GLOSSA psycholinguistics

Vos, M., Gurumukhani, M., Vaidya, A., & Wittenberg, E. (2025). No evidence for culmination inferences based on Hindi ergative marking. *Glossa Psycholinguistics*, 4(1): X, pp. 1–39. DOI: https://doi.org/10.5070/G6011.35449

eScholarship University of California

No evidence for culmination inferences based on Hindi ergative marking

Myrte Vos, The Arctic University of Norway, myrte.v@gmail.com
Mohit Gurumukhani, Cornell University, USA, mgurumuk@cs.cornell.edu
Ashwini Vaidya, IIT Delhi, India, avaidya@hss.iitd.ac.in
Eva Wittenberg, Central European University, Austria, wittenberge@ceu.edu

Prediction, both on the syntactic level and the semantic level, is a central process in language comprehension. For instance, people predict aspects of event structure based on morphosyntactic markers on verbs: hearing *has peeled* directs one's attention towards a culminated event, as opposed to an ongoing event. Here, we ask how general this prediction process is and specifically, whether it extends to cues outside the predicate, using the Hindi split-ergative system as a case study. Ergativity allows properties of an event to be predicted on the subject, which is, notably, a constituent outside the Verb Phrase. In four studies, we map out the role subject marking plays in the prediction of event properties in comprehension. Our results show that in some offline judgments, ergativity is a strong predictor of culminated events; but the cue provided by ergative marking is not taken into account during incremental comprehension, a result that questions accounts of automatically triggered culmination inferences in ergative constructions and provides evidence for a limit on predictive processing.

1. Introduction

Events are central anchorpoints in our human experience, and they shape the way we think about and remember our lives (Eggins and Slade, 2006; Mahr and Csibra, 2018, 2020). Consequently, it is important that when we talk about an event, we convey its main aspects reliably and efficiently by the means that language gives us: who does what to whom, for instance, or whether the event is complete or ongoing. In this article, we focus on how people extract and predict information about this last aspect, event culmination, from the linguistic signal. We study the comprehension of linguistically encoded events, using Hindi ergative marking as a case study to test the limits of prediction in language-guided event comprehension.

Prediction has been a central focus in psycholinguistic work in recent years, since it seems to operate on all levels of cognition.[1] Linguistic prediction ranges from functional elements (DeLong et al., 2005; Otten and Van Berkum, 2009; Wicha et al., 2004; but see Kochari and Flecken, 2019; Nieuwland et al., 2018) to syntactic structures (Dikker et al., 2010; Ferreira and Qiu, 2021; Staub and Clifton Jr., 2006; Yoshida et al., 2013) and specific word choices (Altmann and Mirković, 2009; DeLong et al., 2005; Willems et al., 2016). For Subject-Object-Verb order languages, like Hindi, predictions about verbs seem to be particularly strong and robust (Husain et al., 2014; Levy and Keller, 2013; Vasishth, 2003). Here, we ask whether Hindi grammar also results in robust predictions about event structure: in particular, whether an event is culminated or not.

In Hindi, a language with Subject-Object-Verb word order, information about event culmination is morphologically signaled on the verb through aspect, as in many other languages. However, within the context of transitive predicates, event culmination can theoretically already be predicted at the point of the subject: if the subject is marked for ergative case, the verb has to be perfective, and describe a culminated event. Nominative (zero) marked subjects of transitive verbs, on the other hand, will occur with imperfective aspect. Thus, in Hindi, when viewed from the point of view of real-time language comprehension, case marking on the subject can serve as a very early cue for the grammatical aspect marked on the verb, and trigger inferences about an event's culmination status.

Usually, inferences about aspects of an event's structure, such as culmination status, are guided by verbal morphology (Altmann and Kamide, 2007; Foppolo et al., 2021; Madden and Zwaan, 2003; Minor et al., 2022; Misersky et al., 2021). For instance, an orange is ready to be eaten when the event of peeling has reached its culmination, not before: English marks the difference with *has peeled* versus *is peeling*. Comprehenders of almost all languages expressing this aspectual distinction will look for its morphological marking on the verb describing the event

[1] An ongoing debate focuses on whether some prediction effects are the consequence of feed-forward pre-activation, or of top-down facilitated integration (Ferreira and Qiu, 2021; Huettig et al., 2022). Here, we are interested in whether early grammatical information helps a listener understand a scene before all information has been heard, and, for the purpose of this article, we choose to remain agnostic about the larger debate.

(Dahl and Velupillai, 2013). However, it is unclear whether case marking can trigger inferences about an event's culmination status: While Özge et al. (2017) showed that partitive case marking is used to predict (non-)culmination, this effect again was confined to effects of event prediction based on information contained within the predicate. Thus, it is an open question whether event prediction can be found based on information outside the predicate.

From a theoretical point of view, prediction about event structure based on subject case marking would constitute a deviation from how sentence structure is usually thought about: while the predicate (syntactically, a Verb Phrase containing the verbal complex, any syntactic objects, and adverbials) provides information about the event itself, the subject manifests as its own constituent outside of the predicate as an external object (Husband and Stockall, 2015; Kratzer, 1996). In consequence, if we were to document prediction effects based on the subject alone, this would be the first evidence that predictive inferences about event structure can be triggered outside of the predicate itself, since, so far, case marking has only been found to be used in predictive processes within the verb phrase (Bornkessel-Schlesewsky and Schlesewsky, 2009; Kamide, Altmann, and Haywood, 2003; Kamide, Scheepers, and Altmann, 2003; Özge et al., 2019; Skopeteas et al., 2012).

In what follows, we will briefly discuss the relevant grammatical properties of case marking and word order in Hindi (1.1), summarize previous studies investigating inferences about event structure based on grammatical cues (1.2), and explain the current studies and predictions (1.3). Then, we will present four sets of studies investigating whether, when, and how much ergativity serves as a cue to Hindi native speakers to draw predictive inferences about event structure.

1.1 Word order and case clitics in Hindi

The grammatical system of Hindi has two particularly important properties for the current study. First, Hindi has Subject–Object–Verb (SOV) word order. In consequence, while a Hindi sentence's tense and aspect only appears at the end, case marking on the subject (ergative in (1)) appears already within the first syntactic phrase:

(1) mohan = ne narangi = ko chhil-a hai
mohan = ERG orange = ACC peel-PerfPart Pres.SG
'Mohan has peeled the orange.'

The second important property of Hindi for the purpose of this article is that it is a split-ergative language (Anand and Nevins, 2006; Coon and Parker, 2019; Kachru and Pandharipande, 1978), showing different patterns of case marking that are conditional on two factors: perfectivity and transitivity (Kachru, 1980; Mohanan, 1994). In this pattern, the agent argument in a transitive sentence is assigned ergative case when the verb is inflected for the perfective (e.g., (1)). When the verb does not have perfective morphology, the agent argument will occur with the

nominative (zero) case as seen in (2). Intransitives, on the other hand, will typically occur with the nominative (zero) case on the single argument, regardless of perfective morphology (3), giving rise to a split-ergative system.[2]

(2) Mohan-∅ narangi=ko chhil raha hai
mohan-NOM orange=ACC peel Prog.M Pres.SG
'Mohan is peeling the orange.'

(3) Mohan-∅ roy-a
mohan-NOM cry-PerfPart
'Mohan cried.'

When viewed from the point of view of real-time language comprehension, the case marking on the subject (*mohan* in (1) and (2)) can serve as a very early cue for the grammatical aspect marked on the verb, and trigger inferences about whether an event is culminated or not. As mentioned, this is different from English and many other languages, where this aspectual distinction is only marked morphologically on the verb describing the event (Dahl and Velupillai, 2013).

A consequence of this property is that ergative markers should be strong signals to indicate culminated events (Arunachalam and Kothari, 2011; Butt, 2017). Indeed, a corpus search reveals that the association between ergative marking and perfective tense is just as strong as commonly assumed.[3] Ergative case is an extremely reliable cue for perfectivity, while nominative case is not (**Figure 1**).

---

[2] There are a small number of exceptions to this pattern. Among intransitive verbs in Hindi, the unergative verbs *naach* 'dance' or *nahaa* 'bathe' will take ergative case, as do a number of ingesto-reflexive unaccusative verbs, e.g. *mohan-ne chiinkaa* 'Mohan sneezed'. Mohanan (1994) and Butt (2017) also note that, in perfective transitive complex predicates, the lexico-semantic properties of the light verb dictate the subject marking: e.g., when the light verb is unaccusative, the subject marking must be nominative or dative (as in the examples below, (18a) and (18b) from Butt, 2017).

(i) nadya=ne kahani yad k-i
nadya=ERG story.SG.NOM memory.NOM do-Perf.SG
'Nadya remembered the story (actively)'

(ii) nadya=ko kahani yad a-yi
nadya=DAT story.SG.NOM memory.NOM come-Perf.SG
'Nadya remembered the story' (lit. the memory came to Nadya)

[3] We used the entire corpus from the Hindi Dependency Treebank (Bhat et al., 2017), which was procured from 20522 sentences. We used the Java API provided to extract and analyze the sentences, and classified a sentence as ergative or not depending on whether it contained the ergative marker. We also classified a sentence as perfective if the verb had the suffix *-a* or *-ya* as marked by the Treebank. Based on these two classifications, we classified each sentence into one of the four possible types shown in **Figure 1**.

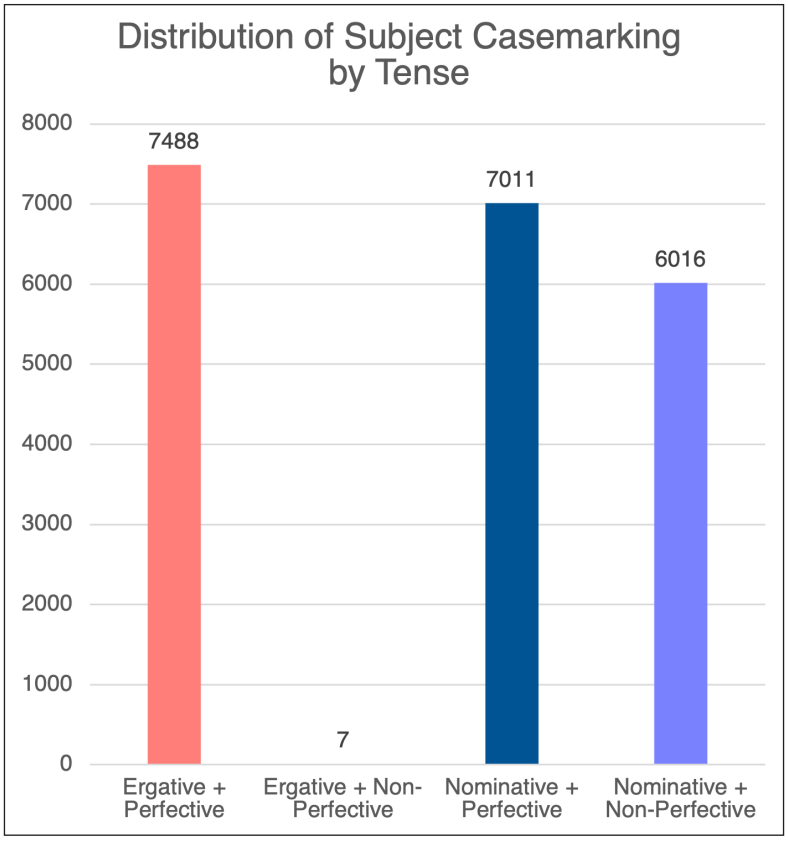


Figure 1: Corpus data: Distribution of case marking in the HindiTreeBank, by tense.

Despite the strength of the association between ergative case and perfective aspect, Arunachalam and Kothari (2011) showed that Hindi perfective sentences are not necessarily always interpreted as culminated events, especially when those sentences describe events with incremental themes (e.g., *peeling an orange*) as opposed to non-incremental themes (e.g., *dropping an orange*; see also Wittek (2002) and Pederson (2007) for similar results in Tamil and German). However, Arunachalam and Kothari (2011) did not study the role of the ergative marker in contrast to null-marked subjects in the nominative; and the main interest in these studies was the defeasibility of the culmination inference, as opposed to the incremental unfolding of predictions about an event interpretation based on grammatical properties.

From an incremental processing perspective, Hindi remains understudied, but several recent studies exploit its link between ergative/nominative case and perfective/imperfective aspect to gain insight into the underlying parsing process. Choudhary et al. (2009) conducted an ERP study of case marking violations in Hindi and found that a mismatch of ergative case with imperfective aspect triggered a stronger N400 than a mismatch of nominative case with perfective aspect, and a late positivity (P600).

This P600 effect was replicated by Dillon et al. (2012) for ungrammatical combinations of ergative markers with future tense; in contrast, a nominative-perfective mismatch did not result in a late positivity, which the authors attribute to a small class of transitive perfective verbs that obligatorily take nominative marking on the agent (Mohanan, 1994), and to the fact that omission of ergative marking is quite common in some Hindi dialects (Butt and King, 2004). This asymmetry between violations of ergative vs. nominative case was, however, not found by Bickel et al. (2015).

That said, the nature of the signaling function of ergative markers in perfective sentences is still subject to debate. An orthogonal hypothesis about what, if anything, ergative case markers may contribute beyond their grammatical necessity claims that ergativity also signals volitionality, or an action done with a certain degree of intention: An event that is described with the ergative marker on the subject is said to convey a sense of 'conscious choice' (Butt, 1995; Longenbaugh and Polinsky, 2017; Mohanan, 1994; also see Maitreyee et al., 2023 for evidence that both children and adults are sensitive to the degree of volitionality in the context of ergative subjects). We will address this hypothesis in Section 2.

1.2 Previous studies of event structure inferences

Our article is one in a growing psycholinguistic tradition asking whether and at which point during incremental sentence comprehension people draw culmination inferences. So far, the evidence has been mixed: While there is some evidence that comprehenders do not commit to culmination inferences as a sentence unfolds incrementally (Misersky, 2022; Pickering et al., 2006; Proctor et al., 2004), other work has found that aspectual information affects ongoing sentence interpretation (Baggio et al., 2010; Kaplan et al., 2021; Misersky et al., 2021; Paczynski et al., 2014).

Several studies have shown that the rapid calculation of culmination inferences can be detected using the Visual World Paradigm. For instance, Altmann and Kamide (2007) showed that when people heard sentences containing verbs in the past perfect tense (*The man has drunk…*), they rapidly look to an object compatible with a culminated action, e.g., an empty glass. However, whenever people heard sentences containing verbs in future tense (*The man will drink…*), they were more likely to look to an object compatible with a non-culminated action, e.g., a full bottle.

These rapid inferences of event culmination based on aspectual information are documented in different populations, and across languages. For instance, Zhou et al. (2014) found that young Mandarin-speaking children used aspectual verbal morphology to rapidly recognize event structure: when they heard the durative suffix *-zhe*, they looked towards a picture of an ongoing version of the event, and when they heard the perfective suffix *-le*, they looked towards the completed version – and these children did so as rapidly as adults. A similar study in Russian,

which has a highly grammaticized distinction between imperfective and perfective aspect, found that adult Russian listeners used aspectual morphology to identify a target picture even before verb offset (Minor et al., 2022). Finally, Turkish children can draw culmination inferences based on partitive case marking before the verb is heard (Özge et al., 2017).

Our present studies are built on a recent article that used a visual world paradigm in Italian to ask when a verb's morphosyntactic telicity[4] cues (perfective aspect) are used to make inferences about event structure (Foppolo et al., 2021). Importantly, Foppolo et al. (2021) separated telicity and culmination by using sentences containing perfective verbs that were either durative (e.g., *has peeled the apple*) or punctual (e.g., *has dropped the ball*): Unlike durative verbs, the culmination, or end point, is lexically encoded in punctual[5] verbs. Thus, the authors were able to ask not only whether people draw telicity inferences based on the verb, but also whether the lexical information they glean from the verb is used to draw culmination inferences during incremental sentence processing, and at which point.

In their studies, Foppolo et al. (2021) showed two pictures containing different objects, one with a completed and one with a non-completed action (e.g., an open newspaper and a closed newspaper). Participants listened to sentences like *Valeria has opened the newspaper* while their eye-movements were measured. The results show that when listening to sentences containing perfective tense, participants launched anticipatory looks towards the picture of the event in its completed state before the offset of the predicate. Crucially, participants were equally fast to look towards the target picture in both the punctual and the durative verb conditions. In other words, perfective aspectual morphology on the durative verbs was enough to trigger a culmination inference.

What is an open question, however, is whether any morphological cue could be a potential trigger for successful culmination inferences: the previous literature has relied entirely on morphosyntactic marking within the verb phrase, that is, case or aspectual marking either on the verb or the object. Across the world's languages, however, there are many ways to grammatically convey information about properties of an event's structure. One cross-linguistically frequent way is ergative marking on subjects, specifically, split ergativity, as found in Modern Standard Hindi. As reviewed above, several studies have shown that Hindi speakers do use case marking on the subject to predict upcoming verbal morphology (Arunachalam and Kothari, 2011; Choudhary et al., 2009; Dillon et al., 2012; but see Bickel et al., 2015). But do speakers also use this abstract information to predict event structure, such as an event's degree of culmination?

---

[4] The need to differentiate *telicity* from *telos*/culmination is an ongoing debate in aspectual semantics, and beyond the scope of this article; with regards to the studies presented here, we refer only to 'culmination inferencing'.

[5] Note that the term *punctual*, like many in aspectual semantics, is not without controversy (see, for instance, Verkuyl, 1989). We take *punctual* here to mean events with an extremely short duration, as opposed to durationless achievements such as *win a race* or *reach the summit*.

1.3 Current studies and predictions

In this article, we ask whether Hindi speakers draw predictive inferences about event culmination based on case marking on the grammatical subject. In four studies, we ask whether Hindi native speakers draw semantic inferences about events, and when they do so, based on the presence or absence of an ergative marker on the subject.

One prediction for the current studies is that if Hindi speakers indeed use case marking as a predictive cue, they should theoretically be able to do so at the very first noun in a sentence, given Hindi's SOV structure. This prediction is supported by data from Choudhary et al. (2009) and Dillon et al. (2012): comprehenders appear to generate expectations about upcoming input based on subject case marking, and mismatches produce an ERP signature that has been interpreted as indexing surprisal or reanalysis.

However, anticipation of grammatically licensed input is not equivalent to inferencing event properties before the verb has been encountered: although there is robust evidence (Foppolo et al., 2021; Minor et al., 2022; Zhou et al., 2014) that grammatical aspect can trigger rapid inferences of event culmination, an alternative prediction is that Hindi case marking does not function as a predictive cue for event structure at all.

This prediction is rooted within a theoretical linguistic literature in which the verb phrase is the domain of aspectual interpretation: under this view, aspectual interpretations are crucially dependent on features within the verb phrase (e.g., Husband and Stockall, 2015; Kratzer, 1996).[6] Located outside of the event-encoding predicate, the subject's case marking may be superfluous to aspectual composition in the verb phrase, or unavailable for integration with the event semantics until the verb phrase constituent is fully parsed. Under this account, ergative case mostly fulfills a structural requirement, and the semantic reading, be it of event culmination or conscious choice, or both, emerges from the complex interaction between the lexical affordances of the main verb, the semantic contribution of the light verb (if present),[7] and the ergative marker (Bashir, 1999; Butt, 1995; Ramchand, 1990; Ranjan, 2016).

A third possibility is one of partial prediction: that, since perfective marking on the verb does not reliably entail event culmination, the presence of an ergative marker does not serve as a predictive cue to a completed event – but that the *absence* of one, i.e., nominative case, does

[6] Martin (2015, 2020) reports evidence that degree of agentive control influences non-culminated readings of causal events. However, the model proposed there operates over conceptual distinctions that pertain to a subject referent's agentivity, specifically, whether the syntactic subject is introduced by Voice_{ag} (for human agents) or Voice_{c} (for non-volitional causers, such as events), following Pylkkänen (2008). Since, in our study, all syntactic subjects denote agentive humans, this purported difference will not play a role.

[7] In this article, we limit ourselves to simplex perfective verbs, keeping in mind that the sentence-final semantic contribution of the light verb should not bear on the question of whether comprehenders predict aspect from subject case; and that simplex verbs, though they do not strictly entail event culmination, still do imply it (Arunachalam and Kothari, 2011).

allow comprehenders to predict an *incomplete* event. Such an absence could be interpreted and integrated as rapidly as a realized affix (Minor et al., 2022). A weaker version of this scenario is one where, in Hindi, neither case marker is a predictive cue, but there is an overall preference for interpreting an event as incomplete unless and until there is evidence not just of an event *telos*, but also of a result state. This preference could be a cross-linguistic strategy, as found in Foppolo et al. (2021). However, partial prediction could also be driven by language-specific affordances changing participants' processing strategies, which are adapted to the contrasts available in particular languages (e.g., Bott and Gattnar, 2015; Bott and Hamm, 2014; Flecken et al., 2014; von Stutterheim et al., 2012).

These predictions about culmination inferences are somewhat orthogonal to another question: whether Hindi ergative case marking is primarily associated with "conscious choice" or intentionality (Butt, 2017; Butt and King, 2004; Maitreyee et al., 2023; Mohanan, 1994; also see Demirdache and Martin, 2015; Mathis and Papafragou, 2022; van Hout et al., 2017 for arguments that knowledge about the goals of an agent, or the degree of an agent's control over an event, influences inferences about event construal). If this is the case, then culmination and intentionality may interact in a way to mask any culmination inferences.

Thus, our first experiment here will address the question of the role of intentionality inferences, before we move to the question of event culmination inferences in Experiment II–IV.

2. Experiment I: Inferences about intentionality

This experiment served as a context-free baseline to ask whether case marking on a specific noun alone, without any event information except for agent and patient, will lead to different inferences about whether an action was done intentionally or not.

2.1 Participants

We paid 48 participants on Amazon Mechanical Turk to participate in this experiment. All participants were self-reported Hindi native speakers and had an IP address in India. As an attention check, and to ensure that they were indeed Hindi native speakers, they had to solve two riddles in Hindi. 17 participants failed this test, so 31 participants were included in the final analysis.[8]

[8] The English translation of the riddles:

i. A boy and a doctor were walking down the street. The boy is a doctor's son while the doctor is not the boy's father. Then who is the doctor? (a) The boy is the doctor (b) The doctor is the boy's mom (c) The mom is the boy (d) The doctor is the mom's son.

ii. A mom and a dad have 6 sons, and each of the sons has one sister. How many people are there in the family? [Answer: 9.]

These are hard riddles, which explains our high exclusion rate.

2.2 Methods and Materials

In this and the following study, we presented 31 sentences without the verb, containing only the subject and an inanimate direct object (e.g., *Mohan(-ne) orange-ko* …). In order to see whether subject case marking alone triggers inferences on event completeness, the sentence fragments either contained an ergative-marked subject (*Mohan-ne*) or a nominative unmarked subject (*Mohan*), as a within-participants manipulation. The direct object was generally definite, singular, and inanimate, and marked with the accusative marker *-ko*,[9] ensuring that the fragment could not be interpreted as passive and nominative marking could not be confused for an unmarked patient role. Participants then answered a forced-choice task about whether they thought this fragment described an action that was done on purpose (see **Figure 2** and **Table 1**).

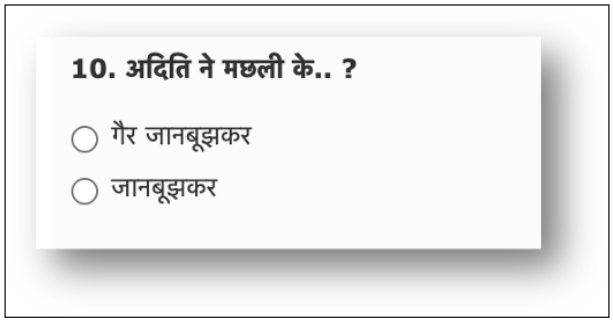


Figure 2: Example of a forced choice task. The task was the same for Experiments I and II, and differed only in whether it asked for intuitions on intentionality (Experiment I) or event culmination (Experiment II).

Table 1: Transcription and translation for an intentionality judgment.

kya yaha kriyaa jaan-bujhkar ki gayi hai ? (Has this action been done on purpose?) **mohan kaagaz = ko …?** (Mohan-NOM paper = ACC … ?) ◦ *gair jaan-bujhkar* (not on purpose) ◦ *jaan-bujhkar* (on purpose)

In addition to the critical sentences, we also included 24 fillers, in which random portions of the sentence were eliminated.

2.3 Results

All our data, with the exception of Experiment IV, were analyzed using mixed-effects models with maximal effects structure on contrast-coded conditions, with random slopes and intercepts for participants and items. In Experiment I, the fixed effect was case marking.

[9] In a handful of items, the direct object was marked with the (locative) adessive case *-pe*. In one item, the direct object, *chane* 'chickpeas', was unmarked. These alternative markings did not affect the telicity of the predicate.

As **Table 2** and **Figure 3a** show, people drew intentionality inferences for fragments with nominative and ergative-marked subjects at the roughly the same rate: Fragments containing an ergative subject were rated to describe an intentional action around 74% of the time, and fragments containing a nominative subject, 76% of the time.

Table 2: Intentionality ratings, summary statistics.

	case marking	N	Intentional	SD	SE
1	ergative	359	74%	0.44	0.02
2	nominative	357	76%	0.43	0.02

The two conditions of the fixed effect, case marking, were contrast-coded as ergative = 1 and nominative = –1. Within this model, the effect of case was statistically non-significant and negative (beta = –0.05, p = 0.575; see **Table 3**), meaning that the likelihood of participants judging an item to be intentional was very slightly, but not meaningfully, higher in the nominative condition.

Table 3: Intentionality ratings, test statistics.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.0993	0.0864	12.73	0.0000
case	–0.0484	0.0864	–0.56	0.5749

2.4 Discussion of Experiment I

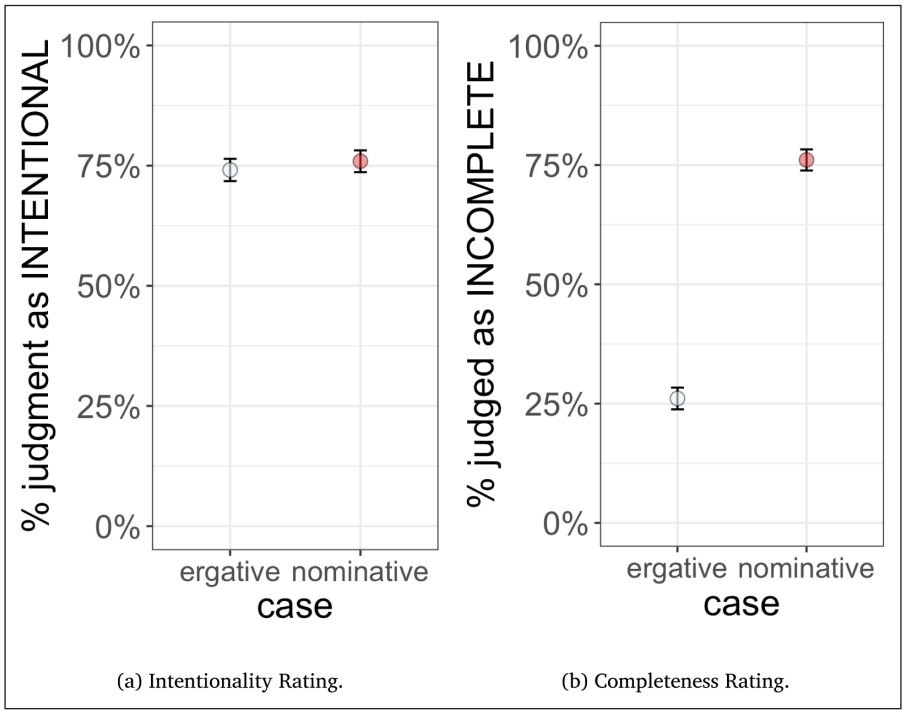
This binary choice task served to determine whether ergative case marking by itself, i.e., in the absence of further semantic information about the predicate, carries a meaning of intentionality with it, a question that has repeatedly arisen throughout the previous literature (e.g., Butt, 1995; Longenbaugh and Polinsky, 2017; Maitreyee et al., 2023; Mohanan, 1994). However, our data show a similar proportion of intentionality inferences for both types of case marking, nominative and ergative.

One possibility for this lack of difference could be that the wording of our task was ill-chosen, given that similar notions are employed throughout the literature, from intentionality to volitionality to “conscious choice”. To the extent that all of these notions draw from a closely related conceptual space, however, these data show that if there is any intentional meaning that arises from a sentence containing ergative marking, this meaning is likely to be inferred

from either the predicate alone (which was absent in our study), or by the combined effect of ergativity and predicative meaning.[10]

Another possible explanation is that the semantic contrast of a volitional vs. involuntary agent conferred by ergative vs. nominative marking only exists in a (small) subset of intransitive verbs (De Hoop and Narasimhan, 2005; Maitreyee et al., 2023; Mohanan, 1994). Given a sentence fragment containing first an animate noun, followed by an inanimate (and accusative-marked) noun, a comprehender has no reason to assume that circumstance applies; they may infer that the nouns form two arguments of a transitive verb, of which the first is judged to have a higher likelihood of being volitional.

For the purposes of this article, however, we lay the intentionality hypothesis aside: We have no reason to assume that subject case marking alone plays any role in inferences on volition or "conscious choice" in event structure. But does it for event culmination?



(a) Intentionality Rating. (b) Completeness Rating.

Figure 3: Results of Experiments I and II: Completeness and intentionality inferences, based on subject and object information alone.

[10] Butt (1995) also notes that "in an unmarked transitive, perfective sentence like […] 'He sang the song by mistake' (ergative subject, no light verb), the ergative case is fulfilling a structural requirement, and does not necessarily indicate a meaning of conscious choice. […] A conscious choice reading is therefore not directly tied to the appearance of ergative versus nominative case on the subject, but depends on a complex interaction between the semantic contribution of the light verb, the specification of conscious choice on main verbs, and the appearance of the ergative case marker."

3. Experiment II: Inferences about event completeness

This experiment, like Experiment I, served as a context-free baseline to determine whether case marking on a specific noun alone, without any event information except for agent and patient, will lead to different inferences about whether an event was completed or not.

3.1 Participants

We paid 48 participants on Amazon Mechanical Turk to participate in this experiment. All participants were self-reported Hindi native speakers and had an IP address in India. As an attention check, and to ensure that they were indeed Hindi native speakers, they had to solve two riddles in Hindi. 18 participants failed this test, so 30 participants were included in the final analysis.

3.2 Methods and Materials

We used the same materials as in Experiment I, but instead of asking people to rate whether they thought the event described by the sentence fragment was intentional, we asked whether they thought the action described by the event was completed.

3.3 Results

As **Table 4** and **Figure 3b** show, people had clear intuitions about whether the event expressed by the fragment was completed or not, depending on case marking: Fragments containing an ergative subject were rated to be incomplete only 26% of the time, but fragments containing a nominative subject, 76% of the time.

Table 4: Completeness ratings, summary statistics.

	case marking	N	incomplete	SD	SE
1	ergative	372	26%	0.44	0.02
2	nominative	372	76%	0.43	0.02

As in Experiment I, the two conditions of the fixed effect were contrast-coded as ergative = 1 and nominative = –1. Within this model, the effect of case was statistically significant and negative (beta = –1.10, $p < .001$; see **Table 5**), meaning that the likelihood of participants judging an item to be incomplete was much higher in the nominative condition.

Table 5: Completeness ratings, test statistics.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.0574	0.0847	0.68	0.4983
case	–1.0994	0.0847	–12.98	0.0000

3.4 Discussion of Experiment II

This binary choice task established that ergative case marking by itself, without any lexical or contextual information about the predicate, gives rise to event completion or culmination inferences: The recorded proportion of culmination inferences is significantly higher than in the nominative-marking condition.

This means that Hindi speakers privilege an event culmination inference when they encounter ergative subject marking, even in the absence of any further information about the kind of event itself. Taking these results as a starting point, the following studies ask whether we can trace when during sentence comprehension these inferences arise: From the earliest point on, that is, right after case marking on the subject is heard? Or do people hold off on their interpretation until they have all available information, as has been argued for other phenomena (e.g., Pickering et al., 2006)?

4. Experiment III: Timed picture-matching

This experiment asks whether participants use case marking on the noun as an early cue to event culmination, given the strong association between ergativity and event culmination, as shown in Experiment II. Following Foppolo et al. (2021), we use a picture-sentence matching task, measuring both accuracy rates and response times. A strong prediction hypothesis would expect that sentences starting with an ergative-marked subject will be matched to pictures showing culminated events as soon as the ergative marker is available; alternatively, one might predict no such effect based on the special structural or semantic status of the subject, which may lead comprehenders to withhold commitment to an interpretation until the whole predicate is heard.

4.1 Method & materials

We created 24 sets of stimuli like (4–7). Each set described two events involving the same object (i.e., an orange) in a Patient role. One event was durative ('peel'), and the other event was punctual ('drop').[11] Each event either described a completed event like (4) or (6), or an event in progress, like (5) or (7). Each sentence was preceded by the command *hame ye bataein* ('Tell us…').

(4) mohan=ne narangi=ko kahan chhila hai
mohan=ERG orange=ACC where peel-PerfPart do
'…where Mohan has peeled the orange.'

(5) mohan-∅ narangi=ko kahan chhil raha hai
mohan-NOM orange=ACC where peel Pres do
'…where Mohan is peeling the orange.'

[11] In compiling the set of events to include in the experiment, we followed Foppolo et al. (2021) as closely as possible, and also used the *for*-adverbial diagnostic to test whether events were sufficiently (non)-durative.

(6) mohan = ne narangi = ko kahan giraya hai
mohan = ERG orange = ACC where drop-PerfPart do
'…where Mohan has dropped the orange.'

(7) mohan-∅ narangi = ko kahan gira raha hai
mohan-NOM orange = ACC where drop Pres do
'…where Mohan is dropping the orange.'

We paired the sentences above with a pair of pictures showing an event either in its completed or its incomplete state. Each participant saw each object (for instance, an orange) only paired with one event; so, for a given participant, Example (4) would be paired with the two pictures in **Figure 4a**, and for the next participant, Example (5) would be paired with the two pictures in **Figure 4a**. A third and fourth set of participants saw Examples (7) and (6), respectively, paired with **Figure 4b**, such that each participant saw 24 trials, six of each condition.

4.2 Procedure

As with Experiments I and II, all studies were deployed on Amazon Mechanical Turk. The norming studies (see below) were also built using the survey tools of Mechanical Turk; the main experiment was built using *jsPsych* (De Leeuw, 2015).

In preparation for this experiment, we conducted a series of norming studies to make sure that any potential effects would not be due to confounds introduced by the stimuli.

4.2.1 Norming studies

First, in order to check the visual complexity of the pictures, we calculated the Shannon entropy of each picture by losslessly compressing each picture file. We then compared the entropy values, that is, the size of the encoded picture file, in a regression, predicting Shannon entropy by completeness and event type, and their interaction (Lin, 1991; Martiniani et al., 2019). The pictures in each condition were visually equally complex, according to this measure, and there was no significant interaction ($ts < 1.6$, $ps > 0.11$).

We then conducted three norming studies: we normed the sentences for grammaticality (4.2.2), and we normed how well our sentence stimuli mapped onto the pictures (4.2.3) and vice versa (4.2.4).

4.2.2 Norming Study I: Sentence Stimuli

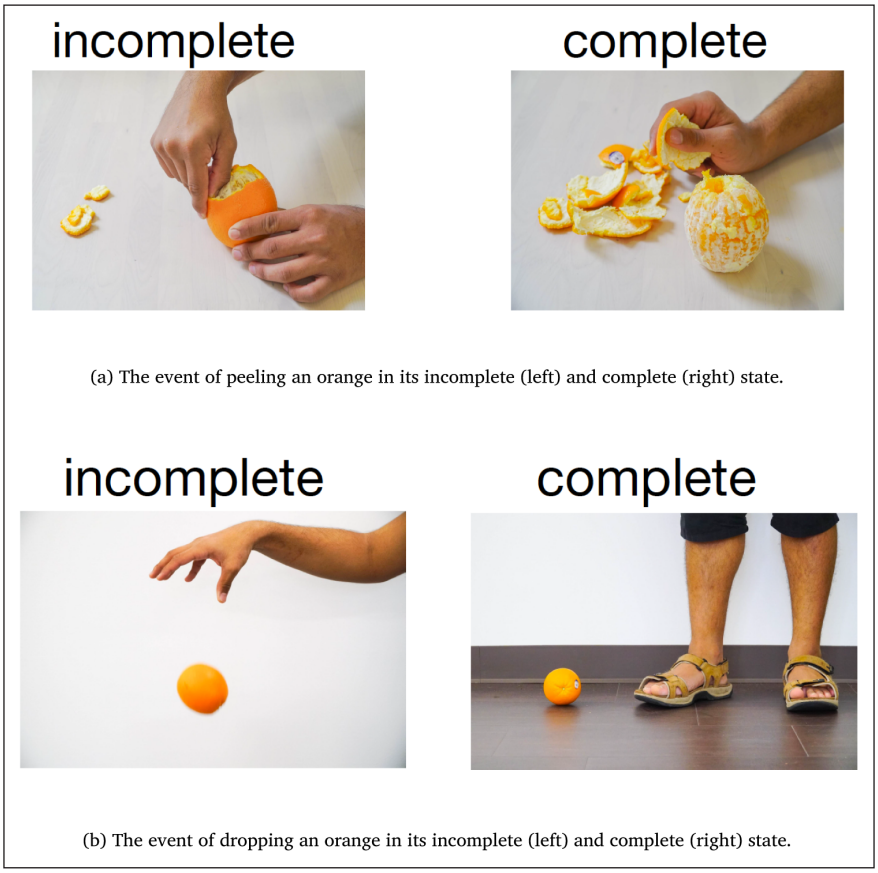
4.2.2.1 Participants

We paid 96 workers for their participation on Amazon Mechanical Turk. In all experiments, all participants were self-reported Hindi native speakers and had an IP address in India. 33 workers participated twice, so we only included their first submitted dataset in the analysis. We excluded

28 workers based on their performance on the riddles, leaving us with 35 individual participants' datasets.

4.2.2.2 Methods and results

Participants rated all sentences on a scale from 1 (ungrammatical) to 7 (grammatical). All factors were contrast-coded and the z-scored responses were analyzed using generalized linear mixed-effects regression with maximal effects structure. To determine significance, we used model comparisons.



(a) The event of peeling an orange in its incomplete (left) and complete (right) state.

(b) The event of dropping an orange in its incomplete (left) and complete (right) state.

Figure 4: Pictures of stages of events associated with the same object (orange): durative **(a)** and punctual **(b)**.

As **Table 6** and **Figure 5** show, there were no significant differences between conditions: Neither the main effect of event type, nor the main effect of case, nor their interactions were significant (main effects χ^2s (2, 25) $<$ 0.33, ps $>$ .36; interaction χ^2 (1, 25) $=$ 0.8092, ps $>$ 0.37). These results make it more likely that any difference we find in the sentence-picture matching task is not due to the varying grammaticality of the sentence conditions.

Table 6: Average grammaticality ratings of sentences in each condition, with Standard Deviations, Standard Errors, and Confidence Intervals. N indicates number of measurements per condition.

Case	Event	N	rating	SD	SE
ergative	durative	369	5.39	1.63	0.09
ergative	punctual	374	5.28	1.66	0.09
nominative	durative	363	5.39	1.65	0.09
nominative	punctual	371	5.40	1.59	0.08

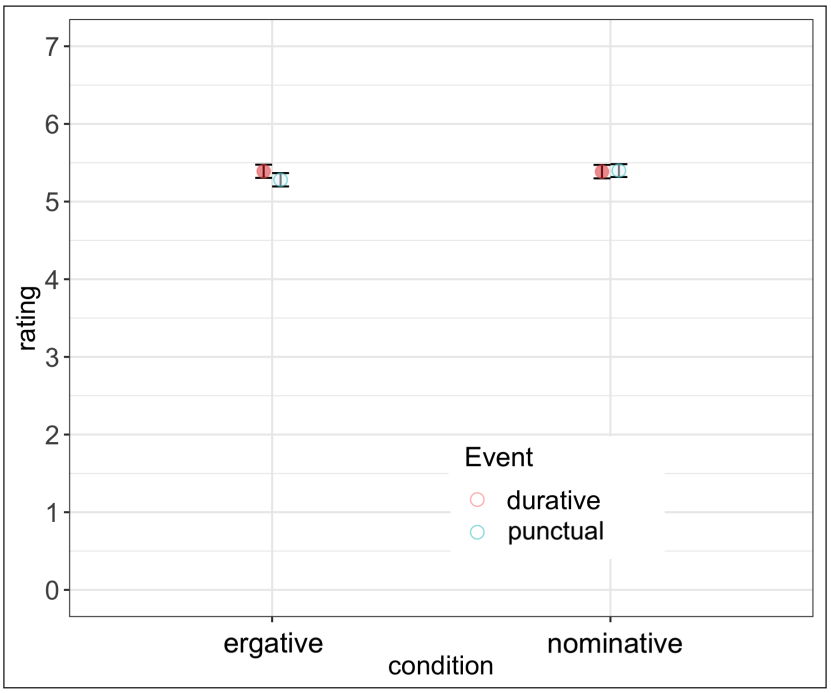


Figure 5: Norming Study I: Sentence ratings on a scale from 1 (bad) to 7 (good). Error bars show standard errors.

4.2.3 Norming Study II: Matching sentences to pictures

This norming study was carried out to ensure that our sentences and pictures were well matched, that is, that the pictures represented the sentences accurately.

4.2.3.1 Participants

We paid 96 workers for their participation. This time, 50 workers participated twice, so we only included their first submitted dataset in the analysis. We excluded an additional 21 workers based on their performance on the riddles, leaving us with 25 individual participants' datasets.

4.2.3.2 Methods and Results

In each trial, participants saw one written sentence and two pictures: the incomplete and completed versions of the event described by the sentence. We manipulated case/aspect as a within-subjects factor: for each object in the set of stimuli (e.g., an orange), participants saw either the two progressive sentences with nominative case (i.e., 'peeling' and 'dropping'), or the two perfective sentences with ergative case (i.e., 'peeled' and 'dropped'). Participants were asked to choose which picture of the two best matched the sentence. We coded the responses as 'correct' when a completed event picture was matched to a sentence with ergative case, and when an incomplete picture was matched to a sentence with nominative case.

Table 7: Average correct matching of sentences to pictures in each condition, with standard error. N indicates number of measurements per condition.

Case	Event	N	% correct	SE
ergative	durative	147	78%	0.03
ergative	punctual	146	75%	0.04
nominative	durative	152	89%	0.02
nominative	punctual	155	86%	0.03

We used logistic mixed models to determine whether the log-odds of selecting the correct picture were above chance level in each of the case/event type conditions, with random intercepts for items and participants.

As **Table 7** and **Figure 6a** show, people reliably matched pictures to the correct sentences (>50% correct in all conditions). The log-odds of selecting the incomplete picture to match the nominative sentence were significantly higher than 0, i.e. chance level (Intercept = 2.97, SE = 0.58, $Z = 5.07, p < 0.001$), and likewise for selecting the complete picture to match the ergative sentence (Intercept = 1.8, SE = 0.42, $Z = 4.3, p < 0.001$).

We then compared the conditions using binomial mixed-effects models, with sum-coded case and event type as fixed effects. We only used random intercepts for items and participants, because the model failed to converge with a maximal random-effects structure. To determine significance, we used model comparisons.

The percentage of correct matches did not depend on event type: There was no difference between punctual and durative events (χ^2 (2, 6) < 1.84, $p > .40$), and there was no interaction between event type and case, either (χ^2 (1, 6) < 0.52, $p > .47$). However, there *was* a main effect of case: Sentences were more likely to be matched to the correct picture in the nominative condition than in the ergative condition (χ^2 (2, 6) < 14.02, $p < .001$). So even though people demonstrated an above-chance preference both for matching ergative sentences with complete

pictures, and nominative sentences with incomplete pictures, that preference appears to be more flexible for ergative sentences – despite the (in)completeness inference tested in Experiment II being about equally strong for both cases.

4.2.4 Norming Study III: Matching pictures to sentences

This last norming study was designed to ensure that the sentences represented the pictures accurately, as a cross-check to Norming Study II.

4.2.4.1 Participants

We paid 96 workers for their participation. This time, there were 49 duplicate datasets. We excluded 19 workers based on their performance on the riddles, leaving us with 26 individual participants' datasets.

4.2.4.2 Methods and results

This study was the mirror image of the previous one: in each trial, participants were now shown one picture and two written sentences. Completedness was manipulated as a between-subjects factor, so that for each object (e.g., an orange), participants saw either the completed picture for an event (i.e., a dropped or fully peeled orange), or the incomplete picture (i.e., an orange being peeled or in mid-air). Paired with each picture were the two sentences (i.e., one with ergative case, and the other with nominative case, as in Examples (4) and (5)). Participants were asked to choose which sentence of the two matched the picture better. Parallel to the previous norming study, responses were coded as 'correct' when a picture showing a completed event was matched to a sentence containing ergative case, and when a picture showing an event in progress was matched to a sentence containing nominative case.

As in the previous norming study, we used logistic mixed models to determine whether the log-odds of selecting the correct picture were above chance level in each of the four conditions.

Table 8: Average correct matching of pictures to sentences in each condition, with standard error. N indicates number of measurements per condition.

Completeness	Event	N	% correct	SE
complete	durative	152	68%	0.04
complete	punctual	156	69%	0.04
incomplete	durative	155	75%	0.03
incomplete	punctual	161	70%	0.04

As **Table 8** and **Figure 6b** show, people again reliably matched pictures to correct sentences (>50% correct in all conditions). The log-odds of selecting the nominative sentence to match the incomplete picture were significantly higher than 0, i.e., chance level (Intercept = 1.21, SE = 0.29, $Z = 4.15$, $p < 0.001$), and likewise for selecting the ergative sentence to match the complete picture (Intercept = 0.95, SE = 0.26, $Z = 3.65$, $p < 0.001$).

As before, the data were then analyzed using binomial mixed-effects models, with contrast-coded event type and completeness as fixed effects, using only random intercepts for items and participants, and determining significance through model comparisons.

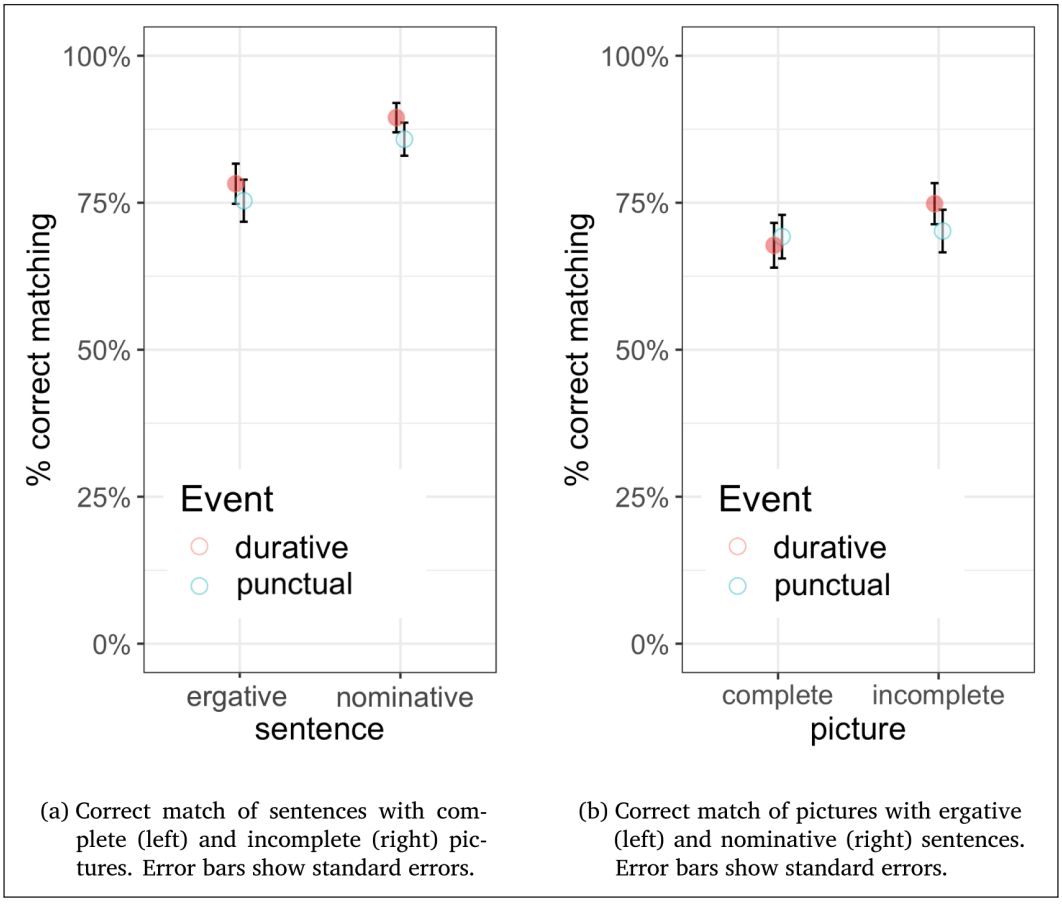
Unlike in the sentence-to-picture norming study, the main effect of completeness was not significant when people matched pictures to sentences (χ^2 (2, 6) = 2.04, $ps > 0.36$). Neither the main effect of event type, nor the interaction (event: χ^2 (2, 6) = 0.33, $p > 0.56$; interaction: χ^2 (1, 6) = 1.9, $ps > 0.30$) was significant. In short, people preferentially matched complete pictures to ergative sentences, and incomplete pictures to nominative sentences – without the latter being given significantly greater preference, as we saw in the sentence-to-picture study. One possible reason for that is that in the 'complete' condition, when the mental representation of the event is constrained to only its culmination, a nominatively-marked sentence is not as good a match as it is in the reverse situation (with two pictures spanning the whole event representation, and only one sentence).

4.3 Timed picture-matching experiment

In summary, to prepare for this experiment, we ran a visual complexity analysis, a grammaticality norming task, and two sentence/picture norming tasks. We found that the visual complexity of the picture stimuli was comparable across conditions; that the grammaticality of the sentence stimuli was, likewise, rated highly and very similarly across conditions; and that participants reliably matched sentences to the correct picture, and vice versa.

In this timed picture-matching experiment, we paired four kinds of sentences (crossing ergative or nominative case with durative or punctual events) with a pair of pictures showing an event either in its completed or its incomplete state. In each trial, the two pictures were presented side-by-side on the screen, without preview; and the sentence (created using Google's speech-to-text API) was presented auditorily. The experiment included no fillers.[12] Participants were instructed to choose the picture that best matched what they heard, and to try to respond as quickly as possible. We analysed participants' accuracy rates (i.e., how often they matched the sentence with the correct picture) and response times. As before, we coded matches of ergative sentences with completed-state pictures, and nominative sentences with incomplete-state pictures, as 'correct'.

[12] Our intent was to conceptually replicate Foppolo et al. (2021), who reported no filler items, only critical and warm-up trials. In addition, fillers, while conventionally included, can play a role that may exaggerate effects instead of hiding cues (see Arehalli and Wittenberg (2021) for discussion).



(a) Correct match of sentences with complete (left) and incomplete (right) pictures. Error bars show standard errors.

(b) Correct match of pictures with ergative (left) and nominative (right) sentences. Error bars show standard errors.

Figure 6: Results of Norming Studies II and III.

4.3.1 Participants

We paid 88 workers for their participation. No workers participated twice, and 16 workers were excluded based on performance on the riddles. In preparing the data, trials with a response time below 2300 ms, or higher than 8000 ms, were filtered out as outliers; as a result, two more workers were excluded.

4.3.2 Results

Accuracy. As before, we used logistic mixed models to determine whether the log-odds of selecting the correct picture were above chance level (zero) in each of the case/event type conditions. Random intercepts for items and participants were included. The results are presented in **Table 9**.

Of the two ergative conditions, only the punctual events had marginally significantly higher-than-chance odds of being matched with the completed picture. Both nominative conditions showed clearly significantly higher-than-chance odds of being matched with the incomplete picture.

Table 9: Average percentage of correct picture matches in each condition, with standard error; and model intercepts, standard errors, *Z*-scores and probability values for the logistic mixed models run on the accuracy data for each of the four conditions separately.

Case	Event	% correct	SE	Intercept	SE	*Z*	*p*
ergative	durative	55%	0.02	0.26	0.28	0.93	0.35
ergative	punctual	58%	0.02	0.48	0.2	2.32	0.02 *
nominative	durative	76%	0.02	2.45	0.6	4.05	$<$0.001 ***
nominative	punctual	61%	0.02	0.59	0.2	2.94	0.003 **

We then fitted a binomial mixed model with contrast-coded event type and case as fixed effects, and random intercepts for items and participants. We used pairwise comparisons to determine significance. Here, we found significant effects for both event type (χ^2 (2, 6) = 11.54, p = 0.003) and case (χ^2 (2, 6) = 22.2, p < 0.0001), as well as their interaction (χ^2 (1, 6) = 8.5, p = 0.003). In other words, participants were significantly more accurate on sentences with nominative case marking on the subject and a durative verb.

Response times. We analysed the response time data (see **Table 10**) for this experiment using a linear mixed model, with reaction time post-subject offset as the response variable and case and event type (and their interaction) as fixed effects. We again used random intercepts for items and participants because the model did not converge with a maximal random-effects structure. We found a significant effect for case (χ^2 (2, 7) = 17.8, p = 0.0001), but not for event type (χ^2 (2, 7) = 3.33, p = 0.18). The interaction between case and event type was also not significant (χ^2 (1, 7) = 0.2, p = 0.65).

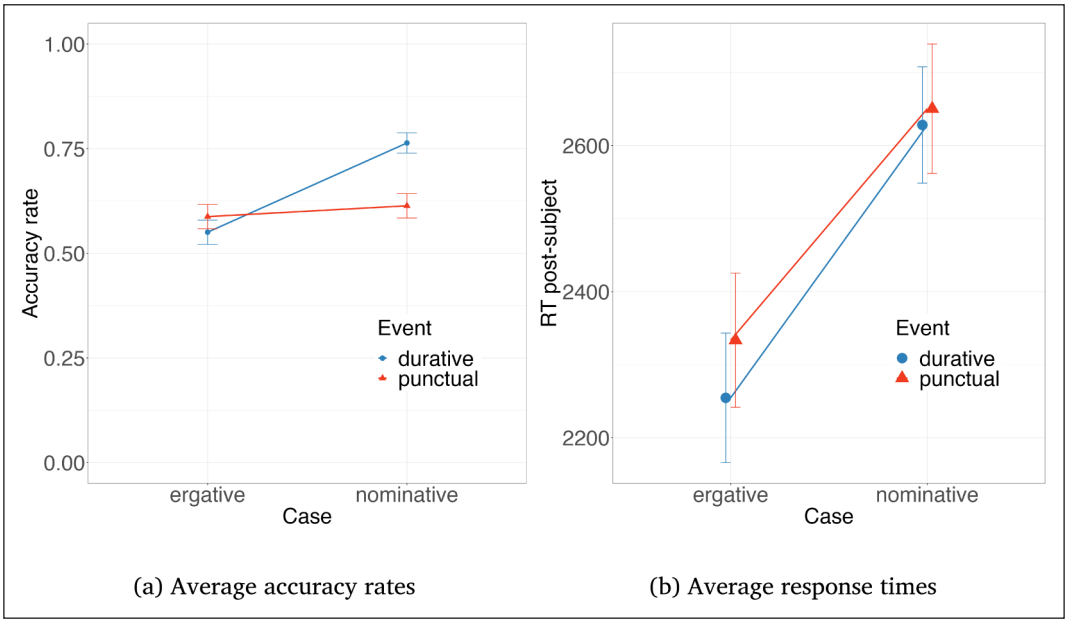
Table 10: Average correct matching of pictures to sentences in each condition; and average picture-matching response time, in milliseconds after the subject offset, with standard error. Verb onset is in milliseconds after subject offset.

Case	Event	% correct	RT	SE	Verb onset
ergative	durative	55%	2255	88.6	1550
ergative	punctual	58%	2334	91.9	1585
nominative	durative	76%	2628	79.6	1627
nominative	punctual	61%	2651	88.5	1627

4.4 Discussion of Experiment III

In our data (see **Figure 7**), even though Case and Event Type *do* interact in terms of the accuracy of participants' responses, they do not in terms of speed. Sentences in the nominative condition elicited longer response times regardless of event type – but it was only when the verb was durative that this longer response time also resulted in higher accuracy.

The strong prediction hypothesis, wherein ergative-marked sentences should be matched to completed-event pictures as soon as the ergative marker is processed, was, therefore, not borne out: ergative case marking may have resulted in faster response times, but with accuracy rates hovering just above chance level, that speed did not translate to the kinds of high accuracy rates expected, based on Experiment II, and on the rates of correct sentence-to-picture matching in Norming Study II. The two chief differences between that study and this one are the presentation mode of the sentence (audio rather than written) and the instruction to respond as quickly as possible. However, despite that prompt, participants waited to respond until they heard the verb: the average response time for ergative sentences was roughly 700 ms after verb onset, and roughly 1000 ms after verb onset for nominative sentences. What may explain this discrepancy between the response times of the ergative and nominative conditions is the object case marking: since the direct object in the nominative condition is the first constituent with overt case marking, the parser needs to disambiguate thematic roles and forecast a transitive



(a) Average accuracy rates (b) Average response times

Figure 7: Average proportion of correct responses **(a)** and average response time in milliseconds after subject offset **(b)**.

predicate, resulting in a slower reaction. A corpus check in the Hindi Dependency Treebank (Bhat et al., 2017) further supports this: the likelihood of encountering a transitive verb after hearing ergative *-ne* is very high, compared to other verb valencies.[13] The sentence type that was, by some margin, most reliably matched to the correct picture was nominative and durative – in other words, an in-progress event with null subject case marking, where the matching picture was a prototypical representation of the verb content.

There is a possibility that the timing data is somewhat unreliable, because the experiment was run in the browser with remote participants each using different hardware set-ups, internet speeds, etc. Bridges et al. (2020), who ran an exhaustive study cross-testing the precision and accuracy of all the most popular browsers, operating systems, and behavioural experiment software libraries, note that though there are unavoidable timing lags (in the range of tens of milliseconds) that are variable across different (combinations of) hardware and software equipment, whatever lag exists does appear to be generally consistent *within* subjects, which is what matters. In any case, we were able to control for this factor by running Experiment IV in the lab, using the same hardware and software configuration for all participants.

In addition, reaction times alone, as in Experiment III, do not provide sufficient evidence for the unfolding of incremental language comprehension: While reaction times can reflect the speed at which an individual responds to a stimulus, they do not capture the underlying cognitive processes that occur during the decision-making and response selection stages. To gain a comprehensive understanding of incremental comprehension, Experiment IV replicated Experiment III using an eye-tracking paradigm.

5. Experiment IV: Eye-tracking while picture-matching

As discussed above, the results in Experiment III constitute only coarse measures of the inference process. It may be that even though participants drew inferences early, they postponed their behavioral decision until later in the sentence. Experiment IV makes use of eye-tracking to test this explanation.

5.1 Method & materials

We used the same materials and method as in Experiment III, collecting data via an EyeTribe portable eye-tracker and custom-made software based on jsPsych, using a localhost server to record participants' eye-movements. Data were recorded at 30 Hz on a 15" MacBook Air, with participants sitting about 60 cm away, and the eye-tracker mounted on the bottom of the laptop screen (see **Figure 8**).

[13] Of 1186 instances of finite verbs occurring with an ergative subject, 859 were transitive; 163 were ditransitive; and 35 were intransitive.

Figure 8: Participant seated in front of portable EyeTribe eyetracker setup (on bottom of the screen).

5.2 Participants

We paid 46 students of Indian Institute of Technology (IIT) Delhi to participate in this experiment, recruited by an IIT Delhi listserv. All participants were Hindi native speakers. We excluded two participants from the analyses because of calibration issues.

5.3 Results

Picture selection data. As before, we used logistic mixed models to determine whether the log-odds of selecting the correct picture were above chance level in each of the case/event type conditions; with random intercepts for items and participants. Results are presented in **Table 11**.

We then fitted a binomial mixed model with contrast-coded event type and case as fixed effects, and random intercepts for item and participant; using pairwise comparisons to determine significance. As in Experiment III, both the main effects of event type (χ^2 (2, 7) = 36.1, $p <$ 0.0001) and case (χ^2 (2, 7) = 111.6, $p <$ 0.0001) were significant, and so was their interaction (χ^2 (1, 7) = 36.1, $p <$ 0.0001).

Eye-tracking data. To analyze eye-tracking behaviour, we excluded trials whose track-loss proportion was greater than 25%, resulting in a removal of 95 trials. We registered a track loss of 3.6% of all observations, leading to no exclusions of either participants or items. The gaze data are presented in **Figure 9**.

Table 11: Average percentage of correct matches of pictures to sentences in each condition in Experiment IV; and model intercepts, standard errors, *Z*-scores and probability values for the logistic mixed models for each of the four conditions separately.

Case	Event	% correct	SE	Intercept	SE	*Z*	*p*
ergative	durative	48%	0.03	–0.02	0.31	–0.08	0.93
ergative	punctual	65%	0.03	0.83	0.32	2.54	0.01
nominative	durative	88%	0.01	2.68	0.43	6.14	<0.001
nominative	punctual	72%	0.02	1.14	0.25	4.55	<0.001

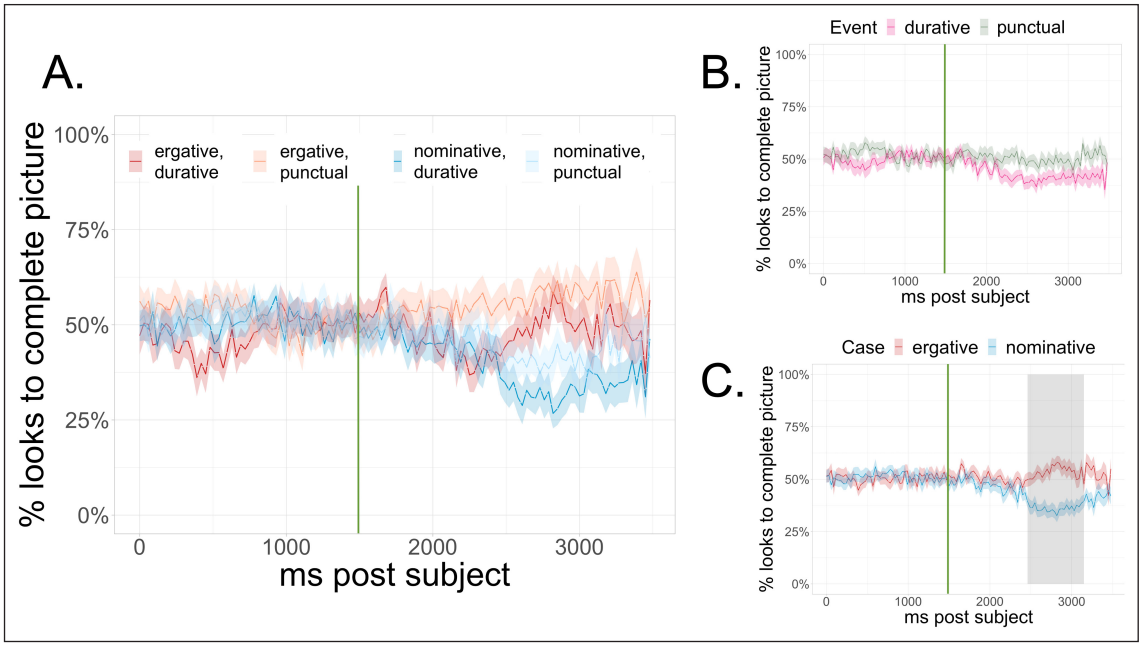


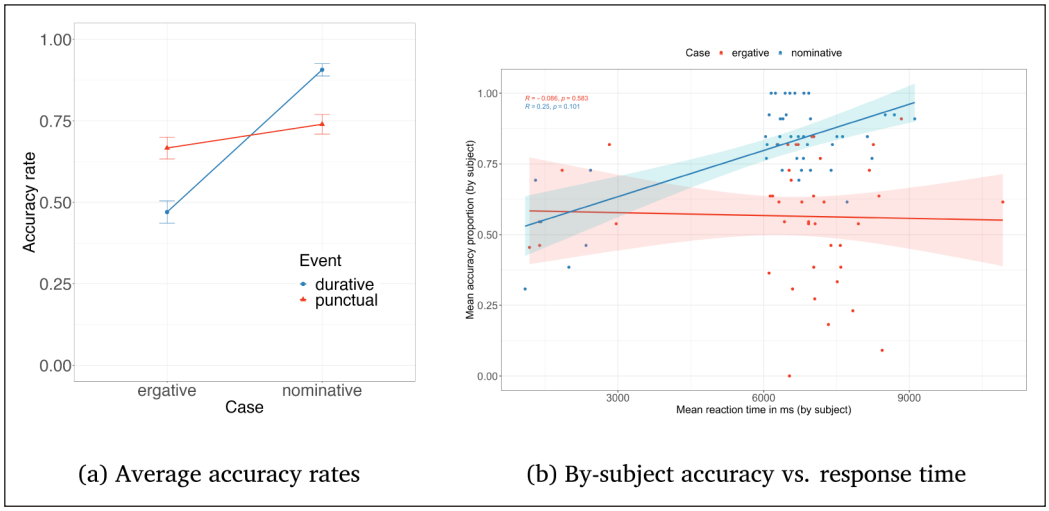
Figure 9: A. Experiment IV, percentage of looks to complete picture, **B.** by event type, **C.** by case. Grey shaded area indicates significant difference; color shaded areas mark confidence intervals. The green vertical line marks the average verb onset (1597 ms after subject offset).

In order to identify in which time windows participants' looking behaviour diverged significantly between conditions, we performed a cluster-based permutation analysis (Maris and Oostenveld, 2007) using the R package eyetrackingR (Dink and Ferguson, 2015). In this analysis, gaze data was binned into 50 ms bins, and each bin was *t*-tested (with Bonferroni correction) to identify clusters of time bins during which looking preference was significantly above chance.[14]

[14] We chose to perform the analysis with alpha = .08, because at the customary alpha = .05 the analysis identified a series of short, near-adjacent clusters which might represent one contiguous effect. This initial relaxation of the

To estimate the likelihood that these clusters represent genuine effects, this analysis was then repeated 1000 times with permuted predictor variables, resulting in a sum z-score distribution against which the original cluster sum z-scores were compared.

We performed separate analyses for the predictor variables Case and Event type. No significant clusters were found for Event type; for Case, we found one significant cluster, between 2460 and 3150 ms (sum $Z = 80.78$, $p < 0.0001$).[15] In other words, between 2490 ms and 3150 ms after the subject was heard, 1500 ms before average sentence end, people looked more towards the incomplete picture in the nominative case.



(a) Average accuracy rates (b) By-subject accuracy vs. response time

Figure 10: Experiment IV: **(a)** mean accuracy by Case and Event Type, and **(b)** scatterplot of mean accuracy vs. mean response time, by subject, with two Spearman correlations (by ergative and nominative condition) overlaid.

However, ergative case marking did not lead to more looks to completed pictures (**Figure 10**), in line with the results from Experiment III, and the picture selection data. To better understand how participants' picture selection speed related to their picture choice, we plotted by-participant mean response time (x-axis) against their mean accuracy rate, in a scatterplot with Spearman correlations for the ergative and nominative conditions. By doing so, we can see a bimodal distribution in the participants: some that respond early, before they have fully heard the verb phrase, and a larger group that responds later. The first group's accuracy rate is around chance level. In the latter group, there is a clear divergence between the two Case conditions: a relatively tight cluster of high-accuracy responses in the nominative condition, and a wide range of responses in the ergative condition, with some participants almost never

significance threshold was compensated later in the analysis, because the permutation procedure against which the original clusters are compared was also obtained with the same alpha.

[15] Note that this method of analysis does not identify a particular moment of effect *onset*.

matching ergative-marked sentences with their completed-event picture; some preferentially matching ergative sentences with the completed picture; and the majority falling somewhere in the middle.

5.4 Discussion of Experiment IV

This experiment confirmed the timed picture-matching data from Experiment III: Hindi speakers do not use ergative case marking for drawing inferences about event structures at the time the case marking would allow them to. Instead, we observe a delay in interpretative commitment until the lexical verb is heard.

In plotting by-subject mean offline accuracy against mean reaction time, we could see that only a handful of participants chose a picture early (i.e. before the verb was fully processed), and that their accuracy was at chance in both the ergative and nominative conditions. The majority of participants chose a picture well after verb offset, where a cluster of above-chance mean accuracy could be observed in the nominative condition, and a much more dispersed scattering of high-to-low accuracy in the ergative condition.

Case marking alone is clearly not a predictive cue, but the time window in which looks towards the completed event diverge significantly between conditions is found *before* verb offset. Assuming about 200 ms processing time to plan and execute an eye movement (Huettig et al., 2011; Saslow, 1967), that suggests that comprehenders in the nominative condition heard, at most, the onset of the verb before looking towards the incomplete picture. This looking preference is brief: immediately after verb offset, looks return to chance level. But it is notable for the fact that it emerges before the comprehender encounters any potential aspectual morphology (or lack thereof).

The preferential looks towards the incomplete event in the nominative condition align with higher offline accuracy in the picture choice (89% for durative verbs, and 74% for punctual verbs). The comparatively low average offline accuracy scores in the ergative condition confirm that, (as observed by Arunachalam and Kothari, 2011) perfective simple verb predicates do not entail event culmination. Event type does appear to have an effect, with choice of the completed event picture marginally above chance in the ergative punctual condition in both Experiments III and IV – again, consistent with Arunachalam and Kothari (2011), who found generally lower acceptability ratings for partially completed punctual events. However, we found no evidence for any effect of Event type in sentence processing: this is likewise consistent with Foppolo et al.'s, 2021 results. Their reason for including the durative/punctual event contrast was that for punctual events, culmination is lexically encoded in the verb, and should, therefore, lead to faster, more decisive culmination inferences during incremental processing. They, like us, found no difference between the two event types – perhaps that is because this experimental manipulation is awkward to realize with punctual events. It requires that an event that, by its

lexical aspectual definition, should not have an in-progress stage, or be split into its incomplete and completed components. The combination of static pictures with the sentence preamble ('show me where…'), implying a frozen snapshot of an event the participant can peruse at leisure, may have coerced the punctual events towards a more progressive reading; or the combination of lexical and grammatical aspect may simply be more 'elastic' (in the words of Pederson, 2007: about Tamil) in Hindi, as suggested by the fact that punctual events with nominative case (i.e., imperfective aspect) normed just as high as the other three sentence types in Norming Study I.

All this suggests that, while there might not be any *aggregated* looking preference towards either picture in the ergative condition, there is considerable variation among individual participants in their ultimate interpretation of perfective sentences. What could explain the degree of freedom Hindi speakers have to decide whether an ergatively-marked subject has finished doing something or not?

6. General discussion

In this article, we asked whether people immediately integrate all available cues to make predictive inferences about event structure during language processing. Specifically, we used Hindi, a verb-final language with a split ergative system, to ask whether volitionality or event culmination inferences are predicted by subject case marking. This question is based on the reliable structural linguistic association between ergative case marking and the perfective aspect; that is, the grammar leaves little room for uncertainty that an event described by the verb in a sentence containing an ergative subject is culminated. However, our data suggest that Hindi native speakers do not reliably draw this inference.

In the first experiment, we found no evidence that volitionality inferences are drawn based on case marking alone. This is in line with recent research indicating that children keep track of both the statistical occurrence of ergative markers given a verb, and clause-level semantics indicating volition (Maitreyee et al., 2023). That is, the intuition that ergative interpretations involve volitionality may be a classic case of correlation, not causation: Ergative marking often occurs in volitional contexts, but it does not carry volitional meaning; instead, there is independent evidence that agent control and volitionality are closely linked to event culmination itself (Demirdache and Martin, 2015; Mathis and Papafragou, 2022; van Hout et al., 2017). While this is a correlation worthy of further investigation, for the purposes of this article, which focuses on ergative markers as cues for event inferences, we can put the volitionality hypothesis aside.

What is interesting in this context is that Experiment II, which used the exact same stimuli, but asked for completeness judgments instead of volitionality judgments, found a strong effect of subject case marking: People drew completeness inferences significantly more given ergative subjects than nominative subjects. That is, in the face of incomplete information about the event itself, ergative case marking was a strong cue for event culmination inferences.

However, these inferences do not percolate to decisions in the often-employed picture-matching task given full sentences (Experiment III): The overall accuracy was lower in ergative trials, but the response time (in the "correct-response" trials) was faster. So, it is possible that encountering ergative case facilitates the retrieval of a transitive predicate, for which it is a reliable cue (barring a few exceptions among the ingesto-reflexive intransitives). However, it may be possible that this does not then also deliver an inference about event completion.

Finally, when tracking participants' eye-movements while they were performing a picture-matching task, it is evident that only some cues of case marking influence real-time processing, but their effect also dissipates by sentence end (Experiment IV). In particular, no preference emerges in the ergative condition – both in terms of looks and picture choice – even after the perfective aspectual morphology is processed. In nominative-marked sentences, on the other hand, participants looked towards the incomplete picture more; but by sentence end, this effect also disappeared; in general, participants were slower, but more accurate, in reacting to sentences containing nominative case marking. Ergative case marking, in contrast, led to faster, but less accurate, chance-level decisions.

Three broad classes of explanations could account for this quite subtle pattern of results. The first class is rooted in a structural account of sentence processing: Perhaps our data are evidence that, in order to draw inferences about event structure based on a linguistic structure, one needs to parse the piece of structure that usually corresponds to event encoding, at least in what Levin (1999) calls *core transitive* verbs: the predicate, that is, the Verb Phrase, with all its (internal) arguments (Husband and Stockall, 2015). Since the subject is not part of this verb phrase, or simply because it is not sufficient to build an event representation, any cues it may potentially carry can usually be safely ignored.

This class of explanations would hold cross-linguistically, but in languages without split ergativity or a system like Hindi, its effect could not be seen. In order to test this hypothesis more systematically, one may want to conduct a similar experiment in a language that has more reliable correlations of aspect marking and subject marking, if such a language exists.

The second class of explanation concerns language-specific affordances; here, the overall, language-wide reliability of case marking patterns in Hindi. For case marking to facilitate sentence processing, there needs to be an unambiguous case-to-argument association (Skopeteas et al., 2012) – and while there is clearly a strong enough association between ergative case and perfective aspect to allow Hindi speakers to infer the latter from the former (Experiment II, corpus data), perhaps there are too many exceptions and caveats to the pattern (Butt, 1995; Butt and King, 2004; De Hoop and Narasimhan, 2005) to make ergative case a reliable predictive cue.

Perfective aspect itself is also associated with *default* agreement. For example, in Example (8), which contains with two case marked feminine nouns, the agreement on the verb is the default masculine singular. As Hindi speakers track the realization of case on subjects and objects, any

early completion judgements made on encountering ergative case can be diluted on encountering the object case marking. These facts further weaken the association between perfective aspect and culmination or completion.

(8) ladkii = ne rotii = ko khaa-ya th-aa
girl.F = ERG bread.F = ACC eat-Perf.M.SG be-Past
'(the) girl ate the bread.'

Imperfective sentences are, then, comparatively easier to parse: an ergative case marker is *blocked* for sentences with imperfective aspect (recall that the reverse is not always true; see Mohanan, 1994), and so by the time a comprehender has encountered null marking on the subject followed by a direct object, the likelihood of the event being incomplete is very high. In contrast, the likelihood of the event being fully culminated remains relatively low, even after the lexical verb has been processed.

Even with this advantage, the in-progress pictures corresponding to sentences with nominative case did not attract more looks until the verb had been processed: Inferencing of event structure during online sentence processing was localized to the verb. Previous evidence suggests that the integration of aspectual information is both highly incremental and quite conservative (Bott, 2013; Bott and Hamm, 2014; Minor et al., 2022): Comprehenders will wait for an unambiguous trigger before they commit to an aspectual interpretation of an event. So even though we have robust evidence that case facilitates early (preverbal) thematic interpretation (Kamide, Altmann, and Haywood, 2003; Kamide, Scheepers, and Altmann, 2003; Özge et al., 2019, 2022), a strong co-occurrence of case with aspect seems to be not enough to induce prediction on the level of event structure.

That said, any effects observed here were primarily driven not by the marked, but by the unmarked structure: the nominative. Participants were more accurate classifying nominative sentences as denoting events in progress, and eye-tracking data reveal that participants looked significantly more to events in progress before verb end when they heard sentences containing nominative structures. That is, our data show that the presence of an asymmetrical case contrast can decrease the completeness bias found by Madden and Zwaan (2003) (but see Minor et al., 2022 and Misersky et al., 2021 for similar lack of evidence on completeness biases).

At first glance, this is unexpected: Nominative structures are not exclusively associated with events in progress – in fact, they often occur with perfective aspect. However, they only occur in perfective aspect if the sentence is intransitive, not transitive (Anand and Nevins, 2006; Coon and Parker, 2019; Kachru, 1980; Kachru and Pandharipande, 1978; Mohanan, 1994). Thus, within the context of transitive events, nominative case marking is as reliable a cue as ergative case marking as soon as a direct object is heard.

Why, then, should there be an asymmetry between ergative and nominative cues during processing? Our interpretation is based on accounts arguing that Hindi perfective simple verbs have only a weak (or pragmatic) implicature of event culmination, and that the culmination

entailment is conferred by a subsequent light verb (Arunachalam and Kothari, 2011; Butt, 1995; Pederson, 2007; Singh, 1998). In that case, the comprehender cannot know until after the lexical verb whether the predicate will be simple or complex, and, therefore, whether the event is definitely culminated or not. Perfective morphology by itself may only lend temporal boundedness to the event, but not culmination: The event may have reached an arbitrary endpoint, but not necessarily its natural endpoint (Martin, 2019; Singh, 1998).

The asymmetry hypothesis may explain why the rate of completedness judgments in the ergative condition of Experiment II was *higher* than in Experiment IV, where the majority of participants could, and mostly did, wait to hear the entire sentence before choosing a matching picture. Given only an ergative sentence fragment, Hindi speakers seem to err on the side of culmination interpretations – a preference that is diminished, not strengthened, by the predicate.

This interpretation is also compatible with evidence that case-aspect mismatches in Hindi generate N400 effects (Choudhary et al., 2009; Dillon et al., 2012). However, the mismatches in these studies were detected when there was a grammatical mismatch between case marking and aspect, as would be expected from corpus data (**Figure 1**); our studies, in contrast, probed event construal from perfectly grammatical structures.

A third, complementary way to explain our findings is that ergative case marking is not strongly associated with event culmination. In fact, our data are squarely in line with Arunachalam and Kothari (2011), who showed that the culmination inference is defeasible in ergative-marked sentences. Our experiments, then, take Arunachalam and Kothari (2011) one step further: Not only is the culmination inference defeasible when probed; our evidence suggests that it is not reliably calculated in the first place.

If this interpretation is correct, then several questions are open for further investigation. One such question is whether perfective aspect in combination with a light verb construction would induce stronger culmination effects in Hindi. Second, to better understand the role of external subject arguments for calculating event structure inferences, it would be highly relevant to test cross-linguistic phenomena related to event structure that are marked on the subject.

In conclusion, in four experiments, we have not found evidence that people immediately integrate all available cues to predict event structure when these cues occur outside of the predicate, and we laid out theoretical explanations that could account for this lack of prediction. We hope that further studies are inspired to test these theoretical accounts in depth.

Abbreviations

∅ = null or non-overt case marking

ACC = Accusative

DAT = Dative

ERG = Ergative

NOM = Nominative

M/F = Male/Female (grammatical gender)

Past = Past tense

Perf = Perfective

PerfPart = Perfective participle

Pres = Present

Prog = Progressive

SG = Singular

Data accessibility

All raw data files, R data processing and analysis scripts, and stimuli are made available through the Open Science Framework.

(https://osf.io/h8zfs/).

Ethics and consent

All experiments detailed in this paper were approved by the Institutional Review Board of University of California San Diego (project number 161583S), and were carried out in accordance with the Declaration of Helsinki.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Conceptualization, E.W. and A.V.; data curation, M.G., A.V. and E.W.; formal analysis, M.G. and E.W.; funding acquisition, E.W.; investigation, M.G.; methodology, E.W., A.V. and M.G.; project administration, E.W.; software, M.G. and M.V.; supervision, E.W.; visualization, M.G. and M.V.; writing – original draft, E.W., A.V., M.G. and M.V.; writing – review & editing, M.V., E.W. and A.V.

ORCiD IDs

Myrte Vos: https://orcid.org/0000-0003-3896-2742

Mohit Gurumukhani: https://orcid.org/0009-0007-8808-2846

Ashwini Vaidya: https://orcid.org/0009-0003-0537-110X

Eva Wittenberg: https://orcid.org/0000-0002-3188-6145

References

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Anand, P., & Nevins, A. (2006). The locus of ergative case assignment: Evidence from scope. In A. Johns, D. Massam & J. Ndayiragije (Eds.), *Ergativity* (pp. 3–25). Springer. https://doi.org/10.1007/1-4020-4188-8_1

Arehalli, S., & Wittenberg, E. (2021). Experimental filler design influences error correction rates in a word restoration paradigm. *Linguistics Vanguard*, *7*(1). https://doi.org/10.1515/lingvan-2020-0052

Arunachalam, S., & Kothari, A. (2011). An experimental study of Hindi and English perfective interpretation. *Journal of South Asian Linguistics*, *4*(1), 27–42.

Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, *22*(9), 2131–2140. https://doi.org/10.1162/jocn.2009.21303

Bashir, E. (1999). The Urdu and Hindi ergative postposition *Ne*: Its changing role in the grammar. In *The yearbook of south asian languages and linguistics* (pp. 11–36). De Gruyter. https://doi.org/10.1515/9783110245240.11

Bhat, R. A., Bhatt, R., Farudi, A., Klassen, P., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., Vaidya, A., Vishnu, S. R., & Xia, F. (2017). The Hindi/Urdu Treebank Project. In N. Ide & J. Pustejovsky (Eds.), *The handbook of linguistic annotation* (pp. 659–697). Springer. https://doi.org/10.1007/978-94-024-0881-2_24

Bickel, B., Witzlack-Makarevich, A., Choudhary, K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *Plos One*, *10*(8). https://doi.org/10.1371/journal.pone.0132819

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). *Processing syntax and morphology: A neurocognitive perspective* (Vol. 6). Oxford University Press. https://doi.org/10.1093/oso/9780199207817.001.0001

Bott, O. (2013). The processing domain of aspectual interpretation. In B. Arsenijević, B. Gehrke & R. Marín (Eds.), *Studies in the composition and decomposition of event predicates* (pp. 195–229). Springer. https://doi.org/10.1007/978-94-007-5983-1_8

Bott, O., & Gattnar, A. (2015). The cross-linguistic processing of aspect – An eyetracking study on the time course of aspectual interpretation in Russian and German. *Language, Cognition and Neuroscience*, *30*(7), 877–898. https://doi.org/10.1080/23273798.2015.1029499

Bott, O., & Hamm, F. (2014). Cross-linguistic variation in the processing of aspect. In B. Hemforth, B. Mertins & C. Cathrine Fabricius-Hansen (Eds.), *Psycholinguistic approaches to meaning and understanding across languages* (pp. 83–109). Springer. https://doi.org/10.1007/978-3-319-05675-3_4

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*. https://doi.org/10.7717/peerj.9414

Butt, M. (1995). *The structure of complex predicates in Urdu*. CSLI Publications (Stanford University).

Butt, M. (2017). Hindi/Urdu and related languages. In J. Coon, D. Massam & L. Demena Travis (Eds.), *The Oxford handbook of ergativity*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198739371.013.33

Butt, M., & King, T. H. (2004). The status of case. In V. Dayal & A. Mahajan (Eds.), *Clause structure in South Asian languages* (pp. 153–198). Springer. https://doi.org/10.1007/978-1-4020-2719-2_6

Choudhary, K. K., Schlesewsky, M., Roehm, D., & Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia*, *47*(13), 3012–3022. https://doi.org/10.1016/j.neuropsychologia.2009.05.009

Coon, J., & Parker, C. (2019). Case interactions in syntax. In *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.315

Dahl, Ö., & Velupillai, V. (2013). Tense and aspect (v2020.4). In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Zenodo. https://doi.org/10.5281/zenodo.13950591

De Hoop, H., & Narasimhan, B. (2005). Differential case-marking in Hindi. In M. Amberber & H. de Hoop (Eds.), *Competition and variation in natural languages* (pp. 321–345). Elsevier. https://doi.org/10.1016/B978-008044651-6/50015-X

De Leeuw, J. R. (2015). Jspsych: A Javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Demirdache, H., & Martin, F. (2015). Agent control over non culminating events. In E. Barrajón López, J. L. Cifuentes Honrubia & S. Rodríguez Rosique (Eds.), *Verb classes and aspect* (pp. 185–217). John Benjamins. https://doi.org/10.1075/ivitra.9.09dem

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, *21*(5), 629–634. https://doi.org/10.1177/0956797610367751

Dillon, B., Nevins, A., Austin, A. C., & Phillips, C. (2012). Syntactic and semantic predictors of tense in Hindi: An ERP investigation. *Language and Cognitive Processes*, *27*(3), 313–344. https://doi.org/10.1080/01690965.2010.544582

Dink, J. W., & Ferguson, B. (2015). Eyetrackingr: An R library for eye-tracking data analysis. http://www.eyetrackingr.com

Eggins, S., & Slade, D. (2006). *Analysing casual conversation*. University of Toronto Press.

Ferreira, F., & Qiu, Z. (2021). Predicting syntactic structure. *Brain Research*, *1770*, 147632. https://doi.org/10.1016/j.brainres.2021.147632

Flecken, M., Von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition*, *6*(1), 45–78.

Foppolo, F., Bosch, J. E., Greco, C., Carminati, M. N., & Panzeri, F. (2021). Draw a star and make it perfect: Incremental processing of telicity. *Cognitive Science*, *45*. https://doi.org/10.1111/cogs.13052

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *Plos One*, *9*(7). https://doi.org/10.1371/journal.pone.0100986

Husband, E. M., & Stockall, L. (2015). Building aspectual interpretations online. In R. G. de Almeida & C. Manouilidou (Eds.), *Cognitive science perspectives on verb representation and processing* (pp. 157–186). Springer. https://doi.org/10.1007/978-3-319-10112-5_8

Kachru, Y. (1980). *Aspects of Hindi grammar*. Manohar.

Kachru, Y., & Pandharipande, R. (1978). On ergativity in selected South Asian languages. *Studies in the Linguistic Sciences*, *8*(1), 111–126.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Kaplan, M. J., Raju, A., & Arunachalam, S. (2021). Real-time processing of event descriptions for partially- and fully-completed events: Evidence from the visual world paradigm. *Proceedings of the Linguistic Society of America*, *6*(1), 118–132. https://doi.org/10.3765/plsa.v6i1.4954

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. https://doi.org/10.1080/23273798.2018.1524500

Kratzer, A. (1996). Severing the external argument from its verb. In J. Rooryck & L. Zaring (Eds.), *Phrase structure and the lexicon* (pp. 109–137). Springer. https://doi.org/10.1007/978-94-015-8617-7_5

Levin, B. (1999). Objecthood: An event structure perspective. *Proceedings of CLS*, *35*(1), 223–247.

Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. https://doi.org/10.1016/j.jml.2012.02.005

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151. https://doi.org/10.1109/18.61115

Longenbaugh, N., & Polinsky, M. (2017). Experimental approaches to ergative languages. In J. Coon, D. Massam & L. Demena Travis (Eds.), *The Oxford handbook of ergativity* (pp. 709–736). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198739371.013.29

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*(5), 663–672. https://doi.org/10.3758/BF03196106

Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X17000012

Mahr, J. B., & Csibra, G. (2020). Witnessing, remembering, and testifying: Why the past is special for human beings. *Perspectives on Psychological Science*, *15*(2), 428–443. https://doi.org/10.1177/1745691619879167

Maitreyee, R., Saxena, G., Narasimhan, B., Sharma, D. M., Mishra, P., Nair, R. B., Samanta, S., & Ambridge, B. (2023). Children learn ergative case marking in Hindi using statistical preemption and clause-level semantics (intentionality): Evidence from acceptability judgment and elicited production studies with children and adults. *Open Research Europe*, *3*(49), 49. https://doi.org/10.12688/openreseurope.15611.2

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, F. (2015). Explaining the link between agentivity and non-culminating causation. *Semantics and Linguistic Theory*, *25*, 246–266. https://doi.org/10.3765/salt.v25i0.3060

Martin, F. (2019). Non-culminating accomplishments. *Language and Linguistics Compass*, *13*(8), e12346. https://doi.org/10.1111/lnc3.12346

Martin, F. (2020). Aspectual differences between agentive and non-agentive uses of causative predicates. *Perspectives on causation: Selected papers from the Jerusalem 2017 Workshop*, 257–294. https://doi.org/10.1007/978-3-030-34308-8_8

Martiniani, S., Chaikin, P. M., & Levine, D. (2019). Quantifying hidden order out of equilibrium. *Physical Review X*, *9*(1), 011031. https://doi.org/10.1103/PhysRevX.9.011031

Mathis, A., & Papafragou, A. (2022). Agents' goals affect construal of event endpoints. *Journal of Memory and Language*, *127*, 104373. https://doi.org/10.1016/j.jml.2022.104373

Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect processing. *Plos One*, *17*(2), e0264132. https://doi.org/10.1371/journal.pone.0264132

Misersky, J. (2022). *About time: Exploring the role of grammatical aspect in event cognition*. Radboud University Nijmegen.

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition*, *214*. https://doi.org/10.1016/j.cognition.2021.104744

Mohanan, T. (1994). *Argument structure in Hindi*. Center for the Study of Language (CSLI).

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Zu Wolfsthurn, S. V. G., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468. https://doi.org/10.7554/eLife.33468.024

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Özge, D., Vidinli, D., Küntay, A., & Snedeker, J. (2017). *When you eat from the cake, is it all gone? morphosyntax as a cue to partitivity*. [[Poster presentation.] 42nd Annual Boston University Conference on Language Development].

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, *183*, 152–180. https://doi.org/10.1016/j.cognition.2018.10.026

Özge, D., Kornfilt, J., Maquate, K., Küntay, A. C., & Snedeker, J. (2022). German-speaking children use sentence-initial case marking for predictive language processing at age four. *Cognition*, *221*. https://doi.org/10.1016/j.cognition.2021.104988

Paczynski, M., Ray, J., & Gina, K. (2014). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience*, *26*(9), 1905–1917. https://doi.org/10.1162/jocn_a_00638

Pederson, E. (2007). Event realization in Tamil. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 331–355). Taylor & Francis.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and aspectual coercion. *Discourse Processes*, *42*(2), 131–155. https://doi.org/10.1207/s15326950dp4202_3

Proctor, A. S., Dickey, M. W., & Rips, L. J. (2004). The time-course and cost of telicity inferences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26).

Pylkkänen, L. (2008). *Introducing arguments*. MIT Press. https://doi.org/10.7551/mitpress/9780262162548.001.0001

Ramchand, G. (1990). Complex predicate formation in Bangla. *Proceedings of the Ninth Annual West Coast Conference on Formal Linguistics*, 443–458.

Ranjan, R. (2016). *Acquisition of ergative case in L2 Hindi-Urdu*. The University of Iowa.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Josa*, *57*(8), 1024–1029. https://doi.org/10.1364/JOSA.57.001024

Singh, M. (1998). On the semantics of the perfective aspect. *Natural Language Semantics*, *6*, 171–199. https://doi.org/10.1023/A:1008208311200

Skopeteas, S., Fanselow, G., & Asatiani, R. (2012). Case inversion in Georgian: Syntactic properties and sentence processing. In M. Lamers & P. de Swart (Eds.), *Case, word order and prominence: Interacting cues in language production and comprehension* (pp. 145–171). Springer. https://doi.org/10.1007/978-94-007-1463-2_7

Staub, A., & Clifton Jr., C. (2006). Syntactic prediction in language comprehension: Evidence from *either…or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425. https://doi.org/10.1037/0278-7393.32.2.425

van Hout, A., Arche, M., Demirdache, H., del Real, I. G., Sanz, A. G., Gavarró, A., Marzo, L. G., Hommes, S., Kazanina, N., Liu, J., Lungu, O., Martin, F., & Strangmann, I. M. (2017). Agent control and the acquisition of event culmination in Basque, Dutch, English, Spanish and Mandarin. *The 41st annual Boston University conference on language development*, *41*, 323–332.

Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings*. Routledge. https://doi.org/10.4324/9780203500576

Verkuyl, H. J. (1989). Aspectual classes and aspectual composition. *Linguistics and Philosophy*, 39–94. https://doi.org/10.1007/BF00627398

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, *50*(4). https://doi.org/10.1515/ling-2012-0026

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516. https://doi.org/10.1093/cercor/bhv075

Wittek, A. (2002). *Learning the meaning of change-of-state verbs: A case study of German child language*. Mouton de Gruyter. https://doi.org/10.1515/9783110907988

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272–302. https://doi.org/10.1080/01690965.2011.622905

Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, *133*(1), 262–276. https://doi.org/10.1016/j.cognition.2014.06.018