introduction with R*. London: Chapman and Hall/CRC.
- Wróbel, Henryk. 2001. *Gramatyka języka polskiego*. Bielsko-Biała: Spółka Wydawnicza Od Nowa.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/llt-2023-0117>).