# LEARNING FROM USE: AN ERROR-DRIVEN APPROACH TO POLISH ASPECT

by

MACIEJ BOROWSKI

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Modern Languages

School of Languages, Cultures, Art History and Music

College of Arts and Law

University of Birmingham

August 2022

# Abstract

This thesis investigates how speakers learn to use grammatical categories of their native language. While most linguistic categories are useful for descriptive purposes, the questions of their learnability and psychological plausibility are often ignored. Taking Polish grammatical aspect as a case study, we investigate if and how speakers learn to distinguish such general categories as well as master their use based on the usage patterns available in the input.

Grounding our research in usage-based theory of language and the principles of error-driven learning, we conducted a number of computational learning simulations based on a manually annotated corpus sample. These simulations, corroborated with the results of behavioural study show that patterns of usage explain the linguistic behaviour of speakers better than the abstract semantic dimensions that are traditionally used to describe the aspectual classes. We also demonstrate that the patterns of use learned by our models contain enough information to correctly classify verbs into their respective aspectual classes. The results of the studies also indicate an important relationship between tense and aspect and suggest that certain tense-aspect combinations could be considered default. This issue is investigated in more detail, again using a combination of learning simulations and experimental methods.

We argue that the studies presented in this dissertation force us to reflect on the relevance of traditional linguistic distinctions for language cognition and acquisition as well as points us towards a usage-based explanation of the aspectual choice. In addition, we also discuss methodological implications that follow from the work presented here. In particular, we highlight the importance of combining different sources of evidence and the need to corroborate the corpus-based models with behavioural data.

## Acknowledgements

This dissertation was written as a part of a research project led by Professor Dagmar Divjak and Dr Petar Milin, funded by the Leverhulme Trust (Leverhulme Trust Research Leadership Award [RL-2016-001] to Dagmar Divjak). I was very lucky to join this project and the Out Of Our Minds lab, as it provided me with an opportunity to work with, and learn from, many amazing people.

The collaborative nature of the project also means that the list of people I want to mention, and the list of reasons I am thankful for is really quite long. I will try to be brief.

First and foremost, I would like to thank my supervisors and the leaders of the project, Prof. Dagmar Divjak and Dr Petar Milin. Working with you was a truly transformative experience on so many levels. I would like to thank you for all the guidance, comments and for showing me the ins and outs of the academia. Most importantly, however, thank you for helping me realize what my strongest skills are. The path ahead is much more clear and much more interesting because of that.

I would also like to thank the rest of the OOOMinds team:

Dr Srdan Medimorec, for your mentorship and friendship. Your conviction that I was the right person for the job helped me immensely;

Christian Adam, for all the walks, talks, hours spent programming and cooking nights we had both in Sheffield and in Birmingham. I have learned so much from you and with you. I am also a much better cook;

Marta Gasiorowska - not only for your support and input in all the projects we worked on together, but also for your friendship and for being there for me when I really needed a friend;

Dr Adnane Ez-zizi, Celia He, Dr Hui Sun, Dr Laurence Romain, Daisy Collins and Dagmar Hanzlikova for all the time we spent working together and supporting each other as a team. I would not be able to complete this work without you.

Special thanks go to my friend in Krakow, Rafał Blecharz for saying the right words at the right moments. Thank you for having my back.

Finally, I would like to thank the people who always supported me and believed in me, no matter what: my parents, Katarzyna and Dariusz.

Table of contents

## Table of figures



## Table of Tables

# Chapter 1. Introduction

Traditionally, aspect is considered a category that allows to express different points of view on an event. In many Slavic languages for example, speakers must regularly 'choose' between two aspectual variants of a verb – perfective and imperfective. It is assumed that by using the former, speakers highlight the completeness of the event; by using the latter, they signal its ongoingness. However, a closer inspection of actual usage cases quickly reveals that the problem of the meaning of aspect is much more complicated. Those intricacies of temporal semantics are a challenge for linguists who try to model them and any second language learners who try to master them. Native speakers on the other hand, rarely consider aspectual choice a problem at all. Despite exceptions, they seem to readily accept the general notions of 'completeness' and 'ongoingness' as good approximations of meaning, and yet they hardly make any errors when choosing between perfective and imperfective. This suggests that their choices might be based on a different type of knowledge than abstract semantic definitions.

In this dissertation, we ground our research in usage-based theory, and explore to what extent the use of aspect can be explained by the patterns available in language. However, we are not simply interested in whether such patterns exist. Staying true to the cognitive commitment (Lakoff, 1990, 1991), we also want to know if and how any patterns we see in the data are actually exploited by the speakers. In other words, our goal is to model language based on what is known about the human mind, and the general mechanisms we propose for language comprehension and production should be grounded in the study of cognition.

While we cannot make claims about the cognitive workings of the language system based on the traditional methods of linguistics – text analysis and intuition, this situation can be remedied by adopting an interdisciplinary approach and using more appropriate methods of investigation. If language is 'something in our minds', it must be studied accordingly.

This, however, is no small task, and a broad characterization leaves it open to interpretation, especially when it comes to applying the cognitive commitments to research practice. How problematic it proves to be exactly is shown by the fact that almost thirty years after Lakoff's commitments have been published, the methodological problems are still not resolved and that the link between cognitive sciences and linguistics often seems quite tentative. Not only do the arguments come second hand (Divjak et al., 2016) but also cognitive linguists rarely look beyond 'cognitive science' as it is presented in the field (Dąbrowska, 2016) even though we know that the original framework, that still inspires many concepts in research today, was based on a 'common-sense' understanding of human cognition (Stefanowitsch, 2011).

Yet, it would be naive to say that the situation in the field can be amended by 'importing' cognitive disciplines into linguistics. First of all, the study of the mind is not complete, and 'what is known about human cognition' is not a monolith. That is, the theories of different aspects of cognition and the interactions between them are still being developed and as in the case of all sciences – there are many problems and many competing views.

Secondly, even if such an approach was possible, it would not do linguistics justice, as it has an unnecessarily reductionist flavour. Not only does it assume that linguistics could be reduced to (some branch of) psychology, but also that linguists cannot produce any valuable input that could improve the state of affairs in other disciplines. The former is too simplifying: research in linguistics would definitely benefit from psychological grounding, but linguistics has questions of its own. The latter seems counter-intuitive – if language is something of our minds, then by learning about it we can learn about our minds as well; in other words, what we discover about learning language should also hold for other areas of our experience. Therefore, what seems to be necessary is a step towards an interdisciplinary approach to the study of language.

Fortunately, that step is already being taken. The need for a more serious treatment of the cognitive commitment has already been expressed in cognitive linguistic circles (e.g. Stefanowitsch, 2011; Dąbrowska, 2016; Divjak, 2016) an attempt has been made of bringing together linguistic and psychology to explain language acquisition phenomena (e.g. Ellis, 2006; Ramscar & Yarlett, 2007) and studies based on these theoretical assumptions have been conducted (e.g. Milin et al., 2016).

The following dissertation also has the ambition to use such an interdisciplinary approach in practice. For this reason, we turn to Learning Theory, and use it to complement and strengthen the postulates posited by usage-based theory of language. We also show that the two theories, even though conceived separately from each other, strive to describe learning using similar principles. The key aspect in both are the contingencies in the experience.

To determine whether the use of aspect can be at least partially explained by the contingencies in language input or whether the abstract semantic definitions are indispensable, we use corpus data to model three problems. First, we evaluate the predictive power of various semantic distinctions proposed in literature as candidates for the invariant meaning of aspect. Next, we model whether it is possible to acquire the semantic distinctions on the basis of the information available to learners in usage. Then, we compare the performance of a model which uses the abstract distinctions to predict aspectual usage to usage-based models which rely only on the co-occurrence of verb forms or aspects and contextual cues.

Importantly, to model our corpus data, we use a learning algorithm based on the Rescorla-Wagner model of learning (Rescorla & Wagner, 1972)- the Naïve Discriminative Learner (Baayen, 2011; Milin, Divjak, et al., 2017). The advantage of this approach is that, in contrast to other modelling methods, our approach is in fact a computational implementation of a learning mechanism proposed by Learning Theory. In other words, we are not only using machine learning methods to calculate the predictive power of the cues we provide, but rather simulate the learning process based on the data we have available.

As said before, we are not only interested in identifying the patterns in language, regardless of the methods used to extract them, but we also want to know whether speakers really make use of these patterns in a meaningful way. Therefore, to validate our models, we test them against behavioural data. The first study aims to test whether the choices native speakers make can be explained by any of the corpus models we have trained. The second study tests a more specific hypothesis, and aims to find out whether the processing of linguistic input is affected by the usage patterns indicated by the most predictive of our computational models.

We focus on Polish aspect as our case study. Surprisingly, there is still relatively little empirical research on aspect in Slavic languages. This is unfortunate, because compared to the formal simplicity of English, the complexities of Slavic languages make them excellent test cases for any linguistic theory. As should become clear from the following sections, if a usage-based learning model can be successfully applied to the case of Polish aspect, it can certainly be applied to other linguistic problems as well.

# Chapter 2. A usage-based theory of language

We will start the dissertation by presenting two theories on which our investigations are based, namely usage-based theory of language and learning theory. In this chapter, we will discuss usage-based theory of language, focusing mainly on the core concept of how speakers utilize patterns present in everyday language use in order to form a representation of the linguistic system. In the following chapter, we will show that learning theory, and in particular the error-driven learning principle, complements the usage-based theory providing the answer to the important question of how exactly usage patterns can be exploited by speakers in the process of language learning. We will argue that even though the two theories were formulated independently from each other – the former comes from linguistics and the latter, broadly speaking, from psychology – they are in fact complimentary. These theoretical considerations will both inform our approach to the problem of aspectual use, which we will discuss in Chapter 4, as well as serve as a point of departure for the studies we will present in the following chapters.

## 2.1 Dynamic yet stable

Language, from a usage-based perspective, seems a bit paradoxical. On the one hand, it is full of fixed expressions and routinised ways of expressing ideas. Phrases like ‘the problem is’, ‘a little bit’, ‘it depends on’ (Hopper, 1987) are used almost as prefabricated elements (Bybee, 2006) and allow little, if any, modification. But things are set even when we move beyond those uncontroversial cases. Native speakers use only some selected patterns from all theoretically possible ones. For instance, it is more “natural” to say ‘will you marry me’ than ‘would you like to become my spouse’ (Pawley & Syder, 1983).

On the other hand, language is also in a perpetual state of flux. As Barlow and Kemmer (2000: viii) put it: 'the speaker's linguistic system is fundamentally grounded in "usage events": instances of a speaker's producing and understanding language'. Each usage event provides input for the active process of forming representations of units in that system, which may vary in their specificity. Some units, like words or idiomatic expressions will be learned 'by heart' - stored as indivisible wholes. More abstract ones, like patterns or schematic constructions, will result from us gradually abstracting away from concrete examples. But if language is grounded in experience, then it cannot have a 'final' stage. There cannot be a set of abstract rules that exist outside of the interactions between the speakers, and which can be conceived as a goal of learning (Hopper, 1998). In other words, the task the learners of any language face is not to figure out what rules there are in order to be able to communicate, but to actively construct language by engaging in interactions with other speakers in the process of communicating (Hopper, 1998). The system that linguists try to describe emerges from the interactions between multiple users, each with their own, unique history of communicative events (Beckner et al., 2009). How then, can the fixedness of patterns emerge from the flux of experience?

According to usage-based theory, the limitation of the use and the productivity of the patterns is based on distributional properties of the input. If speakers learn language from experience, they must pay attention to the features of the experience, and one of the most characteristic features of language is that it is not random but full of statistical regularities and patterns.

Zipf (1935) showed that, in any text, the frequency of the word is inversely proportional to the rank of the word in the frequency table: the most frequent word appears almost twice as often as the second, three times more often than the third most frequent word, etc. The words are usually distributed in such a way that there are a few very frequent items followed by a long tail of items with similar low frequency. Such distributions are ubiquitous and can be found on many other levels of language. For example, Goldberg et al. (2004) show that verbs that are used by children in 3 different syntactic constructions follow Zipf's law. For instance, in a construction that follows the pattern S V Obj Obl (*He put the milk in the fridge*), *put* accounted for 31% of all the verbs used by children, *get* for 16%, *take* for 10% and *do* and *pick* for 6% each. Ellis (2012) shows that not only verbs that appear in a construction 'V it P' follow a Zipfian distribution, but also the prepositions used with a given verb. For instance, 'put it _' is most frequently complemented by *in* (3620 instances in the Corpus of Contemporary American English), then by *on* (1926) and thirdly by *onto* (745). The same distribution is found for both the subjects in those phrases and the nouns that follow the prepositions.

What is important is that such differences in how often items and patterns are used are not curiosities of language or just interesting facts for linguists to explore but they have real and measurable consequences on language learning and processing. In other words, these distributional patterns are the key feature that allows us – speakers – to acquire a language system. These claims are well supported empirically and there exist a few excellent and comprehensive reviews of the literature on the subject (Ellis, 2002; Diessel, 2007; Ambridge et al., 2015; Divjak & Caldwell-Harris, 2015; Divjak, 2019). In the following sections, which are based on these compilations, we will discuss a selected few studies that are particularly relevant for this dissertation. We will focus on showing that distributional patterns are ubiquitous in language as well as discussing the theoretical implications this evidence has.

## 2.2 Frequency effects in language

Divjak and Caldwell-Harris (2015) point out that the first psycholinguistic evidence for the influence of word frequency goes back as early as 1886. Since then a number of empirical facts has been established. First of all, there is evidence that the frequency of occurrence increases the ease of processing. A classic way of demonstrating this effect is the lexical decision task. In this experimental paradigm (e.g. Forster & Chambers, 1973; Whaley, 1978) participants are presented with words and non-words and are asked to decide whether a sequence of letters they see on a screen is a real word in their native language or not. What the results show is that word frequency is one of the strongest predictors of response times. That is, the more frequent the word, the faster people recognize it as a valid entry in the dictionary.

Interestingly, more frequent items are also more easily recognized even when the speech signal is noisy. Rubenstein et al. (1959) demonstrated that the more frequent the word, the easier it is for participants to understand it, even when they were listening for it while also hearing a background noise. This effect of word frequency has recently been replicated in a visual world search task (Van Engen et al., 2020), during which the participants were asked to click on pictures representing the words that they heard. Word frequency affected the speed with which the participants clicked on images, regardless of whether the words were presented with or without background noise. What is of particular interest is that the frequent words in the noisy condition were recognized faster than the infrequent words without the noise present in the background.

As for the effects of frequency in a different mode of interacting with language, eye-tracking studies offer a number of findings. For instance, a study conducted by Rayner and Duffy (1986) showed that while reading, people tend to fixate for longer and generally spend more time gazing at less frequent words compared to more frequent words. In addition, they also showed that the infrequent words negatively affected the amount of time spent on the next parts of the sentence, specifically the word following the target word. This, as the authors suggest, may indicate that more effort is required to access the less frequent words and to understand how they fit into the sentence.

Frequency of occurrence influences the acquisition of language in other ways as well. As pointed out by Ambridge et al (2015), highly frequent items and structures are generally learned sooner that the less frequent alternatives. Not only are the first words that children utter the ones that occur frequently in the input directed to them, but this effect holds for more complex grammatical structures as well. Ambridge et al. (2015) discuss a number of studies demonstrating this effect. For instance, passive constructions, which are very infrequent in spoken English, are also very rarely produced by English-speaking children. The same findings were reported for other languages in which passive and active voice are used with similar clear biases in distribution. However, in languages where the passive is more frequent than the active, the order of acquisition is reversed.

The example of the passive above also shows that the effects of frequency can be found on other levels of linguistic analysis than single words only. Bannard and Matthews (2008) conducted a study in which they tested the sensitivity of children to multiword sequences. Using a corpus of child directed speech, they extracted four-word sequences. Importantly, the final word frequency and length were controlled for, as were the frequencies of final bigrams and trigrams. Such constraints on stimuli allowed them to determine that children are faster and more accurate when producing the sequences with higher frequencies compared to low-frequency sequences when all other things are held equal. Arnon and Snider (2010) demonstrated a similar effect in a study where participants were adults. Just like in Bannard and Matthews (2008), they used four-word phrases as stimuli, matching the frequencies of the final word, bigram and trigram. The results showed that the participants were faster to respond to high frequency items than they were to low frequency items.

All in all, these studies show that people are sensitive to the frequency with which things appear in the input. However, frequency of use alone cannot explain language. First of all, it is not clear how learners track frequencies. It is unlikely that we have some sort of a counter in our minds that registers all the encountered instances of language items. Even though it is not impossible, the idea that we somehow store all the frequencies of all the items we have ever encountered seems implausible. The amount of information we would need to remember and continuously update, if we track all the frequencies of all the items, would be extremely demanding on memory (Baayen et al., 2013). If we reduce the numbers of stored items to those over a certain frequency threshold, a logical problem arises. Since single exposure does not lead to storage, every exposure must be treated as the first one. The counter would remain at 0. (Divjak & Caldwell-Harris, 2015). Finally, the demonstration that certain phenomena, such as faster retrieval or recognition, are correlated with the high frequency of an item does not explain why we would track frequencies. Frequency effects do not explain its function – frequency itself must be incorporated into a larger conceptual and theoretical framework. To understand the role of frequency, we need to highlight that language users pay attention not only to how often the forms appear but also what they appear with. Context is when the frequency really matters and starts to make sense.

## 2.3 From frequency to predictability

One of the classic studies showing that speakers are sensitive to frequency in the context, and that this type of frequency affects linguistic behaviour is Bybee and Scheibman (1999). Analysing the phonological reductions of *don't*, they found that speakers usually shorten the vowels when *don't* is used in typical phrases, such as *I don't know* or *I don't mean*. And since the authors did not find any phonological reasons for such reductions to occur in these combinations, they concluded that the reason for the differences in production can be ascribed to the frequency with which these items occur with both the preceding pronoun and the verb that follows *don't*.

As for the effects of contextual frequency in comprehension, an early eye tracking study provides interesting insights. Ehrlich and Rayner (1981) measured the fixation times and skipping rates for words appearing in paragraphs. The crucial manipulation involved using two types of contexts – one in which the target word was highly predictable and one where more continuations were possible. In addition, some target words contained misspellings, making the word inappropriate in the context. The results showed that people were more likely to skip the target word when it appeared in a predictable context as well as fixated on the target for less time when they did not skip it. What is more, the participants were also less likely to notice misspellings in the context were the target was highly constrained and thus predictable. These results suggest that the predictability of the target word – the result of the target's frequency in a given context - allows the speakers to form expectations of the incoming input, making their behaviour less dependent on the actual text they are reading.

Arnon and Clark (2011) demonstrate experimentally how context and frequency are related by investigating the error rate in irregular plural production in English-speaking children. In their experiments, the children's task was to provide a plural form of a noun in three conditions. In the first one, only a picture was used to facilitate the form. In the second, a familiar frame, such as 'So many _' was used. In the third one, the noun was easily predictable from the provided context (e.g. 'mice' when the context was 'three blind _'). The results show that children make fewer mistakes in irregular plural forms when they produce them in the context of a known frame than in isolation and perform really well when they produce them in a phrasal context that is highly collocated with those nouns.

Predictability in context, calculated as the strength of reliability between a particular construction and a verb used in that construction, also affects acceptability ratings in adults. Divjak (2017) asked speakers of Polish to evaluate the acceptability of sentences containing verbs complemented with a that-clause. She found that raw frequency did not predict people's ratings. However, a different frequency-based measure – Reliance – defined as the frequency of occurrence of the given verb relative to its overall frequency in the corpus, turned out be a significant predictor of acceptability.

The studies discussed above show that we do not simply count how often the words are encountered in isolation. Tracking frequency in context allows us to form expectations of what should come next in the speech signal. This, as Arnon and Clark (2011) rightly point out, allows us to form a functional interpretation of the frequency effects. We pay attention to frequencies in usage, because it is the basis of reliability.

But the results of Divjak (2017) highlight how complex language learning truly is – to be able to assess the probability of a word in the context we must know not only how likely a given item is in the context, but also have a sense of how likely it is to appear in other contexts as well. Even though there is evidence that speakers do exploit probabilities in language from very early on (Saffran et al., 1996), a usage-based theory must also explain how these relations between linguistic items are learned. We will return to this problem in the next chapter. Before we do, however, we will discuss the evidence for the importance of distributional patterns for category learning.

## 2.4 From distribution to categories of language

On a usage-based account, the most important source of information for category learning is not some invariant semantic features that are shared by all members. Instead, we form categories by noticing that some items appear in the same contexts. If they are used in a similar way, they probably belong to the same category. In other words, the distributional patterns of items combined with information about the frequency of occurrence within these patterns should allow the speakers to build more abstract, grammatical categories from the ground up.

Corpus-based modelling studies provide support for these claims. Redington et al. (1993) show, for instance, that distributional patterns provide enough information for classifying words into grammatical categories. Taking data from a corpus of child directed speech they constructed context vectors for the 1000 most frequent target words. They showed that vectors of the preceding and succeeding bigrams of the 150 most frequent words with values representing the frequency with which each target word occurs with the given context bigram carried enough information to be able to cluster the target words into groups that closely mirrored grammatical categories. Similar results were reported by Mintz et al (2002) who showed that using context vectors allows for classification of the target words from child directed speech corpus into nouns and verbs.

Mintz et al (2002) also note that, while the results suggests that categorization based on distributional information certainly seems possible, the patterns of use are most likely not the only source of information that speakers use. Among other possible cues facilitating grouping items into grammatical categories they list, for instance, the semantics and morphology of the target words. Nonetheless, these studies provide evidence that usage patterns can play a role in category building.

Two important caveats are necessary here. Firstly, it is possible that the speakers do not form abstract categories at all; or – less radically – that not all the categories postulated by linguists overlap with the categories that are actually used by the speakers. The fact that we can cluster items into predefined groups does not necessarily entail that these groups are also relevant for the language users. As Divjak (2016) points out, the abstract categories used in linguistics were created for descriptive purposes. The question of their psychological validity still needs to be addressed.

Secondly, the structure of an abstract category, if there are any, is not necessarily homogeneous. The contexts of *run* and *walk*, for example, are more similar to each other than the contexts of *run* and *sing*. As we said above, if learners arrive at generalizations, they (may) do so in a gradual fashion, building on what is similar. Therefore, categories should be formed on the basis of local groupings of items. Those intermediate levels are what– in theory– connects the concrete to the abstract.

## 2.5 Summary

To sum up, usage-based theory posits that language is acquired by employing a domain general mechanism, which allows to build from experience a system of representations that are probabilistic in nature. We have also discussed evidence that suggests that aspectual categories can be learned on the basis of usage, in the way posited by usage-based theory. We discussed corpus-based evidence that shows that there exist contextual preferences which might guide speakers when using aspect.

What still needs to be discussed is the acquisition mechanism itself. The fact that usage patterns exist, does not yet explain how these patterns are utilized. Simply saying that speakers are sensitive to those patterns falls short of what we set out to do. As we have said earlier, the goal of cognitively oriented research is not to simply to describe language phenomena, but also to describe the processes that lead to the emergence of these phenomena.

The mechanism of learning must be sensitive to the features of experience – especially the frequency of occurrence in the context. It must also allow for the system to become relatively stable, but this fixedness must result from dynamic interactions of speakers. Finally, it must give rise to useful generalizations that allow for patterns to be productive and used in novel situations. In the next chapter, we show that Learning Theory describes exactly such a mechanism.

# Chapter 3. Learning Theory

The following sections present the second theory on which this dissertation is based, namely Learning Theory. We will show here how Learning Theory complements the theoretical tenets of usage-based approaches, and leads to further insights. We will present the principles of association formation, focusing on the Rescorla-Wagner model and how it can be applied to language learning. Next, we will discuss how these principles can be used to model corpus data by employing Naïve Discriminative Learner – a computational implementation of the Rescorla-Wagner model. Finally, we will compare Naïve Discriminative Learner to other models and algorithms used in linguistics and machine learning to explain why it can provide us with meaningful insights despite its architectural simplicity.

## 3.1 From automatic salivating to deliberate behaviour

Let us start with classical conditioning, introduced by Pavlov's famous experiments. In this experimental paradigm, the subject (e.g. a dog) is presented with an unconditioned stimulus (US; e.g. food), that naturally evokes a response (e.g. salivation). The food is then paired with – i.e., presented many times after – a conditioned stimulus (CS; e.g. bell). After a sufficient number of such pairings, the conditioned stimulus starts evoking the response – the dog salivates when it hears the bell (Anderson, 2000, p. 9).

Humans also can be subject to classical conditioning. In a popular eye-blinking paradigm, for example, a light or tone is paired with a gentle puff of air in the eye, which causes blinking. As in the experiments with dogs, after a number of pairings, the light or tone evoke blinking without the puff (Anderson, 2000, p. 39).

However, classical conditioning is not the only type of conditioning. In the instrumental paradigm, a subject must react to a stimulus in a certain way in order to receive the reinforcement. For instance, a rat must press a lever after hearing the tone in order to get food. What is learned is that a certain action, performed when the stimulus is present, leads to receiving a reinforcer.

The reactions elicited in instrumental conditioning are far from mindless and automatic for two reasons: first of all, the subject must perform a (complex) action and has control over its behaviour. Moreover, they perform the action in order to achieve something. Unlike salivating, this deliberate behaviour may change when the subject's goals change (e.g. when a rat's hunger is satiated, it will no longer perform the action that leads to food). Nonetheless, the basic principle of those two paradigms is the same, since both can be explained by the same mechanism.

## 3.2 From associations to knowledge

As we have seen, the co-occurrence of the CS and US leads to association, and evidence for the existence of the association stems from observing the conditioned reaction (CR) of the subject, be it salivating, blinking or something more complex, like pressing a lever. It seems then, as advocated by Aristotle and later Locke (Terry, 2006) that contiguity, the temporal or spatial closeness, is what allows us to learn.

However, Rescorla (Rescorla, 1988) points out that the response elicited in animals is not a result of the fact that two things appear together, but rather of the fact that one thing predicts the other. It is contingency, not contiguity that is the basis of learning, and contingency is probabilistic in nature.

Rescorla presents two experimental arguments for his claim (Rescorla, 1988). The first one comes from a study in which rats that were pressing a lever, were presented with a 2 minute tone during which a shock was administered (Rescorla, 1968). The probability of the shock during the tone varied from 0.1 (the shock appeared in 10% of the cases when the tone was present) to 0.2, to 0.4. More importantly, the probability of receiving the shock when no tone was present varied as well (again: 0.1, 0.2, 0.3). What the results show is that the number of times the rats pressed the lever when the tone was present depended not only on how likely they were to receive the shock during the tone (this is predicted by classical conditioning as well: the more pairings, the better the learning) but also on how likely they were to receive the shock, when there was no tone. When the probability of receiving a shock during the tone was high and the probability of a shock when there was no tone was low, rats learned not to press the lever. But when the probabilities matched the tone no longer predicted that pressing the lever might lead to receiving a shock, so the rats continued pressing the lever.

The second argument that it is contingency, not contiguity, that underpins learning is Kamin's (1968) blocking effect (cf., Rescorla, 1988: 153). In an experiment, two groups of subjects were conditioned to react to a compound stimulus, e.g. light and tone. One group had been trained before to react to the light only, however. This group would later show much poorer conditioning to the tone alone, compared to the other group. This demonstrates that if the light is already a good predictor of X, the presence of another conditioned stimulus is not very informative, and therefore not much is learned about the relation between the second stimulus and X.

Therefore, Rescorla argues 'that conditioning involves the learning of relations among events' (1988: 153) and modern theories of learning 'emphasize the importance of a discrepancy between the actual state of the world and the organism's representation of that state. They see learning as a process by which the two are brought into line' (1988: 153). In other words, learning occurs when an organism discovers that there is a difference between what happens and what it knows about the world and the relations between the events in it. The point of learning, it seems, is to reduce the uncertainty about the world.

## 3.3 Formalizing learning

Formally, the Rescorla and Wagner model (Rescorla & Wagner, 1972; Wagner & Rescorla, 1972) can be expressed as an equation that allows to calculate how the associative strength between cues and outcomes changes during the process of learning:

$$\Delta V = \alpha\beta(\lambda - V)$$

where $\Delta V$ is the change of associative strength. Parameters $\alpha$ and $\beta$ specify the learning rate (how quickly a given cue can be associated with an outcome) and salience (how 'noticeable' the cue is) respectively. $\lambda$ is the maximum strength of association. V specifies how strongly the cue and outcome are currently associated. Therefore, to calculate the change in association strength at one learning step, we first calculate the difference between the current association strength and the maximum of strength that can be achieved, and then multiply that value by the product of the values of salience and learning rate.

Having calculated the value of the change of associative strength at a given learning step, we can now obtain the new value of association strength by adding the value of the change to the previous value of association:

$$V^t = V^{t-1} + \Delta V$$

Where $V^t$ is the value of new association strength between a cue and an outcome, $V^{t-1}$ is the value of the association strength at the previous learning step, and $\Delta V$ is the value of the change of associative strength that occurred at this particular learning step.

Importantly, the Rescorla-Wagner model allows to calculate the strength of association when two or more cues are present at the same time, 'competing' to predict the same outcome. For example, the change of cue A will depend on how well other cues (in our case B) are associated with the same outcome:

$$\Delta V_A = \alpha\beta(\lambda - V_{A+B})$$

As we can see, in order to model cue competition, the value we subtract from the maximum strength of association ($\lambda$) is the current value of association between cue A and the outcome *plus* the current value of association between cue B and the same outcome ($V_{A+B}$). This means that the association strengths of other cues that are present at a given learning step influence the value of association between cue A and the outcome.

Let us consider an example. For simplicity, we will take 0.1 as the product of $\alpha$ and $\beta$, and set the value of $\lambda$ to 1. If the current association weight between cue A and some outcome is 0.5, and the association between cue B and the same outcome is 0.1, the value of change for cue A can be calculated as follows:

$$\Delta V_A = \alpha\beta(\lambda - V_{A+B}) = 0.1(1-(0.5 + 0.1)) = 0.1 * 0.4 = 0.04$$

Therefore, the new value of the association between cue A and the outcome will be:

$$V^t_A = V^{t-1}_A + \Delta V_A = 0.5 + 0.04 = 0.504$$

However, if cue B is more strongly associated with the outcome, say at 0.4, the change of the association strength for A will be smaller:

$$\Delta V_A = \alpha\beta(\lambda - V_{A+B}) = 0.1(1-(0.5 + 0.4)) = 0.1 * 0.1 = 0.01$$

As a result, the updated association strength for A will also be smaller:

$$V^t_A = V^{t-1}_A + \Delta V_A = 0.5 + 0.01 = 0.501$$

This is the nature of cue competition – the value of association strength of one cue depends on and is influenced by what is known about the associative strengths of other cues. This competition and the algorithm's ability to handle the competition between the cues is a crucial point that will allow us to understand how this model can be applied to language learning.

### 3.4 Error-driven learning framework

In language, things are rarely clear-cut. It seems, then, that the task that we – learners– are faced with is more complex. To illustrate how the Rescorla-Wagner model can be applied to language learning, let's consider the following, simplified example. Imagine a learner who is trying to figure out when to use the -s ending with English verbs. To make the presentation clearer, we are not going to consider different phonetic variants of the suffix, and we will ignore the issue of separating the stream of sound into individual words. As for the equation itself, for the purpose of illustration, we are going to assume that $\alpha\beta = 0.1$ and $\lambda = 1$ when the outcome is present and 0, when it is absent.

Our learner hears the first sentence:

*She wants a cookie.*

Our knowledge of English is telling us that what the learner should be interested in is the connection between *she* and the verb form, as it is the pronoun that guides the use of the suffix. However, our learner does not know that – and cannot know it. For her, at this stage, all the words she hears are potentially equally informative cues.

Therefore, the connection weight between all the words and the outcome *-s* will increase. We can calculate them in the same way for each of the cues, since right now, all cues' weights equal **0**.

$$\Delta V = 0.5 * (1-(\mathbf{0}+\mathbf{0}+\mathbf{0})) = 0.5$$

As we can see, for each cue, the association between that cue and the outcome *-s* goes up by 0.5 after this first encounter. Since 0 + 0.5 = 0.5, this is the association weight we have at this stage.

*Table 1. The association weights at the first step of learning*

| cues | -s ending |
|---|---|
| she | 0.5 |

| | |
|---|---|
| a | 0.5 |
| cookie | 0.5 |

The next sentence our learners hears is:

*She wants dolls.*

Now, the weight can change only for *she* and *dolls*, since *a* and *cookie* are not present. We already know something about the connection between *she* and *-s* – their association weight equals 0**.5**, and this is what we are going to put into our equation. The weight for dolls is 0, since the learner encounters this combination for the first time.

$\Delta V = 0.5 * (1-(0.5+0)) = 0.25$

Therefore, the association between *she* and *-s* equals 0.5 + 0.25 = 0.75 after this second encounter.

*Table 2. The association weights at the second step of learning*

| cues/outcomes | -s ending |
|---|---|
| she | 0.75 |
| a | 0.5 |
| cookie | 0.5 |
| dolls | |

What the learner knows about *she* and -*s* is going to influence what she learns about *dolls* and -*s*. The weight will increase by the same amount as for *she*, but since the previous weight for *dolls* was 0, its final weigh equals 0 + 0.25 = 0.25. This is the example of blocking that we have discussed above.

*Table 3. The association weights at the second step of learning (blocking).*

| cues/outcomes | -s ending |
|---|---|
| she | 0.75 |
| a | 0.5 |
| cookie | 0.5 |
| dolls | 0.25 |

Next, our learner hears:

*I want a cookie.*

She might be surprised. After all, she already expected *a* and *cookie* to be used with a verb that ends with *s*. Here, however the expected outcome is not present. The weight needs to be recalculated, and because the outcome is absent, lambda = 0. The weight of *I* is also equal to 0, as it is the first encounter. Therefore:

$\Delta V = 0.5 * (0 - (0.5+0 + 0.5)) = -0.5$

As a result, the association weights that represent our learner's knowledge of when to use *-s* look like this:

*Table 4. The association weights at the third step of learning*

| cues | -s ending |
|---|---|
| she | 0.75 |
| a | 0 |
| cookie | 0 |
| dolls | 0.25 |
| I | -0.5 |

The next sentence the learner comes across contains an error:

*I wants a cookie.*

This experience is going to increase the weight of association between *I*, *a*, *cookie* and the *-s* quite sharply, since their co-ocurrence is very surprising, given what the learner learnt before:

$\Delta V = 0.5 * (1 - (-0.5+0 + 0)) = 0.75$

*Table 5. The association weights at the fourth step of learning*

| cues | -s ending |
|---|---|
| she | 0.75 |
| a | 0.75 |
| cookie | 0.75 |
| dolls | 0.25 |
| I | 0.25 |

However, the next correct sentence will help the learner 'unlearn' the incorrect use:

*I want a cookie*

$\Delta V = 0.5 * (0 - (0.25+0.75 + 0.75)) = - 0.875$

*Table 6. The final association weights*

| cues | -s ending |
|---|---|
| she | 0.75 |
| a | −0.125 |
| cookie | −0.125 |
| dolls | 0.25 |
| I | −0.625 |

What the learner knows then, is that *she* is a good predictor of -*s*, whereas *I* predicts the opposite, since the weight is high, but negative. *A* and *cookie* also have negative weights, but only slightly below 0, and are not likely to sway the learner not to use -*s* on the verb when talking about cookies. Similarly, *dolls* are only somewhat positively associated with -*s*.

As the learning continues, the association between cues that rarely occur with a given outcome, or very often with other outcomes, will quickly decrease, since their presence does not say much about the probability of the outcome. Reliable cues, however, will become more associated with their outcomes and therefore more predictive. The goal is to learn to identify which of the cues are predictive of given outcomes (Ellis, 2006; Ramscar et al., 2010; Milin, Feldman, et al., 2017) – to weed them out of the noisy input.

Language learning may then be understood as learning which elements of the utterance pattern with other elements. The mechanism proposed is very simple, implicit and not stimulus-dependent. The same learning principles apply when learning whether tones signal shocks or whether words signal other words. As a result, a learner arrives at a probabilistic representation of the grammar of a language.

Crucially, this representation emerges from having experience with language without any explicit instruction or correction. Informative cues will be learned even if the input is erroneous, simply because the decrease in associative strength which will result from encountering an incorrect input will be quickly made up for when the cue appears with the correct outcome next time it is heard. Learning Theory, then, counters the poverty of stimulus argument and resonates well with what has been already pointed out by researchers in linguistics (e.g. Pullum, 1996): input alone provides enough evidence to learn what is and what is not possible in language. Therefore, error-driven learning fits perfectly into usage-based theory and offers a way of formalizing and describing the process of learning language from experience.

Finally, it's worth stressing that even in our simplified example, the learning outcomes are quite interesting for language modelling. On the one hand, they confirm the 'grammar book' rule that underlines the importance of the subject's person for the verb form selection. On the other, we see that other elements, which are not discussed in grammar books, might also play a role. The learning model might be, then, particularly useful when modelling those linguistic problems for which clear rules are difficult to formulate, but where speakers rarely hesitate in usage. As we will see later, aspect is one of those problems.

### 3.5 The Naïve Discriminative Learner

To simulate the learning of aspectual usage patterns based on error-driven learning principles, we will use NDL – the Naïve Discriminative Learner (Baayen et al., 2011; Milin, Feldman, et al., 2017). NDL is based on the Rescorla-Wagner model, although it simplifies it slightly. Only one free parameter is kept – the learning rate (written below as $\gamma$). In addition, $\lambda$ is assumed to be either 0 or 1, depending on the presence or absence of the outcome. Formally then, for each learning step *t*, the change of association strength between cue *i* and outcome *j* (represented as $\Delta w^t$) can be calculated using the equations presented in Table 7 below, where $\sum(w.j)$ is the sum of weights of associations between all cues and outcome *j* – an equivalent of $V_{A+B}$ in the equation we discussed in section 3.3.

*Table 7. Equations used in NDL for calculating the changes in association weights (Baayen et al, 2011; adapted from (Milin & Divjak, 2019))*

| | | |
|---|---|---|
| the cue is absent | no change | $\Delta w^t=0$ |
| the cue is present;<br>the outcome is present | positive evidence;<br>association strengthened | $\Delta w^t=\gamma(1-\sum(w.j))$ |
| the cue is present;<br>the outcome is absent | negative evidence;<br>association weakened | $\Delta w^t=\gamma(0-\sum(w.j))$ |

To obtain the value of association strength between cue *i* and outcome *j* for the current learning step ($w^{t+1}{}_{ij}$) we again take the weight obtained at a previous learning step ($w^t$) and add the change calculated for the current step ($\Delta w$):

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

As we can see, by increasing the association weight between a cue and an outcome each time both are present at the same time, NDL is able to learn that that cue is predictive of that particular outcome. On the other hand, each time a cue is present but the outcome is absent, the weight is decreased. As a result, cues that cooccur with a given outcome in a random, and hence insignificant way will have weights close to zero, reflecting the average of the positive and negative changes in weights over time. However, the model can also learn that cues which never occur with a given outcome are negatively associated with it. Such strongly negatively associated cues indicate that the outcome will not occur.

## 3.6 Justifying the error-driven learning paradigm

The choice of the error-driven learning paradigm for the purposes of investigating language phenomena might seem controversial. Linguists interested in language acquisition may object to the fact that the paradigm seems to equate language learning to "mindless" association formation, and ignore its subtleties and complexity. Corpus linguists could say that that there already exists a number of methods of calculating associations between words that are also based on co-occurrences in the input. Finally, machine learning specialists may point out that, as a learning algorithm, NDL is quite unsophisticated, compared to the state-of-the art solutions. In this section, we try to address these points and explain the reasons why we choose the error-driven learning approach.

### 3.6.1 Error-driven learning and linguistics

It needs to be underlined here that the idea that co-occurrences and their frequencies are important for language learning is not new in linguistics. As we have already discussed in Section 2, research in linguistics and learning seems to converge on the following point: what matters is not only how often the cues appear with a given outcome but also how often they appear with other outcomes.

Linguistics offers a wide variety of association measures that have been used throughout years of research on language; they take into account everything from raw frequencies to conditional probabilities (see e.g. Pecina, 2010; Divjak, 2019). There are, however, important differences between the Rescorla-Wagner model and other measures. Firstly, as mentioned above, the Rescorla-Wagner model takes into account not only how often the cues appear with a given outcome but also how often they appear with other outcomes. If two cues, A and B, appear very often with an outcome X, they will become highly associated both in a learning model and models that are based only on positive counts. However, if the cue B also appears with an outcome Y, but less frequently, the value of B will decrease for X only in a learning model that adjusts the associative strengths on the basis of both positive and negative evidence. As a result, A will be correctly identified to be predictive of X. In models that only count how often cues and outcomes appear together, the values of the cues A and B for the outcome X will remain equally high. This will result in learners incorrectly assuming that both A and B predict X equally well (Ramscar et al., 2010).

The need for keeping track of how often elements do and do not appear in a given context, as well as how often they appear in other contexts has also been pointed out by some corpus linguists, and there exist measures that include those counts (Stefanowitsch & Gries, 2003). However, these measures still calculate associative strengths for each cue individually. Without modelling how cues compete, the measures cannot explain phenomena such as blocking. Yet, Ellis (2006) shows how important the concept of blocking is for language learning. For instance, it is more difficult for second language students to learn tense inflectional endings when they already make use of adverbial phrases, since such phrases are already well associated with temporal reference. Therefore, learning other cues that predict the same meaning is slower.

However, the list of differences between the model of learning and corpus measures of associations is longer. In the Rescorla-Wagner model, experience is structured not only in the sense of how its elements are distributed but also how the learning events are distributed. The model takes into account not only the frequencies of co-occurrences of the elements of utterances, but also the order in which the utterances themselves are encountered. Given the same language sample, but ordered differently, learners might arrive at similar yet not identical representations. Such individual differences cannot be modelled by statistical models that are based on overall counts alone. However, NDL is learning in a step-by-step, incremental fashion, adjusting the association weights at each step. Therefore, it necessarily takes the order of learning input into account.

Moreover, unlike other measures of associations, the Rescorla-Wagner model allows for adjusting the parameters of learning such as salience. Some stimuli may be attended to less, and learning about them should be more difficult than about stimuli that are more salient. The same holds for language learning. Ellis (2006) points out that low salience of elements such as grammatical function words or bound morphemes, which are usually less stressed and reduced in speech, influences the pace of their acquisition: the more difficult to perceive they are, the longer it takes to acquire them.

For these reasons, the learning model and its computational implementation used in this dissertation are not merely tools for corpus analysis. Perhaps even more importantly, since the equations presented above are a formalization of a theory of learning, the Rescorla-Wagner model provides a cognitive underpinning that is lacking in many of the computational measures and models (Baayen, 2011; Divjak, 2019).

### 3.6.2 NDL and other machine learning algorithms

From the perspective of model architecture, NDL is very simple. This simplicity becomes more pronounced when we consider the complexity of state-of-the-art neural networks applied to language tasks in the area of Natural Language Processing. For example, GPT-3, a language model that wrote an article for the Guardian in 2020, has 96 hidden layers and 175B parameters used to compute the relation between the input and the output. In addition, Long Short-Term Memory and Transformer networks can process data sequentially, which means that the order of elements is preserved. As such, the state-of-the-art solutions seem to be better suited to deal with complex sequences of language, presented in a linear fashion with many references to items presented earlier, which requires some kind of representation of what we have encountered in the past. In fact, some of the neural networks models available today are so powerful that they outperform humans on certain linguistic tasks. For example, the leader board of the General Language Understanding Evaluation (GLUE) benchmark (Wang et al., 2018), a popular collection of language tasks used for testing computational language models, shows that there are currently 19 models that have a better overall score than humans across various linguistic tasks (accessed on 28.01.2022). It is worth noting here that the tasks included in GLUE are far from trivial. To achieve a high score, a language model must be able, for instance, to evaluate whether a sentence is grammatical or not, determine the sentiment of an utterance (positive or negative), and make a number of inferences, including resolving pronoun reference and deciding whether one sentence entails another. Importantly, GLUE favours models that can generalize from one task to another. That is, rather than training the models for each task separately, the leaders on the leader board use the same underlying representation of language to solve all the tasks with only minor adjustments in the outcomes. The scope of the tasks and the performance of these models far exceeds NDL's possibilities, which clearly pale in comparison.

However our goal as linguists, is not to achieve the best performance, but to explain how speakers learn language. Therefore, we think of Machine Learning as a methodological tool that allows us to specify the language learning principles we hypothesise are at play, and simulate learning on a sample of language data. Then, by comparing the results to the behaviour of speakers, we can test the principles themselves as well as other assumptions that we made (such as generalizations, existence of categories etc). Crucially, however, when comparing the predictions of the model to the behaviour of speakers, we do not strive to outperform people.

Instead we would like our model to perform *exactly* like humans – to succeed and fail in the same places, and *for the same reasons*.

### 3.7 Conclusions: learning theory and usage-based linguistics

We have mentioned in the previous chapter that usage-based linguists postulate that language is learned from experience, using a domain-general mechanism that takes into account properties of experience, such as frequency. As we have seen, learning is such a mechanism. It is necessarily based on experience; that is, no associations exist before the beginning of the learning process. It is domain-general, since the contingency of all kinds of cues and outcomes can be learned on the basis of the same principle. Moreover, the properties of experience are essential to learning, since its outcome– knowledge– is a representation of how likely things are to happen, given other things in the environment.

Learning theory explains how knowledge of language can emerge from the flux of experience and why it is so difficult to capture language in clear rules. As Milin at al. (2017: 2) put it: cues compete for outcomes in a never-ending 'tug of war that resists precise quantification by means of simple counts. Crucially, the association strength of a given cue to a given outcome is co-determined not only by how often this cue and this outcome co-occur, but also by how often this cue co-occurs with other outcomes'. Learning is a dynamic process that nonetheless results in some stable representations because what is experienced is not random but structured. Learning leads to – and is based on – the discovery of contingencies in the flux of experience.

The use of the Naïve Discriminative Learner allows us to apply reinforcement learning principles of animal learning to simulate the process of learning language. On the other hand, the simplicity of our model allows us to understand its underlying representations. In other words, we can inspect and interpret what has been learnt by the model, which is not a straightforward task when it comes to complex neural networks. As such, NDL can serve as a way of answering important theoretical questions. We can ask for example, what patterns in language usage humans– not machines– could detect, learn from, and use.

# Chapter 4. Aspect, its meaning and use

Having discussed the theoretical basis of the dissertation, we will now focus on a linguistic category which will serve as a case study in this dissertation, namely aspect and its use. To respect the tradition in aspectology, this chapter must start with the caveat that the following is by no means a complete or comprehensive overview of the vast literature on aspect. Instead, we will focus on the challenges that native learners of Polish must face when learning how to use aspect as a category. First of all, we will present different perspectives on aspect in order to establish what aspectual classes there are for speakers to learn, and what is the basis for distinguishing them. Then, we will present the experimental evidence which suggests that aspect is in fact a cognitive real category, which has an influence on conceptualization. Finally, we will discuss aspect from the usage-based perspective, focusing on whether there are any reliable patterns of co-occurrence which might serve as a basis of mastering aspectual use.

## 4.1 Grammatical approach to aspect

The grammatical approach to aspect assumes only two classes of verbs, perfective and imperfective, which in a subset of languages, such as Polish, are morphologically marked on each verb. As pointed out by the authors of Gramatyka Współczesnego Języka Polskiego (Grzegorczykowa et al., 1999) some verbs can be said to be inflected for aspect, as in the case of *kupić*- *kupować* (buy.PERF – buy.IMPF), where the imperfective counterpart is created by means of suffixation. There is a number of suffixes available, for example. -a-, as in *rzucić*-*rzucać* (throw.PERF – thow.IMPF), -ywa-, as in *odpocząć*-*odpoczywać* (rest.PERF-rest.IMPF) or -owa-, as in the example above. These suffixes are often accompanied by various changes in the stem (e.g. *obrać* -*obierać*: peel.PERF-peel.IMPF). In other cases, where the perfective verb is created by prefixation of the imperfective, as in *robić- zrobić* (do.IMPF – do.PERF), aspect seems to be a derivational category. The choice of a prefix is motivated lexically, and there is a number of items to choose from. Below, in Table 8, we present a list compiled by Łaziński (2020, pp. 19-20).

*Table 8. Perfective prefixes and their examples (Łaziński 2020:19-20)*

| | |
|---|---|
| z-/ze-/s-/ś- | robić – zrobić (do.IMPF – do.PERF ) |
| za- | głosować (vote.IMPF) – zagłosować (vote.PERF) |
| wy- | pić (drink.IMPF) – wypić (drink.PERF) |
| po- | smarować (spread.IMPF) – posmarować (spread.PERF) |
| u- | gotować (cook.IMPF) – ugotować (cook.PERF) |
| o- | golić (shave.IMPF) – ogolić (shave.PERF) |
| na- | pisać (write.IMPF) – napisać (write.PERF) |
| prze- | czytać (read.IMPF) – przeczytać (read.PERF) |
| roz- | propagować (propagate.IMPF) – rozpropagować (propagate.PERF) |
| w- | szamać (have_a_bite.IMPF) – wszamać (have_a_bite.PERF) |
| od- | restaurować (restore.IMPF) – odrestaurować (restore.PERF) |
| pod- | żyrować (endorse.IMPF) – podżyrować (endorse.PERF) |
| wz- | bogacić się (get_rich.IMPF) – wzbogacić się (get_rich.PERF) |

What the table illustrates is the fact that recognizing grammatical aspect on the basis of morphological markers is not an easy task, given the multitude of ways aspect can be expressed. To complicate matters further, the table includes only the prefixes and examples of so called 'pure' aspectual pairs, where adding the prefix changes only the aspect. Most of the verbs, however, can be prefixed with other items as well, which in addition to changing the aspect, also adds additional semantic information. For example, *pisać* (write.IMPF) forms a pure aspectual pair with *napisać* (write.PERF), but can also be prefixed with *pod-pisać* (sign.PERF), *do-pisać* (add_in_writing.PERF), *za-pisać* (take_a_note.PERF) etc. Each of these perfectives also have their own 'pure' imperfective counterparts: *podpisywać* (sign.IMPF), *dopisywać* (add_in_writing.IMPF), *zapisywać* (take_a_note.IMPF).

What is more, there are verbs for which the aspectual counterpart is suppletive, meaning that it is not morphologically related to the aspectual counterpart (e.g. *brać*:take.IMPF - *wziąć*: take.PERF), and some that do not have counterparts at all, like *mieć* (have.IMPF) or *ujrzeć* (behold.PERF). Finally, there is also a number of biaspectual verbs that are impossible to classify as perfective or imperfective on the basis of their form only (e.g. *amputować*: amputate or *ekshumować*: exhume).

Because of this confusing complexity, aspect is usually described as a classifying category, the features of which are expressed either lexically, by prefixes or by suffixes (Grzegorczykowa, 1999). That means that each verb in the Polish system is either perfective or imperfective – it has aspect but is not inflected by it (Grzegorczykowa, 1999). Similar views have also been expressed by other authors (Wróbel, 2001; Bańko, 2002).

This binary division is based on structuralist work, going back to Roman Jakobson (Bermel, 1997), which dominated Slavic linguistics in the 19th and the 20th century (Sasse, 2002). In addition to morphology, there exist distributional reasons for distinguishing a binary category, the most obvious one being that perfective verbs cannot be used in the present tense[1] or future compound tense. Given those restrictions, the grammatical approach took the existence of an aspectual opposition for granted and most of the work on grammatical aspect

[1] But note that perfectives can be used to express "dispositional habituals" Klimek-Jankowska (2008, 2012) suggesting an atemporal reading and a generalization about the subject, as in "Jan pomoże w potrzebie." (Jan will help.PERF in need).

has been dedicated to establishing invariant meaning of both or one of the aspectual classes (Janda, 2004) – a remnant of the classical approach to categorization in general, according to which all members of a category must meet certain criteria or, in other words, fit the definition to qualify for inclusion.

Given the amount of literature available on the subject it is clear that the invariant meaning is not easy to define. Intuitively, the difference between the imperfective and perfective verbs is such that the latter is related to the completion of an event, whereas the imperfective signals its incompleteness or 'ongoingness'. When we consider the following examples:

1) Wczoraj czytałem książkę.
   Yesterday read.1SG.PAST.IMPF book
   I was reading a book yesterday.

2) Wczoraj przeczytałem książkę.
   Yesterday read.1SG.PAST.PFV book
   I read a book yesterday.

we see that even though both sentences describe a situation that happened in the past, the exact information they convey differs. Sentence 2 informs us that the reading of the book was completed, while sentence 1 presents the situation as ongoing or being in the process of happening.

Aspect can be then said to express a perspective that speakers take on the event. To use the metaphor proposed by Comrie (1976), we can say that sentence 1, where the imperfective was used, presents the situation 'from inside'. On the other hand, sentence 2, where the verb used to describe the situation was perfective, can be said to present the same situation 'from outside'[2].

Another notion that is often used to explain the difference between the perfective an imperfective meaning is totality. Comrie (1976) proposes that the perfective presents the situation 'as a whole' while 'the imperfective pays essential attention to the internal structure of the situation' (Comrie, 1976: p.16). A similar view is also proposed for instance by Forsyth (1970) who sees totality as 'the action as a total event summed up with reference to a single specific juncture' (Forsyth, 1970, p.8). To apply this notion to interpret the examples presented above, we can say that sentence 2 presents the reading as a singular situation, and we are not concerned about each individual step that was needed to take in order to finish reading. Instead, we view the entire reading as one, and are more concerned with the result – the fact that the book has been read. In sentence 1 on the other hand, the process of reading itself is the main focus.

Similarly, we can say – referring to the notion of boundedness (e.g. Smith, 1986)– that sentence 2 includes both the beginning and the end (or the initial and final boundary) of the action, whereas sentence 1 does not include such boundaries.

As we can see these notions are rather abstract and related to each other, if not synonymous (Janda, 2004). A possible explanation of their general nature is that they were conceived to describe the meaning of aspect in *all* its uses. However, as we will see, these notions do not apply to all verbs in exactly the same way.

For example, as pointed out by Laskowski (1999), while verbs in sentences such as:

---

[2] While this and the following notions were originally proposed for other languages, we discuss them here as they illustrate which dimensions of temporal semantics are considered particularly relevant for the aspectual meaning in the literature.

| | | | |
|---|---|---|---|
| 3) | Bardzo | się | postarzał. |
| | Very | REFL | get_old.3SG.PAST.PERF |

He got very old.

can be said to present an action without paying attention to ‘temporal constituency’, it is difficult to interpret them as referring to a single juncture or a point in time, as they clearly describe a process that extends over time. In addition, the final boundary is not signalled by this verb, because *postarzeć* (get_old.PERF) does not indicate that the process of getting old has been completed. Similarly, perfective verbs such as *zachorować* (get_ill.PERF), do not signal a completion either, since it is the beginning of a situation, not the final boundary that is expressed.

What is more, in sentences such as 4:

| | | | | | | |
|---|---|---|---|---|---|---|
| 4) | Asia | przesiedziała | godzinę | nad | zadaniem | i |
| | Asia | sit.3SG.PAST.IMPF | hour | over | homework | and |

| | | | |
|---|---|---|---|
| dalej | nad | nim | siedzi. |
| still | under | him | sits |

Asia sat doing the homework for an hour and is still doing it. (Laskowski 1999: 159) we could say that the verb, even though it refers only to a part of an entire situation, and presents it ‘as a whole’, but at the same time it does not indicate a change of state, as the verb *przeczytać* (read.PERF) does in sentence 2.

What is more, while *czytać* (read.IMPF) – *przeczytać* (read.PERF) form a pair where the imperfective can be said to be unbounded and describing a situation as ongoing, with the focus on the process, while its perfective counterpart is bounded and signals that the process has been completed and a result obtained, we cannot say the same about pairs such as: *zakochiwać* (fall_in_love.IMPF) – *zakochać* (fall_in_love.PERF) or *spotykać* (meet.IMPF) – *spotkać* (meet.PERF), where the imperfective signals iterativity or repeatedness rather than unboundedness. In addition, in pairs where the imperfective carries the iterative meaning, the way the endpoint or the final boundary is reached in their perfective counterparts differs from the pair in sentences 1 and 2. If verbs such as *spotkać* (meet.PERF), *kopnąć* (kick.PERF), *kichnąć* (sneeze.PERF) and *mrugnąć* (blink.PERF) signal reaching a boundary, it is not by a completing a processes but rather the endpoint is achieved instantaneously.

What these examples indicate is that the meaning of aspect – or at least the details of it – depend on the inherent meaning of a verb itself. These intricacies of aspectual meaning seem to be better resolved in the lexical approaches to aspect, for which the meanings of the verbs themselves are the point of departure. However, the exact number of verbal categories that can – and should – be distinguished remains disputed.

## 4.2 Lexical aspect

One of the most basic lexical aspectual distinctions is concerned only with the notion of an endpoint, which, much like the grammatical approach, distinguishes only two classes: telic and atelic verbs (e.g. Garey, 1957). Those classes are based – not too dissimilarly from grammatical aspect – on the notion of limit, endpoint or a change of state. Telic verbs express an action going toward an end goal, whereas atelic verbs lack this notion. As a common test shows, telic verbs such as recover in sentence 5, can only be used with and adverbial phrase *in an hour*, which tells us how much time it took to reach the goal, and not with *for an hour*, which informs us only about the duration of an event.

5) John recovered in an hour

   *John recovered for an hour

6) John swam for an hour

   *John swam in an hour. (Filip, 2012)

However, other researchers show that these two categories can be divided further into more fine-grained subcategories. Zeno Vendler (Vendler, 1957), whose classification of events is probably most influential in the studies of lexical aspect, proposed four classes instead of two: states, activities, achievements and accomplishments. This classification is based not only on the notion of endpoint but also on the way this endpoint is achieved. For those events that lack an endpoint, he distinguished between those that are static (states) and those that are dynamic (activities). On the other hand, the endpoint in events that do have it, can be reached either instantly (achievements) or gradually (accomplishments). Vendler's four-way classification is presented in Table 9 below.

*Table 9. Lexical aspect categories proposed by Zeno Vendler (1957).*

| class | features | example |
|---|---|---|
| states | static<br>no endpoint | I know English. |
| activities | dynamic<br>no endpoint | I'm driving a car. |
| achievements | occur instantly<br>have an endpoint | I noticed my error. |
| accomplishments | happen gradually<br>have an endpoint | I wrote a letter. |

Comrie (1976) and later Smith (1997) added one more category to the Vendlerian classification. Focusing on dynamicity, telicity and the duration of events, they also distinguish s*emelfactives* – those events that are dynamic, and do not have endpoints (which means that they are atelic). In addition, such events are punctual – they occur instantly, such as the action of sneezing, in the sentence below.

7) John sneezed loudly.

The categories of the 5-way model of lexical aspect can be described by the combination of the presence or the absence of three features, as presented in Table 10.

*Table 10. The categories of lexical aspect in the 5-way model and the combinations of the features that define them (Smith 1997: 20)*

| | Dynamic | Durative | Telic |
|---|---|---|---|
| States | - | + | - |
| Activities | + | + | - |
| Accomplishments | + | + | + |
| Achievements | - | + | + |
| Semelfactives | + | - | - |

Even though the works presented above were primarily concerned with English, the lexical approach has also been used to classify events in Slavic languages. One important contribution for Polish is Laskowski (1999). In his analysis, Laskowski distinguishes two main classes – states and dynamic actions, further divided into 8 categories based on the combination of 4 semantic features: dynamicity, change of state, telicity and control (of the subject) (Laskowski 1999: 156). He 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 have all but one feature: they are not controlled) can be said to form the aspectual opposition proposed in grammatical approaches. In other cases pairs can be formed, but the meaning changes. Such lexical-aspectual relations, unlike aspectual relations of telic verbs, are not regular and depend on the semantics of the verbs (Laskowski 1999).

To complicate matters further, the class of a verb cannot be determined without taking into an account its context. For this reason, the prevalent view in literature is that it is better to consider lexical aspect as a property of verbal predicates and to simply verbs alone (Verkuyl, 1972; Krifka, 1989). Verykul (1972) shows for instance that the verbal arguments influence the aspectual class to which the verb can be classified. Let us consider two examples below:

8) Mary ate a roll.

9) Mary ate rolls.

Using categories proposed by Vendler, we would need to classify ‘ate’ in sentence 8 as an accomplishment, since it the situation described in the sentence is clearly bounded and the end-point was reached over time. On the other hand, sentence 9 describes a process without and endpoint, hence ‘ate’ in this sentence should be classified as an activity. This example clearly shows that verbal complements need to be accounted for, one way or another.

The importance of the context holds also in Polish. Laskowski (1996) shows that one verb might be classified into two different categories, depending on the context it appears in. For instance, lie can either be a state (that is, according to his classification of verbs: a static, atelic situation which does not denote any change and is not controlled by the subject), as in sentence 13 or a position (that differs from states in that the subject exercises control over it), as in sentence 14. However, without any context, it is impossible to say what type of situation is actually expressed.

13) Zegarek leżał na stole.

| | | | |
|---|---|---|---|
| Watch | lie.3SG.PAST.IMPF | on | table. |

The watch was lying on the table

14) Jan leżał na łóżku.

Jan lie.3SG.PAST.IMPF on bed

Jan was lying on the bed

Łaziński (2020) also discusses interesting examples of recategorization. He shows for instance, that whereas *zaczynać się* (begin.IMPF) can be classified as an accomplishment, as in sentence 15:

15) Koncert długo się zaczynał, aż w końcu się zaczął.

Concert long REFL started, until in end REFL started.

It took a long time before the concert started. (Łaziński 2020:146)

it can also be interpreted as a state, given certain contextual conditions:

16) Za łąką zaczynał się las.

Behind meadow start REFL forest.

Behind the meadow there was a forest. (Łaziński 2020:146)

Similarly, *kucać* (crouch.IMPF), can be interpreted as an achievement in sentence 17:

17) Musiał kucać, żeby uniknąć uderzenia w głowę.

Must crouch.INF.IMPF to avoid hit in head.

He had to crouch to avoid getting hit in the head.

But in sentence 18, it clearly is a state:

18) Kucał przy ogniu.

Crouch.3SG.PAST.IMPF beside fire.

He was crouching next to the fire.

As we can see, the meaning of aspect is a complex, not easily definable problem, that necessarily takes into account both the semantics of the grammatical aspect as a category, the inherent meaning of the verbs and the contexts in which they are used. Given these complexities, it becomes clear why the invariant meanings and the definitions proposed for the grammatical aspect are abstract and metaphorical in nature. Since they need to encompass various possible interpretations of verbs in context, they must remain general if they are to be applied.

However, in some uncontroversial cases, these notions apply in a rather straightforward way that also aligns with the intuitive way of explaining aspectual meaning. These cases can be described as central, or to use terminology proposed by Rosch (1973) and prevalent in cognitive linguistic – prototypical.

This view is expressed for instance by Łazinski and Wiemer (1995), who take terminativity, a concept introduced by Maslov, as central for the study of aspect. They note however, also referring to work by Rosch, that terminativity should be viewed as a category that has a prototype in the centre. Prototypically terminative members are those transitive verbs, for which perfective expresses a visible change, such as the creation of new object, and the effects of those actions cannot be reversed or cancelled. The perfective variant differs from imperfective only in one feature – the imperfective expresses going towards a goal and the perfective expresses achieving that goal (Łazinski and Wiemer, 1995).

Grzegorczykowa (1997) proposes a very similar view on aspect as a category. She also treats telic verbs that name actions and processes that can be completed or move towards completion as prototypically perfective. They form typical aspectual pairs, in which imperfectives signal either the fact that the action is in progress, or its repeatedness (iterativity). Prototypically imperfective verbs, on the other hand, denote states, relations, actions that cannot be completed (e.g *myśleć*: think.IMPF) and movements (*tańczyć*: dance.IMPF). For some of them, perfectivization is not possible (*umieć*: know.IMPF). For others, perfectives may have an inchoative function (denote the beginning of the action– *pokochać*: fall_in_love.PERF) or specify some temporal or quantitative feature (*poleżeć*: lie_for_a_while.PERF) (Grzegorczykowa, 1997).

What the descriptions based on the concept of prototype have in common is that they show that the 'invariant meaning' can be viewed as a feature of the subgroup(s) of verbs, which may be considered prototypical for the whole aspectual category. That explains why pairs like *gotować* (cook.IMPF) – *ugotować* (cook.PERF) or *pisać* (write.IMPF) – *napisać* (write.PERF) are usually described as 'pure aspectual pairs' and why the generalizations introduced in grammar books for first and second language learners usually underline the opposition between ongoing process and a completed result. Such general notions do tell us something, but because of the structure of the category itself, they cannot tell us everything.

## 4.3 The scope of the dissertation

To sum up the discussion so far, we can say that aspect is a classifying category, the meaning of which is usually described in abstract terms. While these notions can be easily applied in prototypical uses, the details of aspectual meaning clearly depend on the inherent meanings of verbs, for which various classifications have been proposed, as well as the context in which these verbs are used. Given these complexities, we may ask if these abstract semantic notions and taxonomies are in fact useful for native speakers or whether – as posited by usage-based approach – aspectual use can be learned on the basis of patterns available in the input instead. Before we can tackle these questions, however, we must first define the scope of the dissertation.

Based on the what we have discussed so far, the main problem of this dissertation – what aspectual categories speakers of Polish actually know and how they acquire them – could be approached from both lexical and grammatical perspectives. However, we are going to focus only on grammatical aspect, for the following three reasons.

First of all, despite the difficulties in defining the semantics of the two aspectual classes, linguists working on Slavic languages still have to reconcile with the fact that there exist good morphological and distributional reasons for distinguishing a binary category. As mentioned, Polish verbs necessarily express grammatical aspect and there are clear restrictions in use, so the questions of learnability and cognitive plausibility of grammatical aspect categories are particularly important for Slavic linguistics. Secondly, despite the theoretical discussions in linguistics, native speakers seem to be blissfully unaware of the subtleties and complexities of temporal semantics and readily accept the basic grammatical labels and definitions. The question here, of course, is whether it is easy because they already know use the category or because it is something taught at school and never questioned. Thirdly, given the number of lexical categories that have been postulated, it would be inefficient and probably incomplete to try to model them all. However, good performance of the grammatical aspect model would suggest that lexical distinctions – while not irrelevant – are not necessary to predict the usage. Conversely, if a model based on grammatical aspect fails, this would indicate the need to introduce more fine-grained subcategories.

Having defined the scope of the dissertation. we will now review existing studies related to its points of focus. First, we discuss behavioural evidence showing that grammatical aspectual categories are in fact cognitively plausible. Then, we review corpus studies to determine whether there is enough structure in the usage of perfective and imperfective for speakers to learn these two categories from input alone.

## 4.4 Grammatical aspect and behavioural studies

The existing experimental evidence suggesting that perfective and imperfective verbs affect conceptualization of events differently and the distinctions between them seems to be related to the notions of ongoingness and completeness. Madden and Zwaan (2003) show that participants who read English sentences in the past simple tense chose pictures showing a complete event more often and faster than those who read sentences in the past progressive. These results suggest that simple past sentences are interpreted as completed and the progressive sentences are interpreted as ongoing. A similar conclusion can be drawn from a study conducted by Morrow (1985), who shows that people are more likely to place a character along the path rather than at a goal when they hear an imperfective description of the situation. Similarly, in a study conducted by Anderson et al. (2008) participants listened to English sentences describing a person moving along a path either in the simple or progressive past. Then, they were asked to place a human character on an image depicting this path. The results show that, having heard progressive sentences, participants were more likely to place a human character at the centre or the beginning of the path. Having heard past simple sentences, however, participants were more likely to place the character towards the end of the path. These results were reproduced in Anderson, Matlock, Spivey (2013).

Other studies have shown that grammatical aspect influences what mental models of the events speakers form, and demonstrated that these models affect the accessibility of other elements of the event. For example, Ferretti, Kutas & McRae (2007) demonstrated that participants were faster to read the names of locations (e.g. arena) when they were related to the progressive verb phrase primes they saw before the name (e.g. was skiing). The results indicate 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 find similar activation effect for participants and instruments (Carreiras et al., 1997; Magliano & Schleich, 2000; Madden & Zwaan, 2003). Golshaie and Incera (2020) also find that activation of instruments depends on verbal aspect and demonstrate that participants were more likely to erroneously indicate that an instrument was mentioned in the sentence, if the sentence contained an imperfective verb. The results of this study not only contribute to the body of evidence that grammatical aspect evokes different conceptualizations, but also shows that the effect can be found in languages other than English. Salomon et al (2013) show that the differences in what is more accessible and active, evoked by imperfective and perfective aspect, also influence complex cognitive processes such as problem solving.

Unfortunately, comprehension and conceptualization of aspect in Slavic has not been extensively studied using experimental methods (for a comprehensive review of psycholinguistic studies in Slavic linguistics see Sekerina, 2006, 2017) One study that sheds more light on the matter is Bott and Gattnar (2015). In their study, the participants read Russian perfective sentences in which the verb was preceded by an adverbial that either matched or did not match the aspect. The results indicate that the Russian speakers were able to detect the mismatch immediately upon reading the verb, which suggests that the aspectual markers on the verb convey important semantic information required for a judgement task.

Another study worth discussing here is Janda and Reynolds (2019) who show that grammatical aspect categories are useful in describing participants behaviour and that the choices speakers make are dependent on the context in which the verbs are used. The participants were asked to read naturalistic texts in Russian and to mark on a three-point Likert scale to what extent perfective and imperfective variants fit into the context. The results show not only that in most cases (81%) the majority of the participants agree on which aspectual variant of the verb should be used, but also that the usage seems to form a continuum at one end of which aspect, and its meaning is fixed – 'anchored by context, highly redundant and the choice (...) is tightly constrained' (Janda & Reynolds, 2019, p. 432).

Taken together the studies seem to suggest that grammatical aspect is not merely a descriptive category proposed by linguists but also potentially a cognitively plausible distinction made by native speakers. However, the studies discussed above are mainly concerned with differences in conceptualization and comprehension, not with learnability. That is, they already assume some potential categories of aspect, without asking how these categories come to be. In this dissertation, however, we want to approach the problem differently. Therefore, we want to first ask what categories can be learned on the basis of usage, and only then validate models experimentally. However, the final step we need to take before we present the empirical work is to discuss the usage patterns of aspect that might serve as a basis of learning.

## 4.5 Aspect from a usage-based perspective

In this section, we focus on how usage-based theory can be employed to describe Polish aspect. Just like anywhere else in language, patterns, frequency biases and preferences can also be found in aspectual usage. The most obvious example of the constraint of choice due to distribution is the fact that in Slavic languages, including Polish, perfective verbs cannot be used in the present tense or in a future compound tense. But research shows that there is more to the distributional story. First of all, there is a number of studies showing statistical biases of aspectual use with certain elements of context. As many descriptive grammarians have noted, aspectual choice is limited by the use of temporal adverbials. Phrases such as 'for a while' or 'often' cannot be used with perfective verbs and corpus studies show that temporal descriptions can be considered reliable cues. Koranova and Bermel (2008) used the Czech National corpus to investigate the aspectual preference of other tense-aspect markers, such as conjunctions, adverbs and temporal adverbial phrases. They demonstrate that a number of them show 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) shows 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 marker. The availability of adverbial cues is indeed very low. Reynolds (2016) estimates it to be at only 2% of all sentences. Nonetheless, 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.

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 categorization, including perfective and imperfective. Corpus studies on aspect show a number of strong tendencies in aspectual use. Rice and Newman (2004) show 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, shows that the type of reflexive pronoun, semantic role of the subject and object, and even the person and number of the subject may matter for the 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 enough information for the learners to potentially acquire categories of grammatical aspect. Learning what in the context serves as a good cue and what verb it predicts might help learners notice similarities between verbs and facilitate a formulation of more general categories, such as grammatical aspect.

To conclude the theoretical chapters, we can now explicitly state the research questions we will attempt to answer in the remainder of the dissertation. First of all, we ask how well can usage-based models predict aspectual choices compared to models based on abstract semantic labels? In other words: how predictive are the labels usually applied to describe aspectual usage and can the choices people make be better explained by the patterns available in usage? Secondly, we ask if these cues can serve as a basis on which the aspectual categories are formed. Finally, we want to know how sensitive speakers are to the cues deemed most predictive by the usage-based models. Can the inclusion of these cues influence comprehension of perfective and imperfective sentences?

We investigate the first research question in Chapters 5 and 6. In Chapter 5 we present the computational simulations, conducted using the Naïve Discriminative Learner, in which we trained models to learn which cues are particularly helpful in predicting the aspectual use. We will also use the simulations to determine whether aspect as a category is a useful generalization which facilitates learning, or whether the models perform better when only lexical items are provided as outcomes. In Chapter 6 we validate the models experimentally, using behavioural data.

In later Chapters (7 and 8) we analyse the most predictive usage-based model to identify the most important cues, as well as the shortcomings of the model. In Chapter 9 we investigate whether the usage-based patterns can serve as a basis of forming aspectual categories. In Chapters 10 and 11 we move to the third research question and investigate whether speakers of Polish are influenced by the cues that, according to our model, are strongly associated with perfective and imperfective.

# Chapter 5 Corpus-based learning simulations

In this chapter we present the results of the corpus-based learning simulations we conducted using the Naïve Discriminative Learner. As we mentioned in the previous chapter, the goal of these simulations is to establish what information speakers might utilize while learning when to use perfective and imperfective. Can aspectual usage be learned from distributional patterns, as proposed by usage-based theory? Or do semantic distinctions offer a more reliable way of describing the linguistic behaviour of speakers? And if they do, then how are these semantic distinctions themselves learned from what is available to the speakers in the input?

Unlike in experiments with humans, computational modelling allows us to strictly control for the type of information each model is trained on. Therefore, each of the models we will present in section 5.7 can be treated as a separate learner to whom different types of input are available. For example, one of them, trained only on distributional information, with no access to a semantic interpretation of the situation in the sentence, can be said to represent a usage-based learner. The other, with access only to abstract semantic labels, represents a learner who uses the distinctions proposed by the traditional approaches to aspect. It is important to note, however, that the models do not necessarily represent the children's acquisition of aspect. That is, we do not attempt to simulate the learning process from an early age, as that would require a different type of training data, which should more closely represent the input children receive. Nonetheless, by comparing the performance of the models, we can pit theoretical approaches against each other and determine which patterns in the data are reliable enough to serve as a basis of human learning.

In addition, we further investigate the reliability of the traditional approach in two ways. First, by training a model to predict the abstract labels from the distributional patterns in order to test whether the semantic distinctions can be learned from input. Secondly, by measuring the interannotator agreement for a subsample of the data. In the following sections we present the data the models were trained on, as well as the details of how each training was conducted. Finally, we discuss the performance of each of the models and the insights these learning simulations provided.

## 5.1 Data preliminaries- frequency profiles of verbs

As discussed above, previous research shows how important frequency is and how it influences the choices speakers make. Given these findings, it is reasonable to assume that when it comes to the use of aspect, differences in frequencies of aspectual counterparts may also play a role. The speakers might be more likely to use one variant over the other when the latter is used less frequently. Therefore we started with a preliminary investigation of the distribution of the aspectual counterparts in pairs.

First, using the data from Polish Wordnet, we obtained an extensive list of aspectual pairs[3]. Polish Wordnet contains information on the aspectual links between verbs, as well as the type of their aspectual relation. In other words, using this resource, it is possible to both find the aspectual partners of a given lemma, e.g. pisać – napisać (write.IMPF – write.PERF); pisać – podpisać (write.IMPF – sign.PERF), as well as annotate their relationship (i.e. ‘pure’ or ‘secondary’). We filtered the list to keep only 'clear' pairs, using the values of aspectual relation

[3] The list of aspectual pairs along with the additional information that allowed for filtering was extracted from the Polish version of Wordnet (Słowosieć) by mgr.inż. Tomasz Naskręt and dr Agnieszka Dziob, who work on the project. I am very grateful to both of them for their help and this resource.

(‘pure’ or ‘secondary’) for which the pairs were annotated[4]. That is, if the verb had multiple counterparts, we took the one without any additional shades of meaning, so that the difference between the two verbs was only in the semantics of aspect. For instance, we considered *pisać* (write.IMPF) - *napisać* (write.PERF) to be a ‘clear’ aspectual pair, whereas pairs such as *pisać* (write.IMPF)- *podpisać* (sign.PERF) were excluded, because adding the perfective prefix also modified the meaning of the verb.

Next, with the help of the data from the dictionary included in the Morfeusz package (Kieraś & Woliński, 2017), we also excluded pairs for which we could not reliably count the frequencies of verbs by looking at the word form. For instance, *stać* can either be an imperfective meaning 'stand' or a perfective counterpart in the pair that means 'become': *stać-stawać*. Since there is no information in the corpus that would allow this kind of disambiguation, frequency counts would be incorrect.

The frequencies of the verbs forms and lemmas were obtained from the Araneum Corpus of Polish (Benko, 2014). We decided to use the Araneum corpus for two reasons. Firstly, it is a compilation of online texts – websites, blog posts and forum posts. Therefore, it is closer to spoken language than the National Corpus of Polish (NKJP), which consists mainly of more formal texts – books and newspapers. As such, the texts in Araneum should be less restricted by the stylistic requirements of more formal writing and reflect the choices speakers usually make. Secondly, it is freely available, which made it is possible to make frequency calculations and extract the sample using python scripts instead of processing the corpus using web interfaces.

[4] The python scripts used to filter the list, calculate the frequencies of verbs, and to extract the corpus sample were written by or in collaboration with my teammate Christian Adam, who worked on a series of computational studies in the Out Of Our Minds project. I am really grateful for his help and the time he spent explaining the subtleties of the code to me.

We discovered that the lemmas in the tags in the annotated version of the corpus are not always correct. To avoid incorrect counts, we first counted the occurrences of the inflected forms of the verbs and summed them up to obtain frequencies for lemmas

Finally, for each pair of verb lemmas on the filtered list, we calculated a perfective bias, which we defined as a proportion of the perfective lemma in the sum of frequencies of the pair. The final list contained 3324 pairs of lemmas, their frequencies and perfective bias.

We established that verbs in aspectual pairs are indeed used with different frequencies (M=0.58, SD=0.27). Figure 1 below illustrates the distribution of the perfective bias in our sample of pairs.

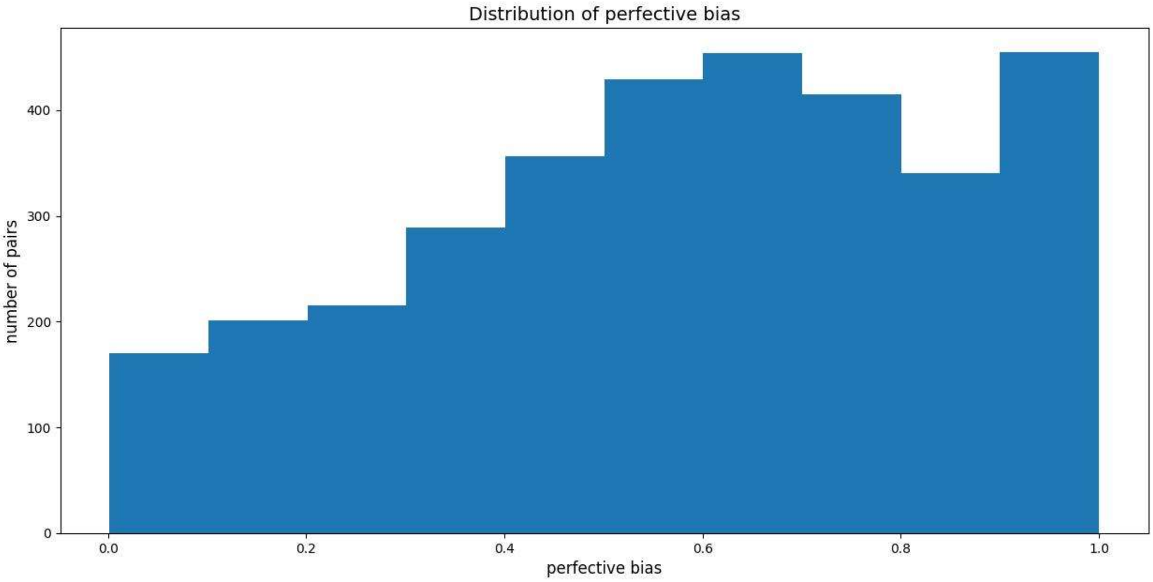


*Figure 1 The distribution of the perfective bias in aspectual pairs*

As we can see, in most pairs, the values of the perfective bias are over 0.5, which indicates that in these pairs the perfective is used more often. This certainly is a valuable insight, as it suggests that the usage patterns of aspect make the perfective more likely, if both perfective and imperfective are available.

Interestingly, 456 pairs had a very strong bias towards the perfective, with the value over 0.90, forming a second peak in the histogram. Out of these, 57 pairs contained a low frequency perfective for which no imperfective counterpart was found in the corpus, and therefore the value of perfective bias for these pairs was 1. However, there is also a number of pairs in which it is the imperfective that is used more frequently. The values of the perfective bias for these pairs are below 0.5. Such a distribution means that in order to control for differences in frequencies between the two aspectual classes in any experiments where we will ask participants to choose between aspectual variants (as we will in Chapter 6), we need to sample verb pairs along the continuum of perfective bias. This will allow us to include pairs where both the perfective and the imperfective are more likely, as well as the pairs where the frequency difference is more balanced.

## 5.2 The selection of verbs

As mentioned in the previous section, we decided to sample the verbs along the continuum of perfective bias to make sure we control the effect of relative frequency of aspect in the pairs. We sampled 18 verbs (9 pairs) stratified sampling over frequency-bands, as implemented in the pyndl bandsample function (similar in application to Divjak et al., 2021). Essentially, this function divides the range of possible values (in our case of perfective bias) into bands and then randomly selects samples from each band. As a result, the sample is evenly distributed along the set of values, while also making sure that the individual samples are randomly selected. The verbs, their frequencies and perfective bias are presented in Table 9 below:

*Table 9. The final sample of 18 verbs*

| lemma | lemma frequency | aspect | competitor | competitor frequency | competitor aspect | perfective bias |
|---|---|---|---|---|---|---|
| rozliczyć (square/appraise.PERF) | 9467 | perfective | rozliczać (square/appraise.IMPF) | 15500 | imperfective | 0.38 |
| urzec (charm.PERF) | 3364 | perfective | urzekać (charm.IMPF) | 3647 | imperfective | 0.48 |
| nabrać (gain/kid.PERF) | 25432 | perfective | nabierać (gain/kid.IMPF) | 21423 | imperfective | 0.54 |
| rozczesać (comb.PERF) | 624 | perfective | rozczesywać (comb.IMPF) | 421 | imperfective | 0.60 |
| wywrócić (topple/fall.PERF) | 2504 | perfective | wywracać (topple/fall.IMPF) | 1382 | imperfective | 0.64 |
| zastąpić (replace.PERF) | 50153 | perfective | zastępować (replace.IMPF) | 22092 | imperfective | 0.69 |
| podmienić (switch.PERF) | 1782 | perfective | podmieniać (switch.IMPF) | 593 | imperfective | 0.75 |
| opracować (develop.PERF) | 69870 | perfective | opracowywać (develop.IMPF) | 14542 | imperfective | 0.83 |
| odszukać (find.PERF) | 44391 | perfective | odszukiwać (find.IMPF | 580 | imperfective | 0.99 |

## 5.3 Discourse chunks

For each of the 18 verbs, we randomly sampled 100 target sentences – i.e. sentences that contained the verbs of interest. However, knowing that some of the examples will have to be excluded due to the nature of the corpus, which – being a collection of texts published online – contained incomplete or otherwise unusable examples, we sampled more sentences than we needed. We extracted one random sentence per document and proceeded to the next document until we collected 400 sentences per verb. These were later manually filtered to weed out gibberish and incomplete sentences, and the first 100 examples for each verb were annotated. In two cases (*podmienić* and *odszukać*) more than 400 examples had to be reviewed in order to obtain a sample of a 100 useable sentences.

In addition, for each target sentence we extracted a context of up to six preceding sentences. This allowed us to capture any possible cues that might trigger a use of one of the aspects, which were not present in the target sentence. The exact number of sentences in the window depended on the amount of preceding sentences available. For instance, 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. The results of Light and Anderson (1985) indicate that the proportion of correct recall of a pronoun referent in the text decreases with the distance and is only slightly above chance level for distances of 6-7 sentences. Given these findings, we decided to include maximum 7 sentences in a chunk: 1 target and 6 preceding sentences.

## 5.4 Sample reliability

As mentioned before, the sample had to be manually processed to remove unusable examples. The amount of chunks we had to discard raised important concerns about the accuracy of frequency counts in the corpus. To investigate if our frequency counts can be trusted, we calculated correlations between frequency counts of the verb lemmas we obtained from Araneum and the frequencies of the same lemmas in other corpora using Spearman's rank-correlation. For comparison we used the counts from the balanced sample of NKJP (300mln tokens), full NKJP (over 1.5bln tokens) and a compilation of corpora prepared by Broda and Piasecki (Broda & Piasecki, 2011.8bln tokens). The results show that the frequency counts were reliable and strongly correlated between different corpora (Table 10)[5].

*Table 10. The correlations between lemma frequency counts in 4 different corpora*

| | Araneum | NKJP balanced | NKJP full | Broda Piasecki |
|---|---|---|---|---|
| Araneum | 1.0000000 | 0.9353383 | 0.9383459 | 0.9338346 |
| NKJP balanced | 0.9353383 | 1.0000000 | 0.9984962 | 0.9699248 |
| NKJP full | 0.9383459 | 0.9984962 | 1.0000000 | 0.9639098 |
| Broda Piasecki | 0.9338346 | 0.9699248 | 0.9639098 | 1.0000000 |

[5] However, the fact that so many examples had to be removed means that any lexical computational model trained on the whole, unfiltered Araneum corpus could be unreliable. We initially planned to conduct an additional learning simulation, training a model to predict target verb forms using all other words that occurred in the chunk as cues. Such a model would not require any manual annotation and could potentially reveal associations between the aspect of the verb and specific words or phrases used in the context. However, since it would require much cleaner training data, we decided not to pursue this avenue.

## 5.5 Annotation

The sample of 1800 discourse chunks was annotated by the author, who is a native speaker of Polish. Another annotator, also a native speaker of Polish, annotated a subsample of 100 randomly selected chunks. The smaller data set was used to calculate inter-annotator agreement, which we discuss later, in section 5.6.

We annotated the sample for 30 'concrete' variables. While some of the variables from this subset have already been shown to be relevant for aspectual usage (we discussed them in Chapter 3), we extended the number of variables to follow the Behavioural Profiles methodology (Divjak & Gries, 2006). This approach advocates for creating a detailed description of the context in which a target item is used, in order to fully capture its usage. This allowed us to not only look beyond the immediate context of the verb, but also to investigate whether there are variables other than the most discussed candidates that are predictive of aspectual usage. Using the Behavioural Profile methodology also ensured that the cues on which we trained our models are theoretically relevant.

In addition, we annotated for the 7 abstract semantic distinctions, most commonly used in the aspectual literature. The full list of variables and their operationalizations is in Appendix 1. We will now discuss the variables in more detail.

*Concrete variables*

The set of concrete variables can be divided into three groups that described the verb, its immediate context and important elements further back in the chunk. The first group captures the properties of the verb, as used in the target sentence. We annotated for the tense, mood, presence of auxiliary verbs and reflexive pronouns.

The second group describes the features of the clause. We marked whether the verb was negated or not, and whether the clause was 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 more traditional labels – subject, direct and indirect object – because the variables were intended to capture features that are distinguishable for native speakers. While these variables are not often shown in literature as particularly relevant for aspectual use (but see Jurkiewicz-Rohrbacher, 2019 discussed in section 4.5), we included them because, as mentioned earlier, we strived to create rich usage profiles of the verbs.

The last group of the contextual variables covers elements that we looked for in the entire chunk we extracted, rather than 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 two grammar books for learners of Polish: Swan (2002) and Sadowska (2012) and can be found in Appendix 2.

*Abstract variables*

The abstract variables captures semantic dimensions presented in literature as possible candidates for the invariant meaning of aspect. Given the abundance of distinctions we decided to focus on the most prominent ones. Based on Janda's (2004) review, we have included the following variables: *boundedness* that describes whether any boundary (i.e. beginning or end) of the action is signalled; *totality* (e.g. Comrie 1976) which indicates if the stages and development of the event were important or whether it was 'presented as a whole' and treated as one indivisible thing; *sequentality/ simultaneity* (e.g. Bondarko 1971) that captures whether the action described happens along with other actions or whether they form a sequence of events; *perspective* that goes back to Comrie's (1976) remark on the imperfective showing an action from 'within' as opposed to the perfective that presents it from a different temporal perspective; *resultativeness* that describes whether the action resulted in a change of state (e.g. Laskowski); *foregrounding* (e.g. Chvany 1990) which evaluates whether the action is presented as a main event on the timeline or rather serves as a backdrop on which other actions move the plot forward. Additionally, we included *specificity* ((Divjak et al., 2015) that captures whether the actions are distinct, performed by identifiable individuals and can be located on the timeline.

## 5.6 Inter-annotator agreement

Although the contextual variables were a layer of generalization over the context, they are definitely less dependent on an annotator's interpretation than the variables that capture the semantic distinctions. This increase in the degree of subjectivity might affect our findings; for this reason we evaluated how reliably these labels could be applied by measuring inter-annotator agreement. A random sample of 100 chunks was sent to a second annotator, also a native speaker of Polish, who, after initial training on a different sample, annotated the data for all 7 abstract variables[6]. We present the agreement in Table 11, using two measures: first, the percentage of agreement, calculated simply as the percentage of cases where the labels applied by both annotators were the same. Since most variables, except sequentiality, were binary choices, a random agreement would be 50%. Secondly, we calculated a kappa score, using the

[6] I am very grateful to Marta Gąsiorowska who annotated the subsample as a second annotator

scikit-learn package for python 3 (Pedregosa et al., 2011). The kappa score is a more sensitive measure as it also takes into the account how likely the labels were to occur in general.

*Table 11. The inter-annotator agreement in the subsample of corpus data*

| **variable** | **percentage of agreement** | **kappa score** | **score description (Landis & Koch, 1977)** |
|---|---|---|---|
| boundedness | 53 | 0.1533546325878593 | slight |
| totality | 60 | 0.007772449148337968 | slight |
| perspective | 65 | 0.27924217462932455 | fair |
| specificity | 69 | 0.3992248062015503 | fair |
| sequentiality | 65 | 0.11359607254875215 | slight |
| resultativeness | 68 | 0.242603550295858 | fair |
| foregrounding | 55 | 0.12756882512601786 | slight |

Overall, the agreement is quite low, with the best scores in the low ranges of 'fair'. In this situation, the first suspect is usually the operationalization. However, there are other studies showing that the more abstract the variable, the more difficult it is to apply consistently. For instance, Divjak et al. (2015) showed that there is very little agreement on types of modality and Dziob et al. (2017) report that they abandoned annotating for more fine-grained verb classes because the low level of agreement. Given this evidence that semantic labels are usually difficult to apply, it is less surprising that the scores are so low.

The fact the definitions were difficult to construct is meaningful too. Abstract distinctions in the literature are rarely, if ever, defined in a concrete, easily applicable way – most often, they are metaphorical and used to describe a couple of constructed examples. It is worth mentioning that disagreement is not necessarily a bad thing. Different speakers 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, we often saw that the two annotators could easily come up with several interpretations on which both could agree. This may explain the abundance of literature on aspectual meaning. It is certainly possible to find many interpretations and come up with labels for them. The question is whether it is necessary.

## 5.7 Modelling

Having described the sample, we will now move on to presenting the learning simulations. As explained in the Chapter 3, we modelled the data using NDL – Naïve Discriminative Learner[7]. Each row in our annotated dataset will be treated as a learning event – an opportunity for the model to adjust the association weights based on the cues and outputs available for a given discourse chunk. We will discuss a total of six models, each based on a different subset of cues and outcomes. The three main models we will discuss first were trained to learn the use of aspect and lemmas based on either the concrete or the abstract variables. Comparing the performance of these models will help us answer the question of whether the usage patterns are reliable enough for speakers to learn aspectual usage, as well as investigate

[7] All models were trained using an internal version of a python package written by Dr Adnane Ez-zizi, a member of the Out Of Our Minds project. I am very grateful for providing this resource to me, as it greatly improved the speed and the ease of training models with NDL.

how reliable the semantic distinctions are. The three other models we will discuss were trained to predict a subset of abstract labels based on the values of concrete variables as cues. These models will allow us to test whether these abstract semantic labels can be learned on the basis to what is available to speakers in usage. The next sections present the models as well as the training procedures in more detail.

### 5.7.1 Training

*Learning aspectual distinctions and the use of lemmas*

In the following sections we refer to the models using the combination of the cues and outcomes used in training. For instance, the model where aspect was the outcome and the contextual variables served as cues is referred to as aspect-concrete model. Table 12 details what combinations of cues and outcomes were used for each model.

*Table 12. Cues and outcomes used for each of the models trained to predict aspectual labels or lemmas*

| **model** | **cues** | **no. of cues** | **outcomes** |
|---|---|---|---|
| lemma-concrete | values of the contextual variables for each chunk (e.g. AgentNumber.plural) | 150 | all 18 lemmas |
| aspect-concrete | values of the contextual variables for each chunk | 149 | perfective, imperfective |

| | | | |
|---|---|---|---|
| aspect-abstract | values of the abstract variables for each chunk (e.g. Boundedness.bounded)<br>[15 cues in total] | 15 | perfective, imperfective |

Each of the models presented above was trained in the following way. First, the corpus dataset was divided into training set, test set and validation set. The validation set – 10% of randomly selected chunks from the full dataset – was used to evaluate how the weights change and whether the model achieves stability. The test set contained 10 randomly selected chunks for each verb – a total of 180 test chunks, amounting to another 10% of the full dataset. The same test set was used to determine the accuracy of all the models.

Each model was then trained for 500 repeated runs, randomizing order each time, with a learning rate of 0.0001[8]. This means that 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 but in a different order, saving the weights it already acquired in the previous round. The procedure was repeated 500 times. This allowed to counter the order effect and to get a stable model.

After training, the models were probed against the test set described above. For aspect-concrete and aspect-abstract models, accuracy was calculated in the same way. First, we

[8] Given the limited number of cues and the fact that the outcome was binary in most of our models, the learning rate was set to a low value to ensure smaller fluctuations in association weights and prevent overfitting the models. A high learning rate could lead to a situation where a frequent initial occurrence of one predictive cue would block learning about other predictive cues, as explained in section 3.2. This decision had an influence on the final association weights, making their values small. However, it is important to note that we did not have any theoretical assumptions regarding the association weights, and as such, their exact values were not relevant to us. What is more, when testing and validating the models experimentally, we compared the activation of one outcome over the other, looking at the relative support of the cues for the outcomes, rather than the values of the activations themselves.

obtained the activations of outcomes by summing the weights of cues for each outcome (perfective and imperfective) that were present in each row. The outcome with the higher activation was taken as the model's preference. Then, we calculated the percentage of cases in which the model correctly preferred the aspect used in the 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.

*Learning abstract semantic labels from context*

In addition, we used the training method described above to evaluate whether it is possible to learn the abstract semantic labels proposed in the literature on the basis of contextual variables (cf. Divjak et al. 2015 for modality). We report here three models we ran for the following variables: perspective, totality and boundedness. We report here three models we ran for the following variables: perspective, totality and boundedness. These variables were selected for two reasons. First of all, they are used in the aspectual literature most often, which makes them theoretically most interesting. Secondly, looking at the distribution of these variables per aspect, we established that their individual levels aligned best with the aspectual use. That is, imperfective sentences were annotated as 'unbounded' in 90% of the cases and signalling lack of 'totality' in 97%. Similarly, perfective sentences were annotated as showing the 'exterior' perspective in 83% of the cases. Since the distributions of other variables per aspect were much more balanced, we decided to focus on boundedness, perspective and totality as potentially most reliable predictors out of all seven abstract labels. The training procedure and the test set were the same as described above. The details of these models are presented in table 13.

*Table 13. Cues and outcomes used for each of the models trained to predict the abstract semantic labels*

| **model** | **cues** | **outcomes** |
|---|---|---|
| perspective-concrete | values of the contextual variables for each chunk | interior, exterior |
| totality-concrete | same as above | yes, no |
| boundedness-concrete | same as above | bounded, unbounded |

## 5.8 Modelling results

### 5.8.1 Predicting aspect/ lemma

Table 14 presents the accuracies of the main models, trained to predict either lemmas or aspect. As we can see, all the models perform quite well. The aspect-abstract model (87.22% of accuracy) performs slightly better than the context-based model of aspect (85%). The model trained to predict lemmas was the least accurate of the three (79.44% of accuracy).

Interestingly, both context-based models seem to perform better predicting the perfective than imperfective. On the other hand, the model trained on semantic abstractions performs really well when predicting imperfective (almost 98% of accuracy), while it correctly predicts perfective only in 76.67% of cases. However, using the two sided Fisher's exact test, we determined that the difference in performance per aspect (perfective vs. imperfective) was significant only in the aspect-abstract model ($p<0.01$). The differences in the context-based models were insignificant ($p=0.14$ for the lemma-concrete model, $p=0.20$ for the aspect-concrete model).

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 study. We have seen that the main focus of the debate on the invariant meaning was on perfective aspect: all the proposed distinctions were said to capture the semantics of perfective, as a marked form. However, the uneven performance of the aspect-abstract models suggest that it is imperfective which meaning can be more reliably described using semantic labels. On the other hand, the concrete models perform slightly (although not significantly) better at predicting perfective aspect, which may suggest that patterns in language make this aspectual class easier to master on the basis of usage.

*Table 14. The performance of the main models*

| **model** | **overall accuracy** | **accuracy per aspect** |
|---|---|---|
| lemma-concrete | 79.44% | perfective: 84.44%<br>imperfective: 74.44% |
| aspect-concrete | 85% | perfective: 88.89%<br>imperfective: 81.11% |
| aspect-abstract | 87.22% | perfective: 76.67%<br>imperfective: 97.78% |

### 5.8.2 Predicting semantic labels

We will now discuss the models trained to predict the semantic variables from the contextual cues. As Table 15 shows, models trained to predict perspective and boundedness of the sentence achieved accuracy of around 80%. The totality model performed worse, correctly predicting slightly more than 76% of cases.

*Table 15. The performance of the models trained to predict abstract semantic distinctions from concrete variables*

| **model** | **accuracy** |
|---|---|
| perspective-concrete | 84.44% |
| totality-concrete | 76.11% |
| boundedness-concrete | 80% |

The results suggest that at least for these variables, there is enough information in the context to learn when each label should be applied. Therefore, the models show that these abstract distinctions can be constructed bottom-up.

However, we must remember how subjective and negotiable the interpretation of the sentence at this level of abstraction is. The low agreement between the annotators, discussed in section 2.6, means that the results could be different if the dataset was annotated by a different person.

## 5.9 Summary

The models we presented in this chapter support both the usage-based and the feature-based approach. The models perform with similar accuracy when predicting the label from the context and when predicting aspect from either context or the set of semantic descriptions of it. In addition, we have shown that the semantic labels themselves can be learned on the basis of the context.

As such, the corpus based learning simulations support both approaches to aspect, and – on its own – are inconclusive. However, any corpus-based study can only tell us about what patterns are or are not present in the data we have. Although the findings are crucial to hypothesis formation, they still need to be confirmed experimentally, since they cannot prove or disprove whether the speakers are sensitive to these patterns, learn them, and use them. Therefore, in the next chapter, we will present a behavioural study designed to further validate the models.

# Chapter 6. Experimental validation

Here we present a gap filling study designed on the basis of our corpus data and the models described in the previous chapter. We will test whether any of the corpus-based models we trained can be used to predict the linguistic behaviour of the speakers. As such, the study will serve as another source of information that will help evaluate the validity of the models.

## 6.1 Method

In order to be able to empirically verify which of the models predicts the actual behaviour of the speakers, we conducted a gap filling study. In the study, participants saw shortened paragraphs taken from our annotated corpus sample, and were asked to choose one of the aspectual variants of the verb provided. Their choices were then compared to the choices 'made' by each of the three models.

The participants were only allowed to choose one option. This design was chosen for two reasons. First of all, since the accuracy for the models was calculated by selecting only one of the options, we wanted to limit participants options in a similar way in order to make comparisons between the models and speakers more reliable. Secondly, we were interested in which of the verbs fits the context best. If both verbs are acceptable, this should be seen from the proportion of answers in each question – the more balanced the answers across respondents, 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, when speaking or writing we do not have the luxury of using both aspects at the same time. The sections below present the details of the study as well as the results.

## 6.2 Obtaining measures from the models

The output of an NDL model is a matrix containing association weights between all the cues and all the outcomes that the model was trained on. Using this matrix, we can calculate an activation of each outcome by summing up the association weights between this outcome and of all the cues that are present in a given chunk. Having done this operation for each outcome, we obtain a vector of activations – a numerical representation of how strongly the model predicts each outcome in a given chunk. The higher the value, the stronger the prediction.

This means, however, that the values of each activation vector depend on the values of the weights. Therefore, a comparison of two models on the same chunk using only the activations is not possible without taking into an account the distribution of values in an activation vector of each model. For example, if the activation of outcome X for model 1 is 0.2 and for model 2 it is 0.8, we could assume that model 2 is more 'certain' that outcome X should be used in the chunk. It is only by looking at the distribution in activations vectors that we realize that 0.2 in model 1 was also the highest value, and that model 1 also strongly predicts outcome X.

Therefore, since we wanted to validate and compare three models – aspect-concrete, aspect-abstract and lemma-concrete, we needed to scale the activations so that they were comparable. Instead of using activations from the models directly, we transformed them to odds ratios. The odds ratios were obtained in the following way: the activations were transformed into probabilities using a softmax function, which turns a vector of values into a vector of the same length that sums up to 1. As such, the resulting vector represents probabilities of all outcomes in a given chunk. Then, the probabilities were used to calculate odds by dividing the probability of an outcome by 1-probability. For instance, if in a chunk the probability of the perfective was 0.4, the odds of the perfective were: 0.4/1-0.4 = 0.66. For the imperfective in this case, the odds would be 0.6/1-0.4 = 1.5. In order to determine how much more likely the original verb was compared to its counterpart, we then calculated the odds ratios by dividing the odds of the original verb (let's say perfective: 0.66) by the odds of the counterpart (in our example, the imperfective 1.5): 0.66/1.5 = 0.44. Odds ratios below 1 indicate that the original aspect was less likely, according to the model; the odds ratios above 1 indicate the opposite. As a result, for each model and a chunk we obtained a comparable measure that tells us how likely the original verb was compared to its aspectual counterpart.

### 6.3 Selecting the stimuli

First, it was necessary to limit the number of lemmas. Otherwise, participants would have to answer too many questions, which would take too long and would most likely result in many incomplete surveys. Therefore we extracted a subset of 10 lemmas (5 pairs), essentially taking every second pair on the list we presented in Table 9. This allowed us to make sure that the pairs are still sampled along the perfective bias continuum, and control for potential biases in choices that could be attributed to the frequencies with which aspectual counterparts are used.

Secondly, some chunks were discarded. Because the study investigates aspectual choice, chunks where there cannot be any choice were excluded. The potential pool of stimuli was limited to those chunks in the annotated corpus that are not in the present tense, since only imperfective can be used in the present tense.

In addition, chunks where the target was either a perfective or present adverbial participle (e.g. *nabierając* or '*nabrawszy*) were also discarded since a counterpart in the same form but different aspect cannot be formed (i.e. *nabrając* is not possible).

As for adjectival participles, it was that established that both perfective and imperfective variants were only used in the passive voice with the auxiliary 'be' (not *zostać*). Therefore, chunks in the passive voice with the auxiliary *zostać* were discarded.

In sentences in the future tense where the imperfective was used originally, *być* (be.IMPF) served as a cue for the model and, because perfective is never used in such context, it is a very predictive cue. However, *być* cannot be left in the target sentence for participants, because it makes the choice obvious. Therefore we excluded and replaced such chunks. Future tense sentences where perfective was used originally could still be used and '*być* + imperfective' was introduced as an option to choose.

Finally, perfective verbs cannot be used with phasal verbs such as *zacząć* (begin.PERF) or *skończyć* (finish.PERF). The chunks that contained these auxiliaries were also excluded.

Chunks that contained other 'aspectual triggers' were not excluded because during annotation we observed that, despite their classification in grammar books, perfective triggers were used with imperfective verbs and vice versa.

In the last step of stimuli selection, one randomly chosen chunk per lemma was obtained for low (between 0.125 and 0.375 quantiles) and high (0.625 and 0.875 quantiles) values of each predictor. This ensured that the sample contains at least one case of each lemma where each model strongly predicts the original verb and at least one where it strongly predicts the counterpart.

As should be clear from the description of the stimuli inclusion criteria, we chose a limited set of possible uses, focusing on the contexts where the use of both aspects was possible. It is worth noting here, however, that both the models and the speakers ‘knew’ more about the aspectual usage – the models were trained on a dataset containing additional contexts where only one aspectual option was possible, and the linguistic experience of native speakers necessarily includes such cases as well. However, we decided to exclude such cases as they offer little insight into which model predicts the behaviour of the speakers better, simply because there is no choice. On the other hand, by selecting the stimuli in which both perfective and imperfective could be used, we were able to establish how strongly each option was preferred by the speakers and determine for which model we observe an expected trend, that is the stronger agreement between the participants for the stronger support of the model.

### 6.3.1 Resampling

The sample was examined to check to make sure all the chunks can be used in the experiment. We decided to replace 3 chunks. In the first one, the text before the target sentence seemed too incoherent. The second chunk contained a direct object in the genitive after the verb *odszukiwać* (find.IMPF), which could be treated by some participants as an error. These chunks were replaced to avoid confusion. The third chunk we removed contained swear words, which might seem too strong for some participants and therefore not allowed according to the ethical guidelines.

The first two chunks were the only ones that belong to the 'high' subset of aspect-concrete odds ratios for their respective verbs, so the new chunks were sampled from these subsets. The third chunk was found in the 'low' subset for both lemma odds ratios and aspect-concrete odds ratios. We made sure the new chunk also met these conditions.

Since the decision to exclude sentences where future compound tense was used originally was made after the sample has been extracted, we resampled 3 further chunks.

The first check we performed in order to determine how similar the models as predictors are, was to calculate the Spearman's correlation coefficient between them. As we can see in

Table 16, the correlation was low, suggesting that the effects of each of the model would differ. Interestingly, the aspect-abstract model was negatively correlated with the lemma-concrete model and even more so with the aspect-concrete model. This means that in the chunks for which the concrete models strongly predicted the use of the original aspect, the aspect-abstract model often predicted the opposite. We might then expect different trends in regression models fitted on behavioural data, and that it is unlikely that both concrete and abstract models will be able to explain the answers provided by the participants.

*Table 16. The correlation between the predictors in the final corpus sample*

| | oddsRatiosLemma | oddsRatiosAspectCon | oddsRatiosAspectAbs |
|---|---|---|---|
| oddsRatiosLemma | 1.000000 | 0.398222 | -0.060065 |
| oddsRatiosAspectCon | 0.398222 | 1.000000 | -0.140897 |
| oddsRatiosAspectAbs | -0.060065 | -0.140897 | 1.000000 |

## 6.4 Preparing the stimuli for the survey

Because we used a corpus containing texts from the internet, the chunks were manually prepared to be ensure that they are in a format that could be read easily. The changes included adding capital letters at the beginning of the sentences and in proper nouns, deleting extra spaces around punctuation marks, and introducing extra blank lines for clarity of reading. We also corrected the typos and added Polish diacritics where necessary – this made texts easier to read but did not influence any of the measures. Finally, the target verbs were replaced by a dotted line of the same length for each sentence.

However, six sentences of context, quite often rather lengthy, was too much to include in the survey. The estimated time of completion was almost 40 minutes, which meant that the dropout rate would be very high, and could result in participants ignoring the context altogether, and focusing only on the target sentence. Therefore, we decided to limit the number of sentences in the chunk. We made sure that the context contains enough information to be able to determine the values of both contextual and abstract variables. For instance, in question 9, we kept one previous sentence, since it was impossible to determine whether the situation described is generic or specific from the target sentence alone.

**Question 9**

*Polish:*

context: Dzięki korzystaniu z telefonii voip za połączenia wychodzące wykonywane przez twoich konsultantów zapłacisz grosze.

target: Wszystkie rozmowy ………………… będą według najtańszych stawek freeconet.

a) rozliczone

b) rozliczane

*English:*

context: Thanks to using voip, you will pay pennies for the calls made by your consultants.

target: All the calls will be ………………… for based on the lowest freeconet rates.

a) accounted.PERF

b) accounted.IMPF

For each stimuli in our sample, we manually limited the window of context. Then, we counted the context sentences that were left in and determined that on average 1-2 sentences of context (mean=1.78, mode=1) provided enough information to annotate the sentences for our set of variables.

## 6.5 Data collection and participants

The study was conducted using Qualtrics software (Qualtrics, 2005). Participants were recruited via social media. All participants answered the full set of 60 gap filling questions. The order of questions and the order of options to choose from were randomized. We received 77 responses. 57 participants were women, two decided not to reveal their gender. 49 had higher education, 26 graduated high school and 2 had vocational education. The age of participants ranged from 18 to 64 years; the mean was 33.61.

## 6.6 Calculating the agreement between respondents

The majority of participants agreed that the original verb fits the gap better in 85% of questions. The proportion of agreement was calculated for all the questions and rounded to two decimal points. For instance, if in Q1 70 out of 77 participants chose answer a, the proportion of agreement for that answer, expressed in percentages equals (70/77)*100 = 90.90%.

The answers were sorted into proportion bins, presented in Table 17 below. The first bin groups all the questions in which one of the answers was chosen by 100-91% of all the participants. The second bin groups all the questions in which one of the answers was chosen by 90-81% of all the participants, etc. The last bin groups all the questions in which both answers were selected equal number of times. In the table, the second column specifies the number of questions in a given bin, and the third column expresses that number in percentages of all questions.

*Table 17. Participants' agreement per proportion of questions*

| | **bin size** | **proportion of all questions** |
|---|---|---|
| **91-100** | 25 | 41.67 |
| **71-80** | 12 | 20.00 |
| **81-90** | 10 | 16.67 |
| **51-60** | 7 | 11.67 |
| **61-70** | 6 | 10.00 |
| **50** | 0 | 0.00 |

To better illustrate the distribution of the agreement between the participants, the data from Table 17 were used to prepare a bar chart in Figure 2.

As we can in Figure 2, a definite majority of participants agreed very strongly (90-100%) in 41.67% of all questions and only in 11.67% of cases both aspectual options were chosen with similar frequency. In other cases, one option was preferred but the strength of agreement varied. The results described above resemble the results obtained from Russian participants and confirm that context limits the choice. However, as in Janda and Reynold's study (2019), this limitation is not categorical. In some contexts, one answer is more likely; in other the likelihoods are more balanced. It seems then, that usage indeed forms a continuum of probabilistic fixedness.

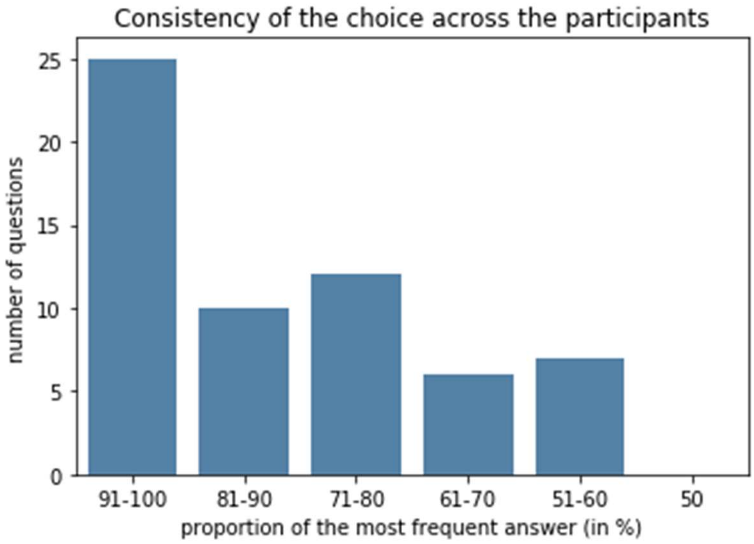


*Figure 2 The consistency of choice across participants*

## 6.7 Individual variance

Interestingly, we also found variance in how likely individual participants were to select the most frequent answer. As Figure 3 illustrates, most of the people only sometimes diverged from the 'correct' answer, (i.e. the one selected by the majority of participants). However, there were also people who disagreed more often, with two participants choosing the other aspectual variant in more than 50% of cases. This shows that it is not only 'when?' that matters for aspectual choice, but 'who?' is an equally important question. This finding fits well into the theoretical framework presented in the first chapters. After all, only individuals can learn.

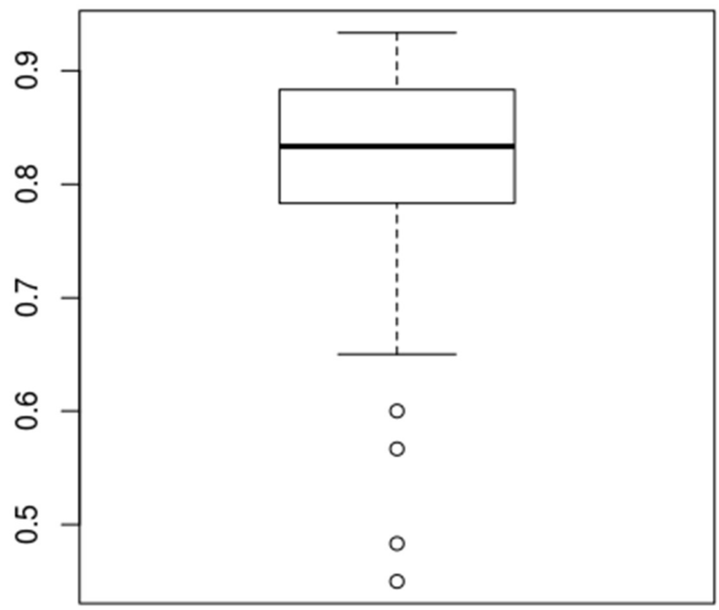


*Figure 3 The distribution of participants' mean score.*

## 6.8 Describing the tools: mixed effects models

Throughout the dissertation, we will use mixed effects regressions to model the experimental data. Therefore, we will explain the principles of these statistical tools as well as explain the reasons for their use in our analysis.

### 6.8.1 Linear regression and mixed effects linear regression

Let us start with the  basic linear model, where we only model fixed effects. Put simply, what these models allow us to do is to model the influence of a predictor on an outcome. Let's say that we are interested in how much work experience influences income. The dataset we would need, should contain the values of years of work experience and the annual income of each participant in our study. Then, we would fit a linear regression model using a following (simplified) formula:

$$y_i = a + bx_i + e_i$$

where y is the value of outcome for each participant, a is the value of the intercept, b is the coefficient of a slope, $x_i$ is the value of the predictor for each participant, and $e_i$ is the residual.

The intercept here is the value of y when x is equal to 0. For example, if the intercept is equal to 1, then the regression line would cross the y axis at point (y=1, x=0). The coefficient of the slope tells us by what value we need to multiply the value of the predictor we have available in our dataset. For example, if the coefficient is 3, we would multiply the years of experience by 3. Finally, the residual is the distance between the actual data point and the value predicted by the model.

The regression algorithm tries to estimate the parameters of the best fitting line – one that lies closest to all data points. This is done by comparing the squared sum of residuals of possible combinations of intercepts and slopes. What it means in practice is that for each data point and each fit, we calculate the residual – the distance between that point and the regression line. Since some points may lie under the line, which makes the residual value negative, we then square all residuals to make their values positive. Finally, we add all the squared residual values together, which gives us the sum of squared residuals. The lower the sum, the better our model fits the data.

Essentially then, regression estimates the values of a – the intercept, and the coefficient – b, using the data from all of our participants. This way, we can learn how much the value of y changes if we change the value of x by one unit. In our example, we estimate how much does the annual income changes for each year of experience. We can then substitute these numerical values in their respective places in the formula to get the value of the outcome given any value of our predictor. Therefore the linear model allows us to generalize beyond the sample of data we have collected and make predictions for any individual, provided that we have data on their years of experience.

### 6.8.2 Logistic regression

Linear regression can only be used when our outcome variable is continuous – that is, we are dealing with numerical values that (theoretically) can extend from negative to positive infinity. However, we must use other modelling techniques when our outcome variable is binary and categorical. One example of such outcome variable could be participants' accuracy of judgement for a given question. Whereas mean accuracy can be expressed as a categorical variable that ranges from 0 (no correct answers) to 100 (all correct answers), the response for each individual question can only be coded as correct and incorrect. As we can see, such an outcome variable can only take two values (and is therefore binary) and these values are categorical rather than numeric. Since computers only understand numbers, we translate these categories as 0 and 1. It is important to remember however that these numbers only encode whether a given example belongs to our reference category (e.g. correct) or not.

In cases where we will use a binary categorical variable because our experimental measure is binary and categorical, we will use a related model – logistic regression. Logistic regression is a generalization of the linear model and is based on similar principles – that is, it also tries to fit a regression line to the data. However, there are important differences between the two, which we will now discuss.

First of all, when using logistic regression we no longer model the change in the outcome y with the change of the value of our predictor – x. Instead, we are modelling the probability of y belonging to a class, given the value of the predictor. To give an example, let's assume that we are using word frequency as a predictor and we are interested in modelling the speed of reading of a given word and whether it was correctly recognized as a word (as opposed to a pseudo-word). For the first problem, we would use linear regression to predict what the reaction time will be given the frequency of a word. For the second, we would use logistic regression to predict the probability of a word being recognized correctly, given the frequency of the word.

The second important difference between linear and logistic regression is that we do not use the Sum of Squared Residuals as a measure of the goodness of the fit. Instead, we use Maximum Likelihood. While this method is less straightforward than the least squared estimation, we can say that in principle, for each fit of the line that we try, we calculate the probability that each datapoint in our dataset belongs to the class encoded as 1. Then, we use the observed status (0, 1) to calculate the likelihood of each status for each datapoint. If the datapoint belonged to the class encoded as 1, its likelihood is simply the probability. If not, then we calculate the likelihood as 1- the probability. Finally, we log transform each likelihood value and sum them together. The resulting value is the likelihood of the fit, and we choose the line with the highest value.

### 6.8.3 Mixed effects models

One of the main assumptions of regression models is that the observations are independent from each other. That is to say, there is no hidden relations or structure between the datapoints that might have an influence on the results. However, as pointed out by Winter (2019) nearly every psycholinguistic study, including the ones presented in this dissertation, are in fact hierarchically structured. This is because the observations are produced by the same participants or, and often – and, measure behavioural response on the same experimental items. Not including such dependencies in our models violates the independence assumption, which may give false positive results (Winter, 2019). Therefore, in order to properly model hierarchically nested data, in addition to fitting fixed effects – i.e. the predictor of interest, we must also fit random effects – i.e. sources of groupings in the data.

The mixed effects can differ in terms of what we allow to vary within the groups. By introducing a random intercept, we allow each random effect to have its own intercept that we add to the fixed intercept in the formula to find the value of the data point. For example, if data is grouped by participant – i.e. each participant answered 10 questions in our study, we should fit a random intercept for each participant. What it means in practice is that we can capture the differences in average responses between the participants. For example, if we measured reaction time, fitting the random intercepts would allow different values of y for y=0 for each participant. So the value of y in point x would be different for each participant, and the fixed effect would capture the influence of the predictor that is the same for all the participants. Therefore, the formula we introduced above would be modified in the following way:

$$y_{ij} = (a + a_j) + x_i + e_{ij}$$

Random slopes allow us to capture the differences in the effect of the predictor. For example, fitting a random slope for each participant would mean that each participant would have their own estimate of the coefficient, which when added to the estimate of the fixed effect, shows how much y changes for each value x of a given participant. This also modifies the formula:

$$y_{ij} = a + (b + b_j)x_i + e_{ij}$$

Finally, fitting both a random intercept and a mixed effect means that we add both the value of each individual intercept to the fixed intercept as well as add the random effect coefficient to the fixed effect coefficient. Therefore, the final formula is:

$$y_{ij} = (a + a_j) + (b + b_j)x_i + e_{ij}$$

As we said before, logistic regression is a generalization of the linear model. Therefore, adding mixed effects to logistic regression works in a similar way to linear regression. In addition to estimating the parameters of fixed effects, we also estimate the intercept and the slope of each random effect.

## 6.9 Model evaluation

In this section we compare how well each of our three NDL-based models predicts the choices the participants made. Using the data from the experiment, we created a new variable – originalChosen, which is a binary categorical variable encoding whether or not participants selected the verb form used originally in the chunk. As for the predictor, in each of the models, we used the values of odds ratios obtained from a given NDL-based model. These odds ratios represent the support of the model for the verb lemma or aspect used originally in the chunk. Therefore, each regression models the relationship between the support from a given NDL-based model and the linguistic behaviour of the speakers. We expected that for an NDL-model that approximates the knowledge speakers have, including the level of representation (aspect or lemma) and the set of cues on which the choices are based, the relationship should be positive. In other words, the stronger the support of the model for the option used originally in the chunk, the more likely should the participants be to choose that particular option as well.

Since we had multiple responses from the same participants as well as multiple responses per individual question, we decided to fit mixed effects logistic regression models. Each of the models was fitted using the glmer function in the lme4 package in RStudio, using R version 3.6.3. Below we present the regression models for each of the NDL-based models and discuss them in more detail.

*Lemma-concrete model*

First, we present the regression model fitted for the lemma-concrete NDL-based model. Here, the odds ratios represent the support of the lemma-concrete model for the lemma of the verb used originally in the chunks. As mentioned before, the dependent variable encodes whether or not the participants chose the same verb.

The formula for the regression model was as follows:

originalChosen ~ oddsRatiosLemma + (1 | question) + (1 | participant)

Table 18 presents the model output, obtained from lme4 summary method. We focus on the fixed effects and include only the parameters for the intercept and the predictor variable.

*Table 18. Regression model output for lemma-concrete NDL model*

| Fixed effects: | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | Z value | Pr(>\|z\|) |
| (Intercept) | 0.9260 | 1.5535 | 0.596 | 0.551 |
| oddsRatiosLemma | 0.6302 | 1.4200 | 0.444 | 0.657 |

As we can see, the estimate for the predictor was positive (0.6302), which means that the overall trend resembled the pattern we expected to see – the stronger the support of the lemma-concrete model, the more likely the participants were to choose the original verb. However, the p-value of 0.657 indicates that the effect was not significant. Therefore, we cannot conclude that the model was validated, despite showing the expected trend.

*Aspect-concrete model*

Next, we present the regression model fitted for the aspect-concrete NDL-based model. In this case, the odds ratios represent the NDL support for the aspect of the verb originally used in the chunk.

The formula used to fit the regression model is presented below:

originalChosen ~ oddsRatiosAspectCon + (1 | question) + (1 | participant)

Table 19 shows the model output. Here we again present only the fixed effects.

*Table 19. Regression model output for aspect-concrete NDL model*

| Fixed effects: | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | Z value | Pr(>\|z\|) |
| (Intercept) | 0.05691 | 0.52048 | 0.109 | 0.91294 |
| oddsRatiosAspectCon | 1.12298 | 0.34464 | 3.258 | 0.00112 |

The estimate for the independent variable oddsRatiosAspectCon was positive, again suggesting that the stronger the support of the NDL-based model for the aspect of the verb used originally, the more likely the participants were to choose the verb. What is more, the p-value of 0.00112 indicates that this effect was significant. From this we can conclude that the aspect-concrete can be successfully used to at least partially explain the linguistic behaviour of the speakers in this particular experiment.

*Aspect-abstract model*

Finally, we present the regression model fitted with the odds ratios obtained from the aspect-abstract model as the predictor. Just like in the aspect-concrete model discussed above, the odds ratios represent the support for the aspect of the verb originally used in the chunk. However, the set of cues the aspect-abstract model was trained on differed. Instead of using the concrete variables, the aspect-abstract model had only access to the semantic labels.

The formula used to fit the model is presented below:

originalChosen ~ oddsRatiosAspectAbs + (1 | question) + (1 | participant)

Again, the table below (Table 20) presents the model output.

*Table 20. Regression model output for aspect-abstract NDL model*

| Fixed effects: | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | Z value | Pr(>\|z\|) |
| (Intercept) | 3.8680 | 0.7307 | 5.294 | 1.2e-07 |
| oddsRatiosAspectAbs | -1.0370 | 0.3174 | -3.267 | 0.00109 |

As we can see, the effect of the activation based on the abstract labels reached significance (p= 0.00109) but had a negative coefficient (-1.0370), which means that the opposite of what we would expect is true – the higher the odds of the original aspect to be used, according to the aspect-abstract model, the less likely the participants were to actually choose it.

These results are quite surprising, given the good performance of the aspect-abstract model on the test set we reported in the previous chapter. However, it is important to note that the performance was evaluated on the basis of accuracy – we only tested whether the most strongly supported option was the same as the aspect of the verb used originally in the chunk. Here, we investigate the problem in a different way, trying to estimate the relationship between the relative strength of activation and the agreement between the participants. Therefore, the conclusion we can draw is that while the abstract labels could perhaps be used to describe the aspectual usage, their actual 'strength' of support does not increase the confidence in choices. This could be taken as an indication, or a further confirmation, of how difficult the abstract labels are to apply in practice.

Another possible explanation of these results, and an interesting avenue for further research, is that the interpretation of the chunks we selected for the study necessarily required the target verb, without which it is difficult, if not impossible, to determine the semantic dimensions captured by the abstract variables we investigated. Since the task was to choose between the imperfective and perfective target verb, the speakers also had to choose which interpretation they preferred, rather than make sure that the aspect fitted the interpretation supported by the context. This suggests that in the cases where both aspects are possible, the abstract semantic distinctions are less contextually salient and therefore less reliable.

Taken together the regression models show that only the aspect-concrete was validated on behavioural data. Therefore, we are now going to explore it in more detail, focusing specifically on the most predictive cues it distinguished as well as the theoretical insights these learning outcomes provide.

# Chapter 7. Further exploration of the aspect-concrete model

## 7.1 Learning outcomes: strong positive and negative cues

As discussed in the theoretical chapters, the goal of error-driven learning is not just to determine which cues frequently cooccur with the outcomes of interest, but to discover which cues are reliably associated when all the cooccurrences are taken into the account. As we can see in the figures 4 and 5 below, only few cues 'survive' the fierce competition of the learning process. The distribution of association weights shows that for the majority of the cues, the weights are close to zero, with a few exceptions that have either strong positive or strong negative association with each outcome. It is worth underlining here that cues with strong negative weight are also important, as they indicate that the outcome we are considering should not be selected.

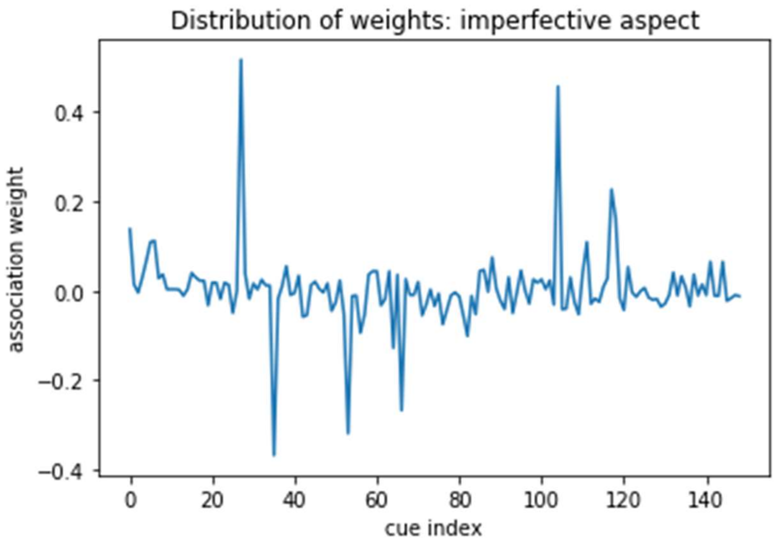


*Figure 4 The distribution of association weights for imperfective aspect*

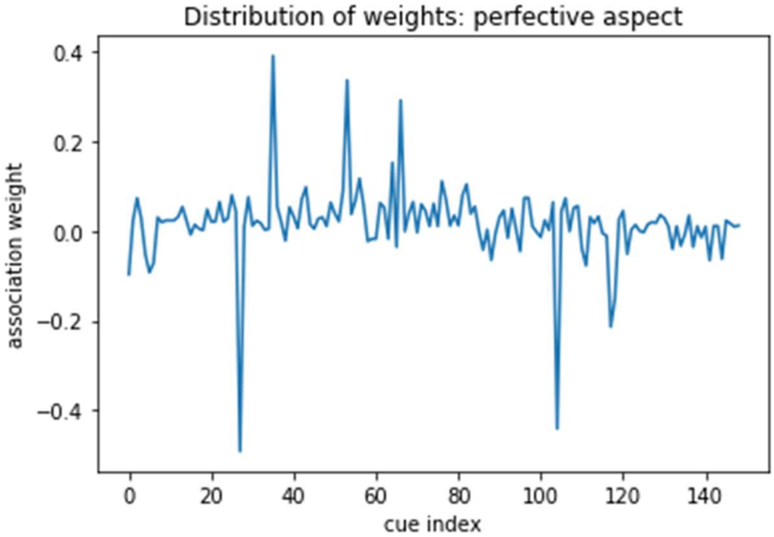


*Figure 5 The distribution of association weights for perfective aspect*

Table 21 below presents 5 strongest negative and positive cues for each of the aspects.

*Table 21. The strongest negative and positive cues per aspect*

| **strongest positive cues** | | | | **strongest negative cues** | | | |
|---|---|---|---|---|---|---|---|
| **perfective** | | **imperfective** | | **perfective** | | **imperfective** | |
| *cue* | *weight* | *cue* | *weight* | *cue* | *weight* | *cue* | *weight* |
| tense: future | 0.392070 | tense: present | 0.516445 | tense: present | -0.492037 | tense: future | -0.367925 |
| cxWord: zostać | 0.337165 | cxWord: być | 0.456216 | cxWord: być | -0.441455 | cxWord: zostać | -0.318909 |
| tense: past | 0.292407 | cx: phasal | 0.226170 | cx: phasal | -0.213583 | tense: past | -0.267503 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| cx: aux | 0.152814 | cxWord: zacząć | 0.161430 | cxWord: zacząć | -0.151636 | cx: aux | -0.127863 |
| cx: modal | 0.117708 | tense: none | 0.137108 | tense: none | -0.096552 | cxWord: udać się | -0.101403 |

Some of the most predictive cues we obtained match the ones already discussed in literature and grammar books. For instance, only imperfective can be used in present tense or with phasal verbs. Therefore, our model passes the "sanity check".

Other important cues for both aspects are the elements of the verb phrase. For imperfective, those are phasal verbs, especially *zacząć*, as well as the verb 'be'. For perfective the presence of modal and other auxiliary verbs in general seems particularly important, especially *zostać* which introduces passive voice[9].

Both future and past tense are reliable predictors of perfective aspect and strongly inhibit the choice of imperfective. Present tense is strongly positively associated with imperfective and, not surprisingly, negatively associated with perfective. It seems then, that according to the model, we can talk about a "default" choice of aspect in each of the tenses: perfective for past and future, imperfective for present. In the next sections we will explore this outcome further. We will focus on how the results we obtained so far link to the cognitive theory of language. In

[9] Note that, as shown in Appendix 1, there were two cues based on auxiliary verbs. We annotated the category of the verb (e.g. modal, phasal and other auxiliaries) as well as which particular auxiliary verb appeared with the main verb. This allowed us to model the influence of the category of the auxiliary in addition to the influence of a specific auxiliary lemma. Given the size of the corpus sample, it was likely that, for instance, individual modal verbs would not occur often enough to emerge as predictive cues themselves. However, including the more general variable of a modal auxiliary verb aggregated their co-occurrences with the target, allowing to model the influence of the presence of a modal verb as a category. On the other hand, the fact that we also annotated for the specific auxiliary lemma, allowed us to correctly distinguish 'be' as a predictive cue for the imperfective, which might not be possible if we only kept a general 'auxiliary' category.

addition, we will present the questions these conclusion lead to, which we will then try to answer in the next chapters.

## 7.2 Canonical vs non-canonical viewing arrangements

The results of the learning simulations we conducted so far, supported by the results of the experimental study described in Chapter 6 allow us to formulate the following conclusion: there are enough distributional biases in the use of the verbs that an error-driven model can learn the main usage patterns of aspectual categories. The analysis of the learning outcomes indicates that the set of the most predictive cues that the model established in training can be divided into two types. On the one hand, we have items such as auxiliary verbs that were strongly and positively associated with one of aspectual categories. However, these cues, even though very predictive, are not very reliable. That is, although there are clear patterns in usage, they are not necessary elements of a correctly formulated sentence and as such, they do not appear very often. Similarly to what we said about the adverbial cues in Chapter 4, these cues are important enough to guide the linguistic behaviour of the speakers, but they cannot do so always. On the other hand, all verbs can always be classified into one of the tense categories we have annotated for. Therefore, tense is not only a very predictive cue – it is also a very reliable one.

This finding is particularly interesting if we take into consideration the fact that when discussed in the literature, the relationship between tense and aspect is often ignored. Instead, the common practice is to present both perfective and imperfective as equally plausible options in the past and future tense. This however seems not to be the case. As illustrated by Table 22, there are significant differences in frequency of use of each aspectual category in different tenses. Imperfective is rarely used when talking about future. Out of 1800 examples, only 7 paragraphs contained an imperfective verb in future tense. The use of imperfective in the past is also limited in our sample and it amounts to 8.3% of all sentences in which imperfective was used. Those findings agree with the results of other corpus studies that looked into the distribution of tense-aspect combinations, which we discussed earlier. Therefore, we can draw the conclusion that these distributional patterns are not limited to our sample only, but resemble the actual patterns of use of the speakers.

*Table 22. The frequency of tense-aspect co-occurrences*

| aspect | tense | count | proportion per aspect |
|---|---|---|---|
| imperfective | present | 519 | 57.66% |
| | past | 75 | 8.33% |
| | future | 7 | 0.77% |
| | none | 299 | 33.22% |
| perfective | past | 337 | 37.44% |
| | future | 115 | 12.77% |
| | none | 448 | 49.77% |

Such distributional biases, and the influence they were demonstrated to have in the learning simulations, suggest that instead of viewing perfective and imperfective aspects as equally plausible options in all tenses, we should rather speak of default and non-default tense-aspect combinations. This conclusion fits well into cognitive linguistic theory. Langacker (2001) discusses how frequency of concurrence results in different types of viewing arrangements. The constructions used often are said to represent a canonical viewing arrangement and reflect how the situation is usually viewed. On the other hand, non-canonical arrangements are deviations from the well-established patterns, and while not incorrect, they offer a different perspective. Adopting this terminology, we can say that the use of perfective is canonical in the past and future tense.

This conclusion, however, leads us to further questions. First of all, given the simplicity of NDL itself as well as a rather coarse-grained model we obtained using it, we can as if there are any other patterns or cues that NDL missed. This will be done in Chapter 8, where we conduct a qualitative analysis of the questions where NDL incorrectly predicted participants' choices.

Secondly, even though the results suggest that perfective and imperfective are important dimensions that are distinguished by the speakers, the question of how these categories are formed based on the input still remains. After all, the aspectual classes are not given but rather they must also be learned. We will explore this avenue in Chapter 9, where we present clustering solutions based on the NDL output matrix to see if the dimensions the model learned to be important for perfective and imperfective can be used to group verbs into aspectual classes.

Finally, we should ask whether the canonical and non-canonical tense-aspect combinations are in fact perceived differently by the speakers. Deviation from expected patterns should have influence on language processing, which we should be able to observe in an experiment. Chapter 11 presents another experimental study designed to test if that prediction holds.

# Chapter 8. Qualitative analysis of errors

In this chapter we will investigate which experimental stimuli proved to be particularly difficult for the aspect-concrete model and try to identify the reasons why. There were two important arguments for conducting an error analysis on the behavioural rather than the corpus data. First of all, the corpus data is not always correct. As we have demonstrated, some choices that the authors of the discourse chunk made go against what the majority of the participants did. Using a subset with many ratings ensures that the choice would be accepted by many speakers. Secondly, the percentage of agreement itself is informative. Errors in cases with strong agreement mean that the model missed something very important that influenced the choice of the majority of the speakers. If we looked at the corpus data set, we would not know which of the errors are 'worse'.

We start the analysis by the tabulating error distribution per aspect, lemma and tense. These tabulations will inform the qualitative part of the error analysis, which we present in section 10.2. The error analysis was a crucial step in identifying cues that the model missed and revealed that in order to increase the explanatory power of our model of aspectual choice we need to extend it and include information from different levels of knowledge of language.

## 8.1 Error distribution

To begin, we will check how many errors there were per aspect. Since we know that in some cases the majority of participants chose a different aspect than used originally, we are going to consider the majority's choice as the preferred use. Therefore, if the verb used in the original chunk was imperfective, but most of the participants chose perfective, we will count this as the model's failure to predict the perfective. As we can see in Table 23, the model performed worse in predicting imperfective aspect. This is consistent with what we have seen in Chapter 5, where we discussed the accuracy of the aspect-concrete model on the test set. However, the difference in performance we are observing here is larger.

*Table 23. The distribution of errors per aspect*

| **aspect** | **number of errors** | **proportion of errors** |
|---|---|---|
| imperfective | 14 | 46.67% |
| perfective | 6 | 20% |

It is worth noting that not all the lemmas appear in Table 24 below, that sums up the error distribution per lemma. One aspectual pair, *rozliczyć – rozliczać* (square/appraise), was particularly easy, and the model achieved an accuracy of 100%. However, the model incorrectly predicted almost all the cases where imperfective *wywracać* (topple/fall.IMPF) was used. It also performed quite poorly, on a chance level, in questions where three other imperfective verbs were used: *odszukiwać* (find.IMPF), *nabierać* (gain/kid.IMPF) and *podmieniać* (switch.IMPF). The model also made an error in two chunks where the majority of the participants chose *odszukać* (find.PERF). In other cases, the model incorrectly predicted people's choice only once.

*Table 24. The distribution of errors per lemma*

| **lemma** | **number of errors** |
|---|---|
| wywracać (topple/fall.IMPF) | 5 |
| odszukiwać (find.IMPF) | 3 |
| nabierać (gain/kid.IMPF) | 3 |
| podmieniać (switch.IMPF) | 3 |
| odszukać (find.PERF) | 2 |
| rozliczyć (square/appraise.PERF) | 1 |
| podmienić (switch.PERF) | 1 |
| nabrać (gain/kid.PERF) | 1 |
| wywrócić (topple/fall.PERF) | 1 |

One potential reason why imperfective was more difficult to predict could be that the model strongly associated past tense with perfective, therefore it was unlikely to correctly predict the use of imperfective in this tense. However, as Table 25 shows, it was actually the cases where tense was annotated as “none” that were much more difficult to predict. This suggests that a closer attention should be paid to the chunks in a “none” tense. Particularly, in addition to reflecting on what cues were missed, we need to consider which of the cues that were present made the model choose an incorrect option.

*Table 25. The distribution of errors per tense*

| tense | number of errors |
|---|---|
| none | 16 |
| past | 4 |

## 8.2 The qualitative analysis of errors

Having discussed the distribution of errors, we will now move on to analysing the individual discourse chunks qualitatively. To better understand the failures of the model we will use the full chunks extracted from the corpus. Each question is presented in the following way: the table contains the original context on the left and its English translation on the right. Below, we present the target sentence along with its gloss and translation. Important cues are marked in bold. The table below the question presents the verbs the participants could choose from, the one that was selected by the majority, the percentage of agreement, and whether the verb chosen by the majority was also the one that was used in the original text. The remaining of this section is divided into subsections, based on the percentage of agreement. We first present the chunks for which the agreement among participants was the highest, and continue through lower bands of agreement. This allows us to better interpret the significance of the model's errors. Arguably, the errors in chunks with a low percentage of agreement, let's say, those where only 60% of participants chose one verb (and 40% the other) are less 'grave', because participants as a group were less certain which answer fits the context best. An incorrect prediction of the model in such case can be said to reflect this uncertainty. However, in cases where the vast majority of participants chose the same answer, the incorrect prediction suggests that the model missed some important cue(s) that swayed the participants toward selecting one of the options. The goal of this analysis is to identify such cues and other potential reasons for disagreement between the participants and the model.

*Errors in questions with high agreement: 100-90%*

Question 33 contained an adverbial trigger that requires imperfective. In addition, there is already another imperfective verb in the sentence.

**Question 33**

| | |
|---|---|
| *'Gdy przyszłam do drużyny miałam 14 lat. Nie miałam określonego wyboru na jakiej pozycji chciałabym grać. Ale któregoś dnia wpadłam na pomysł, że chce spróbować na bramce – wydawało mi się, że to najciekawsza, a zarazem ciężka pozycja w hokeju. Bramkarz to jedna z najważniejszych osób grających na lodzie – to od niego zależy ile wpadnie bramek, a także jak poprowadzi swoją drużynę w grze. Pamiętam dokładnie moje początki: sprzęt ważący ok. 20 kg wydawał się trzy razy cięższy niż naprawdę. Jazda w parkanach (to duże, sztywne ochraniacze na nogi) była zupełnie inna niż* | *When I joined the team I was 14 years old. I didn't know what position I want to play on. But one day I had an idea that I want to try playing as a goalkeeper – I thought it was the most interesting and the most challenging position in hockey. A goalkeeper is one of the most important people playing on ice – it's up to him how many goals are scored, and how he leads the team in a game. I clearly remember my beginnings: the equipment, weighing about 20kg, seemed three times heavier than it really was. Skating in pads (big, rigid guards for legs) was much different that skating without them, and my* |

| | |
|---|---|
| *bez, a ręka bolała od ciągłego trzymania większego, bramkarskiego kija.* | *hand hurt from holding a bigger goalkeeper's stick.* |

Teraz do tego wszystkiego przywykłam, ale wtedy **co chwilę** się potykałam i ………………… .'

Now to this all get.used.1SG.PAST.PERF but then every moment I.REFL stumble.1.SG.PAST.IMPF and …………………

Now I got used to all of it but then I stumbled and ………………… all the time.

| | |
|---|---|
| the answer selected by the majority of participants | wywracać (topple/ fall.IMPF) |
| original chosen | yes |
| agreement | 98.7% |
| aspectual competitor | wywrócić wywracać (topple/ fall.PERF) |
| original lemma | wywracać wywracać (topple/ fall.IMPF) |

In question 21, the model incorrectly selected an imperfective, while a majority of participants chose a perfective. Unlike in other cases we have discussed so far, there are no clear cues in terms of other verbs in the sentence, or adverbials. More interestingly, both verbs could be used in this context, although the intuition of the author (confirmed by the choices the participants made) suggests that imperfective simply sounds odd. One reason for that impression (and the choice that follows) might be a huge frequency bias. In the corpus, 'odszukać' appears 44391 times, while the imperfective counterpart appears only 580 times.

**Question 21**

| | |
|---|---|
| *Kolejną sprawą są mechanizmy związane z hydrauliką, pneumatyką. Sprężarka ustawiona gdzieś w przydomowym magazynie pozwala na wykonywanie wielu prac związanych np. z dobudowywaniem różnych elementów wnętrza, a nawet stworzenie własnego, dobrze funkcjonującego miejsca pracy typowo technicznej. Im lepsze jakościowo sprężarki, tym większa ich efektywność oraz dłuższe zastosowanie. Ich mechanizm działa nawet w zwykłej suszarce, ale tak się składa, że razem z agregatem prądotwórczym i stworzeniem sobie np. automatycznego obiegu wody oraz jej filtracji w mieszkaniu jesteś w stanie żyć praktycznie niezależnie od podstawowych mediów. Wiadomo, że koszty obsługi oraz inne wydatki wejdą w grę, ale pomysł tylko jakie masz dzisiaj możliwości oszczędzania.* | *The next issue are the mechanism related to hydraulics, pneumatics. A compressor set somewhere in a storage next to the house allows to carry out many works related to, for example, building interior elements and even to create your own, well-functioning place for typically technical work. The better quality of the compressor, the more effective they are and the longer they can be used. Their mechanism works even in a regular dryer, but it so happens that with an electric power generator and creating for example, an automatic water circulation and filtration system, you are able to live practically independently from basic utilities. Of course, the costs of service will be an issue, but think about the possibilities of savings. Moreover, the content presents only a few options.* |

| | |
|---|---|
| *Poza tym w treści przedstawione zostały jedynie pewne opcje* | |

Jest ich znacznie więcej i wystarczy odrobina skupienia, aby je ………………… .

Is them much more and is.enough a.bit focus, to them …………………

There are many more and it takes just a bit of focus to ……… them.

| | |
|---|---|
| the answer selected by the majority of participants | odszukać (find.PERF) |
| aspectual competitor | odszukiwać (find.IMPF) |
| original chosen | yes |
| original lemma | odszukać (find.PERF) |
| agreement | 98.7% |

In question 12 we notice a familiar pattern. The target is part of a sequence of actions, of which the first one is expressed with perfective. Perhaps not surprisingly, the majority of participants decided to select perfective for the target as well.

**Question 12**

| | |
|---|---|
| *Usunięcie warstwy nie spowoduje przesunięcia innych warstw do góry. Wciśnięcie klawisza ctrl w momencie kliknięcia kosza, spowoduje usunięcie prócz* | *Removing the layer will not cause moving the other layers up. Pressing ctrl button while clicking on the bin will also remove, in addition to the information about the layer* |

| | |
|---|---|
| *samej informacji o teksturze także danych zawierających jej rozkład na terenie, czyli gdzie i w jakich ilościach pokrywała ona teren. By usunąć wszystkie informacje zawarte na warstwie włączając w to tekstury oraz rozłożenie, wciśnij ctrl shift, a następnie potwierdź operację w okienku. Przenoszenie warstwy. 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.* | *itself, the data on its distribution in the field, or where and in what quantities it covered the field. To remove all the information contained in the layer, including the textures and the distribution, press ctrl shift and then confirm the operation in the pop up window. Moving layers. To move layers use the up and down arrow buttons. Moving a layer moves the layer and the information about its distribution.* |

Przeniesienie samych informacji o rozłożeniu można zatem wykonać w dwóch krokach, najpierw **przenieść** całą warstwę, a na koniec ………………… tylko tekstury.'

Moving alone information about distribution can therefore perform in two steps, first **move.INF.PERF** entire layer, and in end …… only textures.

Moving only the information about the distribution can be therefore done in two steps, first move the entire layer and then …… only the textures.

| | |
|---|---|
| the answer selected by the majority of participants | podmienić (switch.PERF) |
| aspectual competitor | podmieniać (switch.IMPF) |
| original chosen | yes |
| original lemma | podmienić (switch.PERF) |
| agreement | 97.4% |

The same conclusion can be drawn about question 52. The perfective selected by the majority of the speakers follows another perfective with which it forms a sequence of actions.

**Question 52**

| | |
|---|---|
| *Jego skutki w coraz większym stopniu dają o sobie znać. Zadłużanie publiczne rośnie w ekspresowym tempie. Dzięki potężnej machinie manipulacji stara maksyma ‘chleba i igrzysk’ straciła swoje zastosowanie. Dziś wystarczają same ‘igrzyska’. Koszta związane z ich organizacją ponosi społeczeństwo, natomiast zyski trafią do ograniczonej grupy biznesmenów, dyrektorów czy członków rad nadzorczych. Chwała i splendor zarezerwowane są dla elit, reszta musi poddać się niedawno wydłużonemu rygorowi pracy.* | *Its consequences are more and more visible. The public debt is growing rapidly. Thanks to the mighty manipulation machine, the old maxim ‘bread and circuses’ lost its use. Today, ‘circuses’ are enough. The costs of organizing them are paid by the society, but the profits go to a small group of businessmen, directors or board members. The glory and splendour are reserved for the elites, the rest must submit to the recently extended work duty.* |

| | |
|---|---|
| | |

Czas się **wyrwać** z hipnozy i ………………… elity czerpiące zyski z naszego codziennego trudu.

Time REFL **tear.away.INF.PERF** from hypnosis and ………………… elites gaining profits from our daily toil.

It's time to break the spell and ……. the elites profiting from our daily toil.

| | |
|---|---|
| the answer selected by the majority of participants | rozliczyć (square/appraise.PERF) |
| aspectual competitor | rozliczać (square/appraise.IMPF) |
| original chosen | yes |
| original lemma | rozliczyć (square/appraise.PERF) |
| agreement | 96.1% |

Question 20 contains an adverbial strongly suggesting that the action expressed by the target verb was repeated. In combination with the contemporary adverbial participle *wbiegając,* also present in the target sentence, the context suggest that the target verb expresses a repeated action that happened a few times in the process of running up the stairs.

**Question 20**

| | |
|---|---|
| *maeda 2009 - 03- 27, 23:47* | *maeda 2009 - 03- 27, 23:47* |

| | |
|---|---|
| *-a więc chodźmy - dołączył do nich oleg.* <br> *maeda 2009 - 07- 16, 12:35* <br> *wieczorem oleg wybiegł gwałtownie na dach szkoły.* | *- so let's go – oleg joined them* <br> *maeda 2009 - 07- 16, 12:35* <br> *in the evening oleg ran fiercely to the roof of the school* |

Nie przejmował się tym, że wbiegając po schodach **co chwilę** się ………………… .

Not care.3SG.PAST.IMPF REFL this, that run.PRPART on stairs every moment REFL …………………

He didn't care that while running on the stairs, every now and again he …………………

| | |
|---|---|
| the answer selected by the majority of participants | wywracać (topple/fall.IMPF) |
| aspectual competitor | wywrócić (topple/fall.PERF) |
| original chosen | yes |
| original lemma | wywracać (topple/fall.IMPF) |
| agreement | 96.1% |

There is no adverbial in question 3, however the target verb is coordinated with another imperfective verb. This again could serve as a cue for participants, making them select the imperfective as a response.

**Question 3**

| | |
|---|---|
| *Gdy rodzice przynieśli dzieciątko Jezus do świątyni, była tam prorokini Anna, córka Fanuela z pokolenia Asera, bardzo podeszła w latach. Od swego panieństwa siedem lat żyła z mężem i pozostała wdową. Liczyła już osiemdziesiąty czwarty rok życia. Nie rozstawała się ze świątynią, służąc bogu w postach i modlitwach dniem i nocą. Przyszedłszy w tej właśnie chwili, sławiła Boga i mówiła o nim wszystkim, którzy oczekiwali wyzwolenia Jerozolimy. A gdy wypełnili wszystko według prawa pańskiego, wrócili do Galilei, do swego miasta – Nazaret.* | *When the parents brought baby Jesus to the temple, an elderly prophet Anna, daughter of Panuel of the tribe of Aser was there. From her maidenhood, she lived seven years with her husband and remained a widow. She was eighty four years old. She never left the temple, serving God through fasts and prayers all day and all night. Having entered the temple at this very moment, she was praising God and telling everyone who was waiting for the redemption of Jerusalem about him. And when they had done everything required by the Law of the Lord, they returned to Galilee to their town – Nazareth.* |

Dziecię zaś **rosło** i ………………… mocy, napełniając się mądrością, a łaska boża spoczywała na nim.

Child though **grow.3SG.IMPF** and ………………… power, filling REFL wisdom, and grace of.god rest on him.

And the child was growing and …………………… power, filling with wisdom and the grace of God was upon him.

| | |
|---|---|
| the answer selected by the majority of participants | nabierać (gain.IMPF) |
| aspectual competitor | nabrać (gain.PERF) |
| original chosen | yes |
| original lemma | nabierać (gain.IMPF) |
| agreement | 93.5% |

Question 40 is interesting because the participants disagreed with the author and chose perfective. At the same time, there are no additional cues in the sentence that could bias participants towards selecting that particular answer. However, there are two possible reasons for the choice. First of all, the tense was annotated as 'none' since the sentence is in the conditional mood and cannot be said to express future or past action. The form of the verb, however, was past, which may sway participants to the perfective as a default option. What is more, the frequency distribution of the lemmas is again biased towards the option that was selected by the majority. Perfective was counted 2504 times, whereas the imperfective 1382.

**Question 40**

| | |
|---|---|
| *Miałam informację na stronie firmowej na ten temat w odpowiedniej opcji, ale niestety w html-u, a już dawno mam nową stronę na* | *I had an information about the right option on my company's website, but unfortunately in the html version, but I've switched to a* |

| | |
|---|---|
| *joomli i może i mam drobną poligrafię, ale już chyba nie umieszczałam podpopcji nt. zasad feng shui. Więc tylko daję ci do poczytania artykuł, jaki kiedyś na ten temat napisałam do Millionaire magazynu, do działu, który wymyśliłam i prowadziłam. Oczywiście, jak przeczytasz artykuł i cię przekona ta informacja, że też chcesz mieć wg tych zasad zaprojektowane wizytówki, to udostępniam ci go po to, żebyś wiedziała, czego ewentualnie wymagać od firmy, która przygotowuje dla ciebie projekt wizytówki :-) Zrobienie 100 sztuk wizytówek z przesyłką ode mnie byłoby bezsensowne, a nawet 200 czy 300 czy 500 sztuk, biorąc pod uwagę, jakie oferty na druk 500 sztuk wizytówek można znaleźć w sieci. To tylko wskazówki co do projektu. I jak przeczytasz w artykule, nic więcej nie ma na ten temat, nawet w książce, która mi to obiecywała w ofercie, dlatego ją zakupiłam, żeby się dowiedzieć jeszcze więcej.* | *website on joomla long time ago and maybe I have desktop printing, but I don't think I added a suboption on feng shui. So I'm sending you an article to read, which I wrote some time ago for a Millionaire magazine, for a section I came up with and led. Of course, if you read the article and the information convinces you that you also want your business cards designed according to these principles, I'm sharing it for you to know what to ask the company which will design the cards for you. Making a 100 business cards with posting from me wouldn't make sense, and even 200 or 300 or 500 cards, taking into account what offers for 500 business cards you can find on the Internet. These are only guidelines for design. And as you will read in the article, there is not much else on the topic, even in a book that promised me this, which was the reason I bought it, to learn more.* |

A z kolei też uprzedzam, że specjaliści od feng shui to może by w ogóle te koncepcje ………………… do góry nogami, ale swoich innych koncepcji do tej pory nigdzie nie opisali.

But in turn also warn.1SG.PRES.IMPF, that experts from feng shui this maybe would in general these concepts ………………… to up legs, but their other concepts to this time nowhere not describe.3PL.PAST.PERF

And I'm also warning you that the feng shui experts may well ………….. these concepts upside down, but they haven't described their own concepts anywhere yet.

| | |
|---|---|
| the answer selected by the majority of participants | wywrócić (topple/fall.PERF) |
| aspectual competitor | wywracać (topple/fall.IMPF) |
| original chosen | no |
| original lemma | wywracać (topple/fall.IMPF) |
| agreement | 92.2% |

In question 13, we have another verb, this time imperfective, which could have guided the choice. Participants selected an imperfective target.

**Question 13**

| | |
|---|---|
| *Ileż to bezsensownych projektów konsultantów znam z doświadczenia lub* | How many useless consultant's projects do I know from my experience or have heard of |

| | |
|---|---|
| *słyszałem o nich od znajomych. Zdecydowana większość to właśnie 'dupokryjki' dla zarządów… Dobrzy menedżerowie nie potrzebują zewnętrznych konsultantów (może z wyjątkiem IT – ale tu też przekręty odchodzą na maksa) bo potrafią zatrudnić dobrych pracowników. 90% konsultantów należałoby pogonić z firm, wtedy funkcjonowały by one dużo lepiej i taniej… Działalność?? To jest wpychalność do kieszeni!!! A zawsze jest ich malutko ale tyle, że każdemu są potrzebni.* | from my friends. The majority of them are 'bottom-covering' for companies' boards. Good managers don't need outside consultants (maybe apart from IT – but here there are also many swindles) because they can hire good employees. 90% of consultants should be chased away from the companies, they would function better and more cheaply. Entrepreneurship? More like bribery! And there are always so few of them, but just enough that everyone needs them. |

Nawet Tusk z takim poparciem musi ………………… oczami i **bredzić** jaka to przykładowa koalicja.'

Even Tusk with this support must ……………….. eyes and **ramble.INF.IMPF** how it exemplary coalition.

Even Tusk with this support must be ……………. his eyes and ramble on about what an exemplary coalition this is.

| | |
|---|---|
| the answer selected by the majority of participants | wywracać (topple/fall.IMPF) |
| aspectual competitor | wywrócić (topple/fall.PERF) |

| | |
|---|---|
| original chosen | yes |
| original lemma | wywracać (topple/fall.IMPF) |
| agreement | 90.9% |

*Errors in the mid-range of agreement (80-60%)*

In question 30, the majority of participants decided to select the less frequent imperfective. There are two possible reasons for this choice. On the one hand we could say that the use of the imperfective suggests that searching for the candidate's phone number in their CV is a longer and inconvenient process. On the other it is possible that the auxiliary “be”, which in this context forms a compound future tense with the modal, influences the choice because of its strong association with imperfective. This association is much stronger than the association between modals and perfective

**Question 30**

| | |
|---|---|
| *‘Dotyczy to zarówno stanowisk fizycznych jak i umysłowych.' Agnieszka, lat 29, młodszy kierownik projektu w agencji pracy. Wymogi formalne listu motywacyjnego: list motywacyjny nie powinien być zbyt długi, powinien zmieścić się na jednej stronie A4. Powszechnie obowiązuje forma zwięzła i konkretna; bez wodolejstwa. W prawym górnym rogu powinniśmy wpisać datę i* | *‘It concerns both blue- and white-collar workers.’ Agnieszka, age 29, junior project manager in an employment agency. Formal requirements of a motivational letter: a motivational letter should not be too long, and should fit on a single A4 page. As per custom, it should be brief and to the point; no waffling. In the upper-right corner we should put in the date and the place, and below the* |

| | |
|---|---|
| *miejscowość, a poniżej adresata oraz funkcję osoby do której się zwracamy.* | *addressee and the job title of the person we address the letter to.* |

Z lewej strony umieszczamy swoje dane - jest to bardzo ważne, bo jeśli rekruter zdecyduje się na naszą kandydaturę, to nie będzie musiał ………………… numeru telefonu w CV, tylko zadzwoni od razu.

On left side put.3PL.IMPF our details – is it very important, because if recruiter decides REFL on our candidacy, it not be.3SG.FUT must ………………… number phone in CV, only call.3SG.FUT.PERF from once.

On the left hand side we put our details – this is really important, because if the recruiter decides to take matters forward, they won't have to ………………… phone number in the CV, but will call right away.

| | |
|---|---|
| the answer selected by the majority of participants | odszukiwać (find.IMPF) |
| aspectual competitor | odszukać (find.PERF) |
| original chosen | yes |
| original lemma | odszukiwać (find.IMPF) |
| agreement | 79.2% |

The choice in question 25 is difficult to explain. On the one hand there is a slight frequency bias towards the perfective, which was used originally (25432 vs 21423). However, such a small difference might be unlikely to have such an effect. A search of the National Corpus of Polish shows that the imperfective *nabierać* (gain.IMPF) occurs more frequently with *znaczenia* (meaning.GEN) (112 vs 51). It is possible that the choice is guided by the presence of the modal verb, since the presence of a modal is associated with the use of perfective. However, the weight is not strong enough to override the sum of all the cues present in the sentence.

**Question 25**

| | |
|---|---|
| *To jest publiczne forum, więc nikogo nie zdziwi publiczne wyrażanie prywatnego zdania. Przeciwnie. Wszyscy na to czekają. I nie martw się o 'fachowość'. rozmowa jest ważniejsza od fachowości. Rozmawiaj śmiało.* | This is a public forum, so no one will be surprised by the public expression of a private opinion. On the contrary. Everyone's expecting it. And don't worry about 'expertise'. A conversation is more important than expertise. So converse freely. |

Te same słowa na piśmie mogą ………………… zupełnie innego znaczenia niż wypowiedziane bezpośrednio jakimś tonem głosu i z jakimś wyrazem twarzy

These same words on writing may ………………… completely different meaning than said directly some tone voice.GEN and with some expression face.GEN

The same words in writing may ………………… a completely different meaning than when said directly, with a tone of voice and a facial expression.

| | |
|---|---|
| the answer selected by the majority of participants | nabrać (gain/kid.PERF) |

| | |
|---|---|
| aspectual competitor | nabierać (gain/kid.IMPF) |
| original chosen | yes |
| original lemma | nabrać (gain/kid.PERF) |
| agreement | 77.9% |

In question 45, participants chose a less frequent imperfective despite the presence of the modal, which served as a strong cue for the model. The use of imperfective here suggest repeatedness, which is compatible with the quantifier describing the direct object – all graphics and animation can be substituted. Perhaps it was this quantifier that influenced the participants' choice.

**Question 45**

| | |
|---|---|
| *Podczas pracy mamy już wygenerowanego password boota. Teraz nie powinniśmy mieć problemu z wgraniem łatki do naszej słuchawki. Pierwsze, co teraz powinniśmy zrobić, to zaopatrzyć się w jakieś ciekawe nakładki do naszego telefonu i wersji SW, która w nim jest. Skrypty dzielą się na dwie kategorie. Pierwsza, to skrypty graficzne.* | During the work we already have a password boot generated. Now we should have no problems with generating a patch for our earphone. The first thing we should do is to get some interesting overlays for our phone and the SW version that's on it. The scripts can be divided into two categories. The first one is the graphic scripts. |

Dzięki nim możemy ………………… wszelką grafikę i animację w telefonie.

Thanks them can.1PL.IMPF ………………… all graphics and animation in phone

Because of them we can …………….. all graphics and animations on our phone.

| | |
|---|---|
| the answer selected by the majority of participants | podmieniać (switch.IMPF) |
| aspectual competitor | podmienić (switch.PERF) |
| original chosen | yes |
| original lemma | podmieniać (switch.IMPF) |
| agreement | 76.6% |

In question 54, participants chose the original imperfective. Again, the model was most likely guided by the presence of the modal verb. However, the plural subject suggest repeatedness, which is compatible with the imperfective 'wywracać'.

**Question 54**

| | |
|---|---|
| *Stare modele były metalowe, nowe to rodzaj żywicy. Już słyszę obawy części modelarzy, którzy mieli do czynienia z tanimi poliuretanami modelowymi. Moge was jednak uspokoić – nowoczesne i dobrze dobrane żywice to zupełnie inna bajka.* | *The old models were made out of metal, the new ones out of some kind of resin. I can already hear the doubts of some modellers, who have experience with cheap modelling polyurethanes. I can reassure you though – modern and well-chosen resins are a* |

| | |
|---|---|
| *Zdecydowanie mogą konkurować chociażby z wtryskiwanym polistyrenem, z którego wykonane są 'standardowe' plastikowe figurki. Pierwsze co zwraca uwagę to jak niesamowicie lekki jest ten materiał. To zdecydowanie dobra wiadomość dla graczy biorących udział w turniejach – walizki z armią przestaną 'urywać ręce'.* | *completely different story. They can definitely compete with injected polystyrene which the 'standard' plastic figurines are made out of. The first thing that attracts our attention is how incredibly light this material is. This is definitely good news for players who participate in competitions – suitcases with your army will stop 'tearing your arms off'.* |

Myślę jednak, że w niektórych wypadkach niewielkie dociążenie podstawki może okazać się konieczne – inaczej **modele** mogą się ………………… .

Think.1SG.IMPF however that in some cases small load base.GEN can turn.out REFL necessary – otherwise **models** may REFL …………………

I think, however, that in some cases, adding a little bit more weight to the base might turn out to be necessary – otherwise the **models** might ………………….

| | |
|---|---|
| the answer selected by the majority of participants | wywracać (topple/fall.IMPF) |
| aspectual competitor | wywrócić (topple/fall.PERF) |
| original chosen | yes |
| original lemma | wywracać (topple/fall.IMPF) |
| agreement | 76.6% |

In question 56 the participants disagreed with the author of the paragraph and chose perfective. Two factors could have an influence here. The first one is the presence of a modal verb. The second is a significant frequency bias towards the perfective (44391 vs 580).

**Question 56**

| | |
|---|---|
| *Publikacje wydawnictwa św. Wojciecha zostały podzielone na trzy lata kształcenia w klasach 1-3. Wszystkie części serii są integralną całością, a każda z nich kontynuuje wiadomości zebrane w poprzedniej. Celem kształcenia w oparciu o podręcznik 'Bliscy sercu Jezusa' jest przede wszystkim ugruntowanie wiadomości dotyczących wiary i wiedzy na temat religii w drugiej klasie szkoły podstawowej. Autorzy pakietu skupili się przede wszystkim na uświadomieniu dzieciom konieczności indywidualnej pracy nad sobą oraz potrzeby zbliżania się do Boga. Treści zawarte w książkach ukazują Jezusa jako przyjaciela każdego człowieka, który przemawia do niego poprzez karty Pisma Świętego.* | *The publications of Saint Adalbert publishing house were divided into three years of primary education. All parts of the series are an independent whole, and each of them elaborates on the information gathered in the previous one. The goal of education based on the handbook 'Close to Jesus's heart' is to consolidate information on faith and knowledge about religion in the second grade. The authors of the collection focused mainly on making children aware how important individual work on themselves and the need for getting closer to God are. The content of the books presents Jesus as a friend of every man, who speaks to them through the pages of the Bible.* |

Wszystkie treści zawarte w publikacjach z serii 'Bliscy sercu Jezusa' zostały przeredagowane tak, aby dzieci korzystające z książek **mogły** swobodnie ………………… potrzebne informacje.

All content included in publications from series 'Close heart.GEN Jesus.GEN' been edited such, that children using from books **could** freely ……………… necessary information

All the contents of the publications from 'Close to Jesus's heart' publications have been edited in such a way that the children using the books **could** freely ………………… all the necessary information.

| | |
|---|---|
| the answer selected by the majority of participants | odszukać (find.PERF) |
| aspectual competitor | odszukiwać (find.IMPF) |
| original chosen | no |
| original lemma | odszukiwać (find.IMPF |
| agreement | 75.3% |

The presence of another imperfective verb in question 58 might have influenced the participants, who chose the original imperfective.

**Question 58**

| | |
|---|---|
| *Dopiero gdy ręcznie ustawiliśmy skanowanie powierzchni, udało się aplikacji tego* | *Only when we set the scanning of the surface manually, was the application able to do* |

| | |
|---|---|
| *dokonać. Niestety, ten sposób zawodzi w przypadku skasowanych płyt dvd RW. Isopuzzle nie znajduje na nich absolutnie nic, ponieważ nie obsługuje niezbędnego do tego celu trybu raw-scan, koniecznego do analizowania zawartości dysku. A przecież nawet na wyczyszczonym dysku są dane, ponieważ – podobnie jak dzieje się to podczas formatowania dysku twardego – na nowo tworzona jest tylko struktura katalogów. Aby odszukać dane na takich dyskach, programy muszą odczytać zagubione bajty i zidentyfikować typy plików. I znów tylko Isobuster i Cdroller dobrze wypadły w tej konkurencji.* | *that. Unfortunately, this method fails when it comes to destroyed DVD-RW discs. Isopuzzle cannot find anything on them, because it doesn't support a raw-scan mode, which is indispensable for analysing the contents of the disc. And yet, the data is still there on a disc that has been erased, because- just as in the case of formatting a hard drive – only a new structure of catalogues is created. To find data on such discs, software must read lost bytes and identify file types. And again, only Isobuster and CDroller performed well in this category.* |

Badcopy ………………… pliki tylko przy zastosowaniu intensywnej metody skanowania, co **zajmowało** mnóstwo czasu.

Badcopy ……………….. files only when using intensive method scanning, which **take.PAST.IMPF** lots time

Badcopy ……………. files only when using intensive scan mode, which **took** a lot of time.

| | |
|---|---|
| the answer selected by the majority of participants | odszukiwać (find.IMPF) |
| aspectual competitor | odszukać (find.PERF) |
| original chosen | yes |
| original lemma | odszukiwać (find.IMPF |
| agreement | 72.7% |

The target sentence in question 22 contains an adverbial phrase with a verb in imperfective, which suggest simultaneous actions: take a breath in while lifting the weights up.

**Question 22**

| | |
|---|---|
| *Unieś sztangielki bokiem w górę, aż sztangielki znajdą się na wysokości barków. Do tego treningu możesz użyć lustra to pozwoli zobaczyć jak wykonujesz to ćwiczenie i korygować ewentualne błędy. Następnie opuść sztangielki na dół aż do pozycji wyjściowej. Ciężar sztangielek nie powinien być zbyt duży, ponieważ to ćwiczenie wymaga dobrego wykonania technicznego, więc tak dobierz ciężar, aby prawidłowo technicznie wykonać ten trening. Następnie bez przerywania przejdź do* | Raise the weights sideways, until they are at shoulder height. You can use a mirror for this training, which will help you see how you perform this exercise and correct any errors. Next, lower the weights to the starting position. The weights shouldn't be too heavy, because this exercise needs to be well-executed technically, so choose weights that will allow you to correctly perform this training. Next, without breaks, repeat the movement. Remember to breathe correctly. |

| | |
|---|---|
| *kolejnego powtórzenia. Pamiętaj o prawidłowym oddychaniu.* | |

………………… powietrza **podczas unoszenia sztangielek w górę**.

………………… air **while rising dumb-bell.PL in up**.

…………………. air **while raising the dumb-bells**.

| | |
|---|---|
| the answer selected by the majority of participants | nabierać (gain/kid.IMPF) |
| aspectual competitor | nabrać (gain/kid.PERF) |
| original chosen | yes |
| original lemma | nabierać (gain/kid.IMPF) |
| agreement | 62.3% |

The target sentence in question 41 contains two other verbs which could have influenced the choice – *umieć* (can.IMPF) and *ocenić* (evaluate.PERF). Arguably however, the perfective *ocenić* that comes after the target describes the end of the process, while the imperfective *umieć* and the target are describing the process itself. This, in combination with the plural object (*bruzdy mięśniowe* – muscle sulci) could have resulted in the imperfective being the preferred option.

**Question 41**

| | |
|---|---|
| *W technice tej ruchy są wykonywane nadzwyczaj precyzyjnie. Pracuje się bardzo powoli dając tkankom czas na fizjologiczną odpowiedź i przystosowanie się do nowych warunków. Kolejną cechą tego rodzaju masażu jest stosowanie bardzo małych lub w ogóle nie stosowania środków poślizgowych. Pracuje się używając skośnego nacisku na tkanki z uwagi na to że jest on lepiej tolerowany przez osobę masowaną, poza tym tkanka mięśniowa dobrze reaguje na rozciąganie i przesuwanie ale na przyciskanie szczególnie do kostnych elementów nie możemy już sobie pozwolić. Przed zabiegiem trzeba określić cel, który chcemy osiągnąć i wybrać odpowiednią technikę. Pracuj różnymi częściami ciała jak łokcie czy przedramię natomiast absolutnie unikaj pracy swoimi kciukami.* | *In this technique, the movements are performed extremely precisely. We work very slowly giving the tissues the time to respond physiologically and to accommodate to new conditions. The next characteristic of this type of massage is that we use very little or no lubricants at all. We work using a sideways pressure on the tissues because it is better tolerated by the person being massaged, in addition the muscle tissue reacts well to stretching and moving but we cannot allow pressing, especially to the bones. Before the treatment we need to clarify the goal we want to achieve and choose the right technique. Work with different body parts, such as elbows or forearm, but absolutely avoid working with your thumbs.* |

Istotnym elementem w technikach masażu głębokiego jest wrażliwość palpacyjna terapeuty - musi on umieć odnajdować zmiany napięcia w tkankach, ………………… bruzdy mięśniowe oraz ocenić kiedy masaż należy zakończyć by uzyskać efekt terapeutyczny a jednocześnie nie przestymulować tkanki.

Important element in techniques massage.GEN deep is sensitivity palpative therapist.GEN – must he be.able.INF.IMPF find.INF.IMPF changes tension.GEN in tissues ………………… muscle sulci and evaluate.INF.PERF when massage need finish to obtain effect therapeutic and simultaneously not overstimulate tissue

An important element of a deep tissue massage is the palpative sensitivity of the therapist – he must be able to find changes in the tension of the tissues, ………………… muscle sulci and evaluate when a massage should finish, to achieve a therapeutic result while not overstimulating the tissue.

| | |
|---|---|
| the answer selected by the majority of participants | odszukiwać (find.IMPF) |
| aspectual competitor | odszukać (find.PERF) |
| original chosen | yes |
| original lemma | odszukiwać (find.IMPF) |
| agreement | 61% |

*Errors in the low agreement range (> 60%)*

In question 43, we yet again see another imperfective verb, apart from the target verb, which could have biased the participants' choice towards the imperfective.

**Question 43**

| | |
|---|---|
| *Osoby, które uprawiają ten sport z pewnością muszą zainwestować w odpowiednie ubrania, które wspomagają opływ powietrza wokół ciała. Kolarz powinien mieć także kask zabezpieczający głowę przy wypadku, ochraniacze na rękach oraz na kolanach, rękawiczki ze wzmocnieniami na dłoniach, które zabezpieczają najbardziej wrażliwe części dłoni przed urazami oraz okulary do jazdy na rowerze, które chronią twoje oczy przed owadami i różnymi pyłkami. Jeżeli podczas szybkiej jazdy rowerem wpadnie ci coś do oka to bardzo szybko możesz mieć wypadek gdyż chcesz wyciągnąć to z niego a wtedy trudno jest patrzeć się też na drogę i utrzymywać równowagę, dlatego właśnie okulary są tak wspaniałym rozwiązaniem. Dodatkowo nie zapomnij o tym by w twoim* | *People who do this sport definitely must invest in proper clothing which supports air flow around the body. A cyclist should also have a helmet protecting the head in case of an accident, protectors on his hands and knees, padded gloves, which protect the most vulnerable parts of your hands from injuries and glasses, which protect your eyes from insects and dust. If something gets into your eye while cycling at high speed, an accident is very likely, as you will try to take it out and it's then difficult to see where you're going and maintain balance, that's why glasses are such a great solution. Also, make sure your bike has a bottle holder. Cycling makes you sweat a lot, that's why you must remember about drinking, thanks to which you won't run out of steam and will definitely be able to* |

| | |
|---|---|
| *rowerze był uchwyt na napój. W czasie jazdy rowerem pocisz się, dlatego koniecznie trzeba pamiętać o uzupełnianiu płynów, dzięki temu nie opadniesz szybko z sił i z pewnością podołasz reszcie trasy, którą musisz pokonać. Jeżeli bierzesz udział w wyścigu nie radzę wody pić.* | *complete the rest of the route. If you're racing, I don't recommend actually drinking water.* |

Polecam raczej ………………… jej do ust i ją **wypluwać** dzięki temu nie poczujesz potrzeby skorzystania z toalety.

Recommend.1SG.PRES.INF rather ………………… her to mouth and her **spit.out.INF.IMPF** thanks that no feel.2SG.FUT.PERF need use with bathroom.

Rather, ……………….. it in and **spit out**, this way you won't need to go to the bathroom.

| | |
|---|---|
| the answer selected by the majority of participants | nabierać (gain/kid.IMPF) |
| aspectual competitor | nabrać (gain/kid.PERF) |
| original chosen | yes |
| original lemma | nabierać (gain/kid.IMPF) |
| agreement | 59.7% |

Question 55 informs about the possibility of generating an infinite number of poems by replacing each line of the original poem. The quantifier describing the line of the poem (all) and the adverb used (*dowolnie* – "at will") suggest repeatedness, which is compatible with the imperfective chosen by the majority of the participants.

**Question 55**

| | |
|---|---|
| *Wojciech Kalaga w swoim artykule 'Liberatura: słowo, ikona, przestrzeń' objaśnia współdziałanie najistotniejszych elementów dzieła liberackiego. Litera, najmniejszy graficzny składnik słowa, jeszcze zanim z innymi literami utworzy morfem lub wyraz, usamodzielnia się w kreowaniu znaczeń. Rodzajem twórczości, który nam to uświadamia, jest poezja konkretna, w której litera jest nie tylko nośnikiem znaczenia, ale tworzywem tekstu wizualnego – 'sememem' tekstobrazu. Litera jednak nie tylko potrafi być niezależna; potrafi także wspomagać znaczenia lub wskazywać ścieżki interpretacji. Powyższe zdjęcie jest przykładem liberatury – książki, w której tekst, sposób zapisu, forma oraz materiał stanowią nośnik treści. Sto tysięcy miliardów* | *Wojciech Kalaga in his article 'Liberature: a word, an icon, a space' explains the collaboration of the most important elements of a 'liberative' work. A letter, the smallest part of a word, even before forming a morpheme or a word, becomes independent in forming a meaning. A type of art that helps us realize that is 'concrete poetry' in which a letter not only carries meaning but is also a part of a visual text – a sememe of textpicture. A letter is able to be more than independent; it can also support meaning or point towards ways of interpreting. The picture above is an example of liberature – a book in which the text, the method of writing, the form and the content create a medium for the meaning. The 'One hundred thousand billions poem' is a* |

| | |
|---|---|
| *wiersz to kombinatoryczny, permutacyjny cykl 10 sonetów, wydrukowanych na kartkach pociętych na paski.* | *combinatorial, permutational series of 10 sonnets, printed on pages cut into strips.* |

Każdą linijkę wiersza można dowolnie ………………… , co umożliwia czytelnikowi samodzielne wygenerowanie niemal nieskończonej liczby wierszy.'

Each line poem.GEN can freely ………………… , which allows reader independent generating almost infinite number poems.GEN

Each line of a poem can be freely ………………… , which allows the reader to independently generate an almost infinite number of poems.

| | |
|---|---|
| the answer selected by the majority of participants | podmieniać (switch.IMPF) |
| aspectual competitor | podmienić (switch.PERF) |
| original chosen | yes |
| original lemma | podmieniać (switch.IMPF) |
| agreement | 58.4% |

In question 57, the majority of participants chose the imperfective option. The verbs used before the target are imperfective (*kraść* – steal.IMPF and *zwiększa*ć – increase.IMPF). However, the verb after the target is perfective (*zmiejszyć* – decrease.PERF). Each aspect used before and after the target could have served as a cue which guided the choice, which helps explain the low agreement between the participants in this particular question.

**Question 57**

| | |
|---|---|
| *Zasady Slaviki są stosunkowo proste. Gracze kolejno wysyłają do krain swoich bohaterów celem zdobycia skarbu oraz punktów chwały. W swojej turze każdy z graczy wysyła dwóch bohaterów do dwóch różnych krain oraz jednego potwora do kolejnej, nie wybranej wcześniej krainy. Następnie gracz odrzuca jedną z kart potworów z ręki i dobiera tyle tych kart, by mieć ich 5 na ręce. Jeżeli w danej krainie uzbiera się wystarczająca liczba potworów następuje jej rozliczenie, czyli zsumowanie siły potworów oraz bohaterów by przekonać się, która strona wygrała. Naturalnie zarówno bohaterowie, jak i niektóre ze stworzeń posiadają specjalne umiejętności, które w znaczny sposób potrafią komplikować życie naszych przeciwników.* | *The rules of Slavica are relatively simple. Players take turns in sending their heroes to lands to find treasures and points of glory. During their turn each player sends two heroes to two different lands and one monster to an additional land, not previously selected. Next, a player discards one of the monster cards from his hand, and takes as many cards as he needs to have 5 on hand. If there are enough monsters in each land, the strength of the monsters and heroes are tallied to find out which side won. Of course, the heroes and some of the creatures have special powers which can significantly complicate the lives of our enemies.* |

W każdym rodzie znajduje się złodziej, który może kraść skarby znajdujące się w danej krainie, guślarz który zwiększa swoją moc z każdym wojownikiem znajdującym się w tej samej krainie, zwiadowca który pozwoli graczowi ………………… swoich bohaterów w poszczególnych krainach, łucznik który zmniejszy siłę potworów oraz dwóch silnych wojów bez żadnych specjalnych umiejętności.'

In each family there.is thief, who can steal.INF.IMPF treasures found in given land, sorcerer who increase.3SG.PRES.IMPF his power with each warrior found REFL in this same land, scout who allow.3SG.FUT.PERF player ………………… his heroes in each land, archer who decrease.3SG.FUT.PERF strength monsters.GEN and two strong warriors without any special skills.

In each family, there is a thief, who can steal treasures found in a given land, a sorcerer, who increases his power with each warrior in the same land, a scout, who will allow players to ………………… their heroes in each land, an archer, who will reduce the monsters' power, and two strong warriors without any special skills.

| | |
|---|---|
| the answer selected by the majority of participants | podmieniać (switch.IMPF) |
| aspectual competitor | podmienić (switc.PERF) |
| original chosen | yes |
| original lemma | odmieniać (switch.IMPF) |
| agreement | 55.8% |

Finally, in question 17, the majority chose an imperfective. The presence of the modal once again influenced the model, but the fact that the subject (trees) is plural might sway the participants towards interpreting the situation as repetitive and therefore towards choosing the imperfective.

**Question 17**

| | |
|---|---|
| *Na działce czy w ogródku przydomowym, gdzie poza drzewami i krzewami owocowymi uprawia się warzywa i rośliny ozdobne, lepiej sadzić jabłonie i grusze w formach karłowych i półkarłowych, z innych zaś gatunków dla których brak podkładek karłowych - odmiany słabo rosnące. Z orzecha włoskiego i czereśni, drzew bardzo silnie rosnących, w małym ogródku trzeba zrezygnować. Drzewa karłowate i półkarłowate moją te zalety, że ich korony nie rozrastają się nadmiernie i nie zacieniają ogrodu, rosną szybciej i znacznie wcześniej, bo już w drugim i trzecim roku po posadzeniu zaczynają dawać plon oraz na ogół nie owocują przemiennie. Łatwiej je opryskiwać, ciąć i wykonywać wszystkie inne zabiegi pielęgnacyjne, łatwiejszy jest także zbiór owoców. Drzewka karłowe i* | *On an allotment or in a garden next to the house, where in addition to trees and fruit bushes we grow vegetables and ornamental plants, it is better to plant in dwarf or half-dwarf varieties of apple and pear trees, and as for other species that do not have dwarf varieties – these that don't grow as well. We should give up on planting walnut trees and cherry trees, which grow very vigorously, in a small garden. The advantage of dwarf and half-dwarf trees is that their tops do not grow excessively and they don't cast too much shadow over the garden, they mature faster and much earlier, as they bear fruit two or three years after planting, and usually they don't bear fruit in alternate years. It's easier to spray them, cut them and undertake any other maintenance tasks, picking fruit is also* |

| | |
|---|---|
| *półkarłowe otrzymuje się przez szczepienie - 'okulizację' szlachetnej odmiany na podkładach słabo rosnących, powodujących ograniczenie wzrostu drzewa. Dla uzyskania jabłoni karłowych stosuje się w szkółkach wegetatywne podkładki m9 i m26, na których otrzymuje się najmniejsze drzewka, nie przekraczające 1,5 m wysokości.* | *easier. Dwarf and half-dwarf trees are the result of grafting – budding of a premium variant on a base that doesn't support too much growth. To make dwarf apple trees tree nurseries plant them on m9 and m26 foundations, which yield the smallest trees, not exceeding 1.5m in height.* |

Dla dobrego rozwoju wymagają bardzo żyznych gleb, trzeba je też przywiązywać do palików, bo mają słaby system korzeniowy i mogą ………………… się pod ciężarem owoców.

For good development require.3PL.PRES.IMPF very fertile soils, need them also tie.INF.IMPF to poles because have.3PL.PRES.IMPF weak system root.GEN and can.3PL.PRES ………………… REFL under weight fruit.GEN

For their correct development they require very fertile soils, they also need to be tied to poles, because they have a very weak root system and they can ………………… under the weight of their fruit.

| | |
|---|---|
| the answer selected by the majority of participants | wywracać (topple/fall.IMPF) |
| aspectual competitor | wywrócić (topple/fall.PERF) |
| original chosen | yes |

| | |
|---|---|
| original lemma | wywracać (topple/fall.IMPF) |
| agreement | 53.2% |

## 8.3 Summary

The conclusion that can be drawn from the analysis above is that there might be other important cues that could have guided the choice the participants made. One candidate is the presence of another verb, or verbs, used in the same aspect as the target. This cue could potentially explain the choice in 11 out of 20 questions in which the model predicted the answer incorrectly. The model did not have access to this cue, as the set of concrete variables did not contain a variable that would capture occurrences of other verbs. This information was partly coded in one of the abstract variables – sequentiality. However, it is not always the case that the presence of other verbs is 'meaningful'; that is, it does not necessarily signal that two actions are simultaneous or form a sequence. A good example that illustrates this is question 57, where other verbs that occur in the sentence describe actions that different characters in a game can perform. Additionally, we have to remember that the interpretation of the meaning of the cooccurrence of verb is not always easy, as it can differ from person to person. As we discussed in Chapter 5, the kappa score for sequentiality was 0.11 which indicates only a slight interannotator agreement on whether the verbs signal sequence or two actions happening simultaneously. On the other hand, this cue could still be useful, if we assume that speakers are less interested in interpreting the meaning of verbal cooccurrence and more concerned with the cooccurrence itself as well as the aspect of other verbs in the sentence. Therefore, annotating for the aspect of other verbs in the sentence might be a valuable addition in future studies. However, it would require a careful operationalization since, as we have seen, there might be a number of other verbs in the sentence, and the variable would have to capture the aspect values of all them, as well as their proximity to the target verb.

Another factor that might be important is the relative frequency of the verbs. It is possible (and quite likely given the evidence for the importance of frequency we discussed in Chapter 2) that where there are no salient contextual cues, speakers tend to use the verb that is more frequent. Again, the aspect-concrete model had no access to such information.

The number of the object also seems to be a relevant dimension, worth including in the model. Although it did not emerge as a predictive cue in the learning simulation, it could be due to the fact that the model learned about other, more reliable cues, which drove the association weight of object number down. In addition, the analysis has also shown that the use of quantifiers such as "all" could be of importance. This would go hand in hand with the importance of the number of the object – after "all" and "every", such quantifiers suggest plurality, even though it is not grammatically expressed by the number of the objects. For instance, a 'line' in 'every line' is singular grammatically, but we are in fact talking about many lines.

In a few cases we analysed, the presence of an adverbial, such as an adverbial of frequency *co chwilę* (every now and again) could have been the factor that influenced the speakers to choose imperfective in the past tense. These cases support the findings we already discussed in Chapter 4 – adverbials can often be treated as clear indicators of which aspect should be used, as using the other aspect would render the sentence incorrect.

Interestingly, the presence of a modal verb, which had strong a positive association weight with the perfective seems to be a source of the model's errors. NDL learned that the presence of a modal verb serves a cue for the perfective, but a closer inspection of the errors indicate that nine of them contain a modal verb and most of them are followed by an imperfective verb. This suggests some taxonomy of modal verbs would improve the predictions. However, semantic taxonomies and categories tend to be difficult to apply and, as we have already pointed out, the interpretations are subjective and disputable. This is the case for the types modal verbs as well, as Divjak et al (2015) demonstrated. However, Divjak (2011) presents a simpler model of aspectual choice in modal contexts, which focuses on the distinction between specific events, which agents, objects and timelines are concrete and identifiable, and generic situations, which are not. These findings and distinctions were annotated for and captured by one of the variables in the set of 'abstract' variables, namely specificity.

Out of 9 errors in the modal context, all were annotated as presenting a 'generic' situation and only 3 were cases where the majority of participants selected a perfective verb (questions 12, 25 and 56). It can be argued that those 3 cases are in fact less 'generic' than the rest of them and could be interpreted as presenting a specific situation. In question 56, the subject *dzieci korzystające z książek (*children using books*)* the participle could be understood as an element that identifies the agent. In question 12 and 25, the instructions or advice is addressed to a receiver, who asked for them ('you' is used throughout the chunk in both cases), which may suggest that we are in fact talking about a specific problem of an individual, rather than a generic rule. Even if it is not the case that such interpretations were made, it is clear that including specificity in the model would improve the performance – 6 errors would be corrected.

## 8.4 Discussion

What the error analysis has shown are certainly the particular limitations of the NDL based aspect-concrete model, which are a direct result of how 'naive' the Naïve Discriminative Learning algorithm is. First of all, it could not take into account all the complexity of what is present in the experience. For instance, it did not have access to the type of modality, relative frequency and was ignorant of the presence of other verbs in the sentence, all of which seem to be important dimensions that should be taken into account. This, of course, is a result of the choices and decisions made by the author, which– while justified– can be changed. Nonetheless, the failures of the model bring to light the complexity of language knowledge. Learning might be simultaneously happening on different levels of representation (abstract, semi-concrete and lexical) and speakers are learning different things (collocations, grammatical categories, frequencies etc.). Moreover the outcomes of those 'learnings' may compete with each other when it comes to actual behaviour – the more 'abstract' forces may pull the speakers towards one way of communicating (e.g. to use perfective in the past), but knowledge of other things (e.g. relative frequencies) might push them towards another (e.g. using imperfective).

Secondly, because of the way it works, the model treats all learning events equally and is insensitive to conditional dependencies between the cues. As a result, the weights that might turn out to be predictive in a very specific context are driven down by the importance of other, more generally reliable cues. On a psychological level this would mean that learning is more structured and can be fine-tuned to the problem at hand. For instance, having learned that perfective is the default choice in the past, the speakers might be surprised to encounter imperfectives in the same tense. This surprisal might lead to more learning but the problem is now narrowed down to the past tense only, as this is the context where the predictions fail and where the speakers need to tweak them. As a result of the learning process, they might learn, for example, that while the past tense usually goes with a perfective, the presence of a particular adverbial predicts an imperfective. We will explore this avenue in the following chapters.

Finally, we should also point out that the dimensions of experience we have just identified as potential additional cues, may in fact explain only the deviations from the main patterns of use. After all, despite the limitations we have just discussed, we were able to predict at least some of the linguistic behaviour of the speakers in the experiment using our model. What is more, many of the potential additional cues we have discussed in this chapter are either not very reliable – that is associated well but used infrequently (e.g. the temporal adverbials) – or limited to a particular use cases in particular contexts (e.g. the different levels of specificity in the context of the modals).

In the coming chapters, we will focus on the most reliable cues for aspectual use we have identified, and explore how they might be related to the semantics of the perfective and imperfective. We will also investigate the relevance of these dimensions for categorization and model how perfective and imperfective aspect as classes can emerge from learning the similarities between patterns of verbal usage.

# Chapter 9. Building aspectual classes

In Chapter 6, where we discussed the results of the choice experiment, we established the following two findings. First of all, we saw that the aspect-concrete model predicts the choices made by the speakers. Secondly, we saw that the effect of the lemma-concrete model, even though not significant, was similar to the aspect-concrete model. It seems then, that the trend is the same for both models, but there is an advantage in categorizing individual verbs into two aspectual categories. Therefore, we can say that the results suggest that aspect is a useful generalization.

These results fit well into a learning perspective. Establishing which cues are predictive is much more difficult when many outcomes compete for the same, limited set of cues. This is because the competition between the cues is much fiercer, especially if the outcomes share the same important cues. In such cases, for each cue and outcome pair, the delta will change frequently from positive to negative, because we will often observe the same cue with a different outcome. This will result in poorly identified cues, which is exactly what we observed for the lemma-concrete model. As we can see in Figure 6, not only not all verbs have a set of well-differentiated cues, but also the association weights rarely go above 0.2. There is a lot of room to the maximum strength of association level, which – as mentioned in Chapter 3 – is equal to 1 in the Naïve Discriminative Learner.

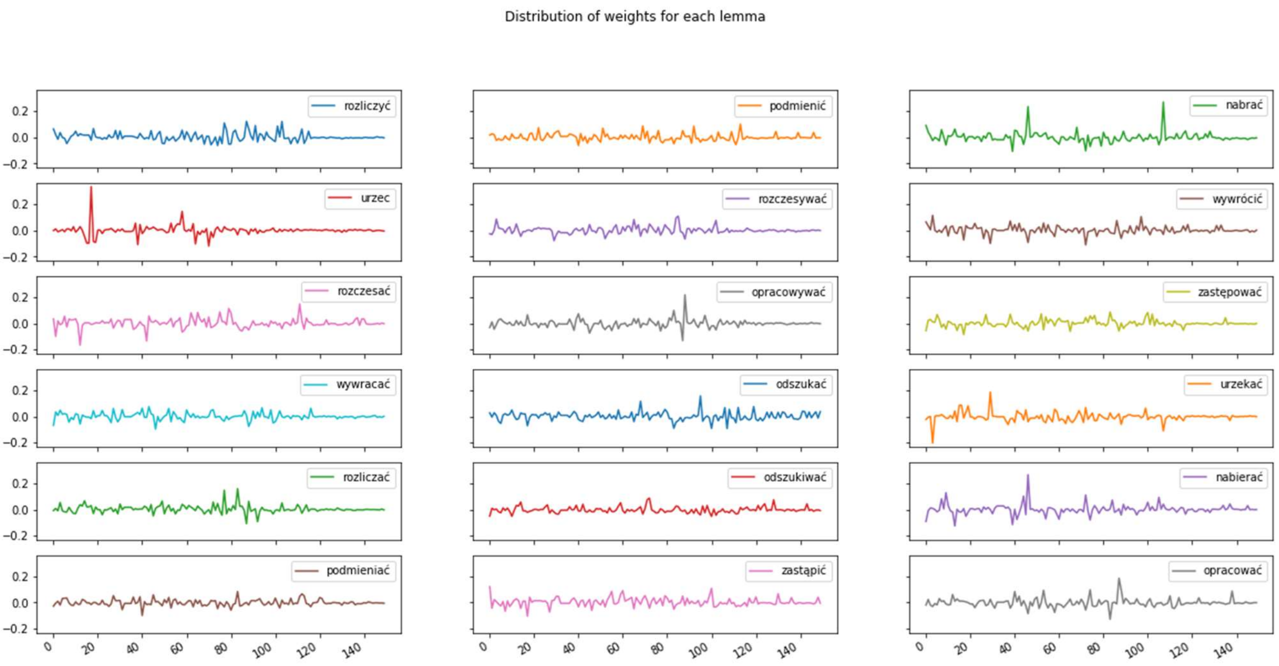


*Figure 6 The distribution of weights for each lemma*

On the other hand, learning becomes easier if we consider fewer outcomes, as in category learning. If predictive cues are rarely shared between the categories, but are shared between the members of each category, then the members will 'work together' to establish the relationship between a given cue and the category to which they belong. Let us consider an example, and say that we have a cue X and two items – A and B, that belong to category z. If we treat A and B as separate outcomes, each occurrence of B and X together will be treated as negative evidence for A and its association with X:

A X – the weight (AX) goes up

B X – the weight (AX) goes down

However, if we take category z as an outcome, then each time either B or A occur with X, the association between z and X is increased:

(z) A  X – the weight (zX) goes up

(z) B X  – the weight (zX) goes up.

This little thought experiment, combined with the results of the aspect-concrete model and the experiment discussed in Chapter 6 suggests that the members of each aspectual category do in fact share important cues. Otherwise, it would be possible to predict each choice of a verb using the lemma-concrete model, where the predictions were based only on the usage patterns of each individual verb. However, the fact that lemma-concrete model and aspect-concrete model both had the same direction of effect, suggests that there are still important differences between imperfective and perfective verbs as classes. There is enough information in the usage patterns of the individual verbs to push the choice in the right direction, but not enough to make the effect of the lemma-concrete model significant.

The question remains, however, how a generalization such as aspect might emerge. As we discussed in the theoretical chapters, usage-based theory assumes that the abstract is always based on the concrete. In other words, what speakers know about categories must be related to what they know about individual members – the linguistic categories must emerge from learning how to use individual verbs. In the following sections, we will attempt to determine the relationships between the usage patterns of the verbs and the categories of aspect. In particular, we will model whether the verbs can be grouped into perfective and imperfective, based on their associations with the set of cues we trained the models on. We used the weight matrices obtained by training the NDL models because the association strengths, as we argued in Chapter 3, provide a better representation of usage patterns than, say, raw frequencies of co-occurrence. This is because while the weights are necessarily based on how often the target appears with a given cue, they also reflect the influence of the co-occurrence with other cues. Therefore, we treat the output of the models as an approximation of the knowledge of the usage of the individual verbs in our sample that speakers may form during learning. The question is, however, whether that knowledge is enough to build a concept of aspectual categories.

## 9.1 K-means clustering

The first analysis we conducted was based on the weight matrix from the lemma-concrete model. As a quick reminder, this matrix consists of a list of outcomes – verbs and their association weight for each of the 'concrete' variables we annotated for. The set of all the weights for a given verb is a distributional vector, which we used to calculate the similarity between the verbs. The first modelling technique we used to group – or to use machine learning nomenclature – cluster verbs together was k-means clustering.

K-means clustering, like other clustering methods, is an example of an unsupervised learning algorithm. This means that, unlike in other methods we used elsewhere in the dissertation, the algorithm is not provided with the outcome labels. Instead, it tries to arrive at a solution on its own. In our case, this means that we will not be using the perfective and imperfective labels to inform the algorithm of which groups each verb should belong to. Instead, k-means will try to group the verbs solely on the basis of the distance between the vectors of the verbs.

The way the k-means algorithm works is conceptually quite simple, and it is based on two main assumptions. The first one is that the centre of a cluster is the arithmetic mean of the points that belong to the cluster. The second is that each point should be closer to the centre of the cluster it belongs to than to the centre of any other cluster. The algorithm assigns the datapoints to clusters in such a way that both assumptions are satisfied. This is done by randomly selecting a centroid – a cluster centre candidate – in the first step. Then, the points are assigned to the centroid that is closest to them, which gives a cluster of points. In the next step, the mean of the distances of all the points that belong to a given cluster is calculated. The centroid is then moved to that mean point. Once we have moved the centroids, the next step involves reclustering, and each point is again assigned to the centroid that is closest. The mean is calculated again and the steps are repeated until the clusters are stable or until we reach the maximum number of iterations – the repetitions of the steps described above. Since the final solution depends on the random points we selected at the beginning, to find the solution that is most optimal globally and not just in a single run, the algorithm should be run multiple times, with different initial centroids. This repeated initialization is implemented in the sklearn.cluster package (Pedregrosa et al., 2011), which was used in the analysis discussed below.

### 9.1.1 K-means clustering using all variables

In this section we present the results of clustering performed using the whole lemma-concrete weight matrix and the KMeans function from the sklearn.cluster package (Pedregrosa et al., 2011) in Python 3.

As we can see in Table 26, clustering on the whole matrix yields unbalanced groups that do not resemble perfective – imperfective classes. In fact, members of the same aspectual pair are grouped together quite often. Cluster 1 contains four aspectual pairs: *podmienić* (switch.PERF) - *podmieniać* (switch.IMPF), *rozczesać* (comb.PERF)- *rozczesywać* (comb.IMPF), *wywrócić* (topple/fall.PERF) - *wywracać* (topple/fall.IMPF) and *zastąpić* (replace.PERF) - *zastępować* (replace.IMPF). In Cluster 2 we can find two other aspectual pairs: *opracować* (develop.PERF) - *opracowywać* (develop.IMPF) and *rozliczyć* (square/appraise.PERF) - *rozliczać* (square.appraise.IMPF). Therefore, the aspectual partners from only three pairs were put into different groups as we expected. However, out of these three pairs, Cluster 1 contains two imperfective verbs and one perfective verb: *nabierać* (gain/kid.IMPF), *urzekać* (charm.IMPF) and *odszukać* (find.PERF). Cluster 2 contains their counterparts, therefore we find there two perfective verbs and one imperfective: *nabrać* (gain/kid.PERF), *urzec* (charm.PERF) and *odszukiwać* (find.IMPF).

*Table 26. Clustering results using k-means on the whole weight matrix*

| Cluster 1 | Cluster 2 |
|---|---|
| nabierać (gain/kid.IMPF)<br>odszukać (find.PERF)<br>podmieniać (switch.IMPF)<br>podmienić (switch.PERF)<br>rozczesać (comb.PERF)<br>rozczesywać (comb.IMPF)<br>urzekać (charm.IMPF)<br>wywracać (topple/fall.IMPF)<br>wywrócić (topple/fall.PERF)<br>zastąpić (replace.PERF)<br>zastępować (replace.IMPF) | nabrać (gain/kid.PERF)<br>odszukiwać (find.IMPF)<br>opracować (develop.PERF)<br>opracowywać (develop.IMPF)<br>rozliczać (square.appraise.IMPF)<br>rozliczyć (square/appraise.PERF)<br>urzec (charm.PERF) |

This result is perhaps not surprising, given that we consider all dimensions of the verbs' usage we annotated for. We could expect that the verbs that are the closest to each other belong to the same aspectual pair, because their meaning, syntactic requirements and contexts should

– in most cases – be the most similar. After all, it is possible to change the aspect of the verb in the sentence without changing anything else.

We would like to argue that these results point to a very important fact. Namely, categorization is a dynamic process and what is considered 'the same' depends on what the reason for categorizing is. If we were asked to group objects in a room together, there would be many solutions available to us at the same time. For instance, we could focus on colour and put together all white and green objects in the room. Another way of clustering objects could be related to their function – the couch and the chairs would go together because we can sit on them, whereas the coffee table and the chest of drawers would form their own category, because they are not used in the same way. Similarly, verbs simultaneously belong to many categories. We could try to distinguish action verbs from verbs denoting states, transitive from intransitive etc. What is more, just like in the case of categorizing the objects in a room, different dimensions of the same items would become important, depending on which categories we want to divide them into. Therefore, we could hypothesize that the verbs could be successfully grouped into perfective and imperfective classes if we take into account only a subset of all the variables.

Fortunately, we already established which dimensions are most useful when predicting aspects. The aspect-concrete model was trained to predict only two classes on the same set of concrete variables, and we already discussed which of them contributed most to the predictions in Chapter 7. We will now test whether the verbs can be clustered into perfective and imperfective, based on these variables exclusively.

### 9.1.2 K-means clustering using aspect-relevant variables

First, we start out by filtering the columns in the weight matrix obtained by training the lemma-concrete model. We kept only the cues that contributed most for predicting aspect in the aspect-concrete model. This subset of variables consisted of the following: tense and a set of cues describing the elements of the verb phrase: CxWord.zacząć, CxWord.zostać, CxWord.być, Cx.aux, Cx.modal, Cx.phasal. Just like before, we used the KMeans function from the sklearn.cluster package.

As we can see in Table 27 below, the clusters obtained when we used only aspect-relevant variables again divide verbs along perfective-imperfective division line and all the verbs are classified correctly. Cluster 1 contains only imperfective verbs while Cluster 2 contains their perfective counterparts. These results seem to confirm the point we arrived at earlier – not all the dimensions are relevant for all possible classes and groups. Instead, what matters for aspect might be different from what matters for transitivity or any other classification of verbs we might want to use. Speakers may be able to categorize verbs into aspectual classes on the basis of what they know about the usage of the verbs, provided that they also know which dimensions are relevant. In next sections we will discuss how they might learn which dimensions of verb usage are relevant for aspect. Before that, however, we will cluster the verbs again, using a different method – hierarchical clustering.

*Table 27. Clustering results using k-means on the subset of the aspect-relevant variables*

| Cluster 1 | Cluster 2 |
| --- | --- |
| nabierać (gain/kid.IMPF) | nabrać (gain/kid.PERF) |
| odszukiwać (find.IMPF) | odszukać (find.PERF) |
| opracowywać (develop.IMPF) | opracować (develop.PERF) |
| podmieniać (switch.IMPF) | podmienić (switch.PERF) |
| rozczesywać (comb.IMPF) | rozczesać (comb.PERF) |
| rozliczać (square.appraise.IMPF) | rozliczyć (square/appraise.PERF) |
| urzekać (charm.IMPF) | urzec (charm.PERF) |
| wywracać (topple/fall.IMPF) | wywrócić (topple/fall.PERF) |
| zastępować (replace.IMPF) | zastąpić (replace.PERF) |

## 9.2 Hierarchical clustering

One important characteristic of k-means clustering is that the number of clusters has to be defined in advance. This has to be done, because we need to know how many centroids we are supposed to fit. However, this means that we explicitly told the algorithm how many clusters it should expect. In the modelling attempts above, we set the k – the number of clusters to two, as this was how many groups we expected theoretically. This choice however, could have influenced the results, and we might have found two clusters exactly because we were looking for two clusters. In order to make sure that this was not the case, we decided to confirm the findings using a different clustering method. While there are ways of determining optimal number of clusters for k-means, we decided to switch to hierarchical clustering as it also allows us to visualize the data. Therefore, in addition to gaining more confidence in the solution we arrived at before, we will also be able to discuss the similarities of the verbs in more detail. It is worth noting that there are two ways of performing hierarchical clustering: divisive clustering and agglomerative clustering. Since we will be using the more common agglomerative clustering, we will only discuss the details of this approach.

Just like k-means clustering, agglomerative hierarchical clustering is an unsupervised learning method. Therefore, the algorithm we used did not know which verbs were perfective and which were imperfective. However, the way the clusters are obtained is different. Instead of randomly selecting centroids and moving them to find the optimal solution, hierarchical clustering builds clusters from the bottom up, by pairing most similar items together. In practice this means that we start with clusters that each contain only one item, and the number of clusters is equal to the number of items. In the next step, we pair clusters (items) that have the smallest distance to each other. In the next iteration, we find more pairs. Since each item is treated as a cluster, at this point a pair can contain individual items, a small cluster and an item, or two smaller clusters. We keep pairing items until we end up with only one pair at the top, called the root.

There are a few ways of calculating distance for hierarchical clustering, called linkage functions. The method we used for our case is called Ward's method or Minimum Variance Method, and it works in the following way. First, at each step, all combinations of clusters are considered – that is, we calculate the distance between all possible pairs. For each merged cluster candidate, we calculate a centroid – a mean point. Then, we measure the distances of all the points in the cluster to the centroid, square and sum them. This gives us a measure of variance, and tells us how much the points deviate from the cluster candidate mean. In the last step, we compare the variance for all possible pairs and merge the one that has the lowest variance. Even though this method is computationally intensive, it is a popular choice for calculating distances between clusters. One of the reasons for its popularity worth mentioning here is that it is less sensitive to outliers than simpler single link or complete link measures, that consider only the distances between the closest or the most distant points.

The optimal number of clusters can be determined by examining a dendrogram – a tree diagram which is a visual representation of the similarities between the clusters. The vertical lines on the plot show us which clusters were linked together and the horizontal lines indicate the distance between them. For instance, we can see that *zastępować* (substitute.IMPF) and *podmienić* (switch.PERF) in figure 7 below were paired first, and the distance between them was roughly 0.37. We can also see that the red cluster and the green cluster were linked into one at a distance of roughly 0.59. When it comes to selecting the number of clusters, a common practice is to draw a horizontal line where the lines in the vertical lines in the dendrogram are the longest. The amount of the lines that are 'cut' by the horizontal line gives us the number of clusters we should select.

### 9.2.1 Hierarchical clustering using all variables

In this section we present the dendrogram representing the clusters obtained when we used all the variables from lemma-concrete weight matrix. To cluster the verbs, we used the scipy.cluster package in python, set the linkage function to Ward's method and used the Euclidian distance as the distance metric.

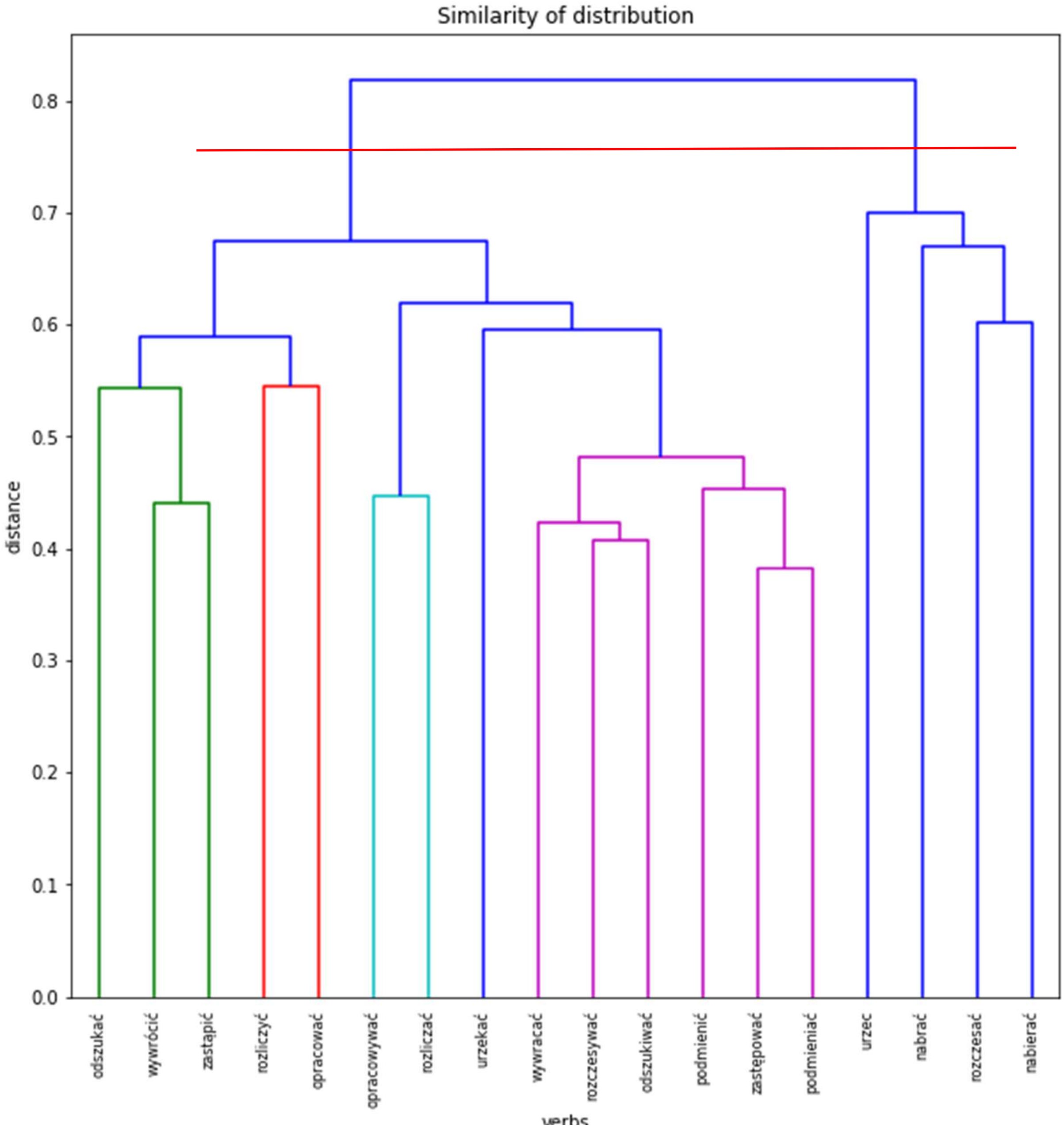


*Figure 7 A dendrogram representing a hierarchical clustering solution using the full set of variables*

As we can see, even though the best choice of the clusters splits the verbs into two groups, their borders do not align with the perfective-imperfective division. The smaller cluster contains only four, rather dissimilar verbs, three of which – *nabrać* (gain/kid.PERF), *udać* *(*succeed/pretend.PERF) and *rozczesać* (comb.PERF) are perfective. Surprisingly, the cluster also contains *nabierać* (gain/kid.IMPF), the imperfective counterpart of *nabrać*.

On the other hand, the larger cluster contains a mix of both perfective and imperfective verbs, and can be further divided into four subgroups. Some of the subgroups could be explained by the semantic similarity of the verbs. For instance, the red cluster contains two perfective verbs: *opracować* (develop.PERF) and *rozliczyć (square/appraise.PERF)*, both of which are often used in more formal contexts related to finance, politics or science. Similarly, the light blue cluster contains the imperfective counterparts of the verbs from the red cluster – *opracowywać* (develop.IMPF) and *rozliczać* (square/appraise.IMPF). Therefore, we can say that at least these two clusters were separated based on their aspectual and semantic similarity. However, the picture is much more complicated when it comes to the remaining two clusters. While the green cluster contains only perfective verbs, the purple one contains both perfective and imperfective. It is also very difficult to find any semantic similarities that the verbs in any of these two clusters share. In summary, we did obtain two clusters of verbs, but these groupings are difficult to explain and contrary to what we expected, they do not divide the verbs into perfective and imperfective.

### 9.2.2 Hierarchical clustering using selected variables

Similarly to what we have done for k-means clustering on a subset of variables, we started by filtering the lemma-concrete weight matrix. Again, we kept the following variables: Tense, CxWord.zacząć, CxWord.zostać, CxWord.być, Cx.aux, Cx.modal, Cx.phasal. Next, just like before, the verbs were clustered together using hierarchical clustering from python's scipy.cluster package, Ward's method as linkage function and Euclidian distance as the distance metric. The dendrogram in Figure 8 visualizes the results.

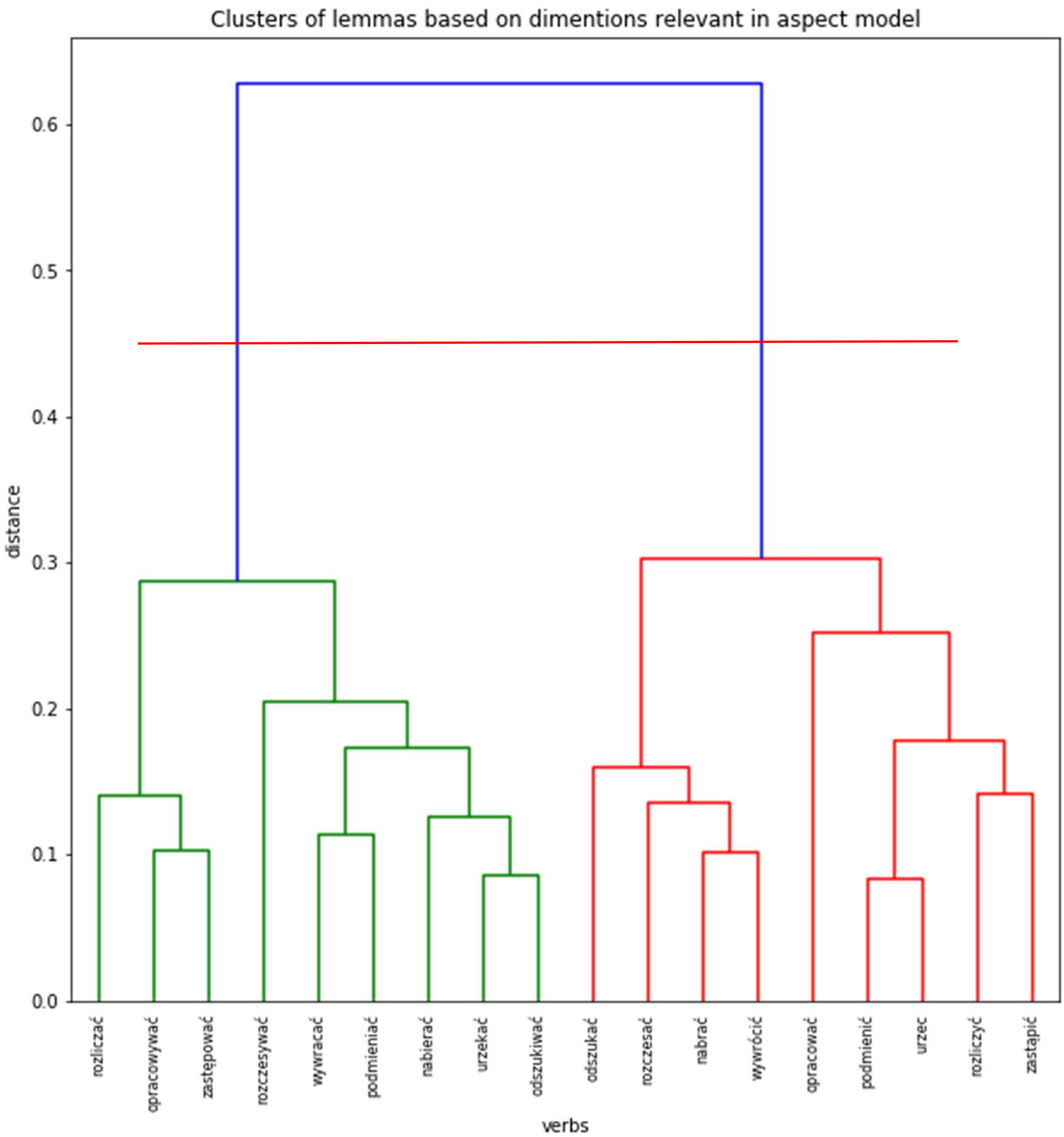


*Figure 8 A dendrogram representing a hierarchical clustering solution using a subset of variables*

As we can see, the two main clusters divide the verbs into perfective (red cluster) and imperfective (green cluster) without any errors. While subgroups can be distinguished, the two main clusters are clearly separated, as the distance between them is large compared to the distances of the subgroups within them. This result confirms the previous findings obtained using k-means clustering and further strengthens our conclusion that categorizing verbs into aspectual classes is indeed possible from the bottom-up, based on what the model learned about the usage patterns of the verbs. The important caveat here is that not all dimensions of verb usage are relevant. Instead, obtaining this solution is possible when we focus only on the dimension that were important for predicting the aspect.

## 9.3 Variable importance

A question we can ask at this point is: which of the variables in the aspect-relevant subset contributed the most to our clustering solution? After all, the variables we selected were the most predictive of aspect as a category. It is possible that their importance is different for the individual verbs and the similarities between them. To give an example, the auxiliary verb be might be one of the most important predictors of the imperfective aspect, because it only appears with an imperfective verb in the future compound tense. However, this dimension might not necessarily important for all the imperfective verbs, as some of them were never used in this construction in the sample. Therefore it is worth inspecting the associations of each individual verb with the subset of the variables that gave us the best clustering solution. This will tell us which of the cues were most consistently important for distinguishing between perfective and imperfective.

The parallel coordinates plot in figure 9 below illustrates the differences in the importance of each variable for each group of verbs. To prepare it, we took cluster assignments obtained from the k-means clustering (the assignments we obtained using the agglomerative hierarchical clustering were the same) and labelled each verb accordingly. For ease of interpretation, we renamed the clusters: Cluster 1 is called imperfective, as it contains only imperfective verbs, and Cluster 2– perfective. Then, using the parallel coordinates function from the pandas package, we generated the plot. Each verb is represented by a line, and the colours signify to which cluster a given verb belongs. The y axis represents the association weights between the verb and each variable.

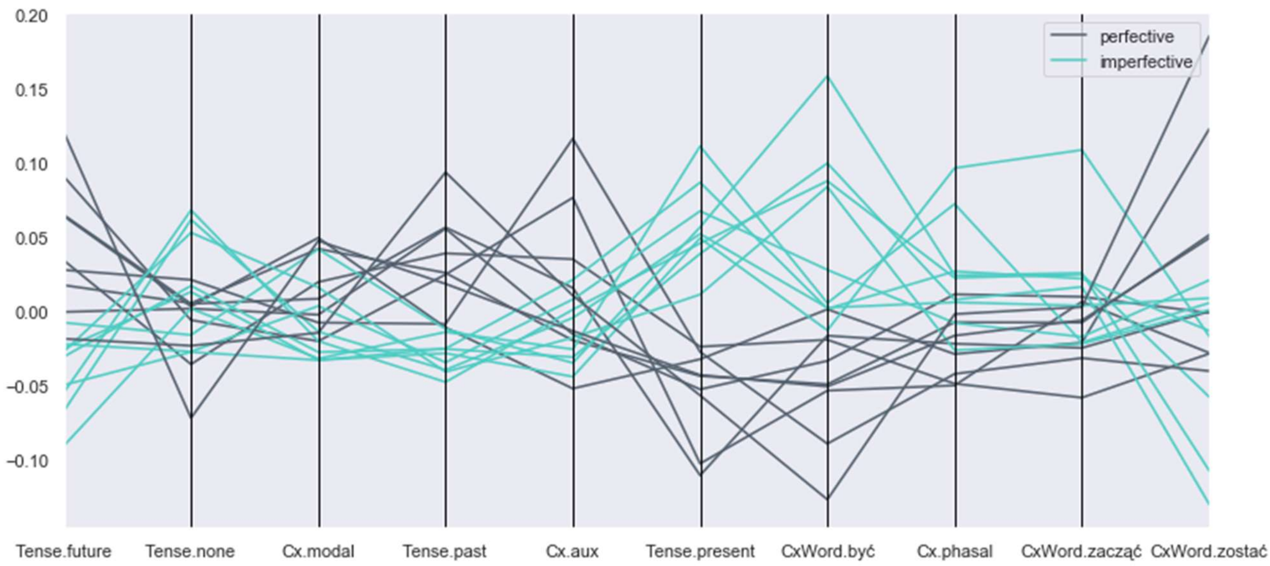


*Figure 9 The importance of each variable for clustering*

As we can see, the dimensions that divide the verbs most clearly are the present tense, the past tense and the presence of the auxiliary "be". For these dimensions, the associations are either positive or negative for most of the verbs belonging to a given aspectual class. We can see for instance, that all imperfective verbs had a positive association with the Tense.present dimension. Another dimension that stands out clearly is the presence of the auxiliary "be", which has a negative association weight with only one of the imperfective verbs. This is perhaps not surprising. As we stated before, only the imperfective verbs can be used in the present tense, and when used in a compound future tense they require the auxiliary "be". Nonetheless, because those dimensions are so uncontroversial, it is important that they were correctly identified.

We can also see that the majority of the imperfective verbs had negative associations with the future tense – this combined with the weights for the auxiliary "be" tells us, that generally, perfective verbs tend to be used in the future more often than the imperfective, but if the imperfective is used, they are accompanied by the auxiliary "be". What is more, we see that the majority of the perfective verbs are positively associated with the Tense.past dimension. This mirrors the results of the aspect-concrete model training and suggests that even though both perfective and imperfective verbs can be used in the past, there is a clear tendency for perfective to appear in this context more often.

As for the other auxiliary verbs, or just the presence of an auxiliary in general, the situation is more complicated and they turned out to be truly relevant for only a handful of verbs. We can also note the mixed lines at the Tense.none and the modal dimensions. This might be due to the fact that those two variables can be further divided into more detailed categories. While we were not interested in such a fine-grained level in this study, it might be worth disentangling these dimensions in future analyses, especially that there exist studies showing for instance patterns of different usage patterns for perfective and imperfective in different modalities (Divjak et al., 2015).

To sum up, the analysis of the parallel plot confirmed that tense seems to be the most important dimension for aspectual use, and there is a tendency for perfectives to be used more in the past and in the future. This is certainly an interesting conclusion – out of all the variables we annotated for, one of the most obvious ones – tense – turns out to be the most important.

Before we conclude this chapter, it is worth highlighting that the clustering analysis provided additional evidence supporting the aspect-concrete model. Not only can the model be used to predict the behaviour of the speakers, but also the variables that the model learned to be the most predictive of aspect can be used to successfully group verbs into perfective and imperfective. This convergence gives us more confidence in the conclusions we arrived at earlier.

## 9.4 Discussion

In the previous chapters we discussed the results of the corpus-based modelling and judgement task, which showed that the behaviour of the speakers can be predicted by the model that assumed a category of aspect as outcome. This suggests that speakers might form some knowledge of perfective and imperfective and their respective distributions. However, the question still remained of how such knowledge may emerge, since speakers are seldom – at least not until formal schooling– explicitly told about the existence of such categories. Instead, what they have at their disposal are the examples of usage of each individual verb.

According to usage-based theory, anything that is abstract is linked to – and based on– what is concrete. Therefore, in order to be able to distinguish aspectual categories, speakers must start from the bottom-up. First they must learn about the usage of verbs and then they would able to categorize verbs into perfective and imperfective aspects by focusing on the similarities along the relevant dimensions. In this chapter we tested whether verbs can be grouped into perfective and imperfective based on the what the model learned about the usage patterns of the verbs.

The results show that at least some of the knowledge of usage patterns speakers may acquire in the process of learning Polish could be employed to form more abstract categories. We demonstrated that even though the lemma-concrete model did not have any access to the perfective and imperfective as a category, it was still possible to cluster the verbs into aspectual classes by comparing the similarities between their usage patterns that the model learned. As such, these findings fit well into usage-based theory, since they suggest that it is in fact possible to form aspectual categories in a bottom-up fashion.

We also saw that clustering on all variables we annotated for did not yield groupings that resemble aspectual categories. However, obtaining this result was much easier when only a subset of variables relevant for aspect was used. This suggests that not all dimensions are simultaneously equally important when it comes to categorization in language. The general notion of 'similarities' between the words as a starting point of abstraction hides the fact that items can form different constellations depending on what dimensions we take into account. It seems that learning which dimensions are relevant for a given purpose is an important part of the process of acquiring more abstract categories.

Importantly, in our modelling, both learning the patterns of associations between the verbs and the concrete cues and establishing which dimensions are relevant for aspect were based on the same learning principle. The underlying mechanism was the same for both levels of learning, and the only difference between them was the degree of abstraction, encoded in the outcome.

It is outside of the scope of this dissertation to answer the question of whether such two-level error-driven learning is in fact the way speakers form categories. It is worth noting however, that there are studies showing that category learning in both humans and non-human animals can be modelled using the principles of associative learning (e.g. Gluck & Bower, 1988; Wasserman et al., 2015). For the time being however, we are satisfied with demonstrating that it is possible to divide the verbs into perfective and imperfective based on what the models have learned in our simulations, as this shows that both the patterns of use and the aspect-relevant dimensions can be learned on the basis of the input alone.

# Chapter 10. Zooming in: modelling aspect in the past tense

The work presented in previous chapters allowed us to formulate a few important conclusions. First of all, we showed that the aspect-concrete model, trained on the aspectual usage patterns, could be used to model the linguistic behaviour of the speakers. In addition, we discussed in the previous chapter that the verbs can be clustered into their aspectual classes when the most predictive cues from the aspect-concrete model are taken into account. Taken together, these findings suggest that aspectual classes might be relevant distinctions that Polish speakers know and use. In addition, the strong relationship between tense and aspect led us to formulate a hypothesis that certain tense-aspect combinations can be considered “default”, or canonical. We discussed in Chapter 7 that while imperfective is used by default in present tense, the model also picked up on a strong bias for perfective to be used in the past and the future tenses. However, it remains to be demonstrated experimentally that speakers are also sensitive to these biases and indeed perceive the same tense-aspect combinations as canonical. If they are, we can expect that the use of non-canonical, less predictable patterns will lead to increase in processing difficulty.

However, as we mentioned in Chapter 7 canonicity is not deterministic. In other words, non-canonical uses are also possible, although less frequent. Because of that, they usually require some sort of contextual support, as shown for instance by Bergs (2010) and more recently by Romain et al. (2022). These findings are perfectly understandable in the light of the learning theory we assumed in this dissertation. Since the non-canonical forms are used less frequently, they are also more surprising. An inclusion of an additional cue would help the listener in processing the message that we uttered, as it would make the non-canonical form less unexpected. This of course, would have an effect on any experimental measure we choose to use in the experiment, potentially resulting in no differences between processing perfectives and imperfectives. Therefore, before we move on the study, we need to first posit the following question: what cues could provide such additional support for the non-canonical use of imperfective in the past and the future tenses?

Unfortunately, the models we trained so far do not allow us to answer this question. As we mentioned earlier, the Naïve Discriminative Learner does not take into account dependencies between the cues, and what it learned about the use of imperfective can be said to be global and independent of local contextual constraints. Therefore, in order to overcome these difficulties and establish whether there are any cues that predict the use of imperfective in its non-canonical tense-aspect combinations, we need to focus on a subsample of the training data. This way we will force the model to "zoom in" and try to distinguish cues that are particularly important in these non-canonical contexts only.

As we have seen in Table 22 in Chapter 7, the imperfective is used in the future tense extremely rarely. Such distribution means that we do not have enough examples of future imperfective sentences to prepare training, test and validation sets. For this reason, we decided to exclude future tense examples completely. In addition, as discussed elsewhere, the "none" category captures a variety of phenomena, such as infinitives in modals, participles or imperatives that might be worth looking at separately. Therefore we will focus exclusively on the use of aspect in the past tense, where despite a huge bias for the perfective, the imperfective is still a valid option.

## 10.1 Training

The training was conducted in the same way as we trained the other NDL-based models. First, we prepared the dataset set by keeping only the subset of the past tense chunks, and then removing the abstract variables and the variable of tense. Then, we split the dataset into training, test and validation sets. The training set contained 372 examples and the validation and test sets contained 20 examples each: 10 for each of the aspects.

The Naïve Discriminative Learner was used to train a model with imperfective and perfective as outcomes and the values of the concrete variables we annotated for as cues. Table 28 below summarizes the model.

*Table 28. Cues and outcomes used for each of the models trained to predict aspectual labels or lemmas*

| **model** | **cues** | **no. of cues** | **outcomes** |
|---|---|---|---|
| pastAspect-concrete | values of the contextual variables for each chunk (e.g. AgentNumber.plural) | 85 | perfective, imperfective |

The model was again trained for 500 epochs, or repeated runs, randomizing the order of the training examples before each run. Just like before, the learning rate was set to 0.0001.

## 10.2 Results

We evaluated the model by comparing the predictions (i.e. perfective or imperfective) to the aspect that was originally used in the discourse chunk. The results showed that the model completely failed to correctly predict imperfective aspect – the model made a correct prediction only for the chunks where perfective was used originally. Interestingly, there was a significant discrepancy between the performance of the model on the training set, compared to the performance on both the validation and test sets. The accuracy of prediction in the training set was high (86.6%) but the model performed at chance level on the test set (50%) and slightly above it on the validation set (55%).

Further investigation revealed that this poor performance on the test set was due to the fact the model always predicted perfective. In other words, it seemed that the model was overfitted and learned that the best option is to always predict perfective. It is perhaps not surprising given the class imbalance – after all, the whole past tense subset contained only 75 imperfective examples, compared to 337 perfective examples. However, the disparity between the performance on the test set and the training set cast a shadow of doubt on the overfitting explanation.

To explore the issue further, we listed the cues that the model learned to be the most predictive for both the perfective and the imperfective outcomes. These cues are presented in Table 29 below.

*Table 29 The strongest negative and positive cues per aspect – training outcomes of the pastAspect-concrete model.*

| **strongest positive cues** | | | | **strongest negative cues** | | | |
|---|---|---|---|---|---|---|---|
| **perfective** | | **imperfective** | | **perfective** | | **imperfective** | |
| *cue* | *weight* | *cue* | *weight* | *cue* | *weight* | *cue* | *weight* |
| AspectualTrigger.none | 0.126465 | AspectualTrigger.imperf | 0.141639 | AspectualTrigger.imperf | -0.126383 | PatientAnimacy.inanimate-abstract | -0.109390 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PatientAnimacy.inanimate-abstract | 0.120327 | Adverbial.frequency | 0.124386 | Adverbial.frequency | -0.105636 | Adverbial.none | -0.093369 |
| Adverbial.none | 0.117247 | AgentPerson.third | 0.098502 | AgentAnimacy.inanimate-substance | -0.070835 | AspectualTrigger.none | -0.075055 |
| AgentPerson.first | 0.100319 | AgentAnimacy.inanimate-substance | 0.076218 | AgentPerson.third | -0.065051 | AgentPerson.first | -0.074152 |
| PatientGender.none | 0.089876 | PatientAnimacy.inanimate-object | 0.063315 | PatientAnimacy.inanimate-object | -0.052122 | PatientGender.none | -0.065830 |

What these weights are showing us is that perfective is often used in first person and in sentences that do not have an indirect object or any adverbials. Interestingly, the relatively small association weights between the perfective outcome and each of the most predictive cues suggest that despite having many learning examples, the model was not able to find anything particularly predictive of the perfective in the past. As for the negative cues, we can see that these are the same as the top most positively associated cues for the imperfective. Given that we only had two classes, competing for the same cues, this is an expected outcome.

As for the imperfective aspect, we see that the weights are also small, but we need to remember that there were far fewer training examples for that class. The fact that the model was able to distinguish them anyway is quite an interesting outcome. What seems to be the most predictive for the use of the imperfective aspect in the past are the adverbials of frequency and other imperfective aspectual triggers. This is an interesting finding, which highlights the strength of a learning approach. Since aspectual triggers occur infrequently, an analysis based on raw frequency counts would suggest that these cues are not particularly useful. However, since we used the association strengths, we were able to identify them as good predictors, in line with what has been described in literature. The other positively associated cues suggest that imperfective is quite often used with third person inanimate agents and in sentences with indirect objects. However, this is a very tentative conclusion given that there were few training examples and the weights are really small. Finally, just like for the perfective, the negative cues are the same as the positive cues associated with the other aspect.

Having conducted this analysis we went back to the test set and discovered that none of the imperfective examples contained either adverbials or any imperfective aspectual triggers. This explains the poor performance of the model – if most predictive cues were not present there, the model chose the default perfective option.

To sum up, these results suggest that the perfective aspect can indeed be taken as a default, canonical option in the past tense. NDL learned that it should always choose perfective, unless the discourse chunk contains particular cues associated with the imperfective. This also supports the conclusions we arrived at earlier. First, this finding further strengthens the view expressed in Chapter 5 that, contrary to the early analyses of aspect, it is the perfective that seems to be the unmarked member of the aspectual opposition. Secondly, we also said that the use of the non-canonical option should often be signposted by additional cues, which is exactly what we see for the imperfective. As we discussed in the previous chapters, temporal adverbials and aspectual triggers are such cues for aspect.

## 10.3 Discussion

In this chapter, we looked into what cues can be said to guide the aspectual use in the past tense. Based on the previous findings in this dissertation, and the theoretical arguments of cognitive linguists, we hypothesized that the perfective should be the default option in the past. At the same time, the imperfective should be a non-canonical option, and as such it would require additional contextual support in the form of additional cues that would help the speakers process the utterance.

In order to test this hypothesis, we conducted a new learning simulation on a subset of data and discovered that the model predicts the imperfective in the test set very poorly. The analysis of the most predictive cues revealed that this poor performance was a consequence of what the model learned to be the most predictive cues for that aspect. NDL learned that the imperfective should be used with either an adverbial of frequency or some other imperfective aspectual trigger, and yet none of the examples in the test set contained such cues. Therefore, the model failed to make correct predictions and defaulted to the perfective aspect.

These results confirm the conclusions we arrived at earlier and suggest that the perfective is indeed the default or the canonical option in the past tense, whereas the non-canonical use of the imperfective calls for some form of additional contextual support. However, this is not to say that imperfective cannot be used in the past without any additional temporal expressions. On the contrary – we have seen for instance, that none of the examples in the test contained such cues despite being correctly formulated chunks of discourse.

It is possible that the choice was guided by other, possibly even non-linguistic cues, which we did not annotate for. Out of practical necessity, our insight into the discourse and the decisions made by the speakers was limited to a window of text. On the other hand, even though additional cues might be helpful, we must note that speakers do not have to use triggers in order to 'justify' their choices. After all, language gives them a lot of freedom, and they might simply choose to not be communicatively helpful. However, what have demonstrated by the learning simulation in this chapter is that while the speakers do not have to use temporal cues, they are more likely to do so when using the imperfective.

Canonicity of use is also worth discussing in the context of the debate on aspectual meaning. As we have seen, the majority of the attempts at defining the meaning of the imperfective and perfective revolves around the notions of ongoingness, incompleteness, being in progress, versus, completeness, boundedness and closure. While not all cases of aspectual use fit those definitions equally well, these approximations certainly reflect the intuitions that linguists and native speakers have. The current results indicate that these intuitions are not unfounded. On the contrary, if we assume that the imperfective aspect is used prototypically in the present tense, while the perfective is used in the past, then it is easy to see how the semantics of tense and aspect overlap in such cases. In other words, the 'ongoingness' of the imperfective can be related to the 'happening now' of the present tense, and the 'completeness' of the perfective can emerge from the 'happened earlier' of the past tense.

What is more, the error-driven learning principle can help explain how such an overlap might emerge. To illustrate its emergence better, let us consider a simplified example with the following set of cues and outcomes. We will treat each verb as a combination of only two cues – aspect and tense. The outcomes we will consider – before and now– will also be binary. Table 30 below describes examples of possible learning steps and the corresponding changes between the associations of the cues and outcomes. Each time we encounter an imperfective present verb the association weight between the 'present' cue and the outcome 'now' will go up. But this will be also true for the 'imperfective' cue and 'now' as an outcome. On the other hand, each occurrence of a past (or future) imperfective verb will decrease the association weight between 'imperfective' and 'now' while not affecting the association between 'present' and 'now'. Similarly, each time we encounter a past perfective verb, the association between 'before' and two cues – 'past' and 'perfective' will increase.

*Table 30. Learning steps illustrating how the associations between tenses, aspects and temporal semantics change for each encounter of a verb*

| verb | cues | outcomes | change |
|---|---|---|---|
| pisze (write.IMPF.3SG.PRESENT) | imperfective, present | now | The association weight between imperfective and now increases.<br>The association weight between present and now increases. |
| pisała (write.IMPF.3SG.FEM.PRESENT) | imperfective, past | before | The association weight between imperfective and now decreases. |

| | | | |
|---|---|---|---|
| | | | The association weight between imperfective and before increases. |
| napisał (write.IMPF.3SG.MSC.PAST) | perfective, past | before | The association weight between perfective and before increases.<br>The association weight between past and before increases. |

Given the distributional patterns of verbal use we have determined previously, we can expect the following learning outcomes. First of all, tenses as cues should have very high association weights with their temporal 'meanings'. That is, 'past' should be strongly associated with 'before' and 'present' with 'now'. However, because of the strong bias for perfective to appear in the past and imperfective in the present, aspects might not be well separated from the temporal meanings. Therefore, they will also be strongly associated with the notions of 'before' and 'now' respectively. The weights should reflect that these are not perfect cues, and as such the aspects should be associated with temporal 'meanings' to a slightly lesser extent than the tenses. Nonetheless, the frequency of co-occurrences of tense and aspect in the usage would ensure that the associations between them are still very important.

In this way, canonicity of use might drive the emergence of the notions of aspectual meaning. The usage patterns may lead to learning that the temporal semantics we usually attribute to tenses cannot be well separated from aspects and might 'bleed into' the aspectual uses in their respective non-canonical tenses. And while it might not be necessary, at least to native speakers, to make a distinction between the aspectual meanings and tense meanings in the canonical cases, the non-canonical uses might require some reinterpretation and semantic detailing. 'Now' is not compatible with the notion of 'past'. 'Ongoingness', on the other hand is. Similarly while 'before' does not fit well into the notion of 'future', 'completeness' does.

As for the future tense, that we left out from the discussion so far, there are reasons to believe that the perfective might be a more canonical aspect in this context than the imperfective. First of all, it is worth noting that similar skewness in distribution and especially the low frequency with which imperfective/progressive forms are used in future tenses can also be found in other languages. Romain et al. (2021) looked at the distributions of tense/aspect combinations in the British National Corpus and show that Future Progressive forms make up only 0.11% of over 7 million sentences, whereas Future Perfect Progressive was only identified 30 times. Dickey (2016), who also points out similar infrequency of imperfective aspect in the future in Russian, offers an interesting perspective on why this is. He suggests that these differences are linked to the way we think about the past and the future. Since the past has already happened, there is more to explore – investigating and enquiring about various relationships between events in the past makes more sense, and often involves making contrasts between the events that occurred and the ones that were happening in the background. On the other hand, the future has not happened yet, which means that there aren't any relations to explore and discuss yet. Instead, we tend to speak of future events in terms of goals that will be achieved and plans that will be realized, which entails the use of perfective rather than imperfective forms.

To sum up, we can say that the modelling we have done so far indicates that the perfective is used canonically in the past tense and the imperfective in the present. However, the fact that such usage patterns can be learned from the input, does not entail that all aspectual choices can be explained by the distribution. Speakers might deviate from these patterns, and such deviation conveys meaning. After all, the use of an imperfective verb in a past sentence does results in a different interpretation of the same sentence than it would when a perfective verb was used. It might be that these non-canonical uses are the cases where construal comes into play; the speakers are free to choose which conceptualization and of the event they want to present, and which aspects of the event they want to highlight.

Therefore, the distributional patterns are not deterministic. However, the strong biases in the frequencies of co-occurrence we have demonstrated might help explain what semantic notions are conveyed by such deviations from the canonical patterns and how they can be learned from usage. The usage examples of the imperfective in the present tense might drive its interpretation as indicating 'ongoingness' in the past, since they strengthen the connection between the morphological markings of the imperfective and the temporal semantics of the present. On the other hand, the canonical uses of the perfective aspect strengthen its association with the temporal semantics of the past tense.

We must note that these conclusions are based mainly on the learning simulations supported by theoretical considerations. However, the methodology we assumed in this dissertation calls for an experimental validation. The next chapter describes an experimental study we conducted to further investigate the relationships between tenses and aspects.

# Chapter 11. Judgement task

In the previous sections we discussed the relation between tense and aspect in the light of canonicity. The results we have obtained so far indicate that present tense and the imperfective can be treated as a canonical or a 'default' combination, whereas past and future tenses typically take perfective verbs. We also suggested that aspectual meaning might come into play in those utterances that deviate from the usual patterns of usage. For instance, in past tense sentences, there is no real need to try to disentangle past tense meaning from notions of completeness when the sentence contains a default perfective verb. On the other hand, an imperfective verb in the past context is less likely and therefore more surprising. It make sense to ponder why our interlocutor has chosen a non-default form and what their communicative intention was.

If the conclusions we reached so far hold, we should be able to find supporting evidence for them by looking at the linguistic behaviour of the speakers in an experimental setting. We might expect that since the deviation from the default, canonical patterns is less frequent, it should increase the difficulty of processing. In other words, we might expect that native speakers would require more time to react to sentences where imperfective verbs are used in the past tense, since this is a non-canonical combination. On the other hand, the processing of the sentences where perfective verbs are used in the past, should be faster[10].

Additionally, we can expect that the inclusion of an additional cue that predicts the aspect of the verb used later on in the sentence, should decrease the surprisal and, in turn, decrease the processing time as well. The corpus studies we discussed in this Chapter 4 (Koranova and

[10] It is worth noting that we cannot directly test a reverse of this set up, where imperfective verbs are used in their default, present tense. This is because perfective verbs are never used in the present tense, which means we would not be able to prepare stimuli containing non-default present-perfective combinations. As for the future tense our results have shown that it is also more strongly associated with perfective aspect, making future-perfective a default combination again.

Bermel, 2008; Janda and Reynolds 2019), and the models we described in the previous chapters suggest that the best candidates for such cues are temporal adverbials.

In this chapter we present a decision study designed to these hypotheses. We start however, by presenting relevant studies that justify the use of the behavioural measures we selected, as well as present evidence that non-canonical and less frequent variants are processed more slowly than their canonical counterparts.

## 11.1 Choosing measures: reaction times and error rates in linguistic studies

In this study, we will be using two measures – reaction times and error rates. In this section, we discuss the measures in more detail, focusing in particular on the evidence showing that the more predictable a word or a grammatical category it represents is, the faster participants react to it in a decision task.

Reaction time is the time required by participants to provide a response. This measure is one of the most commonly used in psycholinguistics (Baayen & Milin, 2010; Jiang, 2012). Based on the assumption that more complex structures or items require more time, reaction times are said to indicate the mental effort required to process an item. Perhaps the most well-known experimental paradigm involving reaction time as a measure is the lexical decision task. In this task, participants hear or see a string of letters and have to decide if it forms an acceptable word in a given language or not. In an early paper, Rubenstein, Garfield and Millikan (1970) show among other things that proper words are recognized faster that pseudowords and that frequency plays an important role – the more frequent items are recognized faster than less frequent ones.

This design can be extended to phrases and full sentences. For instance, Clifton and Frazier (2004) investigate whether using different information structure in dative constructions in

English influenced acceptability ratings and reaction times needed to evaluate these variants. In their experiments, the participants saw four types of experimental stimuli – NP NP constructions (e.g. They gave a woman the report) and NP PP constructions (e.g. They gave the report to a woman). In each variant, the order of known and unknown information, marked by the use of definite or indefinite article, also varied. They show that participants generally preferred NP PP constructions, as indicated by faster reaction times and higher acceptability ratings. In addition, the measures also indicate a preference of given-before-new information structure for NP NP constructions but new-before-given in sentences where NP PP construction was used.

Kleiman (1980) presents an interesting extension of the lexical decision task, very similar in principle to the task we present here. In her experiment, participants first saw sentences in which the last word was removed and were asked to read them at their own pace. Next, a fixation point was shown at the bottom of the screen. After 600ms, the participants were presented with a string of letters and had to decide if it forms an acceptable English word. The results of this study show, among other things, that context is an important source of information. In particular, the participants responded faster to words that were deemed better completions of the sentences (e.g. 'table' when it followed a sentence 'The cup was placed on a ...') compared to either words that were possible but less likely continuations (e.g. 'chair' in the same sentence) or words unrelated to the best completions (e.g. 'floor').

Similar conclusions were reached by Schwanenflugel and Schoben (1985), who show that in high-constraint sentences, in which 78% or more participants agree on what word should follow the context, highly expected items were reacted to faster than their possible, but less expected counterparts. That is, when presented with a sentence 'The bar would not serve drinks to ...' people reacted faster if it was continued with 'minors' rather than 'kids'. Interestingly, they did not find the same effect when a sentence posed fewer constraints. For example, there was no significant difference between reactions to words 'chef' and 'cook' when they followed the sentence 'The lady was a competent ...' . Nonetheless, the study adds to the evidence that in contexts where speakers expect certain words to occur, the reactions are faster than to other words even if they are also acceptable in the same contexts.

Interestingly, the study also shows that increasing cue validity – i.e. the number of times the sentence was actually followed by a highly predictable item, affects the reaction times. This shows that distributions of continuations in the experiment – in addition to the expectations formed on the basis of the usage – affects the behaviour of the participants.

The studies discussed so far present the evidence that people process (highly) predictable items faster. What is more, we have also seen that speakers clearly make use of the available context when making predictions. However, in the current study, the focus is on predictability of grammatical categories, rather than individual words. The question is then whether people make predictions on more abstract levels of linguistic analysis? After all, as pointed out by Pickering and Garrod (2007), in certain contexts, looking at language at a more abstract level leads to greater predictability – for example, English articles are often followed by either adjectives or nouns (Pickering & Garrod, 2007). Therefore, predicting grammatical categories – as opposed to individual words that represent them – should in theory be easier. But is there evidence that speakers actually do learn at such abstract level and make use of this knowledge when processing linguistic input?

In an early study, Wright and Garrett (1984) present evidence that suggests that speakers do in fact predict categories. In their study, participants read sentences word by word. Each sentence ended with a target word on which reaction times were measured. The authors tested how various syntactical contexts influence the reaction times on the syntactically possible and impossible combinations. In one experiment, discussed here to illustrate the design, the experimental items consisted of two types of sentences – in the first one, modal verbs were followed either by a noun or a main verb. In the other, prepositions were followed by either a verb or a noun. Importantly, the target words fitted the context only syntactically and they could not be predicted on the basis of meaning, as illustrated by examples below:

If your bicycle is stolen you must FORMULATE
If your bicycle is stolen you must BATTERIES (Wright and Garrett, 1984)

What the results of the study show is that speakers were significantly slower to respond to ungrammatical combinations of items than they were when they read grammatically acceptable combinations. Since the target words could not be predicted on the basis of semantic properties, the results suggests that what speakers predict is not only the word itself, but a grammatical category to which it belongs.

A similar conclusion was reached by Van Berkum et al. (2005). They investigate whether speakers of Dutch make predictions about the upcoming input based on the linguistic information they have received earlier. The paper presents the results from three experiments, two of which were conducted using event-related brain potential (ERP) as a measure. The third employed a self-paced reading task, in which the participants read a text word by word and the speed at which they pressed the button to proceed to the next word was recorded. The experimental set up was very similar in all three tasks: the participants read chunks of texts up to an indefinite article, which were either continued with a more or less probable adjective-noun combinations. Importantly, the grammatical gender of the adjectives matched the grammatical gender of the noun. However, the gender in the less probable continuations was different from the gender of the highly probable adjective-noun pairs. The ratings of the probability of continuation were evaluated before the study in a spontaneous cloze completion pretest, during which Dutch speakers were asked to complete the chunks of the text with words they saw as the most fitting in the given context. The results show that people do use the information available in the context to anticipate what is coming next. In the EEG experiments, the researchers observed the N400 effect at the adjective, which suggests that the speakers were predicting the noun and its grammatical properties, such as gender. Having encountered an adjective which grammatical properties did not match what they anticipated, the participants were surprised, which was reflected in the activity of their brains. Similarly, in the self-paced reading task the participants slowed down having encountered a less probable continuation. However, they did not do so immediately after encountering the adjective, but still before encountering the noun. This study is relevant to our current study for two reasons. First, it demonstrates that the speakers use the available context to make predictions about possible upcoming input. This is not only consistent with the theory we presented in the earlier chapters, but also with the specific hypotheses in the current study. Just like in the studies we presented above, we expect that the speakers will use the context to make predictions about the word – including its grammatical properties, such as gender or aspect – that should fill the gap in the sentence. Secondly, it demonstrates that reaction time as a measure (albeit used in a slightly different experimental paradigm) indicates similar effects as more technologically advanced measures, such as ERPs.

To summarize the studies discussed above, we can say that the evidence suggests that speakers make predictions based on the context. Additionally, the predictability of a word in the context plays an important role despite the fact that other possible continuations exists – words that are more predictable are read faster than their less likely counterparts. Importantly, the predictions seem to be made on different levels of linguistic analysis. That is, there is evidence that not only words are being predicted but also the grammatical categories they represent. This is particularly important since in the current study, we investigate whether people predict the grammatical category of aspect. In contrast to other studies presented here, however, the likelihood of the words used in the gap were not calculated using a cloze tests. Instead, we use the association weights learned in the NDL based simulations.

## 11.2 Conducting reaction time research online

While choosing widely used measures for the study should not be controversial, it is perhaps worth elaborating on the decision to conduct the study online. In this section, we present evidence that conducting time sensitive research online is a reasonable methodological choice.

The main motivation for conducting this study online was unfortunately related to the particular limitations brought about by the COVID-19 pandemic, as it was not possible to run the study in the lab. Fortunately however, these limitations resulted in a boom in online research and two very recent papers investigated the accuracies and precision of reaction-time experiments conducted online. The findings were very encouraging. Anwyl-Irvine et al. (2020) show that even though there are differences in timing responses between various combinations of experimental platforms, operating systems and devices, the overall performance of online set ups is accurate enough to measure differences in behaviour. For instance, an experiment in PsychoJS (version 3.1.5) – a programming language used on Pavlovia – run on a Windows laptop using a Chrome browser showed an average delay of around 60ms. Changing the browser on the same device to Firefox resulted in 20ms of additional delay. The delay on a macOS laptop with Safari browser showed an average delay of 65ms. Similar conclusions were also reached by Bridges et al. (2020), who report about 44ms lag for PsychoJs (version 2020.1) on Windows using Chrome, and 40ms on Firefox. An important source of differences between the results in the two studies can be a result of the fact that Bridges et al. (2020) used a high-precision button box and a newer version of the software.

What both these studies point at is that while conducting time-sensitive research online necessarily introduces some noise in the data, especially compared to lab-based experiments in which high-precision equipment is used, the platforms that are currently available allow for obtaining reasonable measurements. Therefore, we decided to conduct the study online, taking into account the insights and the recommendations included in both papers. First of all, we used a within-participant design. This meant that each participant provided their answers in both conditions, and the variability caused by the differences in the particular set up used by the user should stay the same across the conditions. Secondly, to further limit the differences in timing due to the browser, we asked the participants to use Chrome.

## 11.3 The judgement task

Having discussed the method of data collection and the behavioural measures we used, we now move on to presenting the details of the study.

### 11.3.1 Method

In the study, the participants saw gapped sentences and had to decide if filling the gap with a provided verb would yield an acceptable sentence. First, only a sentence with a gap was displayed on the screen. The participants were asked to read the sentence carefully, and then to move to another screen, where they saw the same sentence again, and a verb below it. The task was to indicate by pressing buttons on the keyboard, whether the sentence completed with the provided verb is correct or not. The participants were asked to press the 'm' button if they thought that the sentence is correct, and the 'z' button if they felt it is not. We measured the reaction time on the second screen, where the decision was made. The first screen, with a gapped sentence only, was introduced as a way of subtracting the time needed to read and process the context. The time for the decision was limited to 5 seconds to discourage skipping the first screen and reading the sentence on the decision screen. Each trial was separated by a screen with a cross in the centre, which prevented moving to the next sentence accidentally, in case a response button was pressed after 5 seconds. We also included attention checks to make sure the participants are not pressing buttons mindlessly. Three randomly selected sentences were followed by a statement referring to the sentence before it, and the participants had to decide whether the statement matches the information conveyed by the sentence or not. To provide their answers they again used 'm' and 'z' buttons on the keyboard.

### 11.3.2 Materials

We created two sets of 96 sentences in total – 48 experimental items and 48 fillers. Each subset of items consisted of 24 correct and 24 incorrect sentences. In each of these, in 12 sentences a perfective verb was used in the gap, and in the other 12 an imperfective verb was used. This split is summarized in Table 31.

*Table 31. The number of items per set*

| | aspect | correct | incorrect | total |
|---|---|---|---|---|
| experimental items (past tense) | perfective | 12 | 12 | 24 |
| | imperfective | 12 | 12 | 24 |
| fillers (non-past tenses) | perfective | 12 | 12 | 24 |
| | imperfective | 12 | 12 | 24 |

The verbs in experimental items were in the past tense. To make sure the findings generalize across number, person and gender categories, the inflectional categories were counterbalanced: each verb was used in only one number/person/gender combination and each number/person/gender combination was represented by 2 verbs. To limit the number of combinations, we only considered masculine and feminine gender.

The verbs were obtained in the following way. First, we randomly sampled 4 aspectual pairs for each number/ gender/ person combination. Then, the resulting sample of 48 unique verb forms was split into two sets. Set 1 was created by taking every other pair (n, n+2, ...). Set two contains the remaining verbs (n+1, n+3, ...). As a result, in each set, per each aspect, there are 12 singular and 12 plural verbs, 12 masculine and 12 feminine verbs and 8 verbs per each of the three persons, as illustrated in Table 32.

*Table 32. The distribution of experimental items per inflectional categories*

| total | number | gender | person | verb_form |
|---|---|---|---|---|
| 24 | singular | feminine | 1st | verb1<br>verb2 |
| | | | 2nd | verb3<br>verb4 |
| | | | 3rd | verb5<br>verb6 |
| | | masculine | 1st | verb7<br>verb8 |
| | | | 2nd | verb9<br>verb10 |
| | | | 3rd | verb11<br>verb12 |
| | plural | feminine | 1st | verb13<br>verb14 |
| | | | 2nd | verb15<br>verb16 |
| | | | 3rd | verb17<br>verb18 |
| | | masculine | 1st | verb19<br>verb20 |

| | | | | |
|---|---|---|---|---|
| | | | 2nd | verb21<br>verb22 |
| | | | 3rd | verb23<br>verb24 |

The subset of each aspect was then split again, taking every other pair, to create correct and incorrect subsets. The errors were made by shuffling the verbs. As a result, the aspect and tense of the verb used in an erroneous sentence was correct, but its other dimensions (person, gender, number) were not and/or it did not make sense semantically. For instance:

sentence: Wcześniej, zaproszone prelegentki .......... umiejętności dziecka, które nabywa w szkole.

Earlier, invited speaker.3PL.F .......... skills child which acquires in school

Earlier, the invited speakers ....... the skills a child acquires at school.

verb displayed: zalogował

log_in.3.SG.M

The sentences were created by the author, who took adjusted and modified sentences from a sample extracted from the Araneum Corpus of Polish, National Corpus of Polish and various internet sources. Crucially, the experimental sentences were constructed in a way that allowed for both imperfective and perfective aspect to be used. They did not contain any obvious cues, such as triggers or collocations, that could sway the choice one way or another. However, the sentences contained a time adverbial that informed that the verb would be in the past tense. This was done because tense is expressed on the verb, which was removed from the gapped sentences. Therefore, an additional cue was needed for the participants to expect the past – and consequently, to form expectations regarding the aspect.

The fillers were constructed in a similar way. A sample of 96 random non-past verbs in various inflectional patterns was extracted. The sample was split into two sets, and a sentence was created for each of the verbs. Because we did not plan to analyse the responses for these items, they were not restricted in the same ways as the experimental items. This meant that in some of them only one aspect could be used, and that only some contained aspectual triggers.

Finally, each set of stimuli, containing both experimental items and the fillers, was used in two conditions – base and cued. In the cued condition, imperfective sentences contained an additional trigger – an adverbial that can only be used with imperfective verbs. The sentences in the base condition contained only an adverbial indicating that the sentence describes a past situation.

The predictions were different for each of the conditions. In the base condition, we expected to see differences in both reaction times and accuracy between the canonical and non-canonical sentences. In the cued condition however, we did not expect to see any differences, because we predicted that including a cue suggesting which aspect should be used would facilitate processing of sentences containing a non-default aspect of the verb.

This experimental design also allowed us to obtain more insights. Given the results of the studies we reviewed in previous sections, we can also expect that the tense-aspect combinations will influence error detection. Since we assume that the speakers should expect perfective aspect in past sentences, processing imperfective past sentences, which violate both semantic coherence and aspectual expectations, should be more difficult in the base condition. Therefore, we should observe slower reaction times and less accurate evaluations for the incorrect imperfective past sentences.

Alternatively, we could also hypothesise that the more expectations are violated the easier detecting the errors should be. That is, since the incorrect imperfective verbs are both semantically incongruent and aspectually less predictable in the context, evaluating them as errors should be done quicker and more accurately, because they stand out more. However, Wright and Garrett (1984, discussed above) showed that the sentences containing contextually unexpected grammatical categories were still evaluated slower than the items where the category was expected, even if the word in the gap did not fit semantically in either type of item.

Any effect on error detection observed in the base condition should again be alleviated in the cued condition, because the sentences contain an additional cue that helps predict the aspect of the verb. Therefore, we remove the hypothesised difference in the predictability of the grammatical class, making all items equally erroneous only on the level of coherence.

### 11.3.3 Participants and data collection

The experiment was set up in psychopy builder (Peirce et al., 2019) and run on Pavlovia via the Prolific (Prolific, 2014) platform. To allow full counterbalancing, summarized in Table 33 below, the experiment was divided into four versions, one for each condition/set combination: base + set1, base + set2, cued + set1, cued + set2, and the data were collected in two rounds. Each participant saw a different combination of the condition/set combination in each round. The order of the combinations was counterbalanced across participants.

Each version of our study was a separate survey on Prolific, and we opened them one by one, inviting 10 people initially. Having collected 10 responses, we screened out participants who failed 2/3 attention checks and invited new participants so that the total number of approved responses was equal to 10. However, during data collection for the last version of the experiment in round 1, some participants pointed out that it was very easy to accidentally skip the attention check and provide an incorrect response. We subsequently checked the data and determined that it was indeed possible to skip the checks and in fact, almost everyone who failed the checks most likely did so (the response times were often at 300 miliseconds or lower). Therefore, we decided to ignore the attention checks and invite everyone (except the people who timed out or returned their submissions) who took part in round 1 to participate in round 2. In total, there were 7 additional participants. However, in round 2, we opened only 10 places for each study to keep the study counterbalanced. This increased the chances of participants returning and taking part in round 2. For further data analysis we included only the responses from the participants who took part in both rounds of data collection.

Each version of the experiment lasted about 15 minutes. The participants were renumerated via Prolific – the average payment in each study was £7.52 per hour.

The final sample consisted of 38 participants, as two people did not return to participate in round 2. 22 participants (57.9%) were female, 16 (42.1%) were male. The age ranged from 18 to 46 years old (mean: 23.7%). 17 participants (44.73%) had higher education, either postgraduate or undergraduate, and 20 (52.63%) graduated from high school or equivalent. 1 person (2.63%) declared that they were still in high school. All of the participants were Polish nationals as well as native speakers of Polish. 34 of them, or 89.47%, lived in Poland at the time when the experiment was conducted.

*Table 33. A summary of the counterbalancing in the decision task*

| participant group (10 people) | data collection round | condition/ set combination |
|---|---|---|
| 1 | round 1 | base, set 1 |
| | round 2 | cued, set 2 |
| 2 | round 1 | base, set 2 |
| | round 2 | cued, set 1 |
| 3 | round 1 | cued, set 1 |
| | round 2 | base, set 2 |
| 4 | round 1 | cued, set 2 |
| | round 2 | base, set 1 |

## 11.4 Results

### 11.4.1 Data preparation

The data were analysed in RStudio (RStudio Team 2020), using R version 3.6.3 (2020-02-29). Before the analysis, we cleaned the dataset by identifying and removing outliers.

First, we inspected the data by plotting the mean reaction time of each participant on the sentence screen by the mean reaction time of each participant on the decision screen. This way, we identified one outlier – a participant who responded quickly both when reading the provided context (1.15 second on average) as well as when making decisions about the verbs (1.26 seconds on average). We decided that such speed indicates that the participant did not focus enough on the task and that their responses are not trustworthy. Therefore, we removed all responses from this participant, decreasing the final sample to 37 people.

Next, we excluded the data points with missing values in experimental response. Then, we plotted the reaction times on the decision screen and identified 2 outliers where the response times were 0.0378s and 0.0026s. Since these answers should be treated as accidental clicks, we removed these data points from our dataset as well. Additionally, following Schwanenflugel and Schoben (1985), we also removed rows where participants spent less than 400 miliseconds on the sentence screen, as this gave them too little time to familiarize themselves with the context before moving on to the decision screen (cf. Baayen & Milin, 2010). Finally, during the analysis we discovered that in one incorrect experimental item, instead of an incorrect past verb, a non-past verb was erroneously displayed. While this item was still incorrect, the use of a non-past verb introduced a violation of an additional dimension, tense, which we controlled in all other items. Therefore, we excluded that question from the analysis. In total, the final dataset consisted of 6947 datapoints.

### 11.4.2 Mixed effects regression modelling: correct experimental items

*Reaction times*

In this section we will discuss the models fitted on responses we obtained for the correct subset of the experimental items. Since in our study each sentence was evaluated by many participants, and many sentences were evaluated by the same participant, the data points were not independent. This design, as explained in Chapter 6, necessitates the use of a mixed effect model, with a random intercept of participant and sentence.

The analysis of the dependent variable, the reaction time on the decision screen, the showed that the reaction times were not normally distributed. This is often a sign that the residuals – the distances between the actual data points and the fitted regression line – may also not be normally distributed, which in turn violates one of the main assumptions of linear models. This violation was confirmed by the inspection of residuals of a mixed effect model fitted for untransformed response times.

For this reason, we performed a log transformation on the dependent variable. Data transformation is a standard practice in science. What it means in practice is that we apply a function to all data points in our dataset in order to make their distribution more closely resemble a normal distribution. Log transformation, the way it is implemented in Rcore functions, takes a data point from our dataset and returns its natural logarithm. A logarithm is essentially the value of the exponent to which we need to raise the base number, in order to get the value we are calculating the logarithm of. For example, a base-10 logarithm of 100 is 2, because 10 to the power of 2 equals 100. In case of the natural log, the base is Euler's number, or 'e' constant, approximately equal to 2.71828. A natural log of 100 is 4.60517018599, because the 2.71828 raised to the power of 4.60517018599 equals 100.

What it helps us to achieve is make our data more symmetrical, relative to 1. To illustrate that, let's consider this example. On an untransformed axis, the distance between 2 and 1 is greater than the distance between 1 and 0.5. Transforming these values makes the distances equal. A natural log of 1is 0 – a mid-point on our new axis. A natural log of 0.5, a value that is two times less than 1,  is -0.69314718056. A natural log of 2, a value that is two times greater than 1, is 0.69314718056. Therefore, by log transforming the three values we now represent the data in a different way, capturing the distances to a mid-point of our new scale, rather than the actual values from our dataset. The resulting distribution is much closer to normal, which allows us to fit a linear regression model without the risk of violating the assumptions of the model. Applying the log transformation on the reaction times resulted in a distribution that was close to normal, which allowed us to fit a linear regression model without the risk of violating the assumptions of the linear regression.

As for the fixed effects, we fitted aspect and condition, as well as the interaction between them, since we predicted that the changes in response times should depend on both the value of aspect and the condition. In addition, to control for frequency effects of the verb forms, we also fitted log transformed normalized form frequency as a fixed effect. Normalization was a standard recalculation of the raw frequency to frequency per million. This was done by taking each verb frequency, dividing it by the total number of words in the Araneum corpus, and multiplying the result by 1,000,000. The normalized frequency allows for an easier interpretation of how often we can expect a given word. Raw frequencies simply state how often a given word appeared in the corpus but they do not tell us anything about the size of the corpus. 10 occurrences in corpus of 100 words is a lot. However, 10 occurrences in a corpus of one billion words is miniscule. The normalized frequency allows us to compare between corpora since it answers the question of how often we could expect a given word in a chunk of text of size x – in our case, a million words. In addition, we also log transformed the normalized frequency counts, because the distribution of this variable was again skewed.

The model was fitted using the lme4 package (Bates et al., 2015). In addition, we used the lmerTest package (Kuznetsova et al., 2017) to calculate p-values for the effects of the predictors. As we have said, the dependent variable was the log transformed reaction time on the decision screen. The fixed effects were aspect, condition, log transformed normalized verb form frequency and the interaction between the aspect and condition. We also included two random intercepts of participant and sentence. The formula used for the model was:

logtransfRT ~ aspect * condition + formFreqNormLog + (1|participant) + (1|sentence)

Table 34 below presents the model output, obtained from lme4 summary method.

*Table 34. Regression model output: reaction times for correct experimental items*

<table>
<tr><td colspan="5">Fixed effects</td></tr>
<tr><td></td><td>Estimate</td><td>Std. Error</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>Intercept</td><td>0.331469</td><td>0.062100</td><td>5.338</td><td>5.66e-07</td></tr>
<tr><td>aspect|perfective</td><td>0.010778</td><td>0.048637</td><td>0.222</td><td>0.825</td></tr>
<tr><td>condition|cued</td><td>0.010905</td><td>0.045276</td><td>0.241</td><td>0.810</td></tr>
<tr><td>formFreqNormLog</td><td>-0.021278</td><td>0.012879</td><td>-1.652</td><td>0.103</td></tr>
<tr><td>aspect|perfective:condition|cued</td><td>-0.007229</td><td>0.051470</td><td>-0.140</td><td>0.888</td></tr>
</table>

The results show that none of the predictors reached significance. Therefore, the model does not show the trend we expected. As we can see neither the change in condition (0.010905, p=0.810) nor in the aspect of the verb used in the sentence (0.010778, p=0.825) had any significant influence on how fast the participants evaluated the sentence. In addition, we also did not see the expected effect of the interaction between the aspect and the condition (-0.007229, p=0.888). This suggest that the participants responded equally fast to all aspects in all conditions, which again goes against our predictions. Interestingly, the form frequency also turned out to be insignificant (-0.021278, p=0.103), which means that the frequency of the inflected verb form did not affect the speed of evaluation.

To better illustrate the relationship between the dependent variable and our predictors we present a figure showing the distribution of the log transformed reaction times for each aspect and condition. As we can see, the differences between them are miniscule, which was reflected in the model described above.

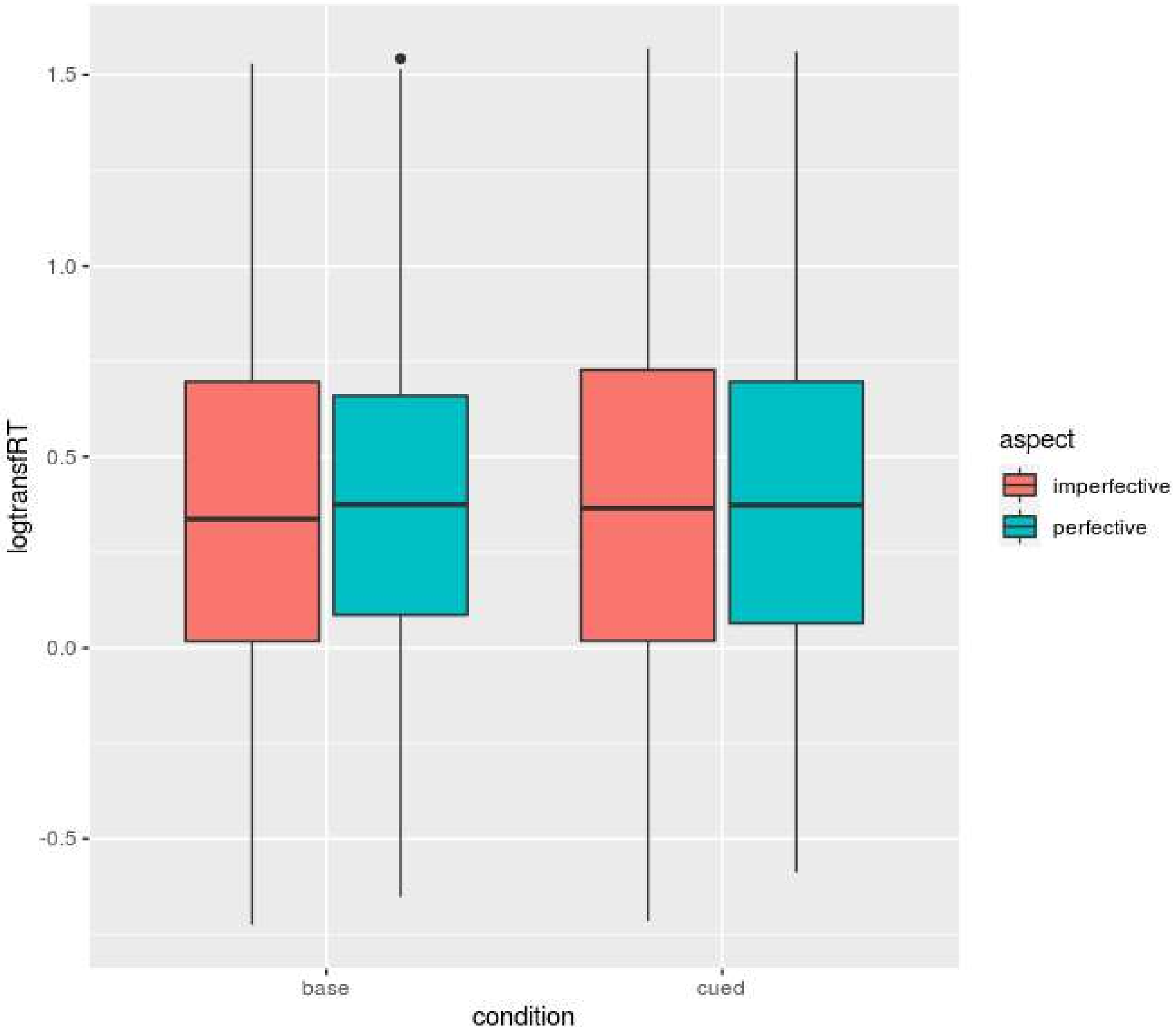


*Figure 10. The distribution of the reaction times for each aspect and condition (correct items).*

*Accuracy*

On average, the participants were very accurate in evaluating the correct experimental items. Table below present mean accuracy achieved across participants for each aspect and condition. As we can see, the values are close to 1, indicating high accuracy.

*Table 35. Mean accuracies for correct experimental items.*

| aspect | condition | mean accuracy |
|---|---|---|
| imperfective | base | 0.99 |
| perfective | base | 0.98 |
| imperfective | cued | 0.99 |
| perfective | cued | 0.97 |

In order to model the relationship between our predictors and the accuracy of evaluations, we used a mixed effect logistic regression model. Just like in the model described in the previous section, we fitted two random effects of participant and sentence. As for the fixed effects, we again used aspect, condition, log transformed normalized form frequency and the interaction between the aspect and condition. As the dependent variable we used participant judgement score (0 or 1 indicating whether the participant evaluated the sentence incorrectly or correctly). The model was fitted using the glmer function from the lme4 package.

Therefore, the formula used for the model was:

participantCorrect ~ aspect * condition + formFreqNormLog + (1|participant) + (1|sentence)

Table 36 below summarizes the model output:

*Table 36. Regression model output: accuracy of judgements for correct experimental items*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | z value | Pr(>\|z\|) |
| Intercept | 5.7668 | 0.7734 | 7.457 | 8.87e-14 |
| aspect\|perfective | -0.4617 | 0.6923 | -0.667 | 0.5048 |
| condition\|cued | 0.2479 | 0.7150 | 0.347 | 0.7289 |
| formFreqNormLog | 0.3053 | 0.1738 | 1.757 | 0.0789 |
| aspect\|perfective:condition\|cued | -0.7133 | 0.8648 | -0.825 | 0.4095 |

These results indicate that the accuracy with which participants evaluated the sentences was not affected by any of our predictors. There were no differences between aspects (-0.4617, p=0.5048) nor conditions (0.2479, p=0.7289). The interaction between these two predictors was also not significant (-0.7133, p=0.4095). This outcome goes against our predictions as it suggests that in both conditions and for both aspects participants responded with similar accuracy. In addition, we can also see the influence of the frequency of the verb form did not reach the significance threshold (0.3053, p=0.0789).

### 11.4.3 Mixed effects regression modelling: incorrect experimental items

In the previous sections we analysed the reaction times and accuracies of responses obtained for correct experimental items. In this section, we will focus on the analysing the responses for the incorrect subset of experimental items. To remind the reader, the incorrect items were created by taking half of the sentences in each aspectual subset and randomly switching the verbs that were displayed with them. Therefore, each incorrect experimental item contained a gapped sentence and a verb that had the same aspect as the original verb, but did not fit the sentence semantically. This allowed us to test whether any observed differences in reaction times and accuracy were related to the violation of expectations of aspect to appear and not just the semantic incoherence of the verb in the sentence.

As we explained at the beginning of this chapter, we expected to observe larger reaction time latencies and less accurate evaluations for the imperfective incorrect sentences, compared to the perfective incorrect sentences, since the former not only violate semantic coherence, but also the hypothesised expectations regarding the aspectual class.

*Reaction times*

As before, to better illustrate the relationship between the reaction times and the two main predictors of interest, we present a figure showing the distribution between the log transformed reaction times on incorrect experimental items for each aspect and condition. As we can see, the differences are again negligible, which was confirmed by the model we discuss below.

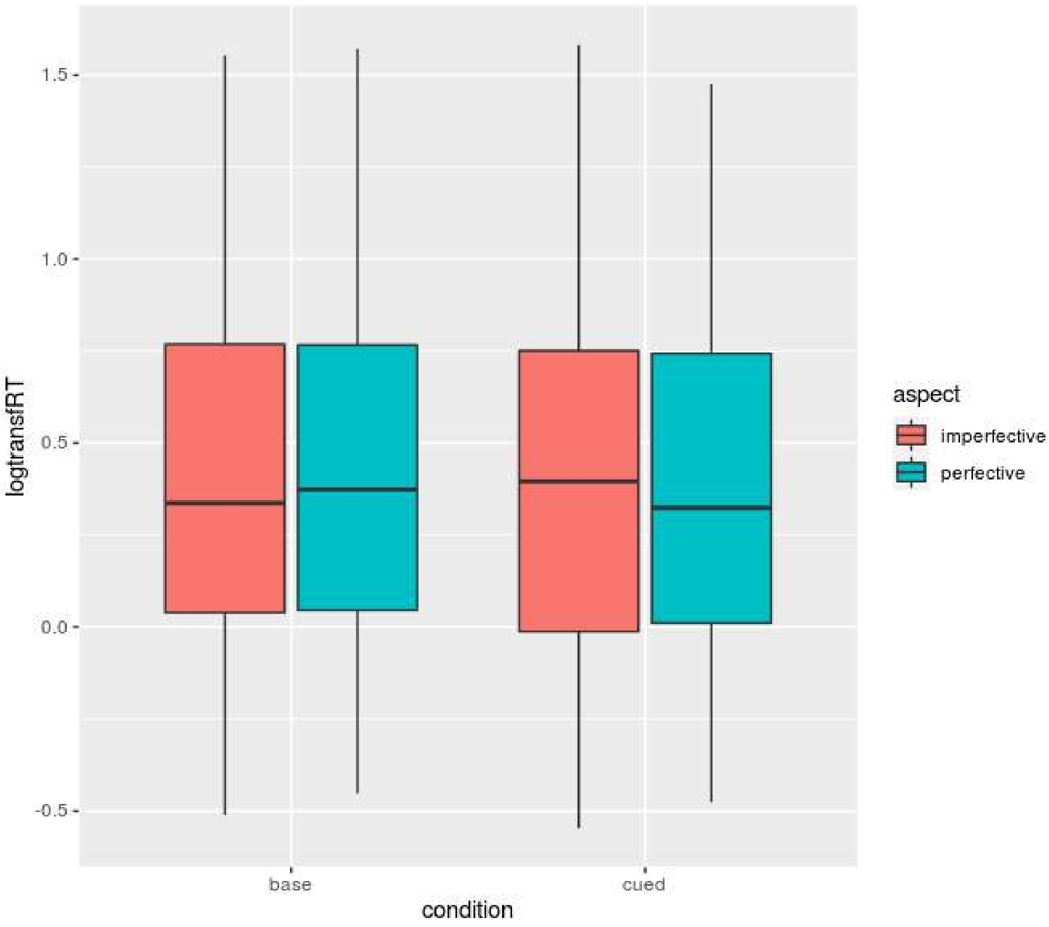


*Figure 11. The distribution of the reaction times for each aspect and condition (incorrect items).*

To model the relationship between the reaction times on the incorrect experimental items and our predictors, we again fitted a mixed effects linear regression model. The dependent variable was the log transformed reaction time on the decision screen. The independent variables included aspect, condition, log transformed normalized verb form frequency of the displayed verb and the interaction between aspect and condition. In addition, we also added two random effects of participant and sentence. The formula used is presented below:

logtransfRT ~ aspect * condition + displayedFormFreqNormLog + (1|participant) + (1|sentence)

Table 36 below presents the model output:

*Table 37. Regression model output: reaction times for incorrect experimental items*

| Fixed effects | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | t value | Pr(>\|t\|) |
| Intercept | 0.287117 | 0.089899 | 3.194 | 0.00191 |
| aspect\|perfective | 0.021436 | 0.055664 | 0.385 | 0.70132 |
| condition\|cued | 0.005823 | 0.055325 | 0.105 | 0.91647 |
| displayedFormFreqNormLog | -0.036007 | 0.018681 | -1.927 | 0.05826 |
| aspect\|perfective:condition\|cued | -0.030241 | 0.060630 | -0.499 | 0.61900 |

As we can see, the reaction times for the incorrect experimental items were not affected by neither aspect (0.021436, p=0.70132) nor the condition (0.005823, p=0.91647), and the interaction between the predictors was also insignificant (-0.030241, p=0.61900). In addition, we can also see that the influence of the form frequency of the displayed verb approached significance but did not cross the significance threshold (-0.036007, p=0.05826). These results again are the opposite of what expected.

*Accuracy*

Evaluating incorrect sentences turned out to slightly more difficult. The mean accuracies across participants, for each aspect and condition, which we present in the table below suggest that for this subset of experimental items, participants were less accurate than for the correct items. The mean scores, however, are still very high (around 90%). This slight drop in performance can perhaps be explained by the fact that the incorrect items, containing a surprising combination of contexts and verbs, were more difficult to process, resulting in a higher error rate.

*Table 38. Mean accuracies for incorrect experimental items.*

| aspect | condition | mean accuracy |
|---|---|---|
| imperfective | base | 0.90 |
| perfective | base | 0.95 |
| imperfective | cued | 0.93 |
| perfective | cued | 0.96 |

Finally, we present the last regression, used to model the influence of the predictors on the accuracies of responses for the incorrect experimental items. Just like before, we fitted a mixed effects logistic regression model with aspect, condition, normalized form frequency and the interaction between aspect and condition as fixed effects. The dependent variable in this case was the participant judgement score. The formula used to fit the model was:

participantCorrect ~ aspect * condition + displayedFormFreqNormLog + (1|participant) + (1|sentence)

The model output is presented in table 37 below:

*Table 39. Regression model output: accuracy of judgements for incorrect experimental items*

<table>
<tr><td colspan="5">Fixed effects</td></tr>
<tr><td></td><td>Estimate</td><td>Std. Error</td><td>z value</td><td>Pr(>|z|)</td></tr>
<tr><td>Intercept</td><td>4.9915</td><td>1.0563</td><td>4.726</td><td>2.29e-06</td></tr>
<tr><td>aspect|perfective</td><td>0.3566</td><td>0.6799</td><td>0.525</td><td>0.5999</td></tr>
<tr><td>condition|cued</td><td>0.7174</td><td>0.6954</td><td>1.032</td><td>0.3022</td></tr>
<tr><td>displayedFormFreqNormLog</td><td>0.4196</td><td>0.2372</td><td>1.769</td><td>0.0769</td></tr>
<tr><td>aspect|perfective:condition|cued</td><td>-0.4158</td><td>0.7763</td><td>-0.536</td><td>0.5922</td></tr>
</table>

Once again we see that none of the predictors can be said to significantly influence the accuracies of the participants' evaluations. Aspect (0.3566, p=0.5999), condition (0.7174, p=0.3022) did not reach significance. The interaction between our main predictors, aspect and condition was yet again insignificant (-0.4158, p=0.5922). Similarly, the influence of the normalized log transformed of the displayed verb form did not cross the significance threshold (0.4196, p=0.0769).

To sum up the findings from all of the models presented in this chapter, we can say neither the accuracy of the participants' judgements nor the reaction times on experimental items – regardless of their correctness – were significantly affected by our predictors. Therefore, we must conclude at this point that we did not observe the hypothesised effects of aspectual categories or the presence of additional adverbial cues on participants' speed and accuracy of judgement in this study.

### 11.5 Bayes Factor analysis

The null results could also suggest that the study was underpowered. That is, the reason we did not find the expected effect could be related to the fact that we did not test enough participants, and therefore, the probability of detecting the effect was too low. In such case, we would make a Type II error and fail to reject the null hypothesis (which, in our case was that there are no differences between RTs for imperfective and perfective sentences) when in fact, we should.

However, since the study has already been conducted, the traditional power analysis is not an appropriate method of answering the question of whether the sample size was too small. There is a number of sources (e.g. Hoenig & Heise, 2001; Dziak et al., 2020) showing that a post hoc power calculation is necessarily related to the p-values obtained in the study, and that the larger the p-value, the larger the obtained sample size is.

Therefore, we decided to conduct a Bayes Factor analysis aiming to answer two questions pertaining to our current study and the results we obtained. First of all, we want to know how confidently we can accept the non-significant result as an indication that the data supports the null hypothesis. Secondly, we want to estimate if collecting more data could change the results and allow to find the expected effect.

#### 11.5.1 Bayes Factor

Bayes Factor is a measure that allows us to estimate the strength of evidence for both the null and the alternative hypotheses, given the data we collected (Dienes, 2008, 2014). What is more, Bayes Factors can be also be calculated for larger sample sizes, thus allowing us to estimate whether collecting more data would increase or decrease our confidence in rejecting the alternative hypothesis.

To calculate the Bayes Factor, we used the equivalent of a calculator by Dienes (2008), implemented in R by Baguely and Kayne (2009), similarly to Vujović (2020). Then, following Wonnacott et al. (2017), we calculated the Bayes Factor for larger sample sizes to evaluate whether adding more participants would provide more support for any of the hypotheses in our study.

The functions require the model of the data and a model of the alternative hypothesis as inputs. Following Vujović (2020), we took the coefficient of the interaction of aspect and condition as a model of the data. To model the alternative hypothesis we used the half-normal distribution with a standard deviation equal to the coefficient of one of the main effects (condition) (Vujović, 2020, p. 67).

The output, i.e. the Bayes factor is the ratio of the likelihood of the alternative hypothesis over the likelihood of the null hypothesis, given the data. The values of the Bayes Factor are interpreted in the following way: those larger than 3 indicate that there is a substantial evidence for the alternative hypothesis, those below .33 suggest the opposite. Values between 0.33 and 3 suggest that no conclusion can be reached (Dienes, 2008; Vujović, 2020). The calculation of the Bayes Factor for larger sample sizes is based on the assumption that the standard error scales with a square root of the sample size (Wonnacott et al., 2017).

Below we present 4 figures showing the estimated Bayes Factor changes for all four models discussed in this chapter.

### 11.5.2 Results

*Reaction times on the correct experimental items*

Figure 12 below presents the values of the Bayes Factor on the y axis and the sample size on the x axis. As we can see, the analysis shows that the current results do not provide enough support for the alternative hypothesis. The interaction between the two main effects – aspect and condition is roughly equal to 1, and does not significantly change, even with a large number of participants. Therefore, we can state that our current study was inconclusive, and that collecting data from more participants would not result in a greater support for either the alternative or the null hypotheses.

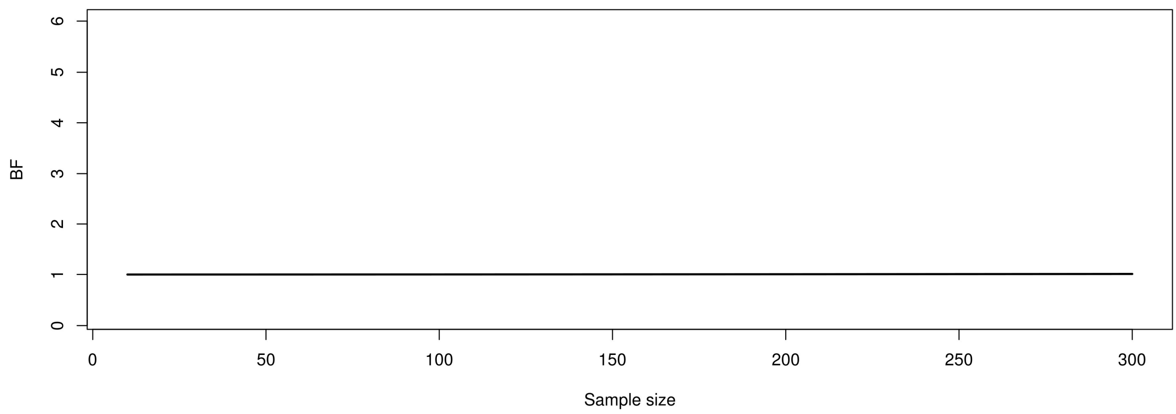


*Figure 12. Bayes Factors for RTs on the correct experimental items*

*Reaction times on the incorrect experimental items*

Similarly, as we can see in Figure 13 below, the analysis suggests that the results of the reaction times model fitted on the incorrect items were inconclusive. What is more, we again see that increasing the sample size would not provide additional support for any of the hypotheses. Although the value of the Bayes Factor does slightly increase with a larger sample, it does not reach a substantial evidence threshold even for a large number of participants.

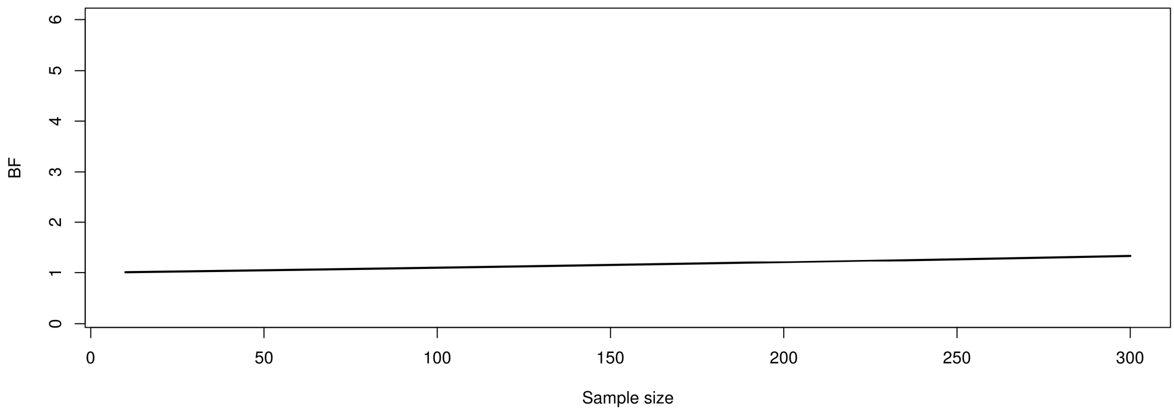


*Figure 13. Bayes Factors for RTs on the incorrect experimental items*

*Accuracy on the correct experimental items*

As for the model of accuracy, fitted for the correct experimental items, the analysis suggests that the substantial level of support for the expected effect could be reached, but only with a very large sample of participants. As shown in Figure 14, the value of the Bayes Factor marginally crosses the threshold of 3 for 235 participants.

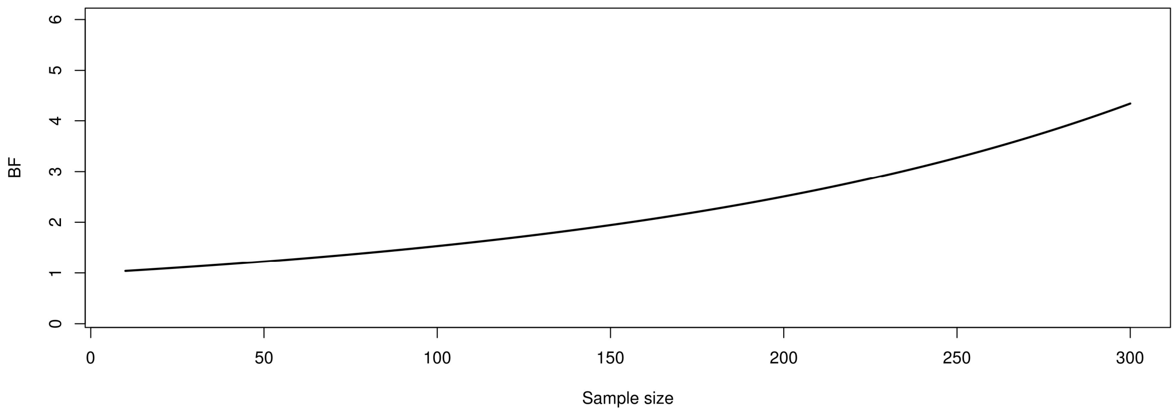


*Figure 14. Bayes Factors for accuracy on the correct experimental items*

*Accuracy on the incorrect experimental items*

Finally, the results of the analysis for the accuracy model fitted on the incorrect experimental items show that while the support for the alternative hypothesis increases slowly, it does not reach the expected threshold even with a large number of participants (Figure 15).

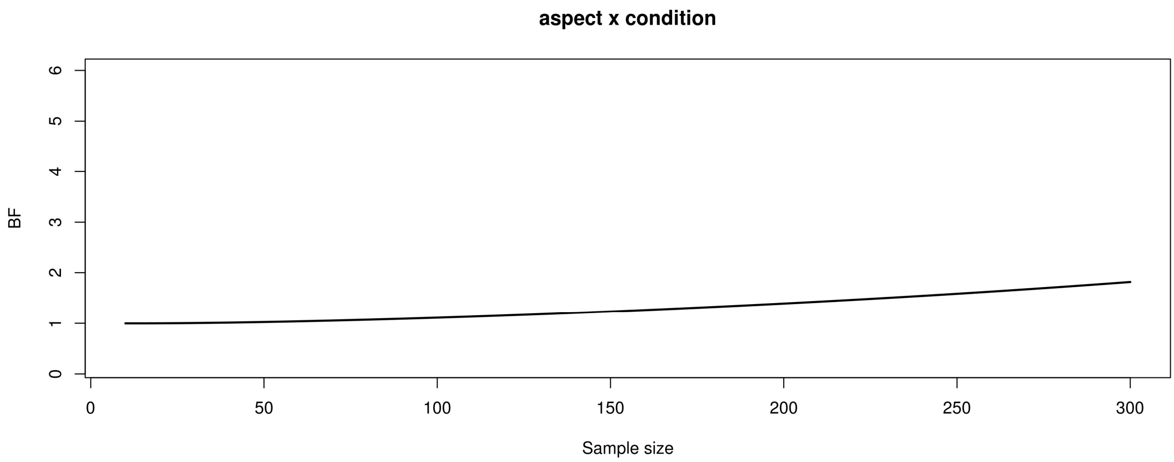


*Figure 15. Bayes Factors for accuracy on the incorrect experimental items*

### 11.5.3 Discussion

Given these outputs, two conclusions can be drawn. First of all, we saw that the current results do not support neither the alternative nor the null hypotheses we formulated. This certainly is an interesting outcome, as the analyses based only on the p-values suggest that it is the null (i.e. no influence of the tense-aspect combination on either reaction times or accuracy) that should be accepted. However, the Bayes Factor analysis does not support this view and suggests that the current study does not falsify the alternative hypothesis. Therefore, more research and perhaps a different experimental approach are needed.

Secondly, as for the question of whether collecting more data would change the results and allow for falsification, we can conclude that even with a large number of participants we would not find the support for the expected effect on reaction times for either the correct or the incorrect experimental items. The effect on accuracy for the incorrect items would also not reach a required level of support. On the other hand, the analysis suggests that collecting more data could provide support for the alternative hypothesis for the model of accuracy on the correct experimental items. However, the estimated required sample size of over 230 participants means that continuing the study would be extremely costly in terms of resources. What is more, given that this finding is only an estimate and that the responses were highly biased towards correct answers, and it is not guaranteed that increasing the sample size would yield significant results. We must also highlight that we arrived at this estimate for only one out of four models, and did not find that a larger sample size would increase the support for the main hypotheses of the study, regarding the influence of tense aspect combinations on the reaction times. Taken together, the results of the analysis presented in this section suggest that much more valuable insights could be gained by testing the hypotheses in a different way rather than simply continuing data collection for the current study.

## 11.6 Discussion

As we have seen, the results of the study go against the hypotheses we stated at the beginning of the chapter. We did not see an effect of aspect on either the reaction times or on the accuracies of the judgements. In addition, we also failed to observe an effect of form frequency. There are a few potential reasons for these outcomes of the study, which we will discuss below.

First of all, it might be argued that the design of the study, especially the decision to conduct it online, affected the results. The differences could simply not be detected, because our instrument – the experimental platform – was not sensitive enough. However, as we discussed at the beginning of the chapter, recent studies investigating the precision of the available tools for conducting research online show that it is not unreasonable to conduct reaction time research online.

Another possible criticism could be that the point of making the decision was not isolated enough. That is, the measures we obtained were reaction times of evaluating how well a given verb fits the sentence, and the fact that participants had to re-read the sentence, which also takes time, may have masked the differences in the speed of reactions. However, we specifically designed the study in such a way, that the participants saw the sentence first, giving them enough time to familiarize themselves with the context, which should eliminate the need to re-read the sentence on the decision screen. They could have, of course, still 'consult' the sentence, looking for the cues allowing them to evaluate how well the verb fits the sentence, but it follows from our hypotheses that it should happen more in the situation where the verb is not in its default aspect. This is not to say that it would not be valuable to test these hypotheses using a different design and a more precise instruments – perhaps a high-precision button box in the lab. After all, the scientific confidence is based on repeated results.

Given what we discussed so far, we could conclude of course, that the reason why the participants were not affected by the aspect of the verb used in the sentence was because the perfective is not the default option. The fact that it appears more often in the past is irrelevant and has no influence on the linguistic behaviour of the speakers. This view, however, is not supported by the Bayes Factor analysis we presented, which suggests that the results do not support the null hypothesis, but rather that the study was inconclusive. What is more, this conclusion goes against the other findings, discussed in Chapter 2, which clearly show that frequency of cooccurrence is perhaps the most important factor affecting both processing and production. In addition, this conclusion would be incoherent with the results of the other study conducted as part of this dissertation. In Chapter 6, we showed that the aspect-concrete model does indeed predict the choices people made. It is definitely worth discussing possible sources of these conflicting findings.

One reason could be that we only used a small subset of all the cues that the model was trained on. However, as we saw in the chapters discussing the aspect-concrete model, other than phasal verbs, tense was the main predictor. In addition, simulations based only on the subset of data that contained exclusively past tense sentences showed that the model fails to predict imperfective, which is most likely related to the fact the main predictors for imperfective in the past were the relatively infrequent adverbial cues. This simulation shows that other cues that we annotated for did not contribute enough for the model to go against the 'default' perfective verb in the past tense.

Another possible explanation is related to the differences between the designs of our two studies – and linguistic dimensions they forced the focus on. In the decision task, we explicitly asked participants to make a choice. Therefore, we can say that the aspect-concrete model predicts the behaviour in a very specific situation – when the participants are forced to choose between the two available options. Since these options differed only in aspect, we can say that the decision task made aspect a more salient dimensions of language and the participants made a conscious choice between two ways of conceptualizing events available in Polish. In the judgement task, this dimension was less prominent if not completely hidden. The participants did not focus on selecting between two available forms. Instead, they only had to evaluate how well a verb fits the sentence, and there are many levels of judgement involved in such a task. The verb could be semantically incoherent (as in our filler sentences), the agreement between the subject and the verb could be violated, the verb could take a wrong number of objects, the verb could be used in a wrong collocation etc.

This in turn points us to a conclusion we also arrived at when discussing possible verb clusterings in Chapter 9. We showed that in our sample, the verbs can be grouped together into aspectual classes based on the concrete variables, but only if a subset of particular variables are taken into account. Trying to cluster verbs based on all the features at once resulted in messy, incoherent groupings. In other words, the findings suggest that what features should be taken into account is necessarily related to the linguistic task at hand. If the task is to make a choice between aspects, we focus on the usage patterns of aspects as categories. If the tasks do not entail a choice of this kind, it might be more helpful to make use of other things we know about how language is used.

To reiterate, the linguistic task we are faced with might be an important reason that affects behaviour and the choices we make in different tasks might be explained by different dimensions of language use. Whereas this is not a direct conclusion from the two studies, and this hypothesis should definitely be tested in a more direct way, this line of reasoning could help explain why the native speakers can 'do away' with aspect in everyday use but when comparing the sentences directly, the patterns of aspectual use become relevant and the notions of completeness and ongoingness that are good approximations of meaning for most of the members are so compelling.

# Chapter 12. Summary and Discussion

We started the dissertation with a description of a very particular linguistic problem – the use of grammatical aspect in Polish. While rarely problematic for native speakers, the choice the between perfective and imperfective aspect is a much more complicated issue for linguists and L2 learners. Inspired by the structuralist school of linguistics, research on aspect has focused mainly on determining the meaning underlying the aspectual classes. After all, we make choices in language to convey meanings – and if we can describe what we mean by using either the perfective or the imperfective, we can answer the question of why speakers choose to use them in particular contexts.

However, we have seen that the attempt to define the meaning of aspects – even though it has a long tradition in literature – has failed to some extent. The abstract semantic labels, such as 'ongoing' vs 'complete', 'internal' vs 'external perspective' or 'bounded' vs 'unbounded', can be said to be useful approximations, but they mask nuances in meaning. Focusing on these nuances on the other hand, leads to more and more granular categories – some even argue that the binary opposition only applies to a certain subset of verbs only.

On the other hand, there are clear distributional reasons to distinguish two aspectual classes. For instance, only the imperfective can be used in the present tense. What is more, there is also experimental evidence suggesting that these two dimensions are distinguished by the speakers and entail different conceptualizations of events. How then, do native speakers of Polish learn to use the perfective and imperfective aspects? And how is usage related to the elusive aspectual semantics? In this dissertation, we attempted to answer these questions by grounding our research in usage-based theory of language as well as theory of learning.

What usage-based theory highlights, first and foremost, is that we as speakers are not born with or given any rules or definitions pertaining to language use. Instead, what is available to us, is the input – thousands of utterances we hear every day. Importantly, it is also assumed that these utterances carry enough information to actually learn how to use a language – including the use of more abstract grammatical categories. It has been shown over and over again that even though languages offer incomputable amounts of possible ways of conveying thoughts, what speakers actually say rarely exploits that freedom. In fact, the analysis of actual usage shows clear distributional patterns – frequency biases in occurrence of certain things in the context of other things.

The conclusion from the literature also points to another very important finding – these patterns are not just statistical curiosities of language. They are in fact the basis of learning. In other words, the fact that language is structured enables us to master its use. However, except for a few notable exceptions, linguists working in the usage-based paradigm have rarely offered an explanation of what cognitive mechanisms are involved in the process, and how exactly speakers use the available usage patterns to construct the grammar of their native language. Usually, they limit themselves to stating that these mechanisms should be domain-general, meaning that the cognitive abilities we use to solve other problems could also be employed to language learning.

In this dissertation however, in addition to investigating the usage patterns of aspect, we also explored the question of the possible mechanisms involved in the acquisition of such patterns. In particular, we focused on error-driven learning – an approach with roots in learning theory. The studies in this area of research show indeed that the basic principles of learning are related to noticing and utilizing the patterns of cooccurrence in the environment. In short, it has been demonstrated that all animals – including humans – are able to form expectations of possible outcomes given a set of cues. In other words, if the contingencies between them are reliable enough, we are able to learn that the presence of a certain cue signals a certain outcome, which in turn can drive behaviour. While the principle itself might sound rather simplistic, the applications of such a simple model of learning are quite impressive. From pigeons learning which light colour signals the occurrence of food, to humans learning how to recognize diseases, the studies have indicated that while most likely not the only principle involved, the exploration and exploitation of distributional patterns of co-occurrence is certainly a key one.

Importantly, this model of learning has been shown to be applicable to language learning as well. If we treat the occurrence of certain elements of linguistic experience as cues, and others as outcomes, then learning the patterns of usage could be modelled using the same principles described above. Such a learning model complements the usage based theory for two reasons. First of all, it is domain-general, since it can be applied to model learning in areas other than language. Second of all, it is based on frequencies of use, but not limited to tallying raw counts. As such, it offers a psychologically plausible explanation of why co-occurrences are so important in language learning and how they are used to form a representation of grammar.

Grounding our research in key findings from linguistics and psychology, we hypothesised that patterns of co-occurrence can also be the driving force behind the use of aspect. The corpus studies we discussed show that there are in fact certain distributional biases that speakers potentially utilize to learn how to use aspect. Therefore, we wanted to establish whether the linguistic behaviour of speakers – the aspectual choices they make – can be modelled using the principles of error-driven learning.

We started out by conducting corpus-based learning simulations. First, we prepared a sample of 18 verbs – or 9 aspectual pairs. For each of the verbs, we annotated 100 discourse chunks taken from a corpus, each containing the target sentence and a window of context. The variables we used can be divided into two categories, which we called concrete and abstract. The concrete categories covered both the elements of the context and the properties of the verbal phrase, for example the tense, the auxiliary verbs, the number, person and gender of the agent in the sentence. On the other hand, the abstract categories were based on the semantic distinctions between the imperfective and the perfective aspect made in the literature – such as boundedness and perspective, to name just a few.

Using this dataset, we conducted learning simulations using a computation implementation of the learning principles we discussed in the theoretical chapters – the Naïve Discriminative Learner. Each of the simulations had a different set of outcomes. First, we wanted to know whether individual verbs can be predicted using the set of concrete variables. A good performance of such a model would suggest that the category of aspect might be superfluous, and that the choices speakers make can be attributed to the usage patterns of the verbs, without the need of ever thinking about the more abstract grammatical categories to which they might belong. Secondly, we tested whether a model is able to learn aspect based solely on the concrete variables. This model directly explored the question of whether the usage contains patterns distinct enough for the speakers to learn how to use perfective and imperfective. A good performance of this model would suggest that the speaker might not usually engage in pondering the details of semantics when making aspectual choices. Instead, it could be claimed that the majority of the choices is more or less automatic and driven by the usage patterns. Finally, we empirically tested how reliable the semantic distinctions are, by modelling the aspectual choices using the abstract variables. A good performance of the final model would suggest that these labels proposed by linguists are in fact quite useful when it comes to describing the aspects.

Interestingly, all three models performed well. In other words, the results of the simulations alone supported not only the view that speakers of Polish can learn how to use aspects by extracting the meaning and the view that they can do so solely based on the patterns of usage, but also the view that they might be able to do away with the category of aspect altogether. This certainly was a curious outcome, and in our interpretation, it entails a very important methodological conclusion. Namely, the results of a corpus-based study need to be supported by other sources of evidence.

As the first piece of evidence against one of the models we can take the analysis of the interannotator agreement scores we calculated for the abstract variables. We asked one more native speaker of Polish to annotate a subsample of the dataset using the set of abstract variables and calculated both the percentage of agreement on each label as well as a more precise kappa score. The results indicate that interpreting the chunks is quite subjective and the labels are difficult to apply. While this disagreement can be taken as a signal that the variables were not defined properly, the comparisons of interannotator agreement in other studies speak against that view. (Divjak et al., 2015; Dziob et al., 2017), in which annotators had to label language data using abstract semantic labels also showed that interpretations often differ, making labelling very unreliable.

However, we need to admit here that the concrete variables we annotated for can also be considered a level of abstraction over the chunks. After all, we did not use the exact words that appeared in the chunks as cues in neither the lemma-concrete nor the aspect-concrete model. Instead, we extracted semantic dimensions such as tense and mood, as well as the number and gender of the agents, recipients and objects in the chunks. We argue however, that this level of abstraction requires much less subjective interpretation, if any at all. What is more, the information captured by the variables – the when, the who and how many– is basic enough that it can be easily extracted by native speakers.

The next step we took to corroborate the models involved conducting a behavioural study in which we asked the participants to make choices between the perfective and imperfective counterparts in smaller excerpts of the discourse chunks we used in the corpus-based simulations. Then, we used the support weight from each of the models as a predictor of whether the speakers would use the aspectual variant of the verb that was originally used in the sample or not. The mixed effects logistic regression models we fitted indicated that only the aspect-concrete model was able to significantly predict the choices that the speakers made. The lemma-concrete model turned out to be insignificant, whereas the aspect-abstract model was not only insignificant but also had a negative coefficients, suggesting that the speakers were in fact less likely to use the aspect the stronger it was supported by this model.

What these results suggest is twofold. First of all, it seems that the abstract semantic labels are not good predictors of the aspectual choices speakers make. Or, to put it less strongly, the individual interpretations of the semantic details of chunks may not generalize to other people. In turn, a model based on such interpretations performs badly when it comes to predicting the linguistic behaviour of a sample of a population of speakers. This outcome is consistent with the conclusions of the comparison of annotations we discussed above.

On the other hand, the good performance of the aspect-concrete model suggests that it is possible to learn how to use the imperfective and the perfective aspects based on their usage-based, distributional patterns. However, these patterns are certainly not deterministic, as the model did not correctly predict all of the answers that the participants provided. This suggests not only that there is a certain degree of freedom when it comes to using the aspects but also that our – after all, quite naive model – was in some ways limited.

As for the question of how much freedom there really is when it comes to using aspects, the analysis of the proportion of agreement for each question revealed that in the most of the contexts the majority of people tend to agree on one option only. There are however contexts where the possibilities are more balanced and both aspects can be used. This finding corroborates the results of Janda and Reynolds (2019), who show that the aspectual contexts in Russian can also be said to form a continuum with certain contexts having an almost deterministic influence on the aspectual choices at the one end and contexts that allow both options at the other.

What then, has the model established to be the most important cues for predicting the aspectual classes? To answer this question, we first listed both the top positively associated cues for each of the outcomes as well as the top negatively associated cues. To put it simply, the positively associated cues are the 'yes' signals – they suggest that we should expect a given outcome. On the other hand, the negatively associated cues are the 'no' signals – according to the model, we should not expect to see the outcome in the context of that cue. It was not surprising to see that most of the top positively associated cues for perfective were also the top negatively associated cues for the imperfective. This is because in the case of binary choice between the outcomes, the increase of the association weight between a cue and one outcome would also automatically decrease the association weight between this cue and the other outcome.

The cues that the model has learned as the most predictive can be divided into two groups: tense and triggers. The perfective aspect was strongly positively associated with the past and the future tense, whereas the imperfective was associated with the present tense. In addition, the model also learned that *zostać* (become.PERF), which introduces passive voice in the past was also strongly positively associated with the perfective. On the other hand, the imperfective was positively associated with the presence of two other cues. First of all, the auxiliary "be", which is an obligatory element for constructing the future compound tense where only the imperfective is used. Secondly, the presence of phasal verbs, such as *zacząć* (begin.PERF), which also can only be followed by imperfective verbs.

As we can see, the cues that the model distinguished are rather coarse-grained and describe very general and quite obvious patterns of use. After all, the perfective cannot be used in the present tense and the fact that speakers use the imperfective with phasal verbs is hardly surprising for any aspectologist. Nonetheless, we were still able to predict the choices people made during the behavioural study using such an uncomplicated model. This certainly is an interesting finding as it suggests that such general rules are enough to master the use of aspectual categories at least to some extent.

However, as we mentioned before, the Naïve Discriminative Learner was not able to correctly predict all the choices that the participants made. The analysis of errors of the aspect-concrete model provided valuable insights into the limitations of the current approach as well as the ones of the model. First of all, we should mention the fact that NDL only had access to the dimensions we provided, which as we have already said, were still some form of generalization. The model had no access to the usage patterns of individual verbs and the individual words as cues. It can be argued however, that since the linguistic experiences of the speakers in the real world are not limited in this way, they learn usage patterns on many levels of granularity. Possibly, this entails that the speakers can also use different cues in different situations to determine which aspect to choose. This of course would need to be confirmed experimentally.

Secondly, the current implementation of NDL treats all cues as equal and therefore cannot learn any dependencies between them. However, certain cues might only become important in some particular contexts. For this reason, learning cue dependencies would be beneficial as it would increase the predictive power of the model. The learning outcomes of the pastAspect-concrete model we discussed in Chapter 10 support this view. We showed that while the model trained on all examples in the dataset did not distinguish adverbials as particularly important for predicting the imperfective, a model trained only on a subset of the past tense sentences did. This suggests that these cues become particularly predictive in the context of past tense. These results can be explained by the distribution of tense-aspect combinations. Because the imperfective occurs in the present much more often than in the past, other elements of the context that are also used in present tense sentences might become positively associated with the imperfective. Since the change of the weights between the cues and outcomes takes into account the sum of weights of all the cues that are present, these other elements may block the learning of the association between the adverbials and the imperfective in the past. The relative infrequency of adverbials in general does not help to remedy this situation. However, if the model is trained only on the past tense sentences, the competition between the adverbials and other cues is more fair.

Despite these limitations, the results of the current studies we discussed so far indicate that it is indeed possible to learn how to use aspectual categories based on their usage patterns, at least to some significant extent. However, the speakers do not have access to these categories at the beginning of language acquisition. If the imperfective and perfective do emerge in the process, they must therefore be based on generalizations over the usage patterns of individual verbs. In other words, since the aspectual classes are not given, the similarities between the ways we use different verbs should serve as the basis of distinguishing the imperfective and perfective as categories. To test whether it would be possible to build these categories using the learning outcomes of the NDL-based model we built two clustering solutions. As input, we used the association weight matrix taken from the lemma-concrete model, which was trained to predict the lemmas of individual verbs based on the subset of the concrete cues. Importantly, the model did not know about any relationships between the verbs, including whether or not they form aspectual pairs. We showed that aspectual classes cannot be obtained using all the cues. On the other hand, using only a subset of cues that were strongly predictive in the aspect-concrete model yielded a perfect separation into imperfective and perfective verbs.

These results provided two important insights. First of all, they suggest that when speaking of building abstractions and generalizations based on the similarities of items, we must be more precise. Items can be similar to each other in many ways and the reason why we group them together may not be irrelevant for which dimensions we take into account to establish the similarity. Secondly, the clustering results further supported the conclusion that the relationship between tenses and aspects is particularly important. Knowing that the perfective is usually used in the past and the future, and the imperfective is mainly used in the present does not only allow to correctly use the aspects in many cases, but also to categorize verbs into aspectual classes.

Taken together, the results of the studies conducted and discussed so far highlight the particular importance of the relationship between tense and aspect. While the fact that only the imperfective can be used in the present tense was a given, both the perfective and imperfective are usually presented in the aspectologist literature as equally likely choices in the past and the future. However, the corpus modelling we have conducted shows something to the contrary. Given the results we have obtained we can propose that the perfective is in fact the default choice in the past and the future tenses. This is of course not to say that the speakers of Polish cannot or do not use the imperfective in tenses other than the present, but rather to underline that they are less likely to do so. What is more, as indicated by the pastAspect-concrete model, the use of the non-default imperfective is often signalled by additional cues, such as temporal adverbials.

These conclusions fit well into cognitive linguistic theory. Langacker (2001) points out similar biases in usage in other languages and constructions. He suggests that if two or more ways coexists in language, the one that is used more frequently should be treated as canonical. That is, they represent the usual ways in which the speakers think and talk about certain situations. The non-canonical forms, while not incorrect, are more rare. He also points out that the non-canonical forms are usually signalled by using additional elements in the sentence, which again resonates well with our findings.

The final study we conducted focuses on providing psycholinguistic evidence for whether perfective can indeed be taken as a canonical option in the Polish past tense. Following the conclusions from the available experimental literature we hypothesised that the default, more frequent variants should be easier to process, since they are more expected. This relative ease of processing should be reflected in the behaviour of the speakers. In the current study, we tested whether native speakers of Polish would evaluate the sentences containing the canonical tense aspect combination faster and more accurately compared to the non-canonical variant. In addition, we also investigated if including an additional temporal cue in the context of the non-canonical variant would make processing less difficult and, as a consequence, decrease the reaction time and increase accuracy.

In order to answer these research questions, we implemented two versions of a grammatical judgement task, where we asked the participants to decide whether a gapped sentence they saw on a screen would be correct if it was filled with an inflected form of a verb provided below. All of the experimental sentences were in the past tense, which was introduced by a time adverbial, which occurred before the gap. In the first task, we measured the time it took to evaluate the fitness of the perfective and imperfective verbs as well as whether the participants answered correctly. In the second task, in addition to the adverbial that introduced the past, we also included adverbials that were associated with each aspect respectively.

The results did not confirm our expectations. Not only did we find no difference between the reaction times on the perfective and the imperfective sentences, but apparently the inclusion of an adverbial also did not have any influence on the reaction times for either sentence type. Interestingly, we also did not observe any influence of frequency of the verb form on the reaction times, which would be the most basic effect we could expect, given the vast amount of literature on the effects of frequency in language processing.

This outcome is difficult to interpret in the light of the findings we discussed throughout the dissertation. One possible interpretation is of course that none of the aspectual forms can be treated as default or canonical. It would certainly be worthwhile to explore the frequencies of tense-aspect combinations in a larger corpus sample – or the entire corpus – and their influence on the association weights. Because we used a manually annotated sample, we were limited to the distributions that were available in this small dataset.

On the other hand, we were able use the aspect-concrete model to predict the choices people made in a previous behavioural study. It is worth noting that there was a qualitative difference between the tasks. In the first one, the participants were given an explicit choice, and the question behind it was: which aspect fits the context better? In the last study we conducted, the participants were only asked if the provided verb fits the sentence. Given that the participants were native speakers, it might be that their expertise thwarted the influence of the distributional patterns when they were simply judging the sentences, but when asked to choose, they followed the usage in most cases.

However, such an interpretation fits the results of other studies looking at the results of frequencies on the reaction times rather poorly. The findings from other studies that we discussed in Chapter 11 suggest that even expert native speakers need more time to process the less canonical linguistic patterns. This forces us to be more suspicious towards the results obtained by us. We also need to highlight that the study was conducted online and not in the laboratory. Even though the state-of-the art experimental platform we used allows for collecting time sensitive data, it is possible that the results obtained on a more sensitive response button box would reveal more detailed patterns in the participants' answers. Finally, the design itself might have influenced the results. The participants were asked to read a gapped sentence first and only then move to the next screen where they saw the same sentence and a verb to evaluate. This was done in order to measure the reaction on the decision only, and to exclude the time needed to read and process the sentence itself. However, it is possible that the participants could have ignored the instructions and moved to the second screen without reading the sentence in the first one out of tiredness or to save time. A possible improvement in the design might involve triggering the change of screens automatically, only after a certain amount of time. This might push the participants to read the sentence more carefully on the first screen. We must note however, that this makes the study longer and more demanding in terms of attention.

All in all, the studies we conducted and discussed in this dissertation point us to a few conclusions. First of all, they indicate that the knowledge of when to use the imperfective and the perfective aspects can be acquired from the patterns of usage, without the need to formulate precise semantic definitions. This of course is not to say that perfective and imperfective verbs do not carry *any* meaning or that they do not affect the conception of events. However, the finding suggest that these semantic contrasts are not always important when uttering a sentence. In other words, even though aspectual categories can be used to highlight different nuances of meaning, they might not always be used to do so. The choices speakers make can be more reliably explained by patterns of use – that is, speakers tend to choose one aspect over the other because it frequently occurs in a given context rather than because it carries a specific meaning. An interesting question arising from this finding is whether deviations from patterns are more meaningful – in cases where speakers consciously choose a dispreferred, less frequently occurring aspect, the semantic contrasts might become particularly important.

As for the question of aspectual meaning itself, it seems that the relationship between the semantics of aspect and the semantics of tense is certainly worth investigating, especially from error-driven learning perspective. Given the distribution of tense-aspect combinations and the principles of learning, we might formulate the hypothesis that some aspects of tense meaning are never fully separated from the aspectual cues within the verbs. This in turn might help explain why the notions of ongoingness and completeness are so prevalent in the aspectual literature and so easily adopted by native speakers.

Our findings also indicate that the error-driven learning is a good candidate for the 'domain-general mechanism' proposed by the usage-based theory of language. As we have discussed, it is necessarily based on frequency of co-occurrence in the context, which – as posited by linguists – also seems to be the basis of human language learning. What is more, because these principles can be translated into computational algorithms, the error-driven learning paradigm offers a psychologically plausible way of modelling the usage patterns in the data. As such, the Naïve Discriminative Learner can be thought of not as much as another corpus method of calculating associations, but as a tool for conducting learning simulations.

However, following Rescorla (1984), 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. Categorization and generalization are the first other mechanisms that immediately come to mind. There are also individual differences in learning strategies and memory capacity, which also influence how individuals tackle the daunting task of mastering a language. The results of our studies indicate, however, that learning usage patterns on the basis of congruency is an important part of language learning, certainly worth exploring further.

Finally, we want to discuss a few methodological implications that follow from the current study. As we have seen, aspect is one of those problems in linguistics that seem to be unsolvable using only traditional tools. The conclusions researchers arrived at based on the analysis of a few, usually carefully selected, examples are definitely valuable, but ultimately empirical methods must be used to determine which of the solutions and models offered – if any – resemble what speakers actually do. Importantly however, as indicated by our results, the evidence we use to evaluate the models should not come from only one source. As we have seen in the first part of this dissertation, our corpus-based learning simulations supported all three models of explaining aspectual use, 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 validating them. Similarly, even though our models, supported by theoretical considerations could be taken as evidence that the perfective should be treated as default choice in the past tense, the experiment we conducted shows us that the case is not that simple and requires additional investigation.

We hope that the work we conducted and the conclusions we reached further amplify the call for an interdisciplinary approach to the study of language, already being expressed (Ellis, 2006; Ramscar & Yarlett, 2007; Stefanowitsch, 2011; Dąbrowska, 2016; Divjak, 2016; Milin, Divjak, et al., 2017). 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 and experiments to allow for formulation and falsification of hypotheses, rather than rely on intuition to merely generate them. 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.

# Bibliography


Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*, 239-273.

Anderson, J. R. (2000). *Learning and memory : an integrated approach* (2nd ed.). Hoboken, NJ : Wiley, 2000.

Anderson, S., Matlock, T., Fausey, C., & Spivey, M. J. (2008). In B. Love, C, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Anderson, S. E., Matlock, T., & Spivey, M. (2013). Grammatical aspect and temporal distance in motion descriptions. *Frontiers in Psychology*, *4*.

Anwyl-Irvine, A., Dalmaijer, E., Hodges, N., & Evershed, J. (2020). Online Timing Accuracy and Precision: A comparison of platforms, browsers, and participant's devices.

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth - children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107-129.

Arnon, I., & Snider, N. (2010). More than words: frequency effects for multi-word phrases. *Journal of Memory & Language*, *62*(1), 67-83.

Baayen, H. R., & Milin, P. (2010). Analyzing reaction times. *International journal of psychological research*, *3*(2), 12-28.

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada*, *11*(2), 295-328.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the Combinatorial Explosion: An Explanation of n-gram Frequency Effects Based on Naive Discriminative Learning. *Language and Speech*, *56*(3), 329-347.

Baayen, R. H., Milin, P., Durdević, D. F., Hendrix, P., & Marelli, M. (2011). An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning. *Psychological Review*, *118*(3), 438-481.

Baguley, T., & Kaye, W. S. (2009). Review of Understanding psychology as a science: An introduction to scientific and statistical inference. http://psychologicalstatistics.blogspot.com/2009/11/r-functions-for-dienes-2008.html

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning: The Effect of Familiarity on Children's Repetition of Four-Word Combinations. *Psychological Science (0956-7976)*, *19*(3), 241-249.

Bańko, M. (2002). *Wykłady z polskiej fleksji*. Wydawnictwo Naukowe PWN.

Barlow, M., & Kemmer, S. (2000). *Usage-based models of language*. Stanford, Cal. : CSLI Pubs, c2000.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1 - 48.

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Christiansen, M. H., Larsen-Freeman, D., Croft, W., & Schoenemann, T. (2009). Language is a complex adaptive system: Postion paper. *Language Learning*, *59*(December), 1–26-21–26.

Benko, V. (2014). Aranea Yet Another Family of (Comparable) Web Corpora. *Proceedings of 17th International Conference Text, Speech and Dialogue.*, 257-264.

Bergs, A. (2010). Expressions of futurity in contemporary English: a Construction Grammar perspective. *English Language and Linguistics*, *14*(2), 217-238.

Bermel, N. (1997). *Context and the lexicon in the development of Russian aspect*. University of California Press.

Bott, O., & Gattnar, A. (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.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ (San Francisco, CA)*, *8*, e9414-e9414.

Broda, B., & Piasecki, M. (2011). Parallel, Massive Processing in SuperMatrix: a General Tool for Distributional Semantic Analysis of Corpora. *International Journal of Data Mining, Modelling and Management*, *5*(1).

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, *37*(4), 575-596.

Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711-733.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, *25*(4), 438-446.

Clifton Jr, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Mem Cognit*, *32*(6), 886-895.

Comrie, B. (1976). *Aspect*. Cambridge University Press.

Dąbrowska, E. (2016). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics*, *27*(4), 479-491.

Dickey, S. (2016). Lexical and Grammatical Aspect. In N. Riemer (Ed.), *The Routledge Handbook of Semantics* (pp. 338-353). Routledge.

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 104-123.

Divjak, D. (2011). Predicting aspectual choice in modal constructions. A quest for the Holy Grail? In M. Grygiel & L. A. Janda (Eds.), *Slavic Linguistics in a Cognitive framework* (pp. 67-85). Peter Lang.

Divjak, D. (2016). Four challenges for usage-based linguistics. *Change of Paradigms - New Paradoxes Recontextualizing Language and Linguistics*.

Divjak, D. (2017). The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence From that-Clauses in Polish. *Cogn Sci*, *41*(2), 354-382.

Divjak, D. (2019). Frequency in language : memory, attention and learning. In. Cambridge: Cambridge University Press.

Divjak, D., & Caldwell-Harris, C. (2015). Frequency and entrenchment. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 53-75). De Gruyter.

Divjak, D., & Gries, S. T. (2006). Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, *2*(1), 23-60.

Divjak, D., Levshina, N., & Klavan, J. (2016). Cognitive Linguistics: Looking back, looking forward. *Cognitive Linguistics*, *27*(4), 447-463.

Divjak, D., Milin, P., Ez-zizi, A., Józefowski, J., & Adam, C. (2021). What is learned from exposure: An error-driven approach to productivity in language. *Language, cognition and neuroscience*, *36*(1), 60-83.

Divjak, D., Szymor, N., & Socha-Michalik, A. (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.

Dziak, J., Dierker, L., & Abar, B. (2020). The Interpretation of Statistical Power after the Data have been Gathered. *Curr Psychol*, *39*(3), 870-877.

Dziob, A., Piasecki, M., Maziarz, M., Wieczorek, J., & Dobrowolska-Pigoń, M. (2017). Towards Revised System of Verb Wordnet Relations for Polish. *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets*.

Ehrlich, S. F., & Rayner, K. (1981). *Contextual effects on word perception and eye movements during reading* [doi:10.1016/S0022-5371(81)90220-6]. Elsevier Science.

Ellis, N. C. (2002). FREQUENCY EFFECTS IN LANGUAGE PROCESSING: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in second language acquisition*, *24*(2), 143-188.

Ellis, N. C. (2006). Language Acquisition as Rational Contingency Learning. *Applied Linguistics*, *27*(1), 1-24.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb Aspect and the Activation of Event Knowledge. *J Exp Psychol Learn Mem Cogn*, *33*(1), 182-196.

Filip, H. (2012). Lexical Aspect. In R. I. Binnick (Ed.), *The Oxford Handbook of Tense and Aspect* (pp. 721-751). Oxford University Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627-635.

Forsyth, J. (1970). A grammar of aspect : usage and meaning in the Russian verb / J. Forsyth. In. Cambridge: Cambridge : Cambridge University Press.

Garey, H. B. (1957). Verbal Aspect in French. *Language*, *33*(2), 91-110.

Gluck, M. a., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227-247.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning Argument Structure. *Cognitive Linguistics*, *3*(2004), 289-316.

Golshaie, R., & Incera, S. (2020). Grammatical Aspect and Mental Activation of Implied Instruments: A Mouse-Tracking Study in Persian. *Journal of psycholinguistic research*, *50*(4), 737-755.

Grzegorczykowa, R. (1997). Nowe spojrzenie na kategorię aspektu w perspektywie semantyki kognitywnej. In R. Grzegorczykowa & Z. Zaron (Eds.), *Semantyczna struktura słownictwa i wypowiedzi* (pp. 25-38). Wydawnictwa Uniwersytetu Warszawskiego.

Grzegorczykowa, R., Laskowski, R., & Wróbel, H. (1999). *Gramatyka współczesnego języka polskiego. Morfologia*. Wydawnictwo Naukowe PWN.

Hoenig, J. M., & Heise, D. M. (2001). The Abuse of Power. *The American Statistician*, *55*(1), 19-24.

Hopper, P. (1987). Emergent Grammar. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139-157.

Hopper, P. (1998). Emergent grammar, the new psychology of language. In M. Tomasello (Ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Linguistic Structure* (pp. 155-157). Erlbaum.

Janda, L. A. (2004). A metaphor in search of a source domain: The categories of Slavic aspect. *Cognitive Linguistics*, *15*(4), 471-527.

Janda, L. A., & Reynolds, R. J. (2019). Construal vs. redundancy: Russian aspect in context. *Cognitive Linguistics*, *30*(3), 467-497.

Jiang, N. (2012). Conducting reaction time research in second language studies / Nan Jiang. In. London: Routledge.

Jurkiewicz-Rohrbacher, E. (2019). *Polish verbal aspect and its Finnish statistical correlates in the light of a parallel corpus*. Helsingin yliopisto.

Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. *M.R. Jones (Ed.). Miami Symposium on the Prediction of Behavior, 1967: Aversive Stimulation. Coral Gables, Florida: University of Miami Press*, 9-31.

Kieraś, W., & Woliński, M. (2017). Morfeusz 2 – analizator i generator fleksyjny dla języka polskiego. *Język Polski*, *XCVII(1)*, 75–83.

Kleiman, G. M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition*, *8*(4), 336-344.

Klimek-Jankowska, D. (2008). On an Analogy between English Nominal and Polish Aspectual Manifestations of Genericity. SuB12, Oslo.

Klimek-Jankowska, D. (2012). Imperfective and Perfective Habituals in Polish: A Bi-Directional OT Account of Variation and Ambiguity. *Journal of Logic, Language and Information*, *21*(1), 31-52.

Koranova, I., & Bermel, N. (2008). From Adverb to Verb: Aspectual Choice in the Teaching of Czech as a Foreign Language. In C. Cravens, M. U. Fidler, & S. C. Kresin (Eds.), *Between Texts, Languages and Cultures: A Festschrift for Michael Henry Heim* (pp. 53-69). Slavica.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. F. A. K. van Benthem, & P. van Emde Boas (Eds.), *Semantics and Contextual Expression*. Foris Publications.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1-26.

Lakoff, G. (1990). The Invariance Hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics*, *1*, 39-74.

Lakoff, G. (1991). Cognitive versus generative linguistics: How commitments influence results. *Language and Communication*, *11*(1-2), 53-62.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *Mar;33*(1), 159-174.

Langacker, R. W. (2001). The English present tense. *English Language and Linguistics*, *5*(2), 251-272.

Laskowski, R. (1996). Aspekt a znaczenie czasowników (predykaty zmiany stanu). In *Studia z leksykologii i gramatyki języków słowiańskich* (pp. 41-50). Instytut Języka Polskiego (PAN).

Laskowski, R. (1999). Kategorie morfologiczne języka polskiego - charakterystyka funkcjonalna. In R. Grzegorczykowa, R. Laskowski, & H. Wróbel (Eds.), *Gramatyka współczesnego języka polskiego* (pp. 150-224). Państwowe Wydawnictwo Naukowe.

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age and memory for discourse. *Journals of Gerontology*, *40*(6), 737-747.

Łaziński, M. (2020). *Wykłady o aspekcie polskiego czasownika*. Wydawnictwa Uniwersytetu Warszawskiego.

Łaziński, M., & Wiemer, B. (1995). Terminatywność jako kategoria stopniowalna. *Prace Filologiczne*, *XL*, 99–126.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*(5), 663-672.

Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, *29*(2), 83-112.

Milin, P., & Divjak, D. (2019). About Language, Learning and Tomatoes. https://outofourminds.bham.ac.uk/blog/8

Milin, P., Divjak, D., & Baayen, R. H. (2017). 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, P., Divjak, D., Dimitrijević, S., & Baayen, R. H. (2016). Towards cognitively plausible data science in language research. *Cognitive Linguistics*, *27*(4), 507-526.

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLOS ONE*, *12*(2), e0171935-e0171935.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393-424.

Morrow, D. G. (1985). Prepositions and verb aspect in narrative understanding. *Journal of Memory and Language*, *24*(4), 390-404.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory : nativelike selection and nativelike fluency. *Language and Communication*, 191-226.

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation*, *44*(1/2), 137-158.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105-110.

Prolific. (2014). In. London, UK.

Pullum, G. K. (1996). Learnability, Hyperlearning, and the Poverty of the Stimulus. *Proceedings of the annual meeting of the Berkeley Linguistics Society*, *22*(1), 498.

Qualtrics. (2005). In. Provo, Utah, USA.

Ramscar, M., & Yarlett, D. (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, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The Effects of Feature-Label-Order and Their Implications for Symbolic Learning. *Cognitive Science*, *34*(6), 909-957.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191-201.

Redington, M., Chater, N., & Finch, S. (1993). Distributional Information and the Acquisition of Linguistic Categories: A Statistical Approach. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*(1), 1-5.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *The American Psychologist*, *43*(3), 151-160.

Rescorla, R. A., & Wagner, A. R. (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, R. J. (2016). Russian natural language processing for computer-assisted language learning. In: UiT The Arctic University of Norway Doctoral Dissertation.

Rice, S., & Newman, J. (2004). Aspect in the making: a corpus analysis of English aspect-marking prepositions. *Language, Culture and Mind. Stanford: CSLI Publications*, 313-327.

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*, *33*(2), 251-289.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and the acquisition of language.* (pp. xii, 308-xii, 308). Academic Press.

Rubenstein, H., Decker, L., & Pollack, I. (1959). Word Length and Intelligibility. *Language and Speech*, *2*(4), 175-178.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, *9*(5), 487-494.

Sadowska, I. (2012). *Polish [electronic resource] : a comprehensive grammar*. Routledge.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928.

Salomon, M. M., Magliano, J. P., & Radvansky, G. A. (2013). Verb aspect and problem solving. *Cognition*, *128*(2), 134-139.

Sasse, H.-J. (2002). Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state?

Schwanenflugel, P. J., & Shoben, E. J. (1985). The Influence of Sentence Constraint on the Scope of Facilitation for Upcoming Words. *Journal of Memory & Language*, *24*, 232-252.

Sekerina, I. A. (2006). Building Bridges: Slavic Linguistics Going Cognitive. *Glossos (Durham, N.C.)*, *8*(fall).

Sekerina, I. A. (2017). Slavic Psycholinguistics in the 21st Century. *Journal of Slavic Linguistics*, *Volume 25*(2), 463-487.

Smith, C. (1986). A Speaker-Based Approach to Aspect. *Linguistics and philosophy*, *9*(1), 97-115.

Smith, C. (1997). *The Parameter of Aspect*. Springer Dordrecht.

Stefanowitsch, A. (2011). Cognitive Linguistics as a Cognitive Science. *Bidirectional Perspective in the Cognitive Sciences*, 296-309.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209-243.

Swan, O. E. (2002). *A grammar of contemporary Polish*. Slavica.

Terry, W. S. (2006). *Learning and memory. Basic principles, processes and procedures*. Pearson.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(3), 443-467.

Van Engen, K. J., Dey, A., Runge, N., Spehar, B., Sommers, M. S., & Peelle, J. E. (2020). Effects of age, word frequency, and noise on the time course of spoken word recognition. *Collabra. Psychology*, *6*(1).

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*, *66*(2), 143-160.

Verkuyl, H. J. (1972). *On the Compositional Nature of the Aspects* (Vol. 12). D.Reidel Publishing Company.

Vujović, M. (2020). *Exploring language learning as uncertainty reduction using artificial language learning* [Doctoral thesis].

Wagner, A. R., & Rescorla, R. A. (1972). Inhibiton in Pavlovian Conditioning: Application of a Theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and Learning*. Academic Press.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding.

Wasserman, E. A., Brooks, D. I., & McMurray, B. (2015). Pigeons acquire multiple categories in parallel via associative learning: A parallel to human word learning? *Cognition*, *136*, 99-122.

Whaley, C. P. (1978). Word—nonword classification time. *Journal of verbal learning and verbal behavior*, *17*(2), 143-154.

Winter, B. (2019). *Statistics for linguists / Bodo Winter*. Milton : Routledge, 2019.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory & Language*, *95*, 36-48.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition*, *12*(1), 31-45.

Wróbel, H. (2001). *Gramatyka języka polskiego*. Spółka Wydawnicza Od Nowa.

Zipf, G. K. (1935). The psycho-biology of language : an introduction to dynamic philology / George Kingsley Zipf. In. Boston, Mass.: Boston, Mass. : Houghton Mifflin.

# Appendices

## Appendix 1: variables used for corpus annotation

| variable main category | variable | variable levels | description |
|---|---|---|---|
| CONCRETE | | | |
| verb | tense | present<br>past<br>future<br>none | is the sentence about past, future or present situation? If not possible to decide (imperatives, infinitives, participles): none |
| | mood | indicative<br>imperative<br>conditional<br>none | indicative: expresses objective, neutral attitude of the speaker (Pisze książkę)<br>imperative: expresses an order, demand or wish (Pisz książkę!)<br>conditional expresses uncertainty, condition: (Pisałby książkę, Byłby pisał, by pisał)<br>Infinitves: none |
| | person | first<br>second<br>third<br>none | |
| | voice | active<br>passive | |
| | cx | aux<br>phasal<br>modal<br>passive<br>no | modal: if the verb is in a construction with a modal verb (móc, trzeba, musieć, powinien itp.)<br>phasal: if the verb is in the construction with a phasal verb (zaczać, skończyć)<br>aux: if the verb is in any other construction<br>passive: if the verb is in passive construction – to capture the auxiliary verb<br>If the verb is embedded in two (e.g. passive + |

| | | | |
|---|---|---|---|
| | | | modal), take the most immediate one, e.g. in Może być napisany values for cx  and cxword are ‘passive’ and ‘być’ |
| | cx_word | word_form<br>none | form of the verb from the constrution (e.g: musi pisać - musi) |
| | reflexiveness | yes<br>no | does the verb have 'się'? |
| sentence/clause | sentence_type | indicative<br>interrogative<br>imperative<br>exclamative | indicative: 'regular' sentences<br>interrogative: questions<br>imperative: pisz!<br>exclamative: hurra, pisze! |
| | clause_polarity | positive<br>negative | is the verb negated?<br>Nie pisze |
| | clause | main<br>dependent | main clause of the sentence or coordinated clause (it can form an independent sentence)<br>dependent: it does not express a complete thought, cannot be an independent sentence, provides additional information or the main clause |
| Agent | Agent_type | noun_phrase<br>pronoun<br>implied<br>none | noun_phrase: pies, pies i kot, ten śmieszny pies z kręconym ogonem<br>pronoun: on, ty, wy, itp.<br>implied: the clause does not have a subject as a separate word, but we can deduce it from the verb form: pisałem wiersz<br>none: the clause has no subject and we can't deduce it from the verb form (e.g.: napisano wiersz or: żeby napisać wiersz) |
| | Agent_person | first<br>second | |

| | | | |
|---|---|---|---|
| | | third | |
| | Agent_number | singular<br>plural<br>none | |
| | Agent_gender | masculine<br>feminine<br>neuter<br>none | we use a 3-way classification. Gender tests:<br>Use singular form and check which demonstrative it fits:<br>ten = masculine<br>ta = feminine<br>to = neuter<br>no subject = none |
| | Agent_case | nominative<br>genitive<br>none | |
| | Agent_animacy | animate_person<br>animate_animal<br>animate_abstract<br>inanimate_object<br>inanimate_substance<br>inanimate_other | animate_person: self-explanatory, e.g.: On, Janek, nauczycielka<br>animate_animal: self_explanatory: pies, kotki<br>animate_abstract: people, but as entities, not individuals, e.g.: firma, mocarstwo<br>inanimate_object: a concrete, countable thing: książka, legitymacja, silnik<br>inanimate_substance: uncountable, fluid thing or similar: wiatr, piasek, fale telepatyczne<br>inanimate_other: other cases, difficult to classify; theoretically animate things that can't act on their own: skóra, roślina |
| Patient | Patient_type | noun_phrase<br>pronoun<br>none | see: Agent_type |
| | Patient _number | singular<br>plural | |

| | | | |
|---|---|---|---|
| | | none | |
| | Patient _gender | masculine<br>feminine<br>neuter<br>none | see: Agent_gender |
| | Patient _case | | |
| | Patient _animacy | animate_person<br>animate_animal<br>animate_abstract<br>inanimate_object<br>inanimate_substance<br>inanimate_other | see: Agent_animacy |
| Recipient | Recipient _type | noun_phrase<br>pronoun<br>none | see: Agent_type |
| | Recipient _number | singular<br>plural<br>none | |
| | Recipient _gender | masculine<br>feminine<br>neuter<br>none | see: Agent_gender |
| | Recipient _case | | |
| | Recipient_animacy | animate_person<br>animate_animal<br>animate_abstract<br>inanimate_object<br>inanimate_substance<br>inanimate_other | see: Agent_animacy |
| adverbials | adverbial | time<br>frequency<br>manner<br>duration<br>none | Look for temporal adverbials that refer to the verb (in the chunk or the sentence)<br>time adverbials situate action in time: in 1999, yesterday, last spring<br>frequency adverbials tell us how often the action happened/is happening: zawsze, często, czasami, nigdy, co trzy lata |
| aspectual triggers | aspectual_trigger | perf<br>imperf<br>none | check if any of the aspectual triggers from the list appears anywhere in |

| | | | |
|---|---|---|---|
| | | | the chunk, provided it is related to the verb.<br>For instance, in:<br>Całą noc pisał. Następnego dnia w ogóle nie mógł się skoncentrować.<br>we would choose none, even though 'całą noc' appears in the chunk. |
| | aspectual_trigger_form | | the aspectual trigger as it appears in the chunk |
| ABSTRACT | | | |
| | boundedness | bounded<br>unbounded | The sentence is bounded if at least one boundary (beginning or end) of the action is signalled.<br>For instance, 'W mieszkaniu było ciemno' is unbounded, since we don't know when the darkness started or whether it has finished.<br>On the other hand, 'Po pięciu latach napisał książkę' is bounded because it is clear that writing process has finished.<br>Note: we evaluate sentences, not verbs! |
| | totality | yes<br>no | yes: The action is presented as a 'whole'. It's not important how it developed over time, e.g. in 'napisał książkę' the process is not important, and all hours of writing, the beginning and the end are merged together and 'writing' is treated as one thing.<br>no: it is important how the action was/is developing |

| | | | |
|---|---|---|---|
| | | | over time, the sentence focuses on the process.<br>Note: we evaluate sentences, not verbs! |
| | specificity | specific<br>generic | A specific sentence describes individual actions, localizable in time, performed by known agents. We assign ‘specific’ label if at least one element of the situation is known.  For instance:<br>Wczoraj Janek poszedł do sklepu.<br>A generic sentence describes general rules or actions performed by agents that cannot be identified. The action itself cannot be localized in time.<br>Do podstawówki chodziło się 8 lat.<br>Note: we evaluate sentences, not verbs! |
| | sequentiality | sequence<br>simultaneous<br>none | sequence: action described in the sentence is a part of sequence of events, e.g.: wstałem, zjadłem śniadaniei poszedłem do pracy<br>simultaneous: action described in the sentence co-occurs with another action: jadłem śniadanie i oglądałem telewizję<br>none: the action is neither a part of a sequence of events nor co-occurs with another action<br>Note: we evaluate sentences, not verbs! |
| | perspective | exterior<br>interior | exterior:action is 'closed', presented from a different temporal perspective, e.g.: napisałem książkę |

| | | | |
|---|---|---|---|
| | | | interior: action is presented 'in the middle of happening', e.g.: pisałem list<br>Note: we evaluate sentences, not verbs! |
| | resultativeness | result<br>no_result | we choose result if there is a change of state, e.g.: Ewa wyszła za mąż. The process of achieving that change can but doesn't have to be expressed by the verb (e.g.: zamieszkać - no process leading to a change is expressed; zbudować - process leading to a change is expressed, but both are resultative)<br>Note: we evaluate sentences, not verbs! |
| | foregrounding | foreground<br>background | foreground: main events on the plot line<br>background: actions that serve as backgrounds, settings, digressions etc. |

Appendix 2: the list of aspectual triggers

| **perfective** | **imperfective** |
|---|---|
| dopiero co | całymi dniami |
| nagle | ciągle |
| nareszcie | czasami |
| wreszcie | czasem |
| natychmiast | długo |
| niechcący | jeszcze |
| nieoczekiwanie | nieraz |
| od razu | niekiedy |
| przed chwilą | nigdy |
| w końcu | od czasu do czasu |
| wkrótce | regularnie |
| wlaśnie | rzadko |
| za chwilę | wciąż |
| zaraz | zawsze |
| znowu | zwykle |
| aż | kiedykolwiek (generalized questions) |
| jak tylko | zacząć |
| zanim | zaczynać |
| już | kończyć |
| cały n/całą n | skończyć |
| od A do Z | przestać |

| | |
|---|---|
| gruntownie | przestawać |
| dokładnie | kontynuować |
| raptem | czas: czas już iść |
| właśnie | pora: pora juz isc |
| kiedy | często |
| zanim (followed by perf) | teraz |
| dopóki nie | codziennie |
| | co godzinę |
| | stale |
| | non stop |
| | bez przerwy |
| | nadal |
| | lubię |
| | nie lubię |
| | można (when used to express general truths and ideas) |
| | trzeba (when used to express general truths and ideas) |
| | warto (when used to express general truths and ideas) |
| | wolno (when used to express general truths and ideas) |

| | |
|---|---|
| | chcieć (when used to express general truths and ideas) |
| | móc (when used to express general truths and ideas) |
| | musieć (when used to express general truths and ideas) |
| | powinien (when used to express general truths and ideas) |
| | woleć (when used to express general truths and ideas) |
| | okazjonalnie |
| | nigdy nie |
| | co tydzień |
| | co miesiąc |
| | co rok |
| | każdego dnia/tygodnia/miesiąca/roku |
| | całymi tygodniami |
| | całymi latami |
| | dopóki |
| | tak długo jak |

Appendix 3: stimuli used in Experiment 1 (Chapter 6)

| Discourse chunk | perfective | imperfective |
|---|---|---|
| Zmiany masy ciała mogą dwojako wpływać na budowę twojej sylwetki. Z jednej strony możesz zeszczupleć z drugiej zaś strony możesz ………………… masy ciała (mięśnie). | nabrać | nabierać |
| Wprowadzono wtedy pilotowy system informatyczny do liczenia i zbierania głosów – zapowiadany przed wyborami z charakterystyczną dla tamtych lat przesadą (tegoroczne wybory samorządowe ze względu na nowatorstwo wykorzystanych rozwiązań, są przedsięwzięciem unikatowym w skali światowej…). A system z hukiem się …………………, nie będąc w stanie przyjąć i przetworzyć danych. | wywrócił | wywracał |
| A gdy wypełnili wszystko według prawa pańskiego, wrócili do Galilei, do swego miasta – Nazaret. Dziecię zaś rosło i ………………… mocy, napełniając się mądrością, a łaska boża spoczywała na nim. | nabrało | nabierało |
| Praktycznym przykładem usług długoterminowych są np. usługi archiwizacji danych - podmioty gospodarcze zobowiązane są do przetrzymywania dokumentacji pracowniczej po okresie prowadzenia działalności przez np. 10, 25 lub 50 lat. Wówczas zlecają usługę wyspecjalizowanej jednostce i opłacają cały okres świadczenia usługi archiwizacji. Przedsiębiorcy przyjmujący takie dokumenty mają wątpliwości, jak ………………… powstały w ten sposób przychód. | rozliczyć | rozliczać |
| W każdym razie zgadzam się, że obecny system filtrowania i weryfikacji komentarzy jest uciążliwy przy dłuższych dyskusjach (ponad 50 postów), bo podając same cytaty, bez numerów postów, trudno będzie je nam potem ………………… (także innym dyskutantom i czytelnikom obserwującym dyskusję). | odszukać | odszukiwać |
| Dopiero jak się takie malce zestawi razem widać jak nasz 'klocek' urósł… Co do naszego prawie 9 kilowego 'klocuszka', to dzisiaj na mamy wachcie 'przypełzł' do pudełka z zabawkami, ………………… je i sam sobie zaczął w nich przebierać. | wywrócił | wywracał |
| W tym czasie napisał też książkę 'Prawda w oczy kole', która krążyła w odpisach. W wersji oryginalnej Mackiewicz tak ………………… sanację za klęskę wrześniową: 'winę główną ponosi naturalnie zdziecinniały kretyn Piłsudski, pasowany na geniusza. | rozliczył | rozliczał |

| | | |
|---|---|---|
| Dzięki korzystaniu z telefonii voip za połączenia wychodzące wykonywane przez twoich konsultantów zapłacisz grosze. Wszystkie rozmowy ………………… będą według najtańszych stawek freeconet. | rozliczone | rozliczane |
| Przykładem – z głowy, na szybko – może być choćby mutacja kodonu stop – zamiast trójki kodującej przerwanie odczytywania RNA. Kiedy ………………… kodon uaa kodonem uag zmiana będzie widoczna tylko w samych kodonach – bez zmiany funkcjonowania kodowanej cząsteczki. | podmienimy | będziemy podmieniać |
| Podchodząc coraz bliżej samochód wydawał mi się mniejszy i jakiś taki niekształtny. Będąc już całkiem blisko, zauważyłam, że ktoś ………………… blachy w jego cacku. | podmienił | podmieniał |
| Rekordzistą i zarazem chyba największym pechowcem tego dnia okazał się Filip Pietrzak. Pływał cały czas w czołówce jednak w pierwszym i drugim wyścigu ………………… swoją omegę tuż przy linii mety, a w trzecim za drugą boją robiąc zwrot przez rufę jego łódka przechyliła się na bok w konsekwencji robiąc 'grzyba'. | wywrócił | wywracał |
| 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 ………………… tylko tekstury. | podmieniać | podmieniać |
| Później jednak, Jogananda uciekł z aśramu Śri Jukteśwara w poszukiwaniu nowego mistrza. Pragnął, aby jego nowym mistrzem został Ram Gopal Mazumdar. Jogananda ………………… go, jednak ten polecił Joganandzie wrócić do Śri Jukteśwara. | odszukał | odszukiwał |
| 90% konsultantów należałoby pogonić z firm, wtedy funkcjonowały by one dużo lepiej i taniej… Działalność?? To jest wpychalność do kieszeni!!! A zawsze jest ich malutko ale tyle, że każdemu są potrzebni. Nawet Tusk z takim poparciem musi ………………… oczami i bredzić jaka to przykładowa koalicja. | wywrócić | wywracać |
| Zamiast palcem serdecznym stuknąć w klawisz 'a', ruszasz okiem w kierunku tego klawisza , a 'inteligentne' oprogramowanie już 'wie' że chodzi ci o klawisz 'a' i reszta już sama się realizuje. Kod klawisza 'a' jest pobierany, przetwarzany, zamieniany w/g tablicy, przeliczany, korygowany, eorowany, podrasowany, raz jeszcze przetwarzany, sprawdzony z wzorcem, potem ………………… na ten z generatora | podmieniony | podmieniany |

| | | |
|---|---|---|
| znaków, przerzucany do szóstej połówki tablicy, wzbogacany o kolory i odcienie, mapowany, w końcu wyświetlany na ekranie. | | |
| Reasumując, należy stwierdzić, że z treści art. 15 ust. 4e updop wynika podstawowa zasada: moment uznania wydatku za koszt podatkowy jest w obecnym stanie prawnym uzależniony od uznania go za taki koszt w ujęciu rachunkowym. Jeśli spółka kierując się zasadami (polityką) rachunkowości, podjęła decyzję o rozliczaniu w czasie poniesionych kosztów na remonty środków trwałych, to powinna je ………………… jednakowo zarówno w odniesieniu do kosztów rachunkowych (bilansowo), jak i podatkowych. | rozliczać | rozliczyć |
| Dla uzyskania jabłoni karłowych stosuje się w szkółkach wegetatywne podkładki m9 i m26, na których otrzymuje się najmniejsze drzewka, nie przekraczające 1,5 m wysokości. Dla dobrego rozwoju wymagają bardzo żyznych gleb, trzeba je też przywiązywać do palików, bo mają słaby system korzeniowy i mogą ………………… się pod ciężarem owoców. | wywrócić | wywracać |
| Niestety, usunięcia cewnika j-j nie można zakwalifikować do procedury 56.031 usunięcie ciała obcego z moczowodu bez nacięcia i rozliczyć grupy l16 lub l17. Hospitalizację można ………………… poprzez produkt z katalogu 1b hospitalizacja z przyczyn nie ujętych gdzie indziej. | rozliczyć | rozliczać |
| Ta ochrona wiąże się oczywiście z wybiciem wszystkich atakujących splicerów, co często nie jest łatwe. Jeśli wszystko pójdzie zgodnie z planem - znów możemy zdecydować czy wrażeń już nam wystarczy i siostrzyczka może się pożegnać z życiem, czy może idziemy ………………… drugie, ostatnie ciało. | odszukać | odszukiwać |
| Wieczorem Oleg wybiegł gwałtownie na dach szkoły. Nie przejmował się tym, że wbiegając po schodach co chwilę się ………………… . | wywrócił | wywracał |
| Ciężar sztangielek nie powinien być zbyt duży, ponieważ to ćwiczenie wymaga dobrego wykonania technicznego, więc tak dobierz ciężar, aby prawidłowo technicznie wykonać ten trening. Następnie bez przerywania przejdź do kolejnego powtórzenia. Pamiętaj o prawidłowym | nabierz | nabieraj |

| | | |
|---|---|---|
| oddychaniu. ………………… powietrza podczas unoszenia sztangielek w górę. | | |
| Wiadomo, że koszty obsługi oraz inne wydatki wejdą w grę, ale pomyśl tylko jakie masz dzisiaj możliwości oszczędzania. Poza tym w treści przedstawione zostały jedynie pewne opcje. Jest ich znacznie więcej i wystarczy odrobina skupienia, aby je ………………… . | odszukać | odszukiwać |
| Sense to jednak nie tylko pulpit, dla wielu aplikacji zostały przecież przygotowane różne skórki. Bardzo wygodne są też skróty na ekranie blokady, za pomocą których już po odblokowaniu możemy przeskoczyć do interesującej nas funkcji. Także i one mogą być dowolnie ………………… . | podmienione | podmieniane |
| Odwrócony kapłan to niejednokrotnie impotent (lub oziębła kobieta), który ze swojej ułomności uczynił (na zewnątrz) cnotę i wymachuje nią, niczym sztandarem. Nienawidzi i tępi każdego, kto ośmiela się myśleć inaczej. Typ fanatyka, gotowego zamęczyć w imię 'dobra' . Zamknięty na tzw. sprawy wyższe. Nie daj się ………………… na jego (jej) rzekomą ascezę lub krystaliczną uczciwość. | nabrać | nabierać |
| Rozmawiaj śmiało. Te same słowa na piśmie mogą ………………… zupełnie innego znaczenia niż wypowiedziane bezpośrednio jakimś tonem głosu i z jakimś wyrazem twarzy. | nabrać | nabierać |
| Inna kobieta zapytała, czy łatwo się zabić na snowboardzie i ilu kolegów dziewczyny już tak zginęło. Staruszki chciały też poznać techniczne szczegóły jazdy na snowboardzie: czy łatwo się …………………, jaką osiąga się prędkość, ile kosztuje snowboard. | wywrócić | wywracać |
| Zgodnie z koncepcją abc koszty pośrednie rozlicza się na produkty w przekroju działań i procesów generujących te koszty, a nie w przekroju podmiotów produkcyjnych (np. wydziałów). Rachunek ten stanowi nową metodę pomiaru i kalkulacji kosztów. Zgodnie z nią koszty pośrednie są ………………… na produkty za pomocą wielu różnych podstaw rozliczeń (cost drivers). | rozliczane | rozliczone |
| A było to tak: kiedy przypadkowo wypadł mu z kieszeni różaniec, esesmani kazali mu go podeptać. Kiedy ks. Władysław odmówił, Niemiec rzucił różaniec w błoto i kazał go księdzu pocałować. Ksiądz ukląkł i ………………… wargami krzyżyk różańca, co wywołało wściekłość strażników. | odszukał | odszukiwał |

| | | |
|---|---|---|
| Staję codziennie rano przed lustrem i widzę twarz promieniejącą sukcesem. W kółko powtarzam sobie, że dam radę, że nie istnieje dla mnie granica możliwości. Powtarzam, bo specjalista mi tak poradził. Powtarzam, bo stwierdził, że dzięki temu ………………… jeszcze większej wiary w siebie i osiągnę perfekcję. | nabiorę | bedę nabierać |
| Pamiętam dokładnie moje początki: sprzęt ważący ok. 20 kg wydawał się trzy razy cięższy niż naprawdę. Jazda w parkanach (to duże, sztywne ochraniacze na nogi) była zupełnie inna niż bez, a ręka bolała od ciągłego trzymania większego, bramkarskiego kija. Teraz do tego wszystkiego przywykłam, ale wtedy co chwilę się potykałam i ………………… . | wywróciłam | wywracałam |
| Każdy kraj jest silny przede wszystkim siłą swych poszczególnych obywateli. Stąd moje zdrowie, wykształcenie, dojrzałość psychiczna i moralna są siłą mojej ojczyzny. Cóż to byłaby za miłość ojczyzny, gdyby większość mieszkańców jakiegoś kraju była niewykształcona, naiwna, bezkrytyczna, gdyby ludzie dawali się stale okłamywać tej samej partii politycznej, dawali się ………………… na obietnice wyborcze w rodzaju gruszek na wierzbie! | nabrać | nabierać |
| Od dłuższego czasu władze gminy zabiegały o zakup samochodu dla jednostki OSP w Świętej. Zadecydowano, iż zostanie zakupiony samochód bojowy średni. Wizja zakupu pożarniczego samochodu ………………… realnych kształtów, kiedy urząd marszałkowski w Poznaniu wygospodarował środki na zakup samochodów zarówno lekkich, średnich, jak i ciężkich. | nabrała | nabierała |
| W prawym górnym rogu powinniśmy wpisać datę i miejscowość, a poniżej adresata oraz funkcję osoby do której się zwracamy. Z lewej strony umieszczamy swoje dane - jest to bardzo ważne, bo jeśli rekruter zdecyduje się na naszą kandydaturę, to nie będzie musiał ………………… numeru telefonu w CV, tylko zadzwoni od razu. | odszukać | odszukiwać |
| Złożoność działań podejmowanych w ramach realizacji projektu Bien pozwoli na stworzenie specjalizacji modelling and visualisation in bioinformatics, która zostanie wpisana w ramy studiów drugiego stopnia i nauczana będzie z wykorzystaniem efektywnych metod nauczania. Wśród przedmiotów wykładanych w ramach nowej specjalizacji znajdą się przedmioty autorskie, a całość będzie '…………………' zgodnie z systemem ETCS | rozliczona | rozliczana |

| | | |
|---|---|---|
| i jasno sprecyzowanymi kryteriami zaliczania przedmiotów. | | |
| Miałam napisać felieton na czasie, o małych warszawskich spotkaniach teatralnych, ale w ostatnich dniach zdarzyły się rzeczy, przez które zmieniłam zdanie - pisze Malina Prześluga dla e-teatru. Po pierwsze, ku mojej olbrzymiej radości, wczoraj do Poznania zawitał Grigore Todiras, a po drugie, po latach poszukiwań, udało mi się znaleźć w antykwariacie na Matejki jego ' Drwię sobie z waszego teatru'. Todiras nie byłby sobą, gdyby nie ………………… wszystkiego do góry nogami… | wywrócił | wywracał |
| Dla skryptu blogowego wordpress, istnieje kilka wtyczek, wklejających kod na stronę. Przetestowałem i polecam dwie z nich. Oczywiście musisz ………………… tekst na polski. | podmienić | podmieniać |
| Na bazie sprzętu z 14, 17 i 75 dr OP powstaje 34 dr OP, sprzęt 74. dr zostanie poddany utylizacji. Z pozostałych pięciu dywizjonów 1. śbr OP (31, 72, 73, 76 i 77 dr OP), formowany jest 35 dr OP, sprzęt jednego dywizjonu ………………… starszy zestaw w kolejnym nowym dywizjonie, sprzęt jednego dywizjonu zostanie poddany utylizacji. | podmieni | będzie podmieniał |
| Termin kwalifikowalności wydatków w ramach po FIO 2009, podobnie jak w poprzednich edycjach FIO, jest spójny z terminami realizacji zadania, wpisanymi do umowy, czyli maksymalnie od 15 maja 2009r. do 31 grudnia 2010r. Data podpisania umowy może być późniejsza niż termin rozpoczęcia zadania. W jaki sposób można ………………… środki przekazywane przez partnera wskazane przez wnioskodawcę jako wkład własny? | rozliczyć | rozliczać |
| Jestem stałą klientką sklepu feromony i nie żałuję żadnego zakupu! Pracuję w marketingu i odkąd zaczęłam używać feromonów ………………… pewności siebie, częściej się uśmiecham, łatwiej mi nawiązać kontakt z ludźmi i do tego co najważniejsze, oni odbierają mnie pozytywnie, są bardziej ufni i sympatyczniejsi. | nabrałam | nabierałam |
| Istotnym elementem w technikach masażu głębokiego jest wrażliwość palpacyjna terapeuty - musi on umieć odnajdować zmiany napięcia w tkankach, ………………… bruzdy mięśniowe oraz ocenić kiedy masaż należy zakończyć by uzyskać | odszukać | odszukiwać |

| | | |
|---|---|---|
| efekt terapeutyczny a jednocześnie nie przestymulować tkanki . | | |
| Wszystko w jej życiu wytyczała tradycja – nawet długość włosów do ślubu. Jednak młoda pani Kujawczak nie wie do końca, jaka ma być jej życiowa droga. Gdy przyjaciółki wyciągają ją z domu na wieczór panieński, nie wie również, że ta noc zmieni jej przyszłość, ………………… całe życie do góry nogami i pozwoli odnaleźć prawdziwą Sylwię. | wywróci | będzie wywracać |
| Skrypty dzielą się na dwie kategorie. Pierwsza, to skrypty graficzne. Dzięki nim możemy ………………… wszelką grafikę i animację w telefonie. | podmienić | podmieniać |
| Eksperci z departamentu energii w narodowym laboratorium w Idaho stwierdzili, że przecież można nauczyć nowych sztuczek stare, wykorzystywane przez wojsko maszyny. W tym celu w robocie typu talon ………………… wyrzutnie rakiet i karabiny maszynowe na zestaw czujników wykrywających promieniowanie. | podmieniono | podmieniano |
| W czasie jazdy rowerem pocisz się, dlatego koniecznie trzeba pamiętać o uzupełnianiu płynów, dzięki temu nie opadniesz szybko z sił i z pewnością podołasz reszcie trasy, którą musisz pokonać. Jeżeli bierzesz udział w wyścigu nie radzę wody pić. Polecam raczej ………………… jej do ust i ją wypluwać dzięki temu nie poczujesz potrzeby skorzystania z toalety. | nabrać | nabierać |
| Chodzi mi o inwestowanie w pożyczki społecznościowe typu kokos przez osoby prowadzące działalność gospodarczą. Jak to rozliczyć ? Czy mogę to ………………… nie jako firma, tylko np. na koniec roku wpisać w zeznaniu przychody kapitałowe (czy jakoś tak) i od tego obliczyć 19% podatku? | rozliczyć | rozliczać |
| Niemcy były powiązane paktem antykominternowskim z Japonią i Włochami, tzw. paktem stalowym z Włochami oraz układem z ZSRR (pakt Ribbentrop-Mołotow) z 23 sierpnia 1939, który de facto dzielił Europę środkową na strefy wpływów obu państw i dawał Niemcom wolną rękę w wojnie z Polską. Formalnie wojna ………………… charakteru światowego z chwilą wypowiedzenia jej przez Wielką Brytanię i Francję 3 września 1939. | nabrała | nabierała |
| Pytasz co teraz robić? Prawdziwy mężczyzna w takiej sytuacji powinien za wszelką cenę doprowadzić do zniknięcia tego głupiego nagrania z internetu. A | odszukać | odszukiwać |

| | | |
|---|---|---|
| później powinieneś tę organistkę ………………… , udać się do niej z kwiatami i przeprosić. | | |
| Zawodnicy pokonują dziennie ponad 30 km trasy. Daje to w sumie 100 km dystans do przebycia przez trzy kolejne dni. Sporym utrudnieniem wyścigu był etap nocny, podczas którego maszerzy prowadzili swoje zaprzęgi przez zimowy las jedynie w nikłym świetle czołowych latarek. Podczas wyścigu psy w zaprzęgu nie mogą być ………………… . | podmienione | podmieniane |
| Dzięki potężnej machinie manipulacji stara maksyma 'chleba i igrzysk' straciła swoje zastosowanie. Dziś wystarczają same 'igrzyska' . Koszta związane z ich organizacją ponosi społeczeństwo, natomiast zyski trafią do ograniczonej grupy biznesmenów, dyrektorów czy członków rad nadzorczych. Chwała i splendor zarezerwowane są dla elit, reszta musi poddać się niedawno wydłużonemu rygorowi pracy. Czas się wyrwać z hipnozy i ………………… elity czerpiące zyski z naszego codziennego trudu. | rozliczyć | rozliczać |
| Warsztat był kameralny. Było siedem osób. Sympatyczne dziewczyny w różnym wieku chętne do odkrywanie swojego indywidualnego przejawu w kosmicznym planie. Prowadząca warsztat była niewątpliwie jednym z największych atutów tego kursu. Wyczuwałam w niej jakby inny wymiar istnienia. Obecność, spokojny oddech, mądrość i głębia wypowiadanych słów sprawiały, że to spotkanie ze sobą ………………… jeszcze głębszego charakteru. | nabrało | nabierało |
| Jeśli okaże się że jednak inwestor wybuduje ten park rozrywki, zapewni pracę tysiącom osób (bezpośrednio lub pośrednio), będzie płacił gminie podatki (a i zatrudnieni przez niego też płacą podatki, i kupują na lokalnym rynku), to wtedy chwała władzom Grodziska, a wstyd i hańba władzom Piaseczna. Chyba że okaże się że to jednak przekręt. Profesjonalizm urzędników (o ile taki jest) powinien jednak zapewnić nam, obywatelom, że podejmowane są najlepsze możliwe decyzje, a każdy inwestor jest zweryfikowany. A powinni podejmować decyzje dobre dla większości mieszkańców gminy, a nie tylko dla kilku osób ze Złotokłosa. Mam nadzieję, że po 2015 ………………… się urzędników w Piasecznie. | rozliczy | bedzie rozliczać |

| | | |
|---|---|---|
| Serdecznie zachęcamy do zapoznania się z naszym katalogiem najróżniejszych stron internetowych. Dla zapewnienia wygody naszych użytkowników został on podzielony na tematyczne kategorie i podkategorie. Ponadto, zamieszczanych jest tutaj wiele informacji odnoszących się do firm, jak również sklepów internetowych. Dzięki naszym wszelkim staraniom, każdy użytkownik może w sposób łatwy i przyjemny ………………… interesujące materiały. | odszukać | odszukiwać |
| Mogę was jednak uspokoić – nowoczesne i dobrze dobrane żywice to zupełnie inna bajka. Zdecydowanie mogą konkurować chociażby z wtryskiwanym polistyrenem, z którego wykonane są 'standardowe' plastikowe figurki. Pierwsze co zwraca uwagę to jak niesamowicie lekki jest ten materiał. To zdecydowanie dobra wiadomość dla graczy biorących udział w turniejach – walizki z armią przestaną 'urywać ręce'. Myślę jednak, że w niektórych wypadkach niewielkie dociążenie podstawki może okazać się konieczne – inaczej modele mogą się ………………… . | wywrócić | wywracać |
| Powyższe zdjęcie jest przykładem liberatury – książki, w której tekst, sposób zapisu, forma oraz materiał stanowią nośnik treści. 'Sto tysięcy miliardów wiersz' to kombinatoryczny, permutacyjny cykl 10 sonetów, wydrukowanych na kartkach pociętych na paski. Każdą linijkę wiersza można dowolnie ………………… , co umożliwia czytelnikowi samodzielne wygenerowanie niemal nieskończonej liczby wierszy. | podmienic | podmieniać |
| Autorzy pakietu skupili się przede wszystkim na uświadomieniu dzieciom konieczności indywidualnej pracy nad sobą oraz potrzeby zbliżania się do Boga. Treści zawarte w książkach ukazują Jezusa jako przyjaciela każdego człowieka, który przemawia do niego poprzez karty Pisma Świętego. Wszystkie treści zawarte w publikacjach z serii 'Bliscy sercu Jezusa' zostały przeredagowane tak, aby dzieci korzystające z książek mogły swobodnie ………………… potrzebne informacje. | odszukać | odszukiwać |
| Naturalnie zarówno bohaterowie, jak i niektóre ze stworzeń posiadają specjalne umiejętności, które w znaczny sposób potrafią komplikować życie naszych przeciwników. W każdym rodzie znajduje się złodziej, który może kraść skarby znajdujące się w danej krainie, guślarz który zwiększa swoją moc z każdym wojownikiem znajdującym się w tej samej krainie, zwiadowca który pozwoli | podmienić | podmieniać |

| | | |
|---|---|---|
| graczowi ………………… swoich bohaterów w poszczególnych krainach, łucznik który zmniejszy siłę potworów oraz dwóch silnych wojów bez żadnych specjalnych umiejętności. | | |
| A przecież nawet na wyczyszczonym dysku są dane, ponieważ – podobnie jak dzieje się to podczas formatowania dysku twardego – na nowo tworzona jest tylko struktura katalogów. Aby odszukać dane na takich dyskach, programy muszą odczytać zagubione bajty i zidentyfikować typy plików. I znów tylko Isobuster i Cdroller dobrze wypadły w tej konkurencji. Badcopy ………………… pliki tylko przy zastosowaniu intensywnej metody skanowania, co zajmowało mnóstwo czasu. | odszukał | odszukiwał |
| Przepisy nie nakładają obowiązku przeznaczenia środków pieniężnych otrzymanych z pomocy restrukturyzacyjnej na pokrycie kosztów restrukturyzacji. W związku z tym, spółka może rozporządzać otrzymanymi środkami w dowolny sposób. Ani ARR, ani żadne inne organy nie zostały upoważnione do kontroli beneficjentów pomocy w zakresie faktycznego wykorzystania środków pomocowych. Otrzymujący pomoc restrukturyzacyjną jest ………………… jedynie z realizacji planu restrukturyzacji wcześniej uzgodnionego i zaakceptowanego przez ARR, a nie z kwot wydatkowanych z uzyskanych środków. | rozliczony | rozliczany |
| Jeśli fotografujesz aparatem cyfrowym, to z całą pewnością wiesz, jak szybko przybywa zdjęć w twoim archiwum. Po przeczytaniu tego tekstu nie będziesz mieć żadnych wątpliwości, że tylko odpowiednia praca z plikami pozwoli ci bez problemu ………………… zdjęcia nawet kilka lat po ich zrobieniu i być spokojnym o oryginały tych fotografii. | odszukać | odszukiwać |
| I jak przeczytasz w artykule, nic więcej nie ma na ten temat, nawet w książce, która mi to obiecywała w ofercie, dlatego ją zakupiłam, żeby się dowiedzieć jeszcze więcej. A z kolei też uprzedzam, że specjaliści od feng shui to może by w ogóle te koncepcje ………………… do góry nogami, ale swoich innych koncepcji do tej pory nigdzie nie opisali. | wywrócili | wywracali |

Appendix 4: stimuli used in Experiment 2 (Chapter 11)

**Set 1: Base**

| **aspect** | **item** | **correct** | **verbDisplayed** | **sentenceDisplayed** |
|---|---|---|---|---|
| perf | expItem | 1 | wypowiedział | Na ostatnim spotkaniu, prezes .......... się o nim krytycznie. |
| perf | expItem | 1 | wskazała | Zlecona wtedy analiza .......... czynniki wpływające na pogorszenie stanu wałów przeciwpowodziowych. |
| perf | expItem | 1 | zastanowili | Po wysłuchaniu świadków, zebrani .......... się komu powinni przyznać rację. |
| perf | expItem | 1 | poszerzyły | W zeszłym roku, polski rząd poszerzył zakres konsultacji dyplomatycznych. |
| perf | expItem | 1 | zaprosiłeś | Wydaje mi się, że w zeszły weekend .......... zbyt wielu gości. |
| perf | expItem | 1 | otrzymałaś | W zeszłym roku, .......... pieniądze z rządowego programu pomocy przedsiębiorcom. |
| perf | expItem | 1 | zdaliście | Jeśli dobrze pamiętam, to w maju .......... naprawdę ważny test. |
| perf | expItem | 1 | zechciałyście | Jest nam miło, że poprzednim razem .......... skorzystać z naszej oferty. |
| perf | expItem | 1 | pozazdrościłem | Przyznam, że po spotkaniu .......... Ani umiejętności językowych i pewności siebie. |
| perf | expItem | 1 | zsunęłam | Po odcięciu pasa bezpieczeństwa, .......... się wolno na ziemię |
| perf | expItem | 1 | podnieśliśmy | W ubiegłym roku .......... ceny, żeby odrobić straty spowodowane inflacją. |
| perf | expItem | 1 | dostałyśmy | W początku zeszłego semestru .......... spore pieniądze od rodziców. |
| impf | expItem | 1 | wypowiadał | Poprzednim razem, przewodniczący .......... się o nim z uznaniem. |
| impf | expItem | 1 | wskazywała | Przeprowadzona wtedy symulacja .........., że samolot nie mógł zerwać linii. |

| | | | | |
|---|---|---|---|---|
| impf | expItem | 1 | zastanawiali | Po wyborach .......... się, czemu ludzie nie chcieli głosować na lewicę. |
| impf | expItem | 1 | poszerzały | W zeszłym roku, firma poszerzała zakres swoich usług internetowych i telefonicznych. |
| impf | expItem | 1 | strzelałeś | W poprzednim sezonie .......... gola w każdym ważnym meczu. |
| impf | expItem | 1 | puszczałaś | Na poprzednich wakacjach w ogóle nie .......... dzieci w okolice wody. |
| impf | expItem | 1 | zdawaliście | W czerwcu zeszłego roku .......... egzamin na prawo jazdy. |
| impf | expItem | 1 | myślałyście | Było mi miło, że .......... o mnie, kiedy byłam w szpitalu. |
| impf | expItem | 1 | zazdrościłem | Wtedy .......... ci takiego drogiego samochodu, własnego mieszkania i wystawnego życia. |
| impf | expItem | 1 | pokonywałam | Pamiętam, że w zeszłym roku .......... tą trasę w dwadzieścia minut. |
| impf | expItem | 1 | podnosiliśmy | W zeszłym roku .......... opłaty, żeby załatać dziurę w budżecie. |
| impf | expItem | 1 | dostawałyśmy | Według dokumentów, w poprzednim roku akademickim, .......... stypendium naukowe. |
| perf | filler | 1 | zdecydujemy | Jeśli .......... zgodnie z sercem, to inwestycja sprawi nam radość. |
| perf | filler | 1 | dołączę | Liczę, że .......... do waszego grona i podejmiemy to wyzwanie razem. |
| perf | filler | 1 | skoczę | Jeszcze tylko .......... szybko do sklepu po cukierki i owoce. |
| perf | filler | 1 | przydadzą | Poniższe wskazówki na pewno się .......... osobom preferującym dietę wegańską. |
| perf | filler | 1 | zdołają | Nie przewidział, że po tej klęsce bolszewicy .......... odtworzyć swoją potęgę. |
| perf | filler | 1 | dotrzymam | Co mi grozi, jeśli nie .......... terminu przeglądu kasy fiskalnej? |
| perf | filler | 1 | nauczą | Nawet jeśli się .........., w pytaniach mogą pojawić się nowe pojęcia. |
| perf | filler | 1 | oddamy | Gdy .......... trochę krwi, nasz organizm szybko się zregeneruje. |
| perf | filler | 1 | zamkniemy | Oczywiście, jeżeli uda się dojść do konkluzji, zakończymy nasze prace. |
| perf | filler | 1 | pożegna | W środę Zakopane .......... młodszego z braci na nowym cmentarzu publicznym. |

| | | | | |
|---|---|---|---|---|
| perf | filler | 1 | zaprzestanie | Mamy nadzieję, że młodzież całkowicie .......... kontynuowania złych przyzwyczajeń. |
| perf | filler | 1 | spocznie | Wiedziała, że chłopiec nie .........., dopóki nie pozna odpowiedzi. |
| impf | filler | 1 | stawiają | Oto najważniejsze pytania, które .......... sobie klienci przed podjęciem decyzji. |
| impf | filler | 1 | mówicie | Więc .......... panowie, że facet, który mnie oszukał właśnie poszedł siedzieć? |
| impf | filler | 1 | doprowadza | Kryzys, który dotyka europejskie państwa, .......... do upadku wielu firm. |
| impf | filler | 1 | kosztuje | Program .......... jedynie pięćdziesiąt złotych ale jest dostępny tylko na iPada. |
| impf | filler | 1 | zamieniają | Nie brakuje tu młodych ludzi, którzy .......... swoje pomysły w biznesy. |
| impf | filler | 1 | przynoszę | Często .......... sobie w słoikach przetwory wykonane przez moją mamę. |
| impf | filler | 1 | grasz | Na jakimkolwiek poziomie .........., zawsze staraj się robić to najlepiej. |
| impf | filler | 1 | skłaniają | Rosnące ceny ropy naftowej .......... producentów samochodów do poszukiwania alternatywnych paliw. |
| impf | filler | 1 | zsyła | Wierzę, że od czasu do czasu los każdemu .......... wskazówki. |
| impf | filler | 1 | kona | Rzeczpospolita .......... raczej w uścisku zadłużenia i szybko rosnącej inflacji. |
| impf | filler | 1 | brakuje | W gospodarce .......... optymizmu, a politykom trudno jest się dogadać. |
| impf | filler | 1 | słyszę | Do dziś gdy .......... ten zwrot, to przypominam sobie mojego profesora. |
| perf | expItem | 0 | strzeliłeś | W zeszłym miesiącu, kiedy .......... się do systemu, sprawdzał dane. |
| perf | expItem | 0 | puściłaś | W czasach komunizmu nie .......... podziemna działalność edukacyjna i wychowawcza. |
| perf | expItem | 0 | ustała | Podczas drugiej wojny światowej, polscy powstańcy .......... się wielką odwagą. |

| | | | | |
|---|---|---|---|---|
| perf | expItem | 0 | zalogował | Wcześniej, zaproszone prelegentki .......... umiejętności dziecka, które nabywa w szkole. |
| perf | expItem | 0 | wyjechałyśmy | Dwa sezony temu strzelałeś gola każdej drużynie, z którą grałeś. |
| perf | expItem | 0 | odznaczyli | Ostatniej zimy nie .......... swoich podopiecznych na stok narciarski. |
| perf | expItem | 0 | zestresowałem | Na pewno wcześniej .......... oferty innych sklepów oferujących komputery osobiste. |
| perf | expItem | 0 | sprawdziliście | Dziękuję, że .......... o mnie, kiedy szukałyście chętnych na wyjazd. |
| perf | expItem | 0 | omówiły | Muszę przyznać, że przez to wszystko naprawdę się wtedy .......... |
| perf | expItem | 0 | rozpowszechniliśmy | W ubiegłym sezonie pokonywałam ten dystans w mniej niż minutę. |
| perf | expItem | 0 | pomyślałyście | W zeszłym roku .......... wśród uczniów materiały edukacyjne dotyczące zdrowia. |
| perf | expItem | 0 | pokonałam | W jej urodziny .......... odpocząć trochę do domku jej rodziców. |
| impf | expItem | 0 | chciałyście | Wczoraj .......... się zdalnie za pomocą specjalnej aplikacji desktopowej. |
| impf | expItem | 0 | zapraszałeś | Wrzawa na trybunach nie .......... nawet kiedy najlepszy zawodnik przestrzelił bramkę. |
| impf | expItem | 0 | sprawdzaliście | W czasach dyktatury, ci nieustępliwi społecznicy .......... się wielką odwagą. |
| impf | expItem | 0 | wyjeżdżałyśmy | Grupy negocjacyjne .......... przedwczoraj dwadzieścia różnych kwestii, które reguluje ustawa. |
| impf | expItem | 0 | rozpowszechnialiśmy | Myślę, że w zeszłym tygodniu .......... naprawdę wiele osób. |
| impf | expItem | 0 | stresowałem | Zapewne .......... wtedy pieniądze w ramach programu wymiany naukowej. |
| impf | expItem | 0 | ustawała | Jeżeli wcześniej nie .......... oferty tej firmy, to zdecydowanie powinniście. |
| impf | expItem | 0 | logował | Było nam miło, że ostatnim razem .......... skorzystać z naszych usług. |
| impf | expItem | 0 | omawiały | Szczerze mówiąc, bardzo się wtedy .......... całą tą sytuacją. |
| impf | expItem | 0 | odznaczali | Po zwolnieniu ręcznego hamulca, .......... się w dół po drodze. |

| | | | | |
|---|---|---|---|---|
| impf | expIte m | 0 | pobiegnie | W latach dziewięćdziesiątych .......... reżyserską wersję filmu na wideo. |
| impf | expIte m | 0 | zsuwałam | W zeszłe wakacje .......... w góry, żeby odpocząć trochę od zgiełku. |
| perf | filler | 0 | poniosą | Dodatkowo, jeśli posiadacz konta .......... nowego klienta, to dostanie prowizję. |
| perf | filler | 0 | przyprowadzi | Spacerkiem .......... również na plażę miejską czy nad liczne jeziora. |
| perf | filler | 0 | złapię | Osoby, które .......... ofiarą wirusa, skarżą się na ból gardła. |
| perf | filler | 0 | wprowadzą | Niestety, pieniądze z odszkodowania nie .......... nigdy strat jakie ponieśliśmy. |
| perf | filler | 0 | uratuje | Kto zapłaci za ten odcinek, jeśli obwodnica .......... jednak inaczej? |
| perf | filler | 0 | otrzymywałaś | Dobra reklama sprawi, że klient .......... do twojej firmy. |
| perf | filler | 0 | dojdziesz | Jak tak dalej pójdzie, to w niedługo .......... zezwolenia na oddychanie. |
| perf | filler | 0 | wyłoni | W sumie to nieważne co .........., zawsze coś da się złowić. |
| perf | filler | 0 | pokryją | Nic już nie .......... tamtych sadzonek, ale mam jeszcze zapas. |
| perf | filler | 0 | padną | Młodość przemija, a to jak ją .......... zależy od nas. |
| perf | filler | 0 | trafi | Jury .......... zwycięzcę na podstawie głosów użytkowników i własnej oceny. |
| perf | filler | 0 | spędzimy | Eksperci sugerują, że Republikanie wkrótce .......... polityczne koszty swej nieustępliwości. |
| impf | filler | 0 | sądzimy | Historycy motoryzacji zgodnie .......... Forda do najlepszych konstrukcji samochodów klasy średniej. |
| impf | filler | 0 | depcze | Zdaje się, że protestujący .......... kolejną demonstrację pod siedzibą firmy. |
| impf | filler | 0 | przypuszczamy | Powstrzymywanie rozwoju pustyń .......... różnorodność roślinną oraz zmniejsza ilość dwutlenku węgla. |
| impf | filler | 0 | lecimy | Każdy biznes jest inny, dlatego .......... przedsiębiorcom chodzenie na konferencje biznesowe. |

| | | | | |
|---|---|---|---|---|
| impf | filler | 0 | chroni | Szkoda, że nie .......... niżej i nie mogę zrobić lepszego zdjęcia. |
| impf | filler | 0 | śmieją | Jak .........., przyczyną katastrofy był wyciek oleju z silnika. |
| impf | filler | 0 | zaliczają | Kobiety, które .......... się zajść w ciążę, powinny unikać kofeiny. |
| impf | filler | 0 | szykują | Nie truj, że ktoś .......... twoją godność albo cię pozbawia szacunku. |
| impf | filler | 0 | dopuszczam | Tymczasem ja .......... poprzez drzewa na jezioro za domem. |
| impf | filler | 0 | zalecam | Jak .........., morderca kręcił się po okolicznych sklepach dzień przed zbrodnią. |
| impf | filler | 0 | starają | Oczywiście .......... możliwość, że gdzieś popełniłem błąd i chętnie go poprawię. |
| impf | filler | 0 | spoglądam | Najpierw cię ignorują, potem się .........., a na końcu szanują. |

**Set 1: cued**

| aspect | item | correct | verbDisplayed | sentenceDisplayed |
|---|---|---|---|---|
| perf | expItem | 1 | wypowiedział | Na ostatnim spotkaniu, prezes .......... się o nim krytycznie. |
| perf | expItem | 1 | wskazała | Zlecona wtedy analiza .......... czynniki wpływające na pogorszenie stanu wałów przeciwpowodziowych. |
| perf | expItem | 1 | zastanowili | Po wysłuchaniu świadków, zebrani .......... się komu powinni przyznać rację. |
| perf | expItem | 1 | poszerzyły | W zeszłym roku, polski rząd poszerzył zakres konsultacji dyplomatycznych. |
| perf | expItem | 1 | zaprosiłeś | Wydaje mi się, że w zeszły weekend .......... zbyt wielu gości. |
| perf | expItem | 1 | otrzymałaś | W zeszłym roku, .......... pieniądze z rządowego programu pomocy przedsiębiorcom. |
| perf | expItem | 1 | zdaliście | Jeśli dobrze pamiętam, to w maju .......... naprawdę ważny test. |
| perf | expItem | 1 | zechciałyście | Jest nam miło, że poprzednim razem .......... skorzystać z naszej oferty. |
| perf | expItem | 1 | pozazdrościłem | Przyznam, że po spotkaniu .......... Ani umiejętności językowych i pewności siebie. |

| | | | | |
|---|---|---|---|---|
| perf | expItem | 1 | zsunęłam | Po odcięciu pasa bezpieczeństwa, .......... się wolno na ziemię |
| perf | expItem | 1 | podnieśliśmy | W ubiegłym roku .......... ceny, żeby odrobić straty spowodowane inflacją. |
| perf | expItem | 1 | dostałyśmy | W początku zeszłego semestru .......... spore pieniądze od rodziców. |
| impf | expItem | 1 | wypowiadał | Poprzednim razem, przewodniczący .......... się o nim z uznaniem. |
| impf | expItem | 1 | wskazywała | Przeprowadzona wtedy symulacja ciągle .........., że samolot nie mógł zerwać linii. |
| impf | expItem | 1 | zastanawiali | Po wyborach codziennie .......... się, czemu ludzie nie chcieli głosować na lewicę. |
| impf | expItem | 1 | poszerzały | W zeszłym roku, firma regularnie poszerzała zakres swoich usług internetowych i telefonicznych. |
| impf | expItem | 1 | strzelałeś | W poprzednim sezonie zawsze .......... gola w każdym ważnym meczu. |
| impf | expItem | 1 | puszczałaś | Na poprzednich wakacjach zwykle nie .......... dzieci w okolice wody. |
| impf | expItem | 1 | zdawaliście | W czerwcu zeszłego roku jeszcze .......... egzamin na prawo jazdy. |
| impf | expItem | 1 | myślałyście | Było mi miło, że często .......... o mnie, kiedy byłam w szpitalu. |
| impf | expItem | 1 | zazdrościłem | Wtedy długo .......... ci takiego drogiego samochodu, własnego mieszkania i wystawnego życia. |
| impf | expItem | 1 | pokonywałam | Pamiętam, że w zeszłym roku zwykle .......... tą trasę w dwadzieścia minut. |
| impf | expItem | 1 | podnosiliśmy | W zeszłym roku, co kwartał .......... opłaty, żeby załatać dziurę w budżecie. |
| impf | expItem | 1 | dostawałyśmy | Według dokumentów, w poprzednim roku akademickim, co miesiąc .......... stypendium naukowe. |
| perf | filler | 1 | zdecydujemy | Jeśli .......... zgodnie z sercem, to inwestycja sprawi nam radość. |
| perf | filler | 1 | dołączę | Liczę, że .......... do waszego grona i podejmiemy to wyzwanie razem. |
| perf | filler | 1 | skoczę | Jeszcze tylko .......... szybko do sklepu po cukierki i owoce. |

| | | | | |
|---|---|---|---|---|
| perf | filler | 1 | przydadzą | Poniższe wskazówki na pewno się .......... osobom preferującym dietę wegańską. |
| perf | filler | 1 | zdołają | Nie przewidział, że po tej klęsce bolszewicy .......... odtworzyć swoją potęgę. |
| perf | filler | 1 | dotrzymam | Co mi grozi, jeśli nie .......... terminu przeglądu kasy fiskalnej? |
| perf | filler | 1 | nauczą | Nawet jeśli się .........., w pytaniach mogą pojawić się nowe pojęcia. |
| perf | filler | 1 | oddamy | Gdy .......... trochę krwi, nasz organizm szybko się zregeneruje. |
| perf | filler | 1 | zamkniemy | Oczywiście, jeżeli uda się dojść do konkluzji, zakończymy nasze prace. |
| perf | filler | 1 | pożegna | W środę Zakopane .......... młodszego z braci na nowym cmentarzu publicznym. |
| perf | filler | 1 | zaprzestanie | Mamy nadzieję, że młodzież całkowicie .......... kontynuowania złych przyzwyczajeń. |
| perf | filler | 1 | spocznie | Wiedziała, że chłopiec nie .........., dopóki nie pozna odpowiedzi. |
| impf | filler | 1 | stawiają | Oto najważniejsze pytania, które .......... sobie klienci przed podjęciem decyzji. |
| impf | filler | 1 | mówicie | Więc .......... panowie, że facet, który mnie oszukał właśnie poszedł siedzieć? |
| impf | filler | 1 | doprowadza | Kryzys, który dotyka europejskie państwa, .......... do upadku wielu firm. |
| impf | filler | 1 | kosztuje | Program .......... jedynie pięćdziesiąt złotych ale jest dostępny tylko na iPada. |
| impf | filler | 1 | zamieniają | Nie brakuje tu młodych ludzi, którzy .......... swoje pomysły w biznesy. |
| impf | filler | 1 | przynoszę | Często .......... sobie w słoikach przetwory wykonane przez moją mamę. |
| impf | filler | 1 | grasz | Na jakimkolwiek poziomie .........., zawsze staraj się robić to najlepiej. |
| impf | filler | 1 | skłaniają | Rosnące ceny ropy naftowej .......... producentów samochodów do poszukiwania alternatywnych paliw. |
| impf | filler | 1 | zsyła | Wierzę, że od czasu do czasu los każdemu .......... wskazówki. |

| | | | | |
|---|---|---|---|---|
| impf | filler | 1 | kona | Rzeczpospolita .......... raczej w uścisku zadłużenia i szybko rosnącej inflacji. |
| impf | filler | 1 | brakuje | W gospodarce .......... optymizmu, a politykom trudno jest się dogadać. |
| impf | filler | 1 | słyszę | Do dziś gdy .......... ten zwrot, to przypominam sobie mojego profesora. |
| perf | expIte m | 0 | puściłaś | W zeszłym miesiącu, kiedy .......... się do systemu, sprawdzał dane. |
| perf | expIte m | 0 | strzeliłeś | W czasach komunizmu nie .......... podziemna działalność edukacyjna i wychowawcza. |
| perf | expIte m | 0 | ustała | Podczas drugiej wojny światowej, polscy powstańcy .......... się wielką odwagą. |
| perf | expIte m | 0 | zalogował | Wcześniej, zaproszone prelegentki .......... umiejętności dziecka, które nabywa w szkole. |
| perf | expIte m | 0 | wyjechałyśmy | Dwa sezony temu strzelałeś gola każdej drużynie, z którą grałeś. |
| perf | expIte m | 0 | odznaczyli | Ostatniej zimy nie .......... swoich podopiecznych na stok narciarski. |
| perf | expIte m | 0 | rozpowszechniliśmy | Na pewno wcześniej .......... oferty innych sklepów oferujących komputery osobiste. |
| perf | expIte m | 0 | sprawdziliście | Dziękuję, że .......... o mnie, kiedy szukałyście chętnych na wyjazd. |
| perf | expIte m | 0 | omówiły | Muszę przyznać, że przez to wszystko naprawdę się wtedy ........... |
| perf | expIte m | 0 | zestresowałem | W ubiegłym sezonie pokonywałam ten dystans w mniej niż minutę. |
| perf | expIte m | 0 | pomyślałyście | W zeszłym roku .......... wśród uczniów materiały edukacyjne dotyczące zdrowia. |
| perf | expIte m | 0 | pokonałam | W jej urodziny .......... odpocząć trochę do domku jej rodziców. |
| impf | expIte m | 0 | chciałyście | Wczoraj co godzinę .......... się zdalnie za pomocą specjalnej aplikacji desktopowej. |
| impf | expIte m | 0 | zapraszałeś | Wrzawa na trybunach zwykle nie .......... nawet kiedy najlepszy zawodnik przestrzelił bramkę. |
| impf | expIte m | 0 | sprawdzaliście | W czasach dyktatury, ci nieustępliwi społecznicy codziennie .......... się wielką odwagą. |

| | | | | |
|---|---|---|---|---|
| impf | expIte m | 0 | wyjeżdżałyśmy | Grupy negocjacyjne długo .......... przedwczoraj dwadzieścia różnych kwestii, które reguluje ustawa. |
| impf | expIte m | 0 | stresowałem | Myślę, że w zeszłym tygodniu często .......... naprawdę wiele osób. |
| impf | expIte m | 0 | rozpowszechnialiśm y | Zapewne .......... wtedy co miesiąc pieniądze w ramach programu wymiany naukowej. |
| impf | expIte m | 0 | omawiały | Jeżeli wcześniej nie .......... regularnie oferty tej firmy, to zdecydowanie powinniście. |
| impf | expIte m | 0 | logował | Było nam miło, że ostatnim razem wciąż .......... skorzystać z naszych usług. |
| impf | expIte m | 0 | ustawała | Szczerze mówiąc, wtedy codziennie bardzo się .......... całą tą sytuacją. |
| impf | expIte m | 0 | odznaczali | Po zwolnieniu ręcznego hamulca, długo .......... się w dół po drodze. |
| impf | expIte m | 0 | otrzymywałaś | W latach dziewięćdziesiątych rzadko .......... reżyserską wersję filmu na wideo. |
| impf | expIte m | 0 | zsuwałam | W zeszłe wakacje często .......... w góry, żeby odpocząć trochę od zgiełku. |
| perf | filler | 0 | poniosą | Dodatkowo, jeśli posiadacz konta .......... nowego klienta, to dostanie prowizję. |
| perf | filler | 0 | przyprowadzi | Spacerkiem .......... również na plażę miejską czy nad liczne jeziora. |
| perf | filler | 0 | złapię | Osoby, które .......... ofiarą wirusa, skarżą się na ból gardła. |
| perf | filler | 0 | wprowadzą | Niestety, pieniądze z odszkodowania nie .......... nigdy strat jakie ponieśliśmy. |
| perf | filler | 0 | uratuje | Kto zapłaci za ten odcinek, jeśli obwodnica .......... jednak inaczej? |
| perf | filler | 0 | pobiegnie | Dobra reklama sprawi, że klient .......... do twojej firmy. |
| perf | filler | 0 | dojdziesz | Jak tak dalej pójdzie, to w niedługo .......... zezwolenia na oddychanie. |
| perf | filler | 0 | wyłoni | W sumie to nieważne co .........., zawsze coś da się złowić. |
| perf | filler | 0 | pokryją | Nic już nie .......... tamtych sadzonek, ale mam jeszcze zapas. |
| perf | filler | 0 | padną | Młodość przemija, a to jak ją .......... zależy od nas. |

| | | | | |
|---|---|---|---|---|
| perf | filler | 0 | trafi | Jury .......... zwycięzcę na podstawie głosów użytkowników i własnej oceny. |
| perf | filler | 0 | spędzimy | Eksperci sugerują, że Republikanie wkrótce .......... polityczne koszty swej nieustępliwości. |
| impf | filler | 0 | sądzimy | Historycy motoryzacji zgodnie .......... Forda do najlepszych konstrukcji samochodów klasy średniej. |
| impf | filler | 0 | depcze | Zdaje się, że protestujący .......... kolejną demonstrację pod siedzibą firmy. |
| impf | filler | 0 | przypuszczamy | Powstrzymywanie rozwoju pustyń .......... różnorodność roślinną oraz zmniejsza ilość dwutlenku węgla. |
| impf | filler | 0 | lecimy | Każdy biznes jest inny, dlatego .......... przedsiębiorcom chodzenie na konferencje biznesowe. |
| impf | filler | 0 | chroni | Szkoda, że nie .......... niżej i nie mogę zrobić lepszego zdjęcia. |
| impf | filler | 0 | śmieją | Jak .........., przyczyną katastrofy był wyciek oleju z silnika. |
| impf | filler | 0 | zaliczają | Kobiety, które .......... się zajść w ciążę, powinny unikać kofeiny. |
| impf | filler | 0 | szykują | Nie truj, że ktoś .......... twoją godność albo cię pozbawia szacunku. |
| impf | filler | 0 | dopuszczam | Tymczasem ja .......... poprzez drzewa na jezioro za domem. |
| impf | filler | 0 | zalecam | Jak .........., morderca kręcił się po okolicznych sklepach dzień przed zbrodnią. |
| impf | filler | 0 | spoglądam | Oczywiście .......... możliwość, że gdzieś popełniłem błąd i chętnie go poprawię. |
| impf | filler | 0 | starają | Najpierw cię ignorują, potem się .........., a na końcu szanują. |

**Set 2: base**

| aspect | item | correct | verbDisplayed | sentenceDisplayed |
|---|---|---|---|---|
| perf | expItem | 1 | zaufał | W ubiegłym sezonie trener chyba za bardzo .......... swojej szczęśliwej gwieździe. |

| | | | | |
|---|---|---|---|---|
| perf | expItem | 1 | uniknęła | W zeszłym roku, ta międzynarodowa firma .......... płacenia za licencję. |
| perf | expItem | 1 | przyswoili | W ubiegłym roku, licealiści .......... sobie wiedzę o istniejących rodzajach podatków. |
| perf | expItem | 1 | zrzuciły | Poprzedniego wieczora, wielkie bombowce .......... na miasto śmiercionośny ładunek. |
| perf | expItem | 1 | zapisałeś | Na ostatnim spotkaniu .......... wszystkie uwagi, jakie mieli pracownicy. |
| perf | expItem | 1 | pokłóciłaś | Powiedziałaś, że wczoraj .......... się z nim przez telefon. |
| perf | expItem | 1 | zniszczyliście | Jeśli wcześniej .......... swoją relację, terapia będzie wymagało dużo ciężkiej pracy. |
| perf | expItem | 1 | napisałyście | Jak .......... wcześniej, gołąb może być interpretowany jako zwiastun dobrej nowiny. |
| perf | expItem | 1 | dorzuciłem | W zeszłym sezonie .......... się do kupna piłek dla naszej drużyny. |
| perf | expItem | 1 | przejechałam | Wydaje mi się, że wczoraj .......... tamtędy wracając z pracy. |
| perf | expItem | 1 | zaprojektowaliśmy | W zeszłym miesiącu .......... nowe produkty, które mają pomóc starszym osobom. |
| perf | expItem | 1 | przyjechałyśmy | W lipcu .......... do naszej babci, żeby pomóc jej w ogródku. |
| impf | expItem | 1 | ufał | Przyznał, że w zeszłym roku za bardzo .......... swojej intuicji. |
| impf | expItem | 1 | unikała | W zeszłym roku, ich firma prawnicza .......... płacenia składek. |
| impf | expItem | 1 | chorowali | W zeszłym miesiącu, nasi chłopcy .......... na zapalenie płuc. |
| impf | expItem | 1 | zrzucały | W zeszłym tygodniu, brytyjskie samoloty .......... tam paczki z żywnością. |
| impf | expItem | 1 | zapisywałeś | Wczoraj .......... jego numer na kartce papieru leżącej obok telewizora. |
| impf | expItem | 1 | kłóciłaś | Z tego, co słyszałem, to wczoraj .......... się z Markiem. |
| impf | expItem | 1 | zachowywaliście | .......... wtedy zimną krew i dlatego jesteśmy z was dumni. |
| impf | expItem | 1 | pisałyście | W trakcie ostatniego spotkania, .......... raport podsumowujący przeprowadzone prace. |
| impf | expItem | 1 | negocjowałem | W maju .......... ze swoimi klientami nowe warunki kontaktu. |

| | | | | |
|---|---|---|---|---|
| impf | expItem | 1 | przejeżdżałam | W zeszłym tygodniu .......... tamtędy, żeby sprawdzić, czy coś się zmieniło. |
| impf | expItem | 1 | projektowaliśmy | W ubiegłym miesiącu, na potrzeby naszego klienta, .......... materiały reklamowe. |
| impf | expItem | 1 | wyruszałyśmy | Potem razem .......... na wyprawę do sklepu po colę i cukierki. |
| perf | filler | 1 | doczekamy | Ciekawe, czy kiedyś .......... napisów w języku angielskim w publicznych miejscach. |
| perf | filler | 1 | naciągają | Oszuści .......... przede wszystkim osoby starsze, więc warto przestrzec seniorów. |
| perf | filler | 1 | opublikuje | Portal .......... artykuł opisujący każdą pracę oraz wywiady ze zwycięzcami. |
| perf | filler | 1 | pomówimy | Dziś .......... o akustyce, która ma duży wpływ na finalny efekt. |
| perf | filler | 1 | pójdzie | Nie oznacza to jednak, że twoje dziecko .......... właśnie tą drogą! |
| perf | filler | 1 | przygotuję | Z takim zapleczem i sprzętem w czterdzieści minut .......... trzydaniowy obiad. |
| perf | filler | 1 | rozegrają | Kolejne spotkanie na wyjeździe stomilowcy .......... dopiero w najbliższą środę. |
| perf | filler | 1 | sięgną | Rządzący nie .......... po takie okulary, przez które widać prawdziwe życie. |
| perf | filler | 1 | usprawni | Ustawa przegłosowana w Parlamencie Europejskim .......... kontrolę i egzekwowanie przepisów prawa. |
| perf | filler | 1 | wyjdziemy | Ze szkolenia .......... zadowoleni, niezależnie od powodów dla których się zgłaszamy. |
| perf | filler | 1 | wypuszczą | Twórcy obiecują, że niedługo .......... serię aktualizacji zwiększającej funkcjonalność tej aplikacji. |
| perf | filler | 1 | zapomną | Nie .......... pierwszego uczucia, lecz po latach będą już innymi ludźmi. |
| impf | filler | 1 | cieszy | Bardzo mnie .......... naprawdę udany debiut wychowanka naszej akademii. |
| impf | filler | 1 | grzebie | Przy barze blondyn .......... wykałaczką w stojącym przed nim drinku. |
| impf | filler | 1 | jeżdżą | Znam wiele osób, które do pracy .......... rowerem zamiast samochodem. |
| impf | filler | 1 | maleje | Wraz z upływem czasu, .......... ostrożna uprzejmość obecna podczas pierwszej rozmowy. |

| | | | | |
|---|---|---|---|---|
| impf | filler | 1 | narzeka | Wiele osób .......... na schorzenia kręgosłupa związane z wykonywaną pracą. |
| impf | filler | 1 | przedkładam | Ja .......... jakość nad ilość – wolę kupować rzeczy droższe, ale lepsze. |
| impf | filler | 1 | rezygnujemy | Sami korzystamy z tego hostingu i .......... z niego wkrótce. |
| impf | filler | 1 | rozróżniamy | W zależności od powierzchni, .......... krioterapię miejscową i ogólnoustrojową. |
| impf | filler | 1 | snują | Niektórzy .......... przypuszczenia, że kometa mogłaby spowodować ogólnoświatowy kataklizm. |
| impf | filler | 1 | udowadnia | Łukasz na każdym kroku .........., że zna się na budowaniu relacji. |
| impf | filler | 1 | wracamy | Jeśli .......... większą grupą osób, warto przemyśleć wynajęcie busa. |
| impf | filler | 1 | zaczynamy | Alergię na soję możemy wykryć dopiero kiedy .......... karmić dziecko, |
| perf | expIte m | 0 | zmartwiły | Przedwczoraj, jej przyjaciel .......... się z kolacji dużą ilością pracy. |
| perf | expIte m | 0 | przypisaliśmy | W trakcie ostrzału, maszyną, którą sterował podporucznik, .......... potężna eksplozja. |
| perf | expIte m | 0 | zachowaliście | U ludzi, którzy niedawno .........., intensywność terapii powinna być szczególnie duża. |
| perf | expIte m | 0 | poradziłyście | W zeszłym tygodniu, .......... mnie wieści od najbardziej doświadczonych pracowników firmy. |
| perf | expIte m | 0 | wykręcił | W końcówce meczu, .......... się i próbowałeś strzelać, ale bez skutku. |
| perf | expIte m | 0 | zachorowali | Dwa lata temu .......... zdecydowanie więcej niż twoi koledzy. |
| perf | expIte m | 0 | wyruszyłyśmy | Jeśli wcześniej .......... postęp gry, możecie kontynuować waszą przygodę. |
| perf | expIte m | 0 | pożarłam | Chciałabym się dowiedzieć, jak sobie wtedy .......... z objawami ospy. |
| perf | expIte m | 0 | zarobiłaś | Zeszłej zimy, zgodnie z radą Barbary, .......... z szefem zmianę umowy. |
| perf | expIte m | 0 | szarpnęła | Dziś popołudniu, pożerałam chipsy słuchając mojej ulubionej audycji radiowej. |
| perf | expIte m | 0 | poderwałeś | Na spotkaniu, do każdego z punktów .......... pytania i propozycje odpowiedzi. |

| | | | | |
|---|---|---|---|---|
| perf | expItem | 0 | wynegocjowałem | Po pracy .......... sprawdzić jeszcze kilka zgłoszeń, które miałyśmy na liście. |
| impf | expItem | 0 | pożerałam | Wczoraj szef .......... się ze spotkania, mówiąc że ma ważne telefony. |
| impf | expItem | 0 | radziłyście | W trakcie ich ostatniej kłótni .......... go za rękę. |
| impf | expItem | 0 | przypisywaliśmy | W ubiegłym semestrze, uczniowie .......... sobie materiał z ostatnich rozdziałów podręcznika. |
| impf | expItem | 0 | przyswajali | W zeszłym roku .......... mnie dolegliwości bólowe mojej mamy. |
| impf | expItem | 0 | przyjeżdżałyśmy | W trakcie koncertu .......... się z miejsca i próbowałeś śpiewać. |
| impf | expItem | 0 | podrywałeś | W zeszłym sezonie, to ty .......... więcej od niego. |
| impf | expItem | 0 | martwiły | Wczoraj, na oczach wszystkich sąsiadów, .......... mój skalny ogródek. |
| impf | expItem | 0 | szarpała | Napiszcie proszę, jak sobie wtedy .......... z opuchniętymi stopami. |
| impf | expItem | 0 | wykręcał | W zeszłym roku .......... się do prezentów urodzinowych dla moich znajomych. |
| impf | expItem | 0 | niszczyliście | Dziś popołudniu .......... chipsy słuchając mojej ulubionej audycji radiowej. |
| impf | expItem | 0 | zarabiałaś | Na ostatnim zebraniu, .......... nowe role każdemu z uczestników. |
| impf | expItem | 0 | dorzucałem | Zeszłej zimy .......... do schroniska dla zwierząt, żeby nakarmić psy. |
| perf | filler | 0 | sprawią | Jeśli tylko .......... szczęście, to można w tym miejscu zobaczyć delfiny. |
| perf | filler | 0 | przeciska | Nie .......... telefonu, bo jego głos zacznie źle na mnie działać. |
| perf | filler | 0 | odbiorę | Myślę, że zawodnik .......... presji ze strony działaczy i kibiców. |
| perf | filler | 0 | dopisze | Jeśli takich działań się nie .........., powiaty zostaną pozbawione swoich praw. |
| perf | filler | 0 | wezwę | Przystawka napędzana przez silnik miksera .......... ciasto przez tarcze formujące. |
| perf | filler | 0 | podoła | Gdy .......... się treściom większości spektakli teatralnych, wnioski nasuwają się jednoznaczne. |
| perf | filler | 0 | wypłaci | Po roku rzępolenia .......... skrzypce w futerale i zostanie zwykłą sekretarką. |
| perf | filler | 0 | powstrzyma | Nie warto upierać się przy zajęciach, które nie .......... dziecku przyjemności. |

| | | | | |
|---|---|---|---|---|
| perf | filler | 0 | przyjrzymy | W ostateczności .......... taksówkę i jakoś sama dotrę do szpitala. |
| perf | filler | 0 | schowa | Na początku sierpnia pracodawca .......... wynagrodzenie za przepracowaną część lipca. |
| perf | filler | 0 | zdążę | Może bandyta .......... się obok budynku i napadnie nas po zmroku? |
| perf | filler | 0 | zaczai | Mam nadzieję, że .......... w weekend obejrzeć ten film. |
| impf | filler | 0 | gwarantuję | Sam .......... komu i ile pieniędzy wysyłam z mojego konta. |
| impf | filler | 0 | ostrzega | .......... miłą i bezstresową atmosferę pracy oraz wysoką pensję. |
| impf | filler | 0 | tracisz | Większość z nas .......... tą markę wyłącznie z klockami do budowania. |
| impf | filler | 0 | rozmawiają | To może oznaczać, że .......... dobry moment na zakup mieszkania. |
| impf | filler | 0 | nadchodzi | Firma .........., że ze względu na harmonogram, dostawy mogą się opóźnić. |
| impf | filler | 0 | decyduję | Kobieta wcale nie musi być tym rodzicem, który .......... dziecko |
| impf | filler | 0 | przewija | O problemie .......... bardzo niechętnie, a jeśli już, to temat bagatelizują. |
| impf | filler | 0 | żegnam | Plenerowe koncerty i widowiska .......... całe miasta i przyjezdnych gości. |
| impf | filler | 0 | wyprowadza | Gotując w tradycyjny sposób, .......... wiele witamin i składników mineralnych. |
| impf | filler | 0 | kojarzy | .......... dalej, bo moderatorzy odpisali, że nie zmienią domeny. |
| impf | filler | 0 | walczę | Paulina codziennie .......... swoje psy na długi spacer do parku. |
| impf | filler | 0 | skupiają | Gdy wychodzę, zawsze czule .......... się ze swoim kotem. |

**Set 2: cued**

| aspect | item | correct | verbDisplayed | sentenceDisplayed |
|---|---|---|---|---|
| perf | expItem | 1 | zaufał | W ubiegłym sezonie trener chyba za bardzo .......... swojej szczęśliwej gwieździe. |
| perf | expItem | 1 | uniknęła | W zeszłym roku, ta międzynarodowa firma .......... płacenia za licencję. |
| perf | expItem | 1 | przyswoili | W ubiegłym roku, licealiści .......... sobie wiedzę o istniejących rodzajach podatków. |

| | | | | |
|---|---|---|---|---|
| perf | expItem | 1 | zrzuciły | Poprzedniego wieczora, wielkie bombowce .......... na miasto śmiercionośny ładunek. |
| perf | expItem | 1 | zapisałeś | Na ostatnim spotkaniu .......... wszystkie uwagi, jakie mieli pracownicy. |
| perf | expItem | 1 | pokłóciłaś | Powiedziałaś, że wczoraj .......... się z nim przez telefon. |
| perf | expItem | 1 | zniszczyliście | Jeśli wcześniej .......... swoją relację, terapia będzie wymagało dużo ciężkiej pracy. |
| perf | expItem | 1 | napisałyście | Jak .......... wcześniej, gołąb może być interpretowany jako zwiastun dobrej nowiny. |
| perf | expItem | 1 | dorzuciłem | W zeszłym sezonie .......... się do kupna piłek dla naszej drużyny. |
| perf | expItem | 1 | przejechałam | Wydaje mi się, że wczoraj .......... tamtędy wracając z pracy. |
| perf | expItem | 1 | zaprojektowaliśmy | W zeszłym miesiącu .......... nowe produkty, które mają pomóc starszym osobom. |
| perf | expItem | 1 | przyjechałyśmy | W lipcu .......... do naszej babci, żeby pomóc jej w ogródku. |
| impf | expItem | 1 | ufał | Przyznał, że w zeszłym roku często za bardzo .......... swojej intuicji. |
| impf | expItem | 1 | unikała | W zeszłym roku, ich firma prawnicza co miesiąc .......... płacenia składek. |
| impf | expItem | 1 | chorowali | W zeszłym miesiącu, nasi chłopcy długo .......... na zapalenie płuc. |
| impf | expItem | 1 | zrzucały | W zeszłym tygodniu, brytyjskie samoloty .......... tam co tydzień paczki z żywnością. |
| impf | expItem | 1 | zapisywałeś | Wczoraj co chwilę .......... jego numer na kartce papieru leżącej obok telewizora. |
| impf | expItem | 1 | kłóciłaś | Z tego, co słyszałem, to wczoraj co chwilę .......... się z Markiem. |
| impf | expItem | 1 | zachowywaliście | Zawsze .......... wtedy zimną krew i dlatego jesteśmy z was dumni. |
| impf | expItem | 1 | pisałyście | W trakcie ostatniego spotkania, wciąż .......... raport podsumowujący przeprowadzone prace. |
| impf | expItem | 1 | negocjowałem | W maju zwykle .......... ze swoimi klientami nowe warunki kontaktu. |
| impf | expItem | 1 | przejeżdżałam | W zeszłym tygodniu codziennie .......... tamtędy, żeby sprawdzić, czy coś się zmieniło. |

| | | | | |
|---|---|---|---|---|
| impf | expItem | 1 | projektowaliśmy | W ubiegłym miesiącu, na potrzeby naszego klienta, całymi dniami .......... materiały reklamowe. |
| impf | expItem | 1 | wyruszałyśmy | Potem zwykle razem .......... na wyprawę do sklepu po colę i cukierki. |
| perf | filler | 1 | doczekamy | Ciekawe, czy kiedyś .......... napisów w języku angielskim w publicznych miejscach. |
| perf | filler | 1 | naciągają | Oszuści .......... przede wszystkim osoby starsze, więc warto przestrzec seniorów. |
| perf | filler | 1 | opublikuje | Portal .......... artykuł opisujący każdą pracę oraz wywiady ze zwycięzcami. |
| perf | filler | 1 | pomówimy | Dziś .......... o akustyce, która ma duży wpływ na finalny efekt. |
| perf | filler | 1 | pójdzie | Nie oznacza to jednak, że twoje dziecko .......... właśnie tą drogą! |
| perf | filler | 1 | przygotuję | Z takim zapleczem i sprzętem w czterdzieści minut .......... trzydaniowy obiad. |
| perf | filler | 1 | rozegrają | Kolejne spotkanie na wyjeździe stomilowcy .......... dopiero w najbliższą środę. |
| perf | filler | 1 | sięgną | Rządzący nie .......... po takie okulary, przez które widać prawdziwe życie. |
| perf | filler | 1 | usprawni | Ustawa przegłosowana w Parlamencie Europejskim .......... kontrolę i egzekwowanie przepisów prawa. |
| perf | filler | 1 | wyjdziemy | Ze szkolenia .......... zadowoleni, niezależnie od powodów dla których się zgłaszamy. |
| perf | filler | 1 | wypuszczą | Twórcy obiecują, że niedługo .......... serię aktualizacji zwiększającej funkcjonalność tej aplikacji. |
| perf | filler | 1 | zapomną | Nie .......... pierwszego uczucia, lecz po latach będą już innymi ludźmi. |
| impf | filler | 1 | cieszy | Bardzo mnie .......... naprawdę udany debiut wychowanka naszej akademii. |
| impf | filler | 1 | grzebie | Przy barze blondyn .......... wykałaczką w stojącym przed nim drinku. |
| impf | filler | 1 | jeżdżą | Znam wiele osób, które do pracy .......... rowerem zamiast samochodem. |
| impf | filler | 1 | maleje | Wraz z upływem czasu, .......... ostrożna uprzejmość obecna podczas pierwszej rozmowy. |
| impf | filler | 1 | narzeka | Wiele osób .......... na schorzenia kręgosłupa związane z wykonywaną pracą. |

| | | | | |
|---|---|---|---|---|
| impf | filler | 1 | przedkładam | Ja .......... jakość nad ilość – wolę kupować rzeczy droższe, ale lepsze. |
| impf | filler | 1 | rezygnujemy | Sami korzystamy z tego hostingu i .......... z niego wkrótce. |
| impf | filler | 1 | rozróżniamy | W zależności od powierzchni, .......... krioterapię miejscową i ogólnoustrojową. |
| impf | filler | 1 | snują | Niektórzy .......... przypuszczenia, że kometa mogłaby spowodować ogólnoświatowy kataklizm. |
| impf | filler | 1 | udowadnia | Łukasz na każdym kroku .........., że zna się na budowaniu relacji. |
| impf | filler | 1 | wracamy | Jeśli .......... większą grupą osób, warto przemyśleć wynajęcie busa. |
| impf | filler | 1 | zaczynamy | Alergię na soję możemy wykryć dopiero kiedy .......... karmić dziecko, |
| perf | expItem | 0 | zmartwiły | Przedwczoraj, jej przyjaciel .......... się z kolacji dużą ilością pracy. |
| perf | expItem | 0 | przypisaliśmy | W trakcie ostrzału, maszyną, którą sterował podporucznik, .......... potężna eksplozja. |
| perf | expItem | 0 | zachowaliście | U ludzi, którzy niedawno .........., intensywność terapii powinna być szczególnie duża. |
| perf | expItem | 0 | poradziłyście | W zeszłym tygodniu, .......... mnie wieści od najbardziej doświadczonych pracowników firmy. |
| perf | expItem | 0 | wykręcił | W końcówce meczu, .......... się i próbowałeś strzelać, ale bez skutku. |
| perf | expItem | 0 | zachorowali | Dwa lata temu .......... zdecydowanie więcej niż twoi koledzy. |
| perf | expItem | 0 | wyruszyłyśmy | Jeśli wcześniej .......... postęp gry, możecie kontynuować waszą przygodę. |
| perf | expItem | 0 | pożarłam | Chciałabym się dowiedzieć, jak sobie wtedy .......... z objawami ospy. |
| perf | expItem | 0 | zarobiłaś | Zeszłej zimy, zgodnie z radą Barbary, .......... z szefem zmianę umowy. |
| perf | expItem | 0 | szarpnęła | Dziś popołudniu, .......... chipsy słuchając mojej ulubionej audycji radiowej. |
| perf | expItem | 0 | poderwałeś | Na spotkaniu, do każdego z punktów .......... pytania i propozycje odpowiedzi. |
| perf | expItem | 0 | wynegocjowałem | Po pracy .......... sprawdzić jeszcze kilka zgłoszeń, które miałyśmy na liście. |

| | | | | |
|---|---|---|---|---|
| impf | expItem | 0 | pożerałam | Wczoraj szef długo .......... się ze spotkania, mówiąc że ma ważne telefony. |
| impf | expItem | 0 | radziłyście | W trakcie ich ostatniej kłótni ciągle .......... go za rękę. |
| impf | expItem | 0 | przypisywaliśmy | W ubiegłym semestrze, uczniowie codziennie .......... sobie materiał z ostatnich rozdziałów podręcznika. |
| impf | expItem | 0 | przyswajali | W zeszłym roku często .......... mnie dolegliwości bólowe mojej mamy. |
| impf | expItem | 0 | przyjeżdżałyśmy | W trakcie koncertu co chwilę .......... się z miejsca i próbowałeś śpiewać. |
| impf | expItem | 0 | podrywałeś | W zeszłym sezonie, to ty często .......... więcej od niego. |
| impf | expItem | 0 | martwiły | Wczoraj, na oczach wszystkich sąsiadów, całymi godzinami .......... mój skalny ogródek. |
| impf | expItem | 0 | szarpała | Napiszcie proszę, jak sobie wtedy zwykle .......... z opuchniętymi stopami. |
| impf | expItem | 0 | wykręcał | W zeszłym roku często .......... się do prezentów urodzinowych dla moich znajomych. |
| impf | expItem | 0 | niszczyliście | Dziś popołudniu, długo .......... chipsy słuchając mojej ulubionej audycji radiowej. |
| impf | expItem | 0 | zarabiałaś | Na ostatnim zebraniu, ciągle .......... nowe role każdemu z uczestników. |
| impf | expItem | 0 | dorzucałem | Zeszłej zimy co tydzień .......... do schroniska dla zwierząt, żeby nakarmić psy. |
| perf | filler | 0 | sprawią | Jeśli tylko .......... szczęście, to można w tym miejscu zobaczyć delfiny. |
| perf | filler | 0 | przeciska | Nie .......... telefonu, bo jego głos zacznie źle na mnie działać. |
| perf | filler | 0 | odbiorę | Myślę, że zawodnik .......... presji ze strony działaczy i kibiców. |
| perf | filler | 0 | dopisze | Jeśli takich działań się nie .........., powiaty zostaną pozbawione swoich praw. |
| perf | filler | 0 | wezwę | Przystawka napędzana przez silnik miksera .......... ciasto przez tarcze formujące. |
| perf | filler | 0 | podoła | Gdy .......... się treściom większości spektakli teatralnych, wnioski nasuwają się jednoznaczne. |
| perf | filler | 0 | wypłaci | Po roku rzępolenia .......... skrzypce w futerale i zostanie zwykłą sekretarką. |

| | | | | |
|---|---|---|---|---|
| perf | filler | 0 | powstrzyma | Nie warto upierać się przy zajęciach, które nie .......... dziecku przyjemności. |
| perf | filler | 0 | przyjrzymy | W ostateczności .......... taksówkę i jakoś sama dotrę do szpitala. |
| perf | filler | 0 | schowa | Na początku sierpnia pracodawca .......... wynagrodzenie za przepracowaną część lipca. |
| perf | filler | 0 | zdążę | Może bandyta .......... się obok budynku i napadnie nas po zmroku? |
| perf | filler | 0 | zaczai | Mam nadzieję, że .......... w weekend obejrzeć ten film. |
| impf | filler | 0 | gwarantuję | Sam .......... komu i ile pieniędzy wysyłam z mojego konta. |
| impf | filler | 0 | ostrzega | .......... miłą i bezstresową atmosferę pracy oraz wysoką pensję. |
| impf | filler | 0 | tracisz | Większość z nas .......... tą markę wyłącznie z klockami do budowania. |
| impf | filler | 0 | rozmawiają | To może oznaczać, że .......... dobry moment na zakup mieszkania. |
| impf | filler | 0 | nadchodzi | Firma .........., że ze względu na harmonogram, dostawy mogą się opóźnić. |
| impf | filler | 0 | decyduję | Kobieta wcale nie musi być tym rodzicem, który .......... dziecko |
| impf | filler | 0 | przewija | O problemie .......... bardzo niechętnie, a jeśli już, to temat bagatelizują. |
| impf | filler | 0 | żegnam | Plenerowe koncerty i widowiska .......... całe miasta i przyjezdnych gości. |
| impf | filler | 0 | wyprowadza | Gotując w tradycyjny sposób, .......... wiele witamin i składników mineralnych. |
| impf | filler | 0 | kojarzy | .......... dalej, bo moderatorzy odpisali, że nie zmienią domeny. |
| impf | filler | 0 | walczę | Paulina codziennie .......... swoje psy na długi spacer do parku. |
| impf | filler | 0 | skupiają | Gdy wychodzę, zawsze czule .......... się ze swoim kotem. |