Title:

Imperfective aspect underspecified for number: Evidence from an eye-tracking during reading experiment[1]

Dorota Klimek-Jankowska (University of Wrocław), Anna Czypionka (University of Constance) and Joanna Błaszczak (University of Wrocław)

Abstract

In an eye-tracking during reading experiment we investigated the processing of ambiguous Polish imperfective verbs in contexts with disambiguating ('frequently' and 'yesterday') and neutral preceding adverbs. Grammatical number of NP objects was also manipulated. Verb regions received significantly longer regression path times when following a neutral compared to 'yesterday' contexts. This implies that in neutral contexts both senses of polysemous imperfective verbs are activated on the verbal region. Post-hoc analyses revealed more regressions from singular objects in neutral contexts, suggesting that a preference for a more frequent plural event sense was created before the first fixations on the object were made. Finally, we observed an effect consisting of longer first pass times on singular objects and more regressions from the following region in contexts with 'frequently', which is consistent with the view that imperfective aspect is underspecified for number. This pattern of results is compatible with Relevance Theory, which posits that the selection of one sense (single ongoing or plural) is an outcome of an inferential process based on frequency, context and world knowledge. However, the fact that sense frequency plays a role in this process indicates that it serves as input to context-based inferential processes suggesting that this information is pre-stored in the memory.

Keywords: imperfective aspect in Polish, polysemy, underspecification, number, eye-tracking

[1] This research was supported by the Polish National Science Center (NCN) grant OPUS 5 HS2 (DEC-2013/09/B/HS2/02763).

1. Introduction

This study brings together insights from theoretical linguistics and psycholinguistics in order to contribute to a growing body of research whose goal is to create a uniform semantics of semantically underspecified imperfective aspect. Our second goal is to understand how imperfective aspect is processed in real time (see, for example, Deo 2009, 2015; Ferreira 2004, 2005; Hacquard 2006; Klimek-Jankowska, Czypionka, Witkowski and Błaszczak 2018). Imperfective aspect is cross-linguistically multiply ambiguous. Languages differ in the range of possible readings of imperfective, but the two most canonical ones are single and plural event readings. In its single event reading the imperfective verb refers to an event which is incomplete at the asserted interval, as in *Kiedy podlewała*^{I} *kwiaty w ogrodzie, zauważyła*^{P} *nadchodzącą sarnę*[2] 'When she was watering flowers in the garden, she noticed an approaching roe deer' (see Willim 2006: 200–201). By contrast, in the plural event reading, the imperfective verb refers to a series of events happening on several occasions, as in *Kiedy podlewała*^{I} *kwiaty w ogrodzie wieczorami, słuchała*^{I} *śpiewu ptaków* 'When she watered flowers in the garden in the evenings, she listened to chirping birds'. While most aspectologists agree on the fact that imperfective aspect is semantically underspecified (cf. Borik 2002/2006; Comrie 1976; Dahl 1985; Filip 1993/1999; Forsyth 1970; Hacquard 2006; Kagan 2008, 2010; Klein 1995; Paslawska and von Stechow 2003; Willim 2006), the exact nature of its underspecification is still unclear.

The persistent question is what exactly imperfective is underspecified for. Hacquard (2006) and Frąckowiak (2015) argue that imperfective aspect is a semantically vacuous morpheme whose distinct meanings are introduced by distinct, phonologically null operators PROG (progressive) or HAB (habitual). One problem for this approach is that in Polish imperfective aspect is used to express plural event readings in contexts whose meaning is not necessarily habitual (more details will be provided in Section 2.2). A more plausible answer to this question is that

[2] The superscript I stands for imperfective aspect (imperfective verb), the superscript P stands for perfective aspect (perfective verb).

imperfective aspect is underspecified for number, as proposed by Ferreira (2004, 2005), who suggests that imperfective aspect is an operator which selects for either singular or plural VPs referring to singular or plural events respectively. However, he does not specify which factors determine whether a singular or a plural VP is selected by imperfective. In the present study we will pursue the view that imperfective verbs are in fact polysemous and we may reach a better understanding of the nature of underspecification of imperfective verbs by paying closer attention to recent psycholinguistic research on the processing of semantically underspecified polysemous words (e.g., Falkum 2010, 2011; Foraker and Murphy 2012; Frazier 1999; Frazier and Rayner 1990; Frisson 2015; Klein and Murphy 2001, 2002; Klepousniotou and Baum 2007; Klepousniotou, Pike, Steinhauer and Gracco 2012; Klepousniotou, Titone and Romero 2008; Pickering and Frisson 2001; Pylkkänen, Llinas, and Murphy 2006;). No such research has been done on the processing of semantically underspecified grammatical categories, imperfective aspect being one of them. This study fills in this niche by reporting the results of a pilot eye-tracking during reading experiment in which we examine the impact of disambiguating context on the time-course of interpretation of imperfective verbs in Polish.

This paper is organized as follows. Section 2 introduces relevant background on grammatical aspect in Polish with a special focus on the semantic underspecification of imperfective aspect. Arguments are presented to support the claim that imperfective aspect is underspecified for number. In Section 3, the formal approach to imperfective aspect is extended with questions inspired on the basis of an overview of earlier psycholinguistic research on the processing of semantically underspecified words. These questions were addressed in a pilot study using eye-tracking during reading to monitor the impact of context on the time-course of processing of imperfective aspect in Polish, which is described in Section 4. Section 5 discusses the implications of the findings and future directions and Section 6 presents our conclusions.

2. Grammatical aspect in Polish

2.1. Perfective and imperfective aspect – general facts

In Polish, almost all verbs[3] (including infinitives) are either perfective or imperfective, as shown in (1) and (2) respectively.

[3] With the exception of biaspectual verbs such as, for example, *anulować* 'to cancel' and *aresztować* 'to arrest'.

(1) *Jan* *jechał*^{I}.
Jan.NOM drove.IPFV
'Jan drove.'

(2) *Jan* *przejechał*^{P} *dziesięć* *mil*.
Jan.NOM drove.PFV ten.ACC miles.GEN
'Jan drove ten miles.' (Willim 2006: 175)

Additionally, most verbs in Polish have both perfective and imperfective variants. Polish perfective verbs are usually morphologically marked by means of a prefix or a suffix, as shown in (3a,b) respectively (cf. Bogusławski 1963; Nagórko 1998; Willim 2006; Wróbel 1999, 2001).[4]

(3) a. *pisać*^{I} – ***na****pisać*^{P} 'to write'
b. *błyskać*^{I} – *błys****nąć***^{P} 'to flash'

Imperfective verbs form two classes: primary imperfectives (bare, i.e., 'unprefixed' Vs) (see (4a), (5a)) and secondary imperfectives usually signaled by the presence of an *-ywa* suffix and its allomorphs or by stem alternation (see (4b), (5b)).

(4) a. *pisać*^{I} 'to write'
b. *podpis****ywa****ć*^{I} 'to sign (imperfective)'

(5) a. *bić*^{I} 'to hit'
b. *wbi****ja****ć*^{I} 'to hammer'

A standard assumption is that a distinction should be made between lexical aspect corresponding to Vendler's (1957) states, activities, achievements or accomplishments, and grammatical aspect, which most typically corresponds in languages of the world to the perfective-imperfective formal opposition. The relation between lexical aspect and grammatical aspect is captured in de Swart's (1998: 348) formal semantic representation [Tense [Aspect* [eventuality description]]], in which Tense scopes over grammatical Aspect, which in turn

[4] There are also underived perfective verbs in Polish, for example, *dać*^{P} 'to give', *wziąć*^{P} 'to take', *kupić*^{P} 'to buy'.

scopes over lexical eventuality description of a verbal predicate. The Kleene star * indicates that there may be more aspectual operators.

As stated in Willim (2006: 202), perfective aspect has a very specific meaning. It is in a vast majority of cases used to refer to a single, well-delimited event occurring on a specific occasion. In contrast, imperfective aspect has a wider meaning in that it can be used to describe unbounded, iterative or habitual eventualities. There is a general consensus that imperfective verbs reflect the perspective of an 'insider' who sees a portion of an event from the inside and is oblivious to its endpoints (see Kazanina and Phillips 2003). In more formal terms, imperfective aspect introduces an inclusion relation between the event time interval and reference time interval, where the former includes the latter leaving the potential endpoints of an event from view (for more discussion see, among others, Borik 2002/2006; Comrie 1976; Kamp and Reyle 1993; Klein 1994; Reichenbach 1947; Smith 1991). What is still unclear is why the imperfective aspect is cross-linguistically multiply ambiguous. Languages differ in the range of possible readings of imperfective. As pointed out in Section 1, the two most canonical ones are single event and plural event readings. For this reason, different linguists treat imperfective aspect as non-aspect, non-perfective, semantically underspecified, or semantically unmarked (cf. Borik 2002/2006; Comrie 1976; Dahl 1985; Filip 1993/1999; Forsyth 1970; Kagan 2008, 2010; Klein 1995; Paslawska and von Stechow 2003; Willim 2006).

2.2. Arguments in favor of imperfective being underspecified

Hacquard (2006) argues that imperfective aspect has no meaning at all and its single ongoing or plural readings are realized by covert operators PROG or HAB. Imperfective marking is then taken to be the reflex of the presence of these covert operators in the syntactic structure. A similar view is proposed by Frąckowiak (2015), who following Hacquard (2006) claims that imperfective is a semantically vacuous morpheme whose distinct meanings are introduced by distinct, phonologically null operators. One problem for the approach proposed by Hacquard (2006) is that in Polish imperfective aspect is used to express plural event readings in contexts whose meaning is not necessarily habitual, as shown in (6):

(6) a. *Jan spotykał dzisiaj ludzi z wielu zakątków świata.*
John meet.IPFV.PST.3SG.F today people from different parts world.GEN
'John kept meeting people from different parts of the world today.'

b. *Jan dwa razy czytał tę książkę.*
John two times read.IPFV.PST.3SG.M this book
'John has read this book twice.'

c. *Zawsze kiedy mężczyźni wracali z łowów,*
always when men return.IPFV.PST.3PL from hunts,
cała wioska zbierała się przy ognisku.
whole village gather.IPFV.PST.3SG.F REFL by fire
'Whenever the men returned from hunting, the whole village gathered by the fire.'

In (6a) there were several occasions of John's meeting people from different parts of the world on a specific day. Imperfective here is used to express a plurality of events but the events in the plural set are distributed over a relatively short temporal interval and do not constitute a habit. In (6b) John read a book twice (not habitually). Finally, in (6c) on every occasion of the men returning from hunting, the whole village gathered by the fire. In (6c), the plural event reading results from the universal quantification over events by means of the adverbial quantifier *zawsze* 'always'. It has been convincingly argued by Ferreira (2004, 2005) that contexts with adverbs of quantification have a different semantic interpretation than bare habitual contexts. This shows that there are several plural event uses of imperfective aspect which cannot be captured by the semantics of the HAB operator. What still remains to be addressed is the question of what exactly imperfective aspect is underspecified for. An answer to this question may be inspired by Ferreira's (2004, 2005) number approach to imperfective aspect. Ferreira (2004, 2005) argues that the singular/plural opposition used by Link (1983) to distinguish between singular and plural individuals in the domain of individuals applies to events as well with plural events being characterizable as mereological sums having singular events as their minimal parts. Ferreira (2004, 2005) argues that imperfective aspect is an operator which selects for either plural or singular VPs: IPFV [VP_{sg}/VP_{pl}]. The single ongoing interpretation of an imperfective verb is derived from the logical form with the Imperfective selecting for VP_{sg}, as presented in (7).

(7) $[\![\text{Impfv}_{sg}]\!] = \lambda P_{sg}.\ \lambda t.\ \exists e: \tau(e) \supseteq t \text{ and } P(e) = 1$

The plural event reading of an imperfective verb is derived from the logical form with the Imperfective selecting for VP_{pl}, as formally represented in (8).

(8) $[\![\text{Impfv}_{pl}]\!] = \lambda P_{pl}.\ \lambda t.\ \exists e: \tau(e) \supseteq t \text{ and } P(e) = 1$

Ferreira (2004, 2005) accounts for the unbounded interpretation of imperfective aspect by assuming Klein's (1995) time relational semantics where the perspective time t is included in the temporal trace of an event τ(e).[5]

What is not clear from Ferreira's (2004, 2005) theory is which factors determine the selection of the single event or plural event reading by the imperfective operator. What is important in our study is the observation that bare imperfective verbs (used in isolation) may have a more dominant singular or plural event sense. For example, such imperfective verbs as *palić* 'smoke', *gotować* 'cook', *sprzątać* 'clean', *uczyć* 'teach', *myć* 'wash', *jeść* 'eat' (and the predicates used in our questionnaire) have a more dominant (more frequent) plural event sense while such imperfective verbs as *rodzić* 'give birth', *umierać* 'die' have a more dominant single event reading. This suggests that the ultimate choice of a plural/singular event interpretation of imperfective verbs may be potentially influenced by these preferences associated with polysemous imperfective verbal forms.

Coming back to our central question, namely, which factors determine the selection of a single event or plural event reading of imperfective verbs, one of the crucial factors affecting the ultimate choice of a singular ongoing and plural event reading of imperfective aspect is the grammatical number of the NP objects, as discussed by Klimek-Jankowska and Błaszczak (2021). Examples in (9) and (10) illustrate this point.

(9)	*Rubens*	*malował*	*kobietę.*
	Rubens	paint.IPFV.PST.3SG.M	woman.SG.ACC
	'Rubens was painting a woman.'		

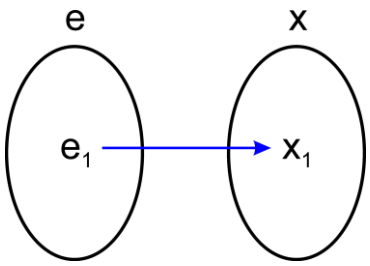
(10)	*Rubens*	*malował*	*kobiety.*
	Rubens	paint.IPFV.PST.3SG.M	woman.PL.ACC

[5]In order to formally capture the fact that under the plural event reading of imperfective each of the events in the plural set is distributed over separate time intervals, Ferreira (2004, 2005) assumes that the domain of intervals D*i* contains singular and plural intervals and there is a homomorphism τ between the structured domain of events and the structured domain of intervals.

'Rubens painted women.'

In (9), in which a singular (indefinite[6]) NP object is used, the imperfective predicate denotes a single ongoing eventuality. Crucially, the plural event reading is blocked in this case. However, when we change the grammatical number of the NP object in (10) to plural, the plural event reading becomes available. The above examples demonstrate that the number of an NP object plays an important role for the interpretation of semantically underspecified imperfective verbs in Polish. Altogether, the grammatical number of NP objects of imperfective verbs plays an important role in the specification process. Under this view, the sentences in (9) and (10) have two possible contextual interpretations each, one with a singular event e and one with a plural event E (note that x is used to represent singular individuals and X is used to represent plural individuals), as presented in (11) and (12) respectively.[7]

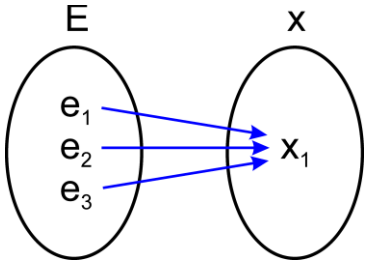
(11) a. **∃e ∃x** [paint(e) ∧ AGENT (Rubens, e) ∧ woman(x) ∧ THEME (e) = x]



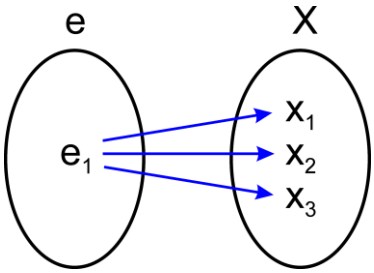
b. **∃E ∃x** [paint(E) ∧ AGENT (Rubens, E) ∧ woman(x) ∧ THEME (E) = x]

[6] In Polish, there is no indefinite marking in NPs but the indefinite/definite reading of bare singular nouns is determined by the information structure. More precisely, under normal intonation the sentence stress falls on the final element, that is, the default placement of the focus exponent in Slavic is in the right periphery of a sentence (see Junghanns 2002, 2003).

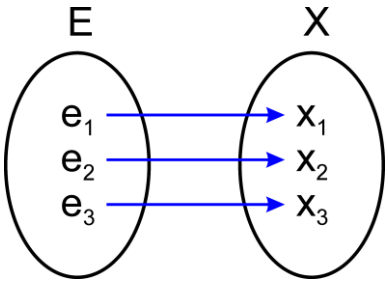
[7] In the following the neo-Davidsonian event semantics is adopted according to which every sentence contains an event argument variable which gets existentially bound by default, verbs are one-place predicates of eventualities, verb arguments are conjuncts bound by the same existential quantifier and they denote relations between an eventuality argument and some participant argument.



(12) a. **∃e ∃X** [paint(e) ∧ AGENT (Rubens, e) ∧ woman(X) ∧ THEME (e) = X]



b. **∃E ∃X** [paint(E) ∧ AGENT (Rubens, E) ∧ woman(X) ∧ THEME (E) = X ∧ f: E↔X][8]



This means that the sentence in (9) with an imperfective verb and a singular NP object is in principle semantically ambiguous between the meanings represented in (11a) and (11b). However, the meaning in (11b) where there is a plural event of Rubens' painting the same woman is pragmatically implausible, therefore the meaning in (11a) is strongly preferred. Similarly, the sentence in (10) is ambiguous between the meanings represented in (12a) and (12b) but the meaning where there is a single event of Rubens' painting multiple women, as indicated in (12a), is pragmatically implausible, therefore the meaning in (12b) is strongly preferred. The configuration in (12b) is the only one in which there are two plural sets and it is possible to establish a bijection (one-to-one) relation between events and individuals in the plural sets E and X giving rise to a dependent reading between pairs of individuals (denoted by an NP_{pl}) and events (denoted by a VP_{pl}). A singular NP object is disfavored in those contexts

[8] $E \leftrightarrow X$ represents a bijection (one-to-one) relation between members of the plural event E (understood as a sum of events) and the members of the plural entity X (understood as a sum of individuals).

in which it is pragmatically more plausible that each plural event involves a different entity. In such contexts plural NP objects are preferred. Under Ferreira's (2004, 2005) number theory of imperfective aspect, we expect the configuration in (11b) to be less acceptable in contexts unambiguously supporting the plural event reading of imperfective verbs combined with singular NP objects in Polish. This should be particularly expected in contexts in which it is pragmatically more plausible that each instantiation of a given plural event involves a different object. We decided to test whether this prediction extends to a larger set of imperfective verbs in Polish in an acceptability rating study described in the following section.

2.3. Imperfective underspecified for number – an acceptability rating study

To examine Ferreira's (2004, 2005) number theory of imperfective aspect, we tested the acceptability of constructions with imperfective verbs involving plural and singular NP objects in sentences with adverbs supporting either the plural event (*często* 'frequently') or single event (*wczoraj* 'yesterday') reading as compared to neutral adverbs (*powoli* 'slowly'). All the tested constructions are listed in Table 1.

Table 1: Example sentences for each of the tested contexts

Condition	**Example sentence**
YESTERDAY-SINGULAR	*Piotrek wczoraj sporządzał raport.* lit. 'Peter yesterday prepared.IPFV report.'
YESTERDAY-PLURAL	*Piotrek wczoraj sporządzał raporty.* lit. 'Peter yesterday prepared.IPFV reports.'
FREQUENTLY-SINGULAR	*Piotrek często sporządzał raport.* lit. 'Peter frequently prepared.IPFV report.'
FREQUENTLY-PLURAL	*Piotrek często sporządzał raporty* lit. 'Peter frequently prepared.IPFV reports.'
NEUTRAL-SINGULAR	*Piotrek powoli sporządzał raport.* lit. 'Peter slowly prepared.IPFV report.'
NEUTRAL-PLURAL	*Piotrek powoli sporządzał raporty* lit. 'Peter slowly prepared.IPFV reports.'

The adverb *wczoraj* 'yesterday' creates a preference for a single ongoing meaning, the frequentative adverb *często* 'frequently'[9] supports the plural event meaning and a neutral adverb *powoli* 'slowly' does not cue any of the meanings. As pointed out above in the discussion of the data in (9) and (11b), the interpretation where there is a plural event of Rubens' painting the same woman is pragmatically implausible. The reading in which each event in the plural set of events involves a different entity in a context with an adverb *często* 'frequently' and an imperfective verb is more naturally expressed with a plural NP object allowing for a bijection relation between the set of events in the denotation of an imperfective verb and the set of individuals in the denotation of a plural NP object. Therefore, we expect FREQUENTLY-SINGULAR to be less acceptable than the remaining conditions. We checked this expectation in a web-based acceptability rating study. Stimuli for the rating study consisted of 600 items in total, divided equally into 6 conditions. The participants were asked to evaluate each sentence on a three-point scale (1 = bad, 2 = OK but could be better, 3 = good). In order to determine the significance of differences between conditions, a Cumulative Link Mixed Model using *clmm* function from *ordinal* package (Christensen 2019) was fitted. The final model included the fixed effect of Condition presented in Table 1; random intercepts for experimental items and random intercepts and random slopes for participants.[10] Significance of the fixed effect of Condition was based on the comparison of the model with intercept only (no effect of interest present) to the model with the fixed effect of interest. Statistical analysis showed a significant result for the fixed effect ($\chi^2(5) = 34.988$; $p < 0.0001$). FREQUENTLY-SINGULAR is rated significantly lower than the remaining conditions, as shown in Figure 1.

[9] According to de Swart (1993), *often* or *frequently* is a frequency adverb and *w*hen it is used in out of the blue contexts, it has a frequency (non-relational) reading and it operates on just one set of events and simply describes the events in question as recurring in time with a certain frequency and regularity. The frequency with which relevant events happen is evaluated as high relative to a contextually determined norm and they are regularly distributed over the relevant part of the time axis (de Swart 1993: 24–25).

[10] An anonymous reviewer asked why the final model did not include the random slopes for items in addition to the remaining random effects. This is because the model with random slopes for items produced a poorer fit (AIC = 811.49) than a model without the random slopes for items (AIC = 809.49).

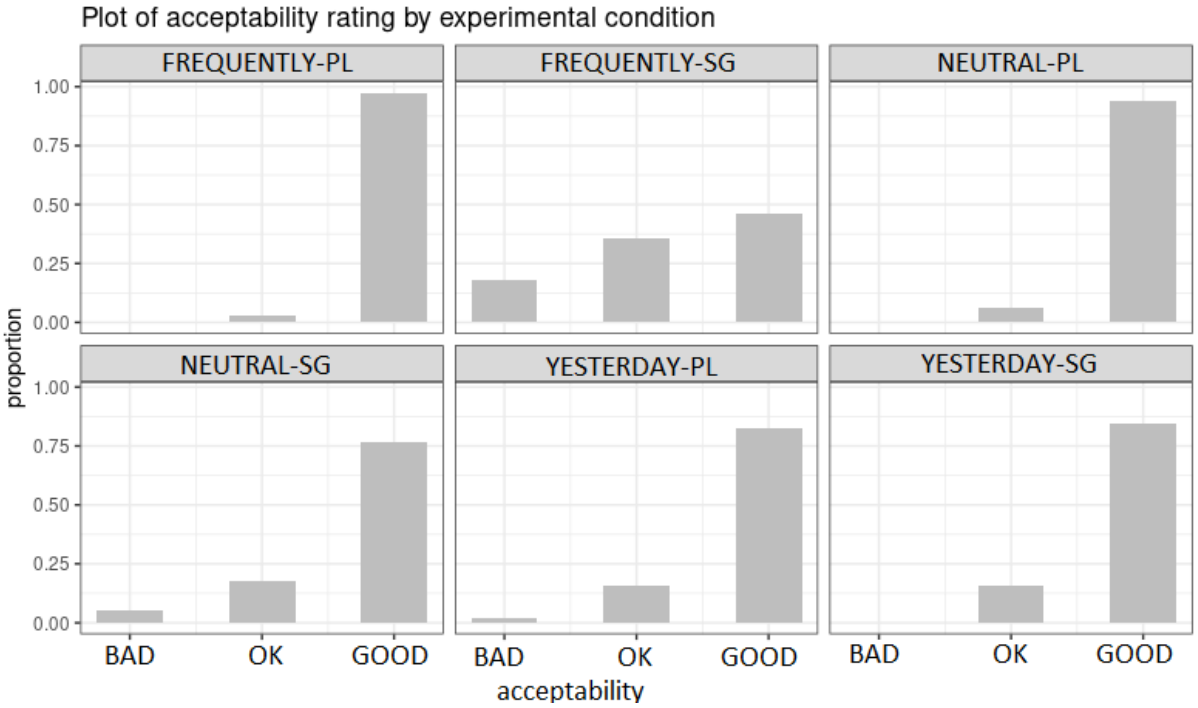


Figure 1: Proportions of absolute counts for acceptability responses (BAD, OK (but could be better), GOOD) for the combinations of aspect and number used in the pretest.

The results of our pilot study showed that contexts in which imperfective verbs are preceded by the adverb *często* 'frequently' and followed by a singular NP object are significantly less acceptable than all the remaining types of the tested contexts. This suggests that frequency adverbs force a plural event reading of imperfective verbs whereas its singular NP object refers to a single entity. This results in the interpretation that each of the events in the plural event in the denotation of an imperfective verb involves the same participant, which in most of the cases is pragmatically implausible. These results are clearly in line with the number approach to the imperfective aspect assumed in this study.

To sum up, inspired by Ferreira (2004, 2005), we adopt the view that imperfective aspect is underspecified for number. In order to understand how this underspecification is resolved in real time, we intend to link insights from Ferreira's (2004, 2005) number theory of aspect and relevant findings of recent studies on the processing of polysemy in our own eye-tracking during reading study. Before we set out to present our experiment, we first offer an overview of the most relevant psycholinguistic approaches to the processing of polysemous words which are essential for formulating our own predictions.

3. Psycholinguistic research on the processing of semantically underspecified words

There are many approaches to the phenomenon of polysemy understood as the instance of a single word which has several related senses. Carston (2019) provides the following examples of the noun *mouth* (adapted from Vicente and Falkum forthcoming).

(13) a. *The whale has a huge mouth*.
b. *The mouth of the cave/tunnel is visible from here*. (an aperture)
c. *The water at the mouth looks green*. (part of a river that enters into an ocean)
d. *I have four mouths to feed*. (whole body/person)
e. *He is a big mouth ('Mack the mouth')*. (someone who talks too much)

The single word *mouth* can be used to refer to the physical mouth of animals/humans in (13a), to the whole body in (13d, e) or to something resembling animal mouths in (13b, c). Polysemy is often contrasted with homonymy in that the senses of a polysemous word are related, while homonymy involves two or more words which have unrelated meanings but their phonological form happens to be the same as in the word *bank* (river bank and a financial institution). The question that is in the center of psycholinguistic research on polysemy is whether different senses of a polysemous word are indeed stored and represented in our mental lexicon or whether we exploit contextual cues to generate new senses of a single lexical entry in online communication. If it is true that different senses of polysemous words are represented in the memory, another relevant question is how they are represented and how they interact with context during language comprehension. There are three main approaches to polysemy which address these questions differently:

(i) Core Meaning hypothesis, which holds that upon encountering a polysemous word, the comprehenders activate its underspecified abstract semantic representation (see Frazier and Rayner 1990; Frisson and Pickering 1999, 2007; Frisson 2009, 2015; Pickering and Frisson 2001). Under this view, comprehenders do not immediately commit to any of the senses. Rather, they delay this decision and use contextual cues to home in on a particular sense. In this sense, it is a top-down model of processing in that context is used to gradually home in on the desired interpretation. Under this approach, sense dominance does not play a role immediately and we should not expect any early effects of sense frequency.

(ii) Sense-Enumeration Lexicon approach according to which different senses of a polysemous word are listed separately but are connected to a joint core representation (see Klein and Murphy 2001, 2002; Pylkkänen, Llinás and Murphy 2006; Foraker and Murphy 2012). Under this view, the processing of polysemous words should proceed in a bottom-up fashion in that different senses of a polysemous word should get automatically activated and the strength of their activation should depend on their meaning frequency. In other words, sense dominance should play a role immediately and early effects of sense frequency are expected.

(iii) Relevance Theory-inspired view was proposed by Falkum (2010, 2011), who assumes that polysemous words are not stored in the lexicon but are rather computed during on-line processing on the basis of context-driven extension (in top-down processing). Falkum (2010, 2011) follows Nunberg's (1979) theory, according to which all that is represented in the lexicon is a Core Meaning of a polysemous word, and its different polysemous extensions are generated on the fly using pragmatic reasoning and world knowledge. This view says that the comprehenders exploit contextual clues to arrive at the intended reading of a polysemous expression. In the absence of such clues, that is in neutral contexts, they choose the most frequent contextual interpretation. Under this approach, sense frequency should not play a role when a preverbal context supports a specific interpretation of a polysemous word, but it should play a role in neutral contexts. In the latter case the most frequent interpretation is chosen.

What is currently lacking is a deeper understanding of whether the same mechanisms which have been observed for polysemous content words apply in the case of a polysemous grammatical category like imperfective aspect. Inspired by these psycholinguistic approaches to the processing of semantically underspecified representations, we set out to answer the following research questions:

(1) How are imperfective verbs processed and what can be inferred from their processing profile about their semantics?
(2) Does the processing of imperfective verbs match one of the views on the polysemous lexicon outlined above?

With these questions in mind, we conducted an eye-tracking during reading experiment. This method offers a number of measures which will allowed us to monitor the impact of context on the incrementality of processing of a semantically underspecified category of imperfective aspect.

4. Eye-tracking during reading experiment (ET)

4.1. Language material

In our experiment, we investigated the impact of preverbal and postverbal context on the time-course of processing of Polish imperfective verbs. In the preverbal context, we used three types of adverbs: *wczoraj* 'yesterday', *często* 'frequently' and a manner adverb such as, e.g., *powoli* 'slowly', creating three types of contexts referred to as YESTERDAY, FREQUENTLY and NEUTRAL respectively. As mentioned earlier in the description of the pretest, the adverb *wczoraj* 'yesterday' creates a preference for the single ongoing meaning, the frequentative adverb *często* 'frequently' cues the plural event meaning and a neutral adverb such as *szybko* 'quickly' or *powoli* 'slowly' does not favor any of the possible meanings of imperfective verbs. Additionally, in each type of tested contexts we used two variants of NP complements differing in grammatical number (SINGULAR or PLURAL), which serves as an important contextual cue pointing to the single or plural event sense of imperfective verbs. While singular NP objects create a strong preference for the single ongoing interpretation, plural NP objects create a strong preference for the plural event reading of imperfective verbs (recall the discussion in Sections 2.2 and 2.3). This gave rise to six groups of experimental conditions: (i) YESTERDAY-SINGULAR, (ii) FREQUENTLY-SINGULAR, (iii) NEUTRAL-SINGULAR, (iv) YESTERDAY-PLURAL, (v) FREQUENTLY-PLURAL, (vi) NEUTRAL-PLURAL. For each imperfective verb we checked its first ten occurrences in the National Corpus of Polish (Przepiórkowski, Bańko, Górski and Lewandowska-Tomaszczyk 2012), in which the verb was used in agentive, affirmative, literal, non-conditional, past tense contexts and in the indicative mood. All the verbs selected for the experiment were more frequently used in a plural event reading. The adverb *wczoraj* 'yesterday' supporting single event readings should lead to a mismatch with a preferred plural event reading of the tested imperfective verbs. Our goal was to examine whether *wczoraj* 'yesterday' leads to an effect of event sense switching (from plural to singular) in imperfective verbs. Regarding the interaction with the grammatical number of NP objects, we investigate whether there is an effect of mismatch between singular objects and contexts in which a commitment is made to a

plural event reading before reaching the object region. In both cases we are interested in whether these effects are immediate or delayed.

The language material consisted of 30 sentences per condition. All sentences were maximally similar. All sentences used in our experiment began with an introductory statement of the type *Mary said that*, followed by an embedded clause (containing the critical imperfective context), followed by a closing statement of the type *and John said so too.* An example of a typical sentence sextet used in our experiment is given in Table 2.

Table 2: Example sentences for experimental conditions divided into Interest Areas

IA1	IA2	IA3	IA4	IA5	IA6	IA7	IA8
Sara	powiedziała, że	Piotr	wczoraj/ często/ powoli	naprawiał	rower/ rowery	i Bartek	też tak powiedział
Sara	said that	Peter	yesterday/ frequently/ slowly	repaired.ipfv	bike/ bikes	and Bartek	said so too

The purpose of the introductory statement was to create a discourse background, that of the closing statement was to create an additional spillover region. Proper names were always female in the introductory statement, and always male in the closing statement. The experimental sentences were divided into eight Interest Areas referred to as IA1, IA2, IA3, IA4, IA5, IA6, IA7, IA8. Conditions begin to differ in IA4; this is the critical Interest Area, containing different adverbs ('yesterday', 'frequently', 'slowly'). IA5 contains imperfective verbs, IA6 contains singular and plural NP complements, IA7 and IA8 are the spillover and wrap-up region. Subjects of the embedded sentences consisted of male proper names, followed by an adverb, an imperfective verb (3-syllable long) and an NP complement. All the words apart from the adverb were kept parallel across conditions for each sentence sextet. NP complements were lexically the same across each sentence sextet but they differed in grammatical number (singular or plural). We selected sixty verbs from the top of a verb frequency list generated by means of the PELCRA search tool for the National Corpus of Polish (Pęzik 2012). This guaranteed that the verbs were comparably frequent and familiar. In the process of selecting verbs, we excluded and replaced those which were lexically punctual (achievements and semelfactives) because these lexical classes combined with imperfective morphology inevitably give rise to a plural

interpretation only. We used only imperfective verbs of accomplishment predicates in our study. For the selected thirty imperfective verbs, the most typical (frequent) NP complements were selected from the National Corpus of Polish (Przepiórkowski et al. 2012) to guarantee the optimal plausibility of verb-complement combinations. Finally, we constructed thirty sentence sextets. The stimuli were arranged in six different versions. Each version contained 5 items per condition (times 6 conditions). We used a Latin Square design to make sure that sentences with the same lexical material were equally distributed across the six versions. In other words, only one sentence from each sextet was used in each version. Each version was divided into four blocks, with a pause between them. In each version, experimental sentences and fillers were randomized. Not more than two sentences from the same condition were displayed one after another. Sentences with the same beginnings up to the verb were equally distributed across versions. A full table of verbs used in the Experiment is given in Appendix A.

4.2. Predictions

There were two specific goals of this study. First, we wanted to examine how early the parser commits to a single ongoing or plural event meaning of imperfective aspect. We also wanted to assess if and how this commitment depends on whether disambiguation cues are present in a preverbal context or not.

There were two crucial manipulations in our study: the manipulation of the preverbal adverb (for which we formulate predictions on the impact of context) and the manipulation of the grammatical number of the NP complement of imperfective verbs (for which we formulate predictions on interaction of context and grammatical number of NP complements).

The employed experimental design (see Section 4.1) was to help us understand whether the processing of imperfective verbs proceeds in a bottom-up fashion (i.e., whether single and plural event senses of imperfective verbs are automatically activated and context helps us select the proper one). Alternatively, it may be the case that the processing of imperfective verbs proceeds in a top-down fashion (i.e., context determines which sense is activated and selected). We will also be able to understand whether the baseline activation of a more frequent plural event sense of the tested imperfective verbs is higher than that of the less frequent singular event sense.

The predictions are made for the verb (IA5), object (IA6), spill-over (IA7) and wrap-up (IA8) regions for the number theory of imperfective aspect and for the three tested models of processing polysemy: (i) the Sense-Enumeration Lexicon model, (ii) the Core-Meaning hypothesis and (iii) the Relevance-Theory hypothesis.

In accordance with the number approach to imperfective aspect, we predict increased processing costs on the singular object resulting from an incompatibility between a preferred plural event reading of an imperfective verb and a singular object in (IA6). This is expected both in FREQUENTLY-SINGULAR as compared with FREQUENTLY-PLURAL and YESTERDAY-SINGULAR and in NEUTRAL-SINGULAR as compared with YESTERDAY-SINGULAR, NEUTRAL-PLURAL. This should be observed under all the three models of processing polysemy. However, only under the Core Meaning Hypothesis, this effect is predicted to show up possibly with a delay (on IA7 and IA8) due to a gradual nature of the process of homing in on the proper interpretation.

Additionally, if the Sense-Enumeration Lexicon approach is correct we expect the effect of sense switching from a dominant plural event sense resulting in higher processing costs for YESTERDAY as compared to NEUTRAL and FREQUENTLY in the verbal region (IA5). No effects are expected on the spillover region (IA7) and wrap-up region (IA8).

Conversely, if the Core Meaning hypothesis is correct we do not expect any effects on the verb (IA5), as in all the contexts it should be equally easy to activate an underspecified abstract semantic representation of polysemous imperfective verbs. However, we expect increased processing costs in the postverbal regions (IA6, IA7 and IA8) in NEUTRAL-SINGULAR as compared with YESTERDAY-SINGULAR and in NEUTRAL-PLURAL as compared with FREQUENTLY-PLURAL since it should be easier to home in on the single and plural event readings in contexts with earlier disambiguation cues than in neutral contexts. This is connected with a gradual (not immediate) nature of the process of homing in on a specific sense under this model.

Finally, the Relevance-Theory inspired view predicts that the comprehenders exploit contextual cues to arrive at the intended reading as soon as they become available. Hence, they should sooner arrive at the contextually appropriate single ongoing or plural event sense in contexts

with *wczoraj* 'yesterday' and *często* 'frequently' as compared to neutral contexts when they read an imperfective verb. This should lead to longer reading times in NEUTRAL on the verbal region (IA5) than in YESTERDAY and FREQUENTLY. No effects are expected on the spillover and wrap-up regions.

Table 3: Effects predicted under different approaches in the tested interest areas

Theoretical approach	IA5 (verb)	IA6 (object)	IA7 (spill-over)	IA8 (wrap-up)
Number theory of imperfective aspect		FREQUENTLY-SINGULAR > YESTERDAY-SINGULAR; FREQUENTLY-SINGULAR > FREQUENTLY-PLURAL; NEUTRAL-SINGULAR > NEUTRAL-PLURAL; NEUTRAL-SINGULAR > YESTERDAY-SINGULAR		
Sense Enumeration Lexicon	YESTERDAY > FREQUENTLY; YESTERDAY > NEUTRAL			
Core Meaning hypothesis		NEUTRAL-SINGULAR > YESTERDAY-SINGULAR; NEUTRAL-PLURAL > FREQUENTLY-PLURAL	NEUTRAL-SINGULAR > YESTERDAY-SINGULAR; NEUTRAL-PLURAL > FREQUENTLY-PLURAL	NEUTRAL-SINGULAR > YESTERDAY-SINGULAR; NEUTRAL-PLURAL > FREQUENTLY-

				PLURAL
Relevance-Theory inspired view	NEUTRAL > FREQUENTLY; NEUTRAL > YESTERDAY			

4.3. Participants

Seventy two Polish native speakers (59 female, mean age 20.3 (SD = 3.08, range 18−23 years) were recruited at the University of Wrocław. Participants received partial course credit. They had no known neurological or reading-related problems. Data from one participant were excluded before the final data analysis because of the poor quality of the recorded material. For the remaining participants, the mean error rate for comprehension questions (see below) for both critical conditions and fillers was 31% (min 25%, max = 37%). Both missing responses and wrong responses were included in the calculation of the mean error rates. A mean error rate of 38% or higher would have led us to the exclusion of the participant data from the analysis. However, none of the participants had an error rate that high.

4.4 Procedure

The experiment was run in eye-tracking lab of the Center for Experimental Research on Natural Language of the University of Wrocław. Participants were tested individually. Eye movements were recorded using an Eye Link 1000 Plus eye-tracker by SR Research interfaced with a compatible PC. The sampling rate was 2000 Hz. Viewing was binocular, but only the dominant eye was monitored. We used a paper roll test to determine eye-dominance. All sentences in this experiment were displayed in a single line with a maximum length of 75 characters. Stimuli were displayed on a 24-inch BenqXL monitor. Participants were seated 61 cm from the computer screen. At this distance, 3.6 characters subtended 1° of visual angle; the eye-tracker has a resolution of <0.01º RMS. Before the experiment, participants were asked to sit comfortably in front of the computer screen. A head support and a chin rest were used to stabilize the head.

After the written instruction, participants received a practice block with 10 sentences and questions, followed by explicit feedback to their answers. Participants were asked to read the sentences for understanding and to read at a normal rate. During the practice session and during the actual experimental session, a calibration routine was performed, and its accuracy was

checked after each sentence. After reading each sentence, the participants pressed a button to remove the sentence. Comprehension questions were asked after each sentence in order to give the participants a task and to keep them attentive. For questions concerning critical sentences, the correct answer was 'no' in 15 questions and 'yes' in the remaining 15 questions. Additionally, for questions concerning filler sentences, the correct answer was 'no' in 45 questions and 'yes' in the remaining 45 questions. In the experimental session, each participant saw 120 sentences divided into 4 blocks of 30 sentences. 30 of the 120 sentences were critical sentences, the remaining 90 sentences were filler sentences. The whole experiment lasted approximately 30 minutes.

4.5. Data preparation and analysis

For each reading time measure, data points longer than 600 ms or shorter than 5 ms were removed as extreme values, together with missing values. Following this, outliers were defined as values that deviated more than two standard deviations from a participant's mean per condition per position, and were removed before the final data analysis. The amount of data removed before analysis is given for each reading time measure below in Section 4.6.

We calculated results for the following reading time measures for all interest areas starting from the position of the verb (i.e., IA5):

- *first pass times* (the sum of all fixations in a region prior to leaving the IA for the first time, either to the left or to the right).[11]
- *regression path durations* (the sum of all fixations from the first fixation in an IA up to but excluding the first fixation to the right of this IA, including any time spent to the left of the IA after a regressive eye movement and any time spent re-reading material in the IA before moving on).
- *total reading time* (the sum of all fixations made within an IA, including those fixations made when re-reading the IA).

[11] We refrain from linking reading time measures with assumptions about early and late processing steps (see Vasishth, von der Malsburg and Engelmann 2013). For the sake of readability, we may occasionally refer to "early" (first pass times) and "late" (regression path durations and total reading time) reading time measures in the running text.

Log-transformed reading times for each measure and position were analyzed in R (R Development Core Team 2019) with linear mixed effects models, using the packages *lme4* (Bates, Mächler, Bolker and Walker 2015, lmer function) and *LMERConvenienceFunctions* (Tremblay and Ransijn 2015, summary function for calculation of *t*- and *p*-values using Satterthwaite approximation). Graphs were prepared using the package *ggplot2* (Wickham 2016).

We specified the factors CONTEXT (three levels: NEUTRAL, FREQUENTLY, YESTERDAY) and NUMBER (two levels: SINGULAR and PLURAL). The information about CONTEXT became available to readers on the adverb, the information about NUMBER became available to readers on the object. For each position and reading time measure, we set out with the most maximal random effects structure warranted by the research question and data, always including random intercepts for participant and item, as well as random slopes for CONTEXT and, where applicable, NUMBER (see also Bates, Kliegl, Vasishth and Baayen 2015 for an outline of a parsimonious mixed modelling approach). We then reduced the random effects structure until the models would converge. Interactions were pursued in a hierarchical fashion. Details and fixed effects tables of the resulting models are given in Appendix B (Table B1 for first pass times, Table B2 for regression path times, Table B3 for total reading times). Analysis of reading time measures progressed in the following way:

- For the verb position (IA5), we analyzed the main effect of CONTEXT on first pass times and regression path times (given that the information about NUMBER only became available when the following word was fixated).

- For the verb position (IA5), we analyzed the main effects and interaction of CONTEXT and NUMBER on total reading times (given that the information about NUMBER would have been available once participants had progressed beyond the verb position).
- For the positions of the object (IA6), spillover (IA7) and wrap-up region (IA8), we analyzed the main effects and interactions of CONTEXT and NUMBER on all reading time measures.

Additionally, we provide a descriptive overview of regression patterns into and out of all IAs from IAs4 to 8. Regressions into IAs give the proportion of items per condition when a regression was made into this IA, averaged over participants. Regressions out of IAs give the proportion of items per condition when a regression was made out of this region, again averaged

over participants. For some descriptively visible differences between conditions, we report the results of post-hoc analyses (see below for more details).

Results are reported if they are statistically significant at the $p < .05$ level, unless explicitly stated otherwise.[12]

4.6. Results

Mean reading time measures and regression probabilities are given in Appendix C. In the following, we provide the results for individual reading times by position.

First pass times

2.1 % of the data were removed as extreme values, and .01% of the data were removed as outliers. A full overview of the outcomes of the statistical analyses is given in Appendix B (Table B1).

Verb (IA5): There were no statistically significant main effects of CONTEXT on the verb (IA5).

Object (IA6): There was a statistically significant interaction of NUMBER and CONTEXT ($t = 2.47$, $p < .05$). The interaction was pursued by assessing the main effect of NUMBER for each level of CONTEXT and the main effect of CONTEXT for each level of NUMBER separately. For individual levels of CONTEXT, there were no statistically significant effects of NUMBER (marginal effect of NUMBER for FREQUENTLY, $t = 2.21$, $p < .06$). For singular, there was a statistically significant difference between FREQUENTLY and NEUTRAL ($t = 3.51$, $p < .01$), and between FREQUENTLY and YESTERDAY ($t = -2.28$, $p < .05$). For plural, there were no statistically significant effects of CONTEXT. Descriptively speaking, first pass reading times for FREQUENTLY-SINGULAR (mean = 248 ms) were longer than those for NEUTRAL-SINGULAR (mean = 201 ms) and those for YESTERDAY-SINGULAR (mean = 213 ms). This effect is visually presented in Figure 2.

[12] Due to the low number of items in our pilot study, we are assuming the traditional *p*-value for statistical significance and are not correcting for multiple comparisons in our analysis (see von der Malsburg and Angele 2017). By doing so, we hope that our data can serve as pointers to inform future studies on aspect processing, and will be replicated in higher-powered settings.

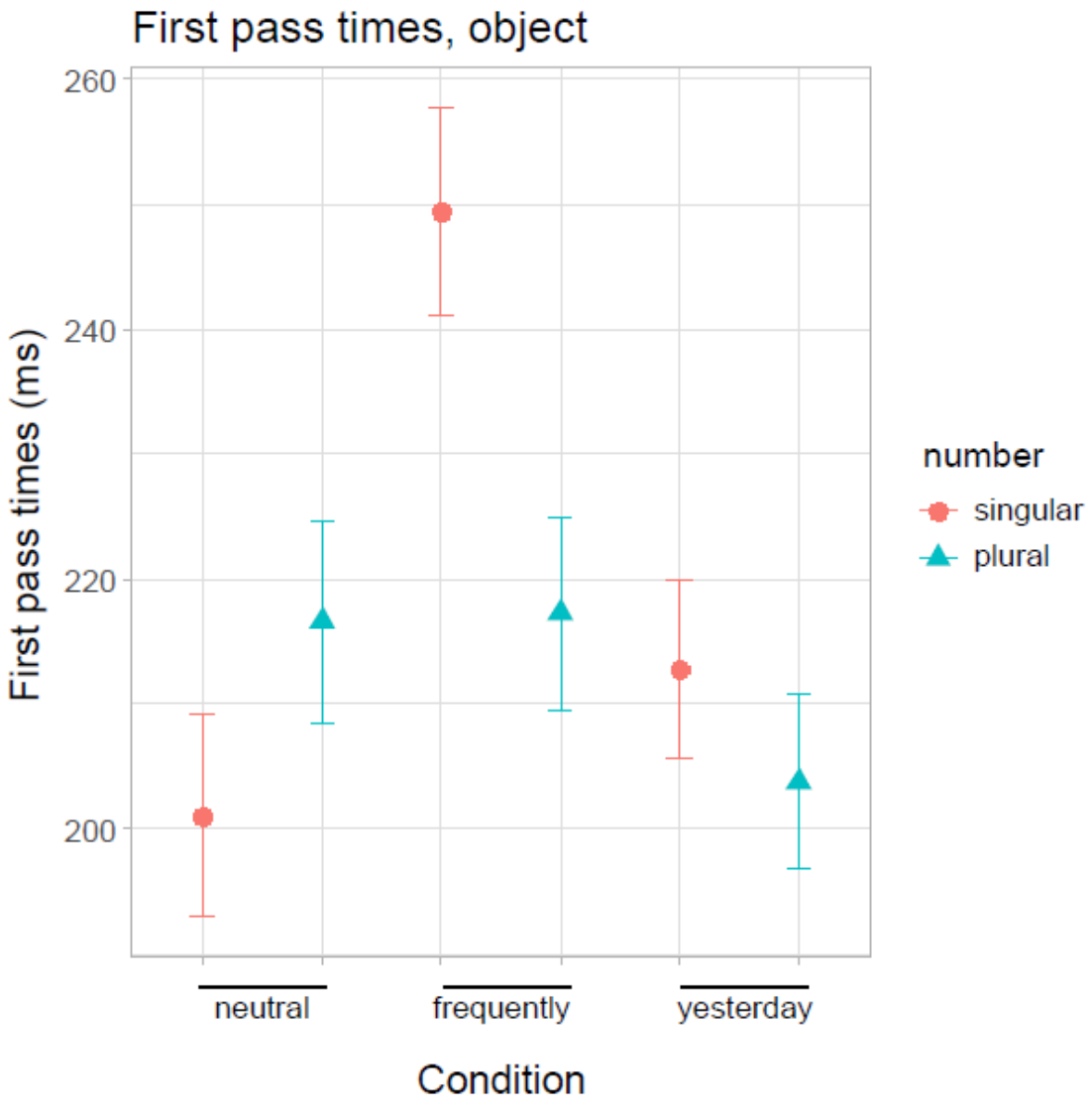


Figure 2: Mean first pass times on the object region (IA6) for each condition, averaged across participants. Error bars depict standard error of the mean.

spillover, wrap-up (IA7): There were no statistically significant main effects or interactions of NUMBER and CONTEXT on the spillover and the wrap-up regions.

Regression path times

2.21 % of the data were removed as missing or extreme values. An additional .01% of the data were removed as outliers. A full overview of the outcomes of the statistical analyses is given in Appendix B (Table B2).

Verb: There was a statistically significant difference between conditions with NEUTRAL and YESTERDAY context (t = -2.30, p < .05). Regression path times were shorter in YESTERDAY than in NEUTRAL context (NEUTRAL-SINGULAR = 526 ms, NEUTRAL-PLURAL = 569 ms, YESTERDAY-SINGULAR = 464 ms, YESTERDAY-PLURAL = 463 ms). There was no statistically significant difference between FREQUENTLY conditions and either NEUTRAL or YESTERDAY conditions. This effect is visually presented in Figure 3.

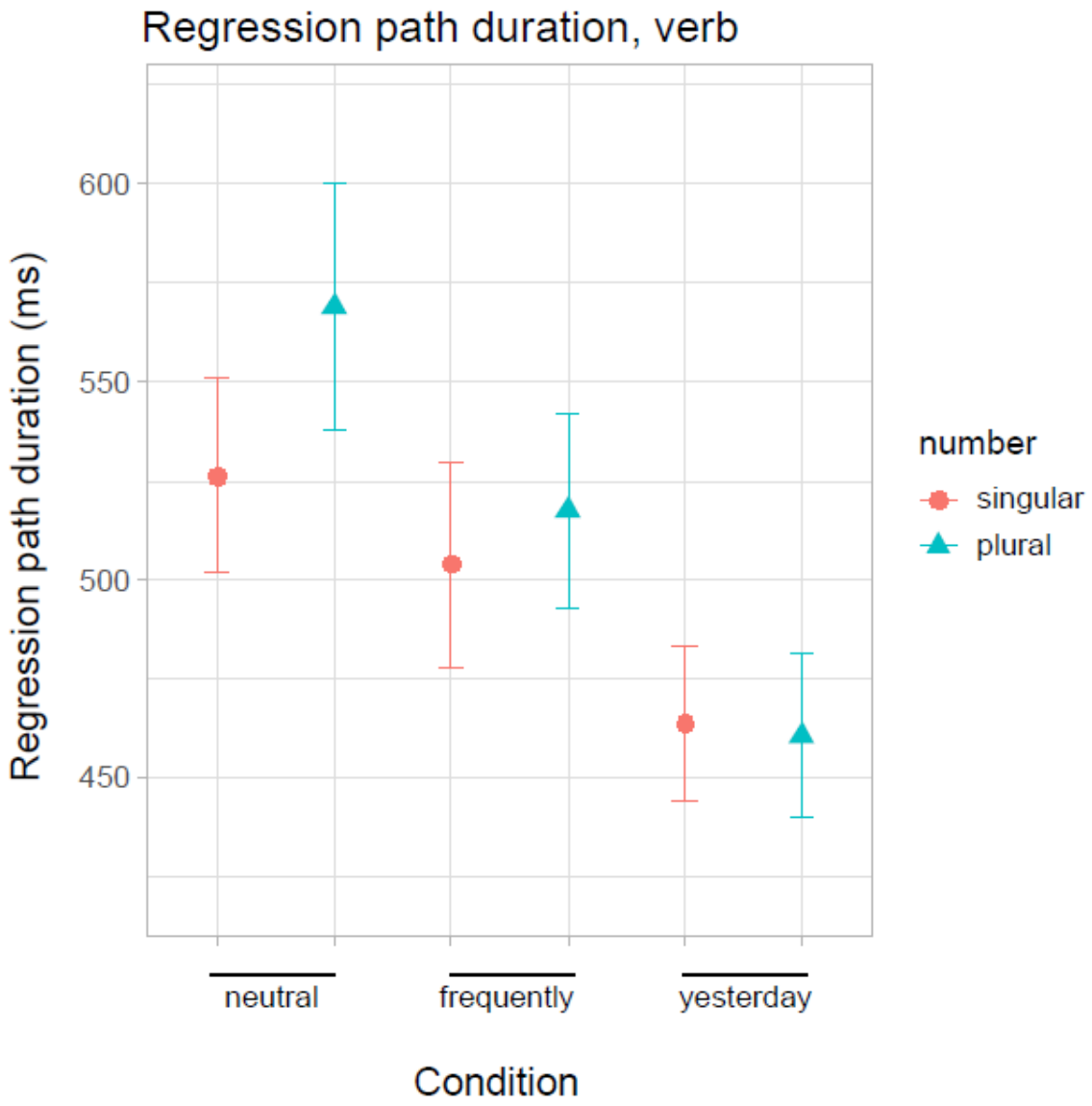


Figure 3: Mean regression path times times on the verb region (IA5) for each condition, averaged across participants. Error bars depict standard error of the mean.

Object (IA6), spill-over (IA7), wrap-up (IA8): There were no significant main effects or interactions of CONTEXT and NUMBER on the object, spillover, and wrap-up regions.

Total reading times

2.08 % of the data were removed as extreme values. An additional .01% of the data were removed as outliers. A full overview of the outcomes of the statistical analyses is given in Appendix B (Table B3).

There were no statistically significant effects on any of the analyzed regions.

Regressions

Due to a low number of items per condition, the pilot dataset does not have enough power to allow for an in-depth statistical analysis of regression probabilities. However, we provide a descriptive overview of the mean regression probabilities for the sake of completeness in Table 3.

For regressions out of object (IA6) and spill-over position (IA7), we provide the results of post-hoc tests that we ran to check if descriptively visible differences are significant. The results of these tests are meant to complement the results of the reading time analysis, and to inform expectations for future studies. A full overview of the results of these post-hoc tests is given in Appendix B (Table B4).

Regressions out of the object position

We ran post-hoc tests on the influence of CONTEXT for singular and plural conditions separately, using the glmer function of the lme4 package. We specified a generalized linear mixed model with CONTEXT as fixed effect, PARTICIPANT and ITEM as random intercepts, and CONTEXT as random slope for PARTICIPANT.

For singular conditions, there was a significant difference between NEUTRAL and FREQUENTLY contexts ($z = -2.33$, $p < .05$), with fewer regressions out of singular objects in FREQUENTLY than in NEUTRAL contexts (FREQUENTLY: 25.3%, NEUTRAL: 41.7%). The difference between YESTERDAY (34.9%) and both other conditions did not reach significance.

For plural conditions, there were no statistically significant effects of CONTEXT.

Regressions out of the spillover region

We ran post-hoc tests on the influence of CONTEXT for singular and plural conditions separately. We specified a generalized linear mixed model with CONTEXT as fixed effect, PARTICIPANT and ITEM as random intercepts, and CONTEXT as random slope for PARTICIPANT.

For singular conditions, there was a significant difference between FREQUENTLY and NEUTRAL (z = -2.42, p < .05), and between FREQUENTLY and YESTERDAY (z = -2.32, p < .05). There were more regressions out of the spillover region in FREQUENTLY-SINGULAR than in NEUTRAL-SINGULAR and YESTERDAY-SINGULAR. There were no statistically significant differences between NEUTRAL-SINGULAR and YESTERDAY-SINGULAR. For plural conditions, there were no statistically significant effects of CONTEXT.

5. General discussion

Table 4 summarizes all the results in the reported eye-tracking study.

Table 4: A summary of the significant results of the eye-tracking study

verb (IA5)	regression path times	NEUTRAL > YESTERDAY (p < .05)
object (IA6)	first pass times	FREQUENTLY-SINGULAR > NEUTRAL-SINGULAR (p < .01) FREQUENTLY-SINGULAR > YESTERDAY-SINGULAR (p < .05)
	regressions out of IA	NEUTRAL-SINGULAR > FREQUENTLY-SINGULAR (p < .05)
spillover (IA7)	regressions out of IA	FREQUENTLY-SINGULAR > NEUTRAL-SINGULAR (p < .05) FREQUENTLY-SINGULAR > YESTERDAY-SINGULAR (p < .05)

We were interested in the impact of context on the time-course of processing of imperfective verbs. We will discuss the findings for each IA separately to allow a closer comparison to the predictions outlined in Section 4.2. This will be followed by a more general discussion related to the tested models of processing polysemy.

Verb (IA5): In the verbal region (IA5), regression path reading times were significantly longer in NEUTRAL contexts than in YESTERDAY contexts. Regression path reading times include any time spent to the left of the IA after a regressive eye movement and any time spent re-reading the material in the IA before moving on. This effect may reflect readers' attempts to resolve the ambiguity of an imperfective verb already on the verbal region (IA5) by inspecting the preceding adverbial. This process appears to be more challenging in neutral contexts because neutral adverbs do not support any of the senses. This in turn would indicate that comprehenders exploit contextual cues as soon as they become available to arrive at a correct sense of a semantically underspecified imperfective verb as predicted under the Relevance-Theory inspired view.

Object (IA7): On the object (IA7), the information about NUMBER first became available. There were significantly longer first pass reading times for FREQUENTLY-SINGULAR than for NEUTRAL-SINGULAR and YESTERDAY-SINGULAR. Regarding regressions, our post-hoc analysis showed that there were more regressions from the object position (IA7) in NEUTRAL-SINGULAR than in FREQUENTLY-SINGULAR. This implies that a preference for a plural event sense had been created by the time the first fixations on the object were made. These results fit the number approach to imperfective aspect, which predicts a mismatch between a singular object and a preferred plural event reading of an imperfective verb.

Spill-over (IA8): Reading time measures did not differ significantly in the spillover region (IA8). However, the post-hoc analysis of descriptive differences in regression showed more regressions from the spillover region in FREQUENTLY-SINGULAR relative to NEUTRAL-SINGULAR and YESTERDAY-SINGULAR. We interpret this as an indication that mismatch detection and resolution (which first led to significant effects on the object region for first pass times) continued on the spillover region.

Taken together, our goal was: (i) to verify if the processing of imperfective verbs matches any of the views on the polysemous lexicon outlined above, (ii) to understand how imperfective verbs are processed and what can be inferred about their semantics from their processing profile.

Concerning the question whether the processing of imperfective verbs matches any of the views on the polysemous lexicon:

(i) There are no effects of sense competition (i.e., effects of switching from a more frequent plural event sense to a less frequent single ongoing sense) on the verbal region (IA5) in contexts supporting the subordinate sense (i.e., YESTERDAY contexts). This does not match the predictions of the Sense Enumeration Lexicon approach under which distinct senses are stored as separate lemmas in the mental lexicon and they compete for selection. Context supporting a subordinate sense should lead to switching from a dominant to a subordinate sense, contrary to fact.

(ii) We report significantly longer regression path reading times on the verbal region (IA5) in NEUTRAL contexts than in YESTERDAY contexts. This implies that the readers did not delay the resolution of the underspecified meaning of imperfective aspect to later regions in neutral context but rather they attempted to resolve it already on the verbal region (IA5). This is not compatible with the Core Meaning hypothesis under which the readers first activate the underspecified abstract semantic representation of a polysemous word and then they home in on one of its senses with a delay. This approach does not predict any effects of resolving underspecification on the region which is underspecified in neutral contexts, contrary to fact.

(iii) The pattern of results obtained in our study is most compatible with the tenets of the Relevance Theory inspired view. According to this view, the interpretation of underspecified words involves the construction of senses on the basis of the interaction of sense frequency, context and world knowledge. The reported longer regression path times on the verbal region (IA5) in a neutral condition indicate that comprehenders exploited contextual cues as soon as they become available.

(iv) The Relevance Theory inspired view is also compatible with the result of our post-hoc analysis which revealed more regressions from the object position (IA6) in NEUTRAL-SINGULAR than FREQUENTLY-SINGULAR. This suggests that a more frequent plural event sense was either selected or received a higher activation than the single ongoing one by the time the first fixations on the object were made. This result implies that the process of resolving an underspecified meaning of imperfective aspect is sensitive to sense frequency.

What is puzzling is that the Relevance Theory inspired approach views the interpretation of polysemy as a purely pragmatic phenomenon with no senses being pre-stored in the memory. Under this view, sense selection is an output of context-based inferential processes. However, the fact that sense frequency plays a role in this process indicates that it serves as input to context-based inferential processes, which in turn implies that information about sense frequency is indeed pre-stored in the memory. This speaks in favour of a less radical version of Relevance-Based approach to polysemy proposed by Carston (2019, 2020), who claims that the interpretation of polysemous words is both semantic and pragmatic. This means that frequency information about senses of polysemous words is lexically pre-stored and it supplies pragmatic inferential processes about which sense is the most likely candidate for selection. Single ongoing and plural event senses of imperfective verbs do not compete for selection because they are part of a single lexical representation. Once this single lexical representation is selected, both senses enter the inferential processes leading to the selection of one of them. However, it is not only sense frequency but also context and world knowledge which collectively determine the output of this process (the output being the most optimal sense). What seems to play a crucial role in this context-dependent inferential process leading to the selection of a single ongoing or plural event sense of an imperfective verb is the grammatical number of an NP object, as supported by the fact that we obtained significantly longer first pass reading times in the object region for FREQUENTLY-SINGULAR than for NEUTRAL-SINGULAR and YESTERDAY-SINGULAR. This fits the number approach of imperfective aspect, which predicts a mismatch between a singular object and a dominant plural event reading of an imperfective verb facilitated by the adverb *często* 'frequently'.

This brings us to the next question, namely what can the processing profile of imperfective verbs tell us about their semantics? As stated in the introduction and in Section 2.2 (see also Section 2.3), we would like to propose that imperfective aspect is an operator which is underspecified for number. This proposal is inspired by Ferreira's (2004, 2005) number theory of imperfective, according to which IMPF selects for either singular or plural VPs referring to singular or plural events respectively. What is not clear in Ferreira's (2004, 2005) theory is which factors determine the selection.

We would like to modify his proposal and suggest that imperfective verbs are in fact polysemous and they enter the interpretation process as a single underspecified lexical

representation with two senses: (i) a single ongoing sense and (ii) a plural event sense. The selection of one of these senses is an outcome of an inferential process based on information about their frequency, context and world knowledge. Additionally, we propose that it is the number of an NP object which is one of the key components determining the ultimate selection of the most optimal sense. We think that significantly longer first pass times on the object position in FREQUENTLY-SINGULAR result from the fact that in the context with the adverb *często* 'frequently' the plural event sense is selected already on the verbal region and a singular NP object leads to a mismatch because it is impossible to create a bijection relation between the members of a plural set of events and a singular entity in the denotation of a singular NP object. The fact that this effect was significant in a larger set of data used in our eye-tracking experiment suggests that it is generally pragmatically implausible to have a plurality of events (distributed over separate time intervals) each of which involves the same entity.[13]

Conclusion

In the reported eye-tracking during reading study we investigated the processing of Polish ambiguous imperfective verbs with plural and singular NP objects in contexts with neutral and disambiguating preceding adverbs. We postulate that verbs marked for imperfective aspect are underspecified for number. More specifically, they are polysemous and they enter the interpretation process as a single underspecified lexical representation with two senses: (i) a single ongoing sense and (ii) a plural event sense. We show that the grammatical number of an NP object plays a crucial role in the sense selection process. We provide evidence that readers do not delay the resolution of the underspecified meaning of imperfective aspect but rather they attempt to arrive at the target reading immediately. This is compatible with Relevance Theory, which (in its strict variant) posits that the interpretation of polysemous words is a purely pragmatic context-based inferential process. We show, however, that this context-based top-down process leading to a selection of an optimal sense relies on sense frequency which is pre-stored in the memory and as such it serves as input to this inferential processes. This speaks in

[13] The reported results come from a pilot study and should be interpreted with caution. More work is needed to further verify our claims. However, we believe that such attempts at reconciling facts from theory and language processing can serve to inform future higher-powered studies investigating more tailored predictions, and are a valuable step in the direction of reaching a better understanding of the puzzling category of imperfective aspect.

favor of a less radical version of Relevance-Based approach to polysemy proposed by Carston (2019, 2020), who claims that the interpretation of polysemous words is both semantic and pragmatic.

Abbreviations

3	third person
ACC	accusative
F	feminine
GEN	genitive
IPFV	imperfective
M	masculine
NOM	nominative
PFV	perfective
PL	plural
PST	past tense
REFL	reflexive
SG	singular

Acknowledgements

Acknowledgement to two anonymous reviewers for their very helpful comments.

This research was supported by the Polish National Science Center (NCN) grant OPUS 5 HS2 (DEC-2013/09/B/HS2/02763).

References

Bates, D., R. Kliegl, S. Vasishth and H. Baayen. 2015. "Parsimonious mixed models." *arXiv preprint arXiv:1506.04967*.

Bates, D., M. Mächler, B. Bolker and S. Walker. 2015. "Fitting linear mixed-effects models using lme4". *Journal of Statistical Software* 67(1). 1–48.

Bogusławski, A. 1963. "Czasowniki dokonane niedokonane i ich stosunki wzajemne" [Perfective and imperfective verbs and their mutual relations]. In: Doroszewski, W. (ed.), *Z polskich studiów slawistycznych* [From Polish studies on Slavic languages]. Warszawa: Polska Akademia Nauk. 153–159.

Borik, O. 2002/2006. *Aspect and reference time*. Doctoral dissertation. Utrecht University. [Published 2006 by Oxford University Press].

Carston, R. 2019. *"*Ad hoc concepts, polysemy and the lexicon". In: Scott, K., B. Clark and R. Carston, (eds.), Relevance, pragmatics and interpretation. Cambridge, UK.: Cambridge University Press. 150–162.

Carston, R. 2020. "Polysemy: Pragmatics and sense conventions". *Mind and Language* 36(1). 108–133.

Comrie, B. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Christensen, R. H. B. 2019. Ordinal – Regression Models for Ordinal Data. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

Dahl, Ö. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Deo, A. 2009. "Unifying the imperfective and the progressive: Partitions as quantificational domains". *Linguistics and Philosophy* 32(5). 475–521.

Deo, A. 2015. "The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift". *Semantics and Pragmatics* 8(14). 1–52.

Falkum, I. L. 2010. "Systematic polysemy and the count-mass distinction". *UCL Working Papers in Linguistics* 22. 16–40.

Falkum, I. L. 2011. *The semantics and pragmatics of polysemy: A Relevance-Theoretic account.* Doctoral dissertation. University College London, UK.

Ferreira, M. 2004. "Imperfectives and plurality". *Proceedings of SALT* 14. 74–91.

Ferreira, M. 2005. *Event quantification and plurality*. Doctoral dissertation. Massachusetts Institute of Technology.

Filip, H. 1993/1999. *Aspect, situation types and noun phrase semantics*. Doctoral dissertation. University of California at Berkeley. [Published 1999 as Aspect, eventuality types and noun phrase semantics. New York/London: Garland Publishing.]

Foraker, S. and G. L. Murphy. 2012. "Polysemy in sentence comprehension: Effects of meaning dominance". *Journal of Memory and Language* 67. 407–425.

Forsyth, James. 1970. *A grammar of aspect*. Cambridge: Cambridge University Press.

Frazier, L. 1999. *On sentence interpretation*. Dordrecht: Kluwer.

Frazier, L. and K. Rayner. 1990. "Taking on semantic commitments: processing multiple meanings vs. multiple senses". *Journal of Memory and Language* 29. 181–200.

Frąckowiak, E. 2015. *Understanding situation and viewpoint aspect in Polish through dative anticausative constructions and factual imperfectives*. Doctoral dissertation. University of Ottawa.

Frisson, S. 2009. "Semantic underspecification in language processing". *Language and Linguistic Compass* 3(1). 111–127.

Frisson, S. 2015. "About bound and scary books: The processing of book polysemies". *Lingua* 157. 17–35.

Frisson, S. and M. J. Pickering. 1999. "The processing of metonymy: Evidence from eye movements". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25. 1347–1365.

Frisson, S. and M. J. Pickering. 2007. "The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard". *Language and Cognitive Processes* 22. 595–613.

Hacquard, V. 2006. *Aspects of modality*. Doctoral dissertation. Massachusetts Institute of Technology.

Junghanns, U. 2002. *Prinzipien der Informationsstrukturierung in slavischen Sprachen*. Habilitation thesis. University of Leipzig.

Junghanns, U. 2003. "Fokussierungsstrategien in slavischen Sprachen". In: Kosta, P., J. Błaszczak, J. Frasek, L. Geist and M. Żygis (eds.), *Investigations into Formal Slavic Linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages (FDSL IV), held at Potsdam University, November 28-30, 2001*. (Linguistik International. 10). Frankfurt am Main: Peter Lang. 181–199.

Kagan, O. 2008. "On the semantics of aspect and number". In: Antonenko, A., J. F. Bailyn and C. Y. Bethin (eds.), *Formal Approaches to Slavic Linguistics. The Stony Brook Meeting 2007*. Michigan Slavic Publications: Ann Arbor. 185–198.

Kagan, O. 2010. "Russian aspect as number in the verbal domain". In: Laca, B. and P. Hofherr (eds.), *Layers of aspect*. Stanford: CSLI Publications. 91–112.

Kamp, H. and U. Reyle, 1993. *From discourse to logic.* Dordrecht: Kluwer.

Kazanina, N. and C. Phillips 2003. "Russian children's knowledge of aspectual distinctions". In: Beachley, B., A. Brown and F. Conlin (eds.), *BUCLD 27 Proceedings*. Somerville MA: Cascadilla Press. 390–401.

Klein, W. 1994. *Time in language*. London and New York, NY: Routledge.

Klein, W. 1995 . "A time-relational analysis of Russian aspect". *Language* 71(4). 669–695.

Klein, D. and G. Murphy. 2001. “The representation of polysemous words”. *Journal of Memory and Language* 45. 259–282.

Klein, D. and G. Murphy. 2002. “Paper has been my ruin: Conceptual relations of polysemous senses”. *Journal of Memory and Language* 47. 548–570.

Klepousniotou, E. and S. R. Baum. 2007. “Clarifying further the ambiguity advantage effect in word recognition: Effects of aging and left-hemisphere damage on the processing of homonymy and polysemy”. *Brain and Language* 103. 148–149.

Klepousniotou, E., G. B. Pike, K. Steinhauer and V. Gracco. 2012. “Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy”. *Brain and Language* 123. 11–21.

Klepousniotou, E., D. Titone and C. Romero. 2008. “Making sense of word senses: The comprehension of polysemy depends on sense overlap”. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34. 1534–1543.

Klimek-Jankowska, D., A. Czypionka, W. Witkowski and J. Błaszczak. 2018. “The time course of processing perfective and imperfective aspect in Polish – Evidence from self-paced reading and eye-tracking experiments”. *Acta Linguistica Academica* 65(2–3). 293–351.

Klimek-Jankowska, D. and J. Błaszczak. 2021**.** “Implications of the number semantics of NP objects on the interpretation of imperfective verbs in Polish“. In: Dočekal M. and M. Wągiel, *Formal Approaches to Number in Slavic and Beyond* (eds.), 99–128. Berlin: Language Science Press.

Link, G. 1983. “The logical analysis of plurals and mass terms: A lattice-theoretical approach”. In: Bäuerle, R., C. Schwarze and A. von Stechow (eds.), *Meaning, use and interpretation of language.* Berlin: de Gruyter. 303–323.

Nagórko, A. 1998. *Zarys gramatyki polskiej (ze słowotwórstwem)* [An outline of Polish grammar (with word-formation)]. Warszawa: Wydawnictwo PWN.

Nunberg, G. 1979. “The non-uniqueness of semantic solutions: Polysemy.” *Linguistics and Philosophy* 3(2). 143–184.

Paslawska, A. and A. von Stechow. 2003. “Perfect readings in Russian”. In: Alexiadou, A., M. Rathert and A. von Stechow (eds.), *Perfect explorations.* Berlin: Mouton de Gruyter. 307–362.

Pęzik, P. 2012. “Wyszukiwarka PELCRA dla danych NKJP” [The PELCRA search engine for the National Corpus of Polish (NKJP)]. In: Przepiórkowski, A., M. Bańko, R. L.

Górski and B. Lewandowska-Tomaszczyk (eds.), *Narodowy Korpus Języka Polskiego* [The National Corpus of Polish]. Warszawa: PWN. 253–279. Available at: http://www.nkjp.uni.lodz.pl/index.jsp

Pickering, M. J. and S. Frisson. 2001. "Processing ambiguous verbs: Evidence from eye movements". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(2). 556–573.

Przepiórkowski, A., M. Bańko, R. L. Górski and B. Lewandowska-Tomaszczyk (eds.). 2012. *Narodowy Korpus Języka Polskiego* [The National Corpus of Polish]. Warszawa: PWN. Available at: http://www.nkjp.uni.lodz.pl/index.jsp

Pylkkänen, L., R. Llinás and G. Murphy. 2006. "The representation of polysemy: MEG evidence". *Journal of Cognitive Neuroscience* 18. 97–109.

R Development Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reichenbach, H. 1947. *Elements of symbolic logic*. New York, NY: Macmillan.

Smith, C. 1991. *The parameter of aspect*. Dordrecht: Kluwer Academic Publishers.

de Swart, H. 1998. "Aspect shift and coercion". *Natural Language and Linguistic Theory* 16(2). 347–385.

de Swart, H. 1993. *Adverbs of quantification: A Generalized Quantifier approach*. Doctoral dissertation. Rijksuniversiteit Groningen.

Tremblay, A. and J. Ransijn. 2015. *LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models.* R Package Version 2.10 Available at: http://cran.r-project.org/package=LMERConvenienceFunctions

Vasishth, S., T. von der Malsburg and F. Engelmann. 2013. "What eye movements can tell us about sentence comprehension". *Wiley Interdisciplinary Reviews: Cognitive Science* 4(2). 125–134.

Vendler, Z. 1957. "Verbs and times". *Philosophical Review* 56. 143–160.

Vicente, A. and Falkum, I. L. forthcoming. "Accounting for the preference of literal meanings in ASC". *Mind & Language*.

Von der Malsburg, T. and B. Angele. 2017. "False positives and other statistical errors in standard analyses of eye movements in reading". *Journal of Memory and Language* 94. 119–133.

Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Willim, E. 2006. *Event, individuation and countability. A study with special reference to English and Polish*. Kraków: Jagiellonian University Press.

Wróbel, H. 1999. "Czasownik" [The verb]. In: Grzegorczykowa, R., R. Laskowski and H. Wróbel (eds.), *Gramatyka współczesnego języka polskiego* [The grammar of contemporary Polish]. Volume 2: *Morfologia* [Morphology]. Warszawa: Wydawnictwo Naukowe PWN. 536–584.

Wróbel, H. 2001. *Gramatyka języka polskiego* [The grammar of Polish]. Kraków: Od Nowa.

Appendix A: The verb phrases used in the experiment

imperfective verb	**singular NP complement**	**plural NP complement**
szacować 'to estimate.ipfv'	stratę 'loss.sg'	straty 'loss.pl'
wygłaszać 'to deliver.ipfv'	wykład 'lecture.sg'	wykłady 'lecture.pl'
naprawiać 'to repair.ipfv'	rower 'bike.sg'	rowery 'bike.pl'
wypełniać 'to fill in.ipfv'	blankiet 'form.sg'	blankiet 'form.pl'
podrywać 'to date.ipfv'	dziewczynę 'girl.sg'	dziewczyny 'girl.pl'
ozdabiać 'to decorate.ipfv'	wnętrze 'interior.sg'	wnętrza 'interior.pl'
sporządzać 'to make.ipfv'	raport 'report.sg'	raporty 'report.pl'
zamiatać 'to sweep.ipfv'	korytarz 'corridor.sg'	korytarze 'corridor.pl'
wyceniać 'to price.ipfv'	działkę 'building plot.sg'	działki 'building plot.pl'
podlewać 'to water.ipfv'	trawnik 'lawn.sg'	trawniki 'lawn.pl'
oceniać 'to grade.ipfv'	sprawdzian 'test.sg'	sprawdziany 'test.pl'
rozliczać 'to settle.ipfv'	wniosek 'application.sg'	wnioski 'application.pl'
montować 'to install.ipfv'	tłumik 'silencer.sg'	tłumiki 'silencer.pl'
wycinać 'to cut down.ipfv'	drzewo 'tree.sg'	drzewa 'tree.pl'
omawiać 'to discuss.ipfv'	problem 'problem.sg'	problemy 'problem.pl'
pakować 'to pack.ipfv'	produkt 'product.sg'	produkty 'product.pl'
nagrywać 'to record.ipfv'	płytę 'disc.sg'	płyty 'disc.pl'
drukować 'to print out.ipfv'	książkę 'book.sg'	książki 'book.pl'
ratować 'to rescue.ipfv'	pacjenta 'patient.sg'	pacjentów 'patient.pl'
poprawiać 'to proofread.ipfv'	esej 'essay.sg'	eseje 'essay.pl'
uszczelniać 'to seal.ipfv'	okno 'window.sg'	okna 'window.pl'
wysyłać 'tp send.ipfv'	paczkę 'package.sg'	paczki 'package.pl'
malować 'to paint.ipfv'	obraz 'painting.sg'	obrazy 'obrazy.pl'
szkicować 'to sketch.ipfv'	budynek 'building.sg'	budynki 'building.pl'
podrabiać 'to fake.ipfv'	podpis 'signature.sg'	podpisy 'signature.pl'
wystawiać 'to give.ipfv'	ocenę 'grade.sg'	oceny 'grade.pl'

rysować ‘to draw.ipfv’	portret ‘portrait.sg’	portrety ‘portrait.pl’
testować ‘to test.ipfv’	maszynę ‘machine.sg’	maszyny ‘machine.pl’
wyłudzać ‘to extort.ipfv’	łapówkę ‘bribe.sg’	łapówki ‘bribe.pl’
usuwać ‘to remove.ipfv’	usterkę ‘flaw.sg’	usterki ‘flaw.pl’

Appendix B: Results of the statistical analyses

Reading time measures

First pass times

Verb	lmer(log(measure) ~ context + (1|vp) + (1|item), data=ia.5)				
	Estimate	**Std. Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
Intercept	5.69	.04	63.44	159.62	<2e-16 ***
YESTERDAY	.06	.04	27.02	1.65	.110
FREQUENTLY	.02	.04	27.02	.55	.586
reference level set to frequently					
Intercept	5.71	.04	63.44	160.20	<2e-16 ***
NEUTRAL	-.02	.04	27.02	-.55	.586
YESTERDAY	.04	.04	27.02	1.10	.280
Object	lmer(log(measure) ~ context*number + (1|vp) + (1|item), data=ia.6)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
Intercept	5.13	.07	27.74	74.76	<2e-16 ***
CONTEXT YESTERDAY	-.06	.09	23.82	-.67	.5081
CONTEXT FREQUENTLY	.01	.09	23.95	.13	.8942
NUMBER	-.13	.09	24.05	-1.35	.1892
CONTEXT YESTERDAY : NUMBER	.18	.13	23.96	1.38	.1791
CONTEXT FREQUENTLY : NUMBER	.33	.13	24.06	2.47	0.0212 *
reference level set to frequently					
INTERCEPT	5.15	.07	27.57	75.06	<2e-16 ***
CONTEXT NEUTRAL	-.01	.09	23.95	-.13	.8942

CONTEXT YESTERDAY	-.08	.09	23.74	-.81	.4276
NUMBER	.20	.09	24.08	2.14	.0431 *
CONTEXT NEUTRAL : NUMBER	-.33	.13	24.06	-2.47	.0212 *
CONTEXT YESTERDAY : NUMBER	-0.14	0.13	23.98	-1.08	.2893
resolve interaction of CONTEXT and NUMBER					
Main effect of NUMBER for each level CONTEXT					
Neutral	lmer(log(measure) ~ number + (1|vp) + (1|item), data=ia.6.neutral				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
Intercept	5.13	.06	8.77	81.54	6.09e-14 ***
Number	-.13	.09	7.98	-1.46	.182
Yesterday					
Intercept	5.07	.08	9.25	67.44	8.93e-14 ***
number	.06	.10	8.00	.55	.595
Frequently					
intercept	5.15	.07	9.46	76.66	1.55e-14 ***
number	.20	.09	8.01	2.21	.0585 .
Main effect of CONTEXT for each level of NUMBER					
singular	lmer(log(measure) ~ context+ (1+context|vp) + (1|item), data=ia.6.singular)				
reference level set to neutral					
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
intercept	5.01	.07	14.75	69.79	<2e-16 ***
frequently	.34	.10	12.45	3.51	0.00406 **
yesterday	.12	.10	12.05	1.25	.23357
reference level set to frequently					
intercept	5.35	.07	12.62	78.05	< 2e-16 ***
neutral	-0.34	0.10	12.45	-3.51	0.00405 **
yesterday	-0.22	0.10	12.22	-2.28	0.04104 *
plural	lmer(log(measure) ~ context + (1|vp) + (1|item), data=ia.6.plural)				
reference level set to neutral					
intercept	5.13	.07	13.41	76.49	<2e-16 ***
frequently	.01	.09	12.04	.14	.889
yesterday	-.06	.09	11.97	-.68	.512
reference level set to frequently					
intercept	5.15	.07	13.32	76.81	<2e-16 ***
neutral	-.01	.09	12.04	-.14	.889
yesterday	-.08	.09	11.93	-.82	.429

Spillover lmer(log(measure) ~ context + (1|vp) + (1|item), data=ia.7)

	Estimate	Std.Error	Df	*t*-value	*p*-value
reference level set to neutral					
INTERCEPT	5.25	.08	25.52	63.33	<2e-16 ***
CONTEXT YESTERDAY	.01	.12	24.16	.04	.970
CONTEXT FREQUENTLY	-.06	.12	23.85	-.50	.623
NUMBER	.04	.12	24.00	.38	.709
CONTEXT YESTERDAY : NUMBER	-.05	.16	24.01	-.31	.756
CONTEXT FREQUENTLY : NUMBER	-.24	.16	23.87	-1.48	.153
reference level set to frequently					
INTERCEPT	5.19	.08	25.43	62.686	<2e-16 ***
CONTEXT NEUTRAL	.06	.12	23.85	.50	.6232
CONTEXT YESTERDAY	.06	.12	24.11	.54	.5976
NUMBER	-.20	.12	23.74	-1.71	.0998 .
CONTEXT NEUTRAL : NUMBER	.24	.16	23.87	1.48	.1529
CONTEXT YESTERDAY : NUMBER	.20	.16	23.88	1.16	.2569

Wrap-up lmer(log(measure) ~ context*number + (1|vp) + (1|item), data=ia.8)

	Estimate	Std.Error	Df	*t*-value	*p*-value
reference level set to neutral					
INTERCEPT	5.93	.05	76.28	115.00	<2e-16 ***
CONTEXT YESTERDAY	.05	.05	23.93	.96	.346
CONTEXT FREQUENTLY	-.04	.05	23.97	-.78	.446
NUMBER	-.01	.05	24.44	-.12	.905
CONTEXT YESTERDAY : NUMBER	-.06	.07	24.21	-.88	.388
CONTEXT FREQUENTLY : NUMBER	-.00	.07	24.28	-.01	.990
reference level set to frequently					
INTERCEPT	5.90	.05	76.85	114.06	<2e-16 ***
CONTEXT YESTERDAY	.04	.05	23.97	.78	.45
CONTEXT FREQUENTLY	.08	.05	24.14	1.73	.0961 .
NUMBER	-.01	.05	24.14	-.14	.8916
CONTEXT YESTERDAY : NUMBER	.00	.07	24.28	.01	.9905
CONTEXT FREQUENTLY : NUMBER	-.06	.07	24.04	-.87	.3935

Table B1: Fixed effects for models used in the analysis of first pass times.

Regression path times

Verb	lmer(log(measure) ~ context + (1+context|vp) + (1|item), data=ia.5)				
	Estimate	**Std. Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
Intercept	6.04	.04	51.54	161.51	<2e-16 ***
YESTERDAY	-.09	.04	29.31	-2.30	.0286 *
FREQUENTLY	-.06	.04	26.81	-1.51	.1430
reference level set to frequently					
Intercept	5.98	.04	53.02	160.60	<2e-16 ***
NEUTRAL	.06	.04	26.81	1.51	.143
YESTERDAY	-.04	.04	28.87	-.95	.349

Object	lmer(log(measure)~context*number + (1+number|vp) + (1|item), data=ia.6)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	6.72	.14	27.99	46.59	<2e-16 ***
CONTEXT YESTERDAY	.04	.20	23.93	.20	.841
CONTEXT FREQUENTLY	-.15	.20	23.98	-.74	.465
NUMBER	.07	.20	24.05	.37	.716
CONTEXT YESTERDAY : NUMBER	-.21	.28	23.98	-.74	.465
CONTEXT FREQUENTLY : NUMBER	-.35	.28	24.02	-1.25	.224
reference level set to frequently					
intercept	6.58	.14	27.90	45.61	<2e-16 ***
CONTEXT NEUTRAL	.15	.20	23.98	.74	.465
CONTEXT YESTERDAY	.19	.20	23.89	.95	.353
NUMBER	-.27	.20	24.05	-1.40	.176
CONTEXT NEUTRAL : NUMBER	.35	.28	24.02	1.25	.224
CONTEXT YESTERDAY : NUMBER	.14	.28	23.98	.51	.617

Spillover	lmer(log(measure) ~ context*number + (1+number|vp) + (1|item), data=ia.7)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	6.17	.18	25.61	35.26	<2e-16 ***
CONTEXT YESTERDAY	-.11	.24	24.07	-.46	.65
CONTEXT FREQUENTLY	.08	.24	23.91	.35	.73

NUMBER	.03	.24	23.98	.10	.918
CONTEXT YESTERDAY : NUMBER	.08	.34	23.98	.24	.815
CONTEXT FREQUENTLY : NUMBER	.43	.34	23.92	1.25	.225
reference level set to frequently					
intercept	6.25	0.18	25.57	35.76	<2e-16 ***
CONTEXT NEUTRAL	-.08	.24	23.91	-.35	.73
CONTEXT YESTERDAY	-.20	.24	24.05	-.81	.426
NUMBER	.45	.24	23.86	1.87	.074 .
CONTEXT NEUTRAL : NUMBER	-.43	.34	23.92	-1.25	.225
CONTEXT YESTERDAY : NUMBER	-.35	.34	23.92	-1.01	.323

wrap-up	lmer(log(measure) ~ context*number + (1|vp) + (1|item), data=ia.8)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	8.16	.02	48.35	367.22	<2e-16 ***
CONTEXT YESTERDAY	-.03	.03	23.65	-1.34	.194
CONTEXT FREQUENTLY	-.02	.03	23.67	-.57	.576
NUMBER	-.04	.03	23.92	-1.43	.165
CONTEXT YESTERDAY : NUMBER	.03	.04	23.80	.69	.500
CONTEXT FREQUENTLY : NUMBER	.02	.04	23.78	0.58	.569
reference level set to frequently					
intercept	8.14	.02	48.50	366.27	<2e-16 ***
CONTEXT NEUTRAL	.02	.03	23.67	.57	.58
CONTEXT YESTERDAY	-.02	.03	23.70	-.77	.450
NUMBER	-.02	.03	23.65	-.62	.542
CONTEXT NEUTRAL : NUMBER	-.02	.04	23.78	-.58	.569
CONTEXT YESTERDAY : NUMBER	.00	.04	23.65	.11	.915

Table B2: Fixed effects for models used in the analysis of regression path durations.

Total reading times

Verb	lmer(log(measure) ~ context*number + (1+number|vp) + (1|item), data=ia.5)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					

intercept	6.53	.09	27.40	72.90	<2e-16 ***
CONTEXT YESTERDAY	-.11	.12	24.02	-.87	.393
CONTEXT FREQUENTLY	-.01	.12	24.02	-.07	.946
NUMBER	-.04	.12	24.13	-.35	.727
CONTEXT YESTERDAY : NUMBER	.13	.17	24.03	.76	.453
CONTEXT FREQUENTLY : NUMBER	.09	.17	24.03	.50	.619
reference level set to frequently					
intercept	6.52	.09	27.41	72.80	<2e-16 ***
CONTEXT NEUTRAL	.01	.12	24.02	.07	.946
CONTEXT YESTERDAY	-.10	.12	24.03	-.80	.431
NUMBER	.04	.12	24.13	.36	.724
CONTEXT NEUTRAL : NUMBER	-.09	.17	24.03	-.50	.619
CONTEXT YESTERDAY : NUMBER	.05	.17	24.03	.26	.797
object	lmer(log(measure) ~ context*number + (1+number+context|vp) + (1|item), data=ia.6)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	6.13	.09	27.47	68.83	<2e-16 ***
CONTEXT YESTERDAY	-.02	.12	24.16	-.14	.893
CONTEXT FREQUENTLY	-.01	.12	24.14	-.07	.944
NUMBER	-.01	.12	24.26	-.56	.582
CONTEXT YESTERDAY : NUMBER	.04	.17	23.98	.24	.810
CONTEXT FREQUENTLY : NUMBER	.04	.17	24.02	.24	.816
reference level set to frequently					
intercept	6.12	.09	27.12	68.95	<2e-16 ***
CONTEXT NEUTRAL	.01	.12	24.12	.07	.944
CONTEXT YESTERDAY	-.01	.12	23.91	-.06	.949
NUMBER	-.03	.12	24.27	-.23	.822
CONTEXT NEUTRAL : NUMBER	-.04	.17	24.00	-.24	.816
CONTEXT YESTERDAY : NUMBER	.00	.17	23.98	.01	.993
spillover	lmer(log(measure) ~ context*number + (1+number+context|vp) + (1|item), data=ia.7)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	5.96	.067	35.66	100.97	<2e-16 ***
CONTEXT YESTERDAY	-.11	.08	25.63	-1.44	.163
CONTEXT FREQUENTLY	-.07	.07	24.13	-.88	.386

NUMBER	-.00	.07	24.19	-.06	.956
CONTEXT YESTERDAY : NUMBER	.17	.11	24.22	1.59	.124
CONTEXT FREQUENTLY : NUMBER	.16	.11	23.93	1.51	.143
reference level set to frequently					
intercept	5.89	.06	34.66	100.57	<2e-16 ***
CONTEXT NEUTRAL	.07	.07	24.07	.88	.3859
CONTEXT YESTERDAY	-.04	.08	26.54	-.56	.5774
NUMBER	.15	.07	23.78	2.09	.0476 *
CONTEXT NEUTRAL : NUMBER	-.16	.11	23.92	-1.52	.1428
CONTEXT YESTERDAY : NUMBER	.01	.11	23.96	.08	.9339
wrap-up	lmer(log(measure) ~ context*number + (1|vp) + (1|item), data=ia.8)				
	Estimate	**Std.Error**	**Df**	***t*-value**	***p*-value**
reference level set to neutral					
intercept	6.27	.07	60.83	91.97	<2e-16 ***
CONTEXT YESTERDAY	.02	.07	23.81	.22	.831
CONTEXT FREQUENTLY	.02	.07	23.83	.33	.744
NUMBER	-.03	.07	24.00	-.43	.672
CONTEXT YESTERDAY : NUMBER	-.05	.10	23.91	-.49	.627
CONTEXT FREQUENTLY : NUMBER	.01	.10	23.94	.07	.944
reference level set to frequently					
intercept	6.29	.07	61.06	92.23	<2e-16 ***
CONTEXT NEUTRAL	-.02	.07	23.83	-.33	.744
CONTEXT YESTERDAY	-.01	.07	23.89	-.12	.909
NUMBER	-.02	.07	23.89	-.33	.746
CONTEXT NEUTRAL : NUMBER	-.01	.10	23.94	-.07	.944
CONTEXT YESTERDAY : NUMBER	-.06	.10	23.85	-.564	.578

Table B3: Fixed effects for models used in the analysis of total reading times.

Post-hoc tests for regression probabilities out of regions

object, singular	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
reference level set to neutral				
Intercept	-.30	.25	-1.23	.2197
YESTERDAY	-.38	.33	-1.14	.2560
FREQUENTLY	-.79	.34	-2.33	.0196 *

reference level set to frequently				
Intercept	-1.09	.25	-4.43	9.41e-06 ***
NEUTRAL	.79	.34	2.33	.0196 *
YESTERDAY	.41	.34	1.21	.2274
object, plural				
reference level set to neutral				
Intercept	-.31	.26	-1.21	.226
YESTERDAY	.06	.36	.16	.877
FREQUENTLY	-.34	.36	-.94	.346
reference level set to frequently				
Intercept	-.65	.26	-2.51	.0122 *
NEUTRAL	.34	.36	.94	.3457
YESTERDAY	.39	.36	1.10	.2732
spillover, singular				
reference level set to neutral				
Intercept	-1.22	.20	-6.00	2.09e-09 ***
YESTERDAY	.03	.29	.11	.9144
FREQUENTLY	.66	.27	2.42	.0153 *
reference level set to frequently				
Intercept	-.56	.19	-2.93	.00337 **
NEUTRAL	-.66	.27	-2.42	.01533 *
YESTERDAY	-.62	.27	-2.32	.02058 *
spillover, plural				
reference level set to neutral				
Intercept	-1.29	.28	-4.62	3.82e-06 ***
YESTERDAY	-.53	.42	-1.27	.206
FREQUENTLY	.17	.39	.44	.658
reference level set to frequently				
Intercept	-1.12	.29	-3.90	9.5e-05 ***
NEUTRAL	-.17	.39	-.44	.6577
YESTERDAY	-.70	.41	-1.71	.0881 .

Table B4. Fixed effects for models used in the analysis of regression probabilities.

Appendix C: Mean reading time measures and probabilities of regressions.

Table 3 provides an overview of mean reading time data over participants for reading time measures for each condition and position. Table 4 provides an overview of mean regression probabilities across participants for each condition and position.

Table 3: Means over participants for reading time measures for each condition and position, with standard deviations in parentheses. Reading times and st.dvs. are given in milliseconds.

First pass times				
	Position			
Condition	**verb**	**object**	**spillover**	**wrap-up**
NEUTRAL-SINGULAR	346 (114)	201 (70)	237 (77)	458 (152)
NEUTRAL-PLURAL	348 (101)	217 (68)	218 (57)	485 (228)
YESTERDAY-SINGULAR	356 (93)	213 (64)	222 (57)	452 (222)
YESTERDAY-PLURAL	353 (106)	203 (61)	227 (62)	474 (185)
FREQUENTLY-SINGULAR	345 (101)	248 (68)	188 (61)	439 (180)
FREQUENTLY-PLURAL	355 (118)	216 (78)	227 (97)	457 (201)
Regression path times				
	verb	**object**	**spillover**	**wrap-up**
NEUTRAL-SINGULAR	526 (222)	1413 (614)	947 (569)	3458 (426)
NEUTRAL-PLURAL	569 (251)	1328 (569)	952 (516)	3592 (473)
YESTERDAY-SINGULAR	464 (205)	1203 (517)	902 (532)	3417 (488)
YESTERDAY-PLURAL	463 (194)	1349 (532)	803 (584)	3444 (406)
FREQUENTLY-SINGULAR	503 (236)	880 (426)	1466 (575)	3485 (450)
FREQUENTLY-PLURAL	518 (232)	1174 (478)	956 (523)	3517 (487)
Total reading times				
	verb	**object**	**spillover**	**wrap-up**

NEUTRAL-SINGULAR	759 (210)	515 (142)	454 (142)	609 (230)
NEUTRAL-PLURAL	798 (237	557 (140)	461 (152)	667 (352)
YESTERDAY-SINGULAR	754 (170)	513 (151)	476 (145)	602 (279)
YESTERDAY-PLURAL	694 (167)	519 (137)	427 (152)	638 (272)
FREQUENTLY-SINGULAR	801 (181)	522 (160)	499 (127)	641 (290)
FREQUENTLY-PLURAL	768 (192	526 (137)	444 (159)	664 (310)

Table 4: Means over participants of regression probabilities for each condition and position. Probabilities are given as the percentage of items per condition (averaged across participants) when a regression was made; separately for each position and for regressions into and out of regions.

Regressions into IA				
	Position			
Condition	**adverb**	**verb**	**Object**	**spillover**
NEUTRAL-SINGULAR	44.35 (3.03)	40.56 (2.86)	14.1 (2.06)	33.93 (2.83)
NEUTRAL-PLURAL	41.49 (3.16)	37.68 (2.66)	15 (2.26)	28.4 (3.66)
YESTERDAY-SINGULAR	39.43 (2.96)	35.14 (2.74)	12.94 (2.09)	30 (3.26)
YESTERDAY-PLURAL	44.35 (2.47)	38.57 (3.13)	14.58 (2.1)	25.23 (2.45)
FREQUENTLY-SINGULAR	41.61 (2.85)	37.43 (3.04)	18.12 (2.35)	29.03 (2.93)
FREQUENTLY-PLURAL	41.61 (2.85)	37.43 (3.04)	18.12 (2.35)	29.03 (2.93)
Regressions out of IA				
	Position			
Condition	**verb**	**object**	**Spillover**	**wrap-up**
NEUTRAL-SINGULAR	4.93 (1.27)	41.67 (2.98)	26 (2.87)	97.97 (.79)
NEUTRAL-PLURAL	6.76 (1.6)	42.3 (3.14)	23.93 (2.36)	98.79 (.59)
YESTERDAY-SINGULAR	3.14 (1.05)	34.92 (2.84)	23.69 (2.59)	98.18 (.71)
YESTERDAY-PLURAL	4 (1.05)	44.12 (3.18)	18.75 (3.23)	96.67 (1.11)

FREQUENTLY-SINGULAR	4.57 (1.16)	25.26 (2.58)	36.83 (3.01)	98.1 (.75)
FREQUENTLY-PLURAL	4.86 (1.18)	35 (2.36)	27.74 (2.77)	98.41 (.69)