The aspectual entailments of telicity markers in German: evidence from non-native and native speakers

Duarte Oliveira [ID]

Centro de Estudos Humanísticos, Universidade do Minho, Braga, Portugal

CAMBRIDGE
UNIVERSITY PRESS

Abstract

In German, it has been shown that the semantic entailments associated with telicity markers are acquired early and that speakers will turn to semantic–pragmatic principles to determine whether an overt culmination is cancellable (e.g., van Hout, 1998, 2008; Richter & van Hout, 2013; Schulz & Penner, 2002; Schulz & Ose, 2008). Here, we test the interpretation of three types of telicity markers by Portuguese L2 speakers of German, as well as Portuguese–German bilinguals and German monolinguals. A Bayesian analysis shows that Portuguese L2 speakers of German have difficulty processing telicity with resultative particles but show target-like performances with bounded DPs and adjectival markers. Our analysis also shows that bilingual and monolingual speakers display no substantial differences in their understanding of telicity entailments, albeit with some variability regarding particle markers. I argue that the existing variation may be due to effects of lexical knowledge and transparency.

1. Introduction

Telicity is an aspectual property of verb phrases that can either be determined lexically – based on the inherent aspectual configuration of the verb – or at the VP level – depending on the boundedness of the selected object and other grammatical components (Borer, 2005; Dowty, 1979; Krifka, 1998; Ramchand, 2008; Tenny, 1994; Travis, 2010; Verkuyl, 1993). Languages like German have specific grammatical means to mark telicity, namely, through the combination of resultative particles with certain durative verbs (e.g., *essen* 'to eat' *versus aufessen* 'to eat up'), as well as in adjectival resultative constructions (e.g., *den Tisch sauber wischen* 'to wipe the table clean'). Contrary to German, resultative constructions with particle or adjectival markers are not productive in European Portuguese (EP), and speakers only resort to more "universal" structures – such as certain direct object constructions and other lexical expressions – to convey telicity. In line with this difference, some authors (e.g., Filip, 2014; van Hout, 1998, 2000; Schulz, Wymann & Penner, 2001) have divided these telicity markers into two types: weak telicity markers (i.e., bounded objects) that imply telicity pragmatically but do not require telic interpretations, and strong telicity markers (i.e., resultative particles and adjectives) that overtly specify the completion of the matrix event, rendering readings of incompletion implausible.

The present study aims to shed some light on the acquisition of these telicity markers by early and late acquirers of German. Firstly, I intend to see whether speakers of German from different acquisitional backgrounds understand the aspectual requirements of strong and weak telicity markers and whether they distinguish between the telicity entailments of resultative particle and adjective constructions. Then, I aim to look into any potential differences between Portuguese-native second language (L2) speakers and native speakers of German – i.e., German–Portuguese bilinguals with German as a minority and a majority language, as well as German monolinguals – in their interpretations of the three types of telicity markers. I decided to include monolingual as well as early bilingual acquirers of German as baseline groups to account for the real-world variability of native speakers of German, who, in the current globalized world, comprise diverse speaker profiles (Davies, 2003; Dewaele, 2018; Rothman & Treffers-Daller, 2014; Rothman et al., 2022; Wiese et al., 2022).

Another distinctive element of this study is the fact that L2 acquisition of telicity marking has – to the best of my knowledge – never been tested for Portuguese-native speakers, nor has any such study ever featured a group of Portuguese–German bilinguals with German as a minority language, which in itself can serve as a baseline for further research on previously overlooked phenomena between typologically distinct language groups. The inclusion of two groups of German–Portuguese simultaneous bilinguals with contrasting heritage languages can open the doors to much debate about the nature and extent of bilingual competence and how different language systems interact with each other in the bilingual mind, especially under a completely different set of social circumstances and linguistic experiences. It can also be considered a novel aspect of the present study, given that such comparisons are not particularly common in research.

Still, the reasons to conflate such an eclectic sample of language users are not just research-based and should also be seen as a gateway for the introduction of a new paradigm in acquisition research, in which speakers' knowledge and competence is tested against a myriad of very diverse language backgrounds, representing the real state of current linguistic communities. Identifying the actual development of the speakers' acquisitional processes and the factors leading towards a more or less target-like acquisition of these structures are not the key purposes of this study. For now, our main goal is to detect and interpret potential differences between these types of speakers and hopefully lay the groundwork for new studies targeting the same or similar properties in different languages. To that end, I will be making use of Bayesian inference techniques to derive some distributional tendencies that can be used as prior information in forthcoming research.

2. Theoretical background

2.1. Compositional telicity and telicity markers

One of the primordial features of telicity – that dates back to the seminal work of Verkuyl (1972, 1993) – is its compositionality. In some cases, telicity is inherent to the semantic denotation of the verb (see [1a]), but it can also arise from the combination of the verb with its direct object (see [1b]), in which case the VP becomes telic if the object (e.g., *the book*) is "bounded" (Ramchand, 2008) or "quantized" (Krifka, 1989, 1998), otherwise it remains atelic (e.g., *books*).

1. a) John stole the wallet/ ∅ wallets.
   b) John read the book/ ∅ books.

This phenomenon motivated the distinction between inherent (see [1a]) and compositional telicity (see [1b]), depending on whether the aspectual framing of the situation is determined lexically or grammatically. Compositional telicity can also be obtained via derivation with specific markers. In German, certain resultative particles (e.g., *ab* 'off', *auf* 'up' and *aus* 'out') can be used to emphasize the culmination of an event. These particles serve as overt telicity markers, reinforcing the implication that the event has reached its endpoint when this is not explicit (see [2]).

2. Hans trank Das Bier **aus**.
   John drank the.ACC beer out.PART
   'John drank up the beer.'

This begs the question of why some languages have derivational means to mark telicity if a similar result could be achieved with bounded objects. The answer is twofold: (i) first, resultative particles are endpoint markers in the sense that they merely delimit the event, i.e., any information about the goal or the undergoer of the event must be explicitly determined by an object; (ii) on the other hand, studies have shown that, although both bounded objects and resultative particles are telicity markers in their own sense, they do not have the same aspectual requirements (c.f. Filip, 2014; Jeschull, 2007; Schulz, 2018). In principle, bounded objects signal the culmination of an event, however this culmination can be cancelled based on world knowledge and pragmatic-contextual inference, making them weak telicity markers. Conversely, resultative particles are semantically robust and do not allow endpoint cancellation (see [3]), which makes them strong telicity markers.

3. Hans trank das Bier (#**aus**), aber es ist noch
   John drank the.ACC beer (out.PART) but it is still
   ein bisschen Bier im Glas.
   a bit beer in-the Glass
   'John drank (up) the beer, but there is still a bit of beer in the glass.'

In German, as well as in other Germanic languages, the resulting state associated with the endpoint of a given event can be lexically specified in secondary predication by an adjective. These structures have been commonly known as resultative adjective constructions (Kratzer, 2005; Müller, 2002; Richter & van Hout, 2013). Both resultative particles and adjectives are strong indicators of telicity, the main difference being lexical in nature, i.e., resultative adjectives make the resulting state overt, while particles merely imply it. In adjectival resultatives, the lexical specification of the resulting state nullifies any conversational implicature of endpoint cancellation (see [4]).

4. Heidi hat ihr Glas Wein (#**leer**) getrunken, aber es
   Heidi has her glass wine empty drunk but it
   ist noch Wein in dem Glas.
   is still wine in the glass
   'Heidi drank her glass of wine (#empty), but there is still wine in the glass.'

It is generally understood that these constructions rely on the interaction between several modules of grammar and language use. Most cases of telicity are not determined at a lexical, but rather at a syntactic level, and stem from the combination of syntactic, semantic and even pragmatic properties. Ramchand (2008), for example, has attempted to explain these phenomena from the perspective of event decomposition. The author posits that any given event consists of a projection of an event phrase with varying levels of complexity, according to the event's own selectional and aspectual features. Under this view, accomplishment predicates – consisting of a verb and a quantifiable direct object – are the result of two subevent projections, i.e., an Initiation-Phrase (InitP) and a Process-Phrase (ProcP), in which the target verb selects a PATH argument that can be further specified as [±bounded], depending on the semantic nature of the object (Figure 1).

On the other hand, resultative adjectives and particles are projected in a third subevent phrase, i.e., the Result-Phrase (ResP), and – in case a bounded object cooccurs – this will no longer be the PATH argument of ProcP, but rather takes the semantic roles of the UNDERGOER of ProcP and the RESULTEE of ResP (Figure 2).

The existence of a latent result phrase in the event's description might explain why resultative particles and adjectives are not as aspectually flexible as bounded objects, turning infelicitous any attempt to pragmatically neutralise the event's inherent endpoint.
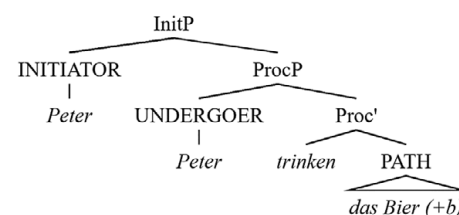

Figure 1. Simplified event structure of *Peter trank das Bier* ('Peter drank the beer'), according to Ramchand's (2008) event-decomposition theory.
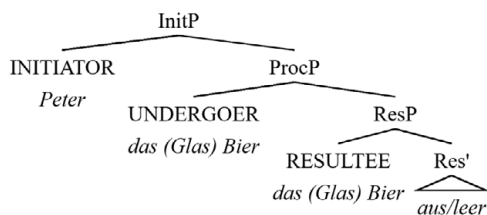
Figure 2. Simplified event structure of *Peter trank das (Glas) Bier aus (leer)* 'Peter drank up the (glass of) beer (empty)', according to Ramchand's (2008) event-decomposition theory.

In practical terms, speakers will only effectively use these strong telicity markers when they aim to make apparent that the event's culmination has been reached, in which case they require the specification of a branched-out result subevent.

In a separate framework, Pustejovsky (1995) categorised predicates with bounded objects as process-oriented – because their process subevent is more prominent – and resultative constructions as endstate-oriented – since they put most emphasis on the event's resulting state. Although methodologically distinct, Pustejovsky's explanation does not deviate considerably from Ramchand's decomposition approach. Noticeably, researchers have resorted to different techniques to try to explain this phenomenon, but they have mostly come to similar conclusions. All in all, there is something in the lexical-semantic composition of events that make them more or less pliable to noncanonical aspectual readings.

Most of the examples we covered in this section pertain to categories of telicity marking in German. Despite the lack of research on telicity markers in European Portuguese (EP), we assume that the patterns of pragmatic telicity observed for bounded objects in other languages also hold for EP. In that sense, weak telicity markers are also very productive in this language, allowing for the same type of conversational implicatures that cancel the event's natural endpoint (as in [5]).

5. O   Paulo comeu **o   bolo**, mas ainda deixou uma
    the   Paulo ate   the cake   but still left   a
    fatia.
    slice
    'Paulo ate the cake, but he still left a slice.'

However, particle markers are non-existent and adjectival resultatives are also not typically licensed by Portuguese grammar, except for very select cases of participial secondary predication (as in [6]; Duarte & Oliveira, 2010).

6. O   pintor pintou a   paisagem **esfumada**.
    the   painter painted the   landscape smoky
    'The painter painted the landscape??smoky.'
                                              (Duarte & Oliveira, 2010, p. 404)

In most cases, the culmination of a given event can be accentuated by introducing certain adverbials or quantifiers to the event's description (as in [7]).

7. A   criança bebeu A   água **até ao   fim/ toda**.
    the   child   drank The water until the end all
    'The child drank up the water/the whole water.'

In the next section, we will look at some of the findings of first and second acquisition of telicity markers, with a focus on bounded object constructions, as well as German verb particles and adjectival resultatives.

2.2. The acquisition of telicity marking

Research has shown that German children acquire the notion of telicity quite early, mainly due to the robust German verb particle system (Schulz & Ose, 2008; Schulz & Penner, 2002; Schulz & Wittek, 2003). Resultative particles, such as *auf* 'up, open' and *aus* 'out, off', are very productive in the early stages of L1 acquisition, even before children start tackling simple verbs and verbal morphology (Behrens, 1998; Dimroth, 2009; Schulz, 2018). This is in line with the findings of many studies suggesting that the notions of completion and boundedness are primordial in L1 acquisition (Andersen & Shirai, 1996; Bardovi-Harlig, 2000; Weist, 2002).

Van Hout (1998) found that the aspectual role of resultative particles is acquired by Dutch and English children even before that of direct objects or determiners, with most 4- to 5-year-olds correctly assigning them a telic interpretation. The author provides support for the *Transparency Principle*, which states that overt linguistic mappings are in principle easier to acquire than "covert" ones (van Hout, 1998, p. 406)[1]. These findings were corroborated for German too (Schulz & Ose, 2008; Schulz & Penner, 2002). As for the role of weak telicity markers, van Hout (1998) found that English-speaking adults allow both telic and atelic readings of bounded object constructions when no opposing contextual cues are provided, while Dutch speakers restrict them to telic interpretations. Schulz and Penner (2002) showed that German children and adults behave in a similar manner regarding bounded objects, fluctuating between telic and atelic interpretations. In a follow-up study, Schulz and Ose (2008) found no substantial differences in interpretation between German children and adults, contrary to previous findings for Dutch (van Hout, 1998, 2000), which suggested that Dutch children are more lenient to accept atelic readings with bounded objects than adults. For some reason, even though German and Dutch are quite similar in structure, German children seem to behave in a more adult-like manner regarding bounded objects than Dutch children, which may indicate that there are certain language-specific factors related to overall language development that modulate these pragmatic-ontological differences between speakers.

Testing 6- to 9-year-old children's interpretation of adjectival resultatives in L1 German, Richter and van Hout (2013) showed that 6-year-olds are sensitive to the syntactic framing of resultatives; however, they are not yet able to fine-tune the selectional restrictions imposed by verbs in these constructions. Conversely, although the younger German children seem to understand the aspectual implications associated with resultatives, their reduced "lexical knowledge" prevents them from correctly deriving the form-meaning combinations allowed by these constructions (Richter & van Hout, 2013, p. 139). The differences in interpretation between children and adults, therefore, do not rely on aspectual mismatching, but rather on their knowledge of semantic verb properties, that develops gradually over time. For example, the property of object affectedness is crucial for determining which verbs can occur in adjectival resultatives. As pointed out in Richter and van Hout (2013, p. 119), in a sentence like *der Tierarzt macht*

---

[1]The role of transparency is not exclusive to the acquisition of aspectual markers. Indeed, a number of studies have shown that transparency plays a decisive role in the acquisition of complex categories across various linguistic areas (Ellis, 2017; Lempert, 1990; Rodina & Westergaard, 2017).

die Tiere gesund 'the veterinary cures the animals', the verb semantics of *machen* 'make' does not express affectedness of the direct object, making it pliable for secondary predication with an adjective. Acquiring this property is an arduous process that takes some time.

In L2 acquisition research, the study of telicity is still relatively sparse. Slabakova (2001) tested Bulgarian speakers of L2 English in their interpretation of telicity markers (i.e., resultative particles, secondary adjective resultatives and double object constructions). The study showed that telicity marking was consistently acquired in the higher proficiency groups, while low intermediate speakers still had some difficulty judging telic situations. Their interpretation of atelic situations, in contrast, was fairly competent. The study also showed that an increase in the L2 speakers' performance was correlated with an increase in their grammatical competence. The author argued that the performance of the low intermediate speakers might be explained by language transfer and a delayed resetting of the "telicity parameter" of their native language (Slabakova, 2001, p. 198). Since Bulgarian – as well as other Slavic languages – does not rely on bounded objects as telicity markers, most learners with lower proficiency automatically assumed that a verb without a particle meant the given situation was atelic, which led to off-target interpretations. This study strongly suggests that L2 speakers whose native language has different telicity parametrizations will start out by accommodating telicity marking in the L2 according to the principles of their L1. This accommodation, however, does not seem fastidious, given that high intermediate to advanced learners show very solid target-like performances.

As for native speakers of Portuguese, there are – to the best of my knowledge – no studies targeting their acquisition of telicity markers in other languages. I assume, however, that Portuguese behaves in a similar way to German and English, in that it allows for both telic and atelic interpretations of bounded object constructions, provided the appropriate pragmatic-contextual cues are present, making such events pliant to aspectual reconfigurations. I also expect Portuguese-native speakers to be more flexible in their interpretations of resultative particles, mainly because the specification of the event's endstate is not as transparent as in resultative adjective constructions.

In short, the properties addressed in this study involve a variety of processes that stem from different linguistic areas, both grammar-internal (e.g., semantics, syntax and morphology) and grammar-external (e.g., pragmatics and discourse)[2]. The existing research on telicity acquisition brings forward some preliminary considerations: (i) L1 speakers are aware of the semantic implications associated with telicity markers from early on, manifesting a clear understanding of the notion of event completion; (ii) they also understand bounded objects as ambiguous and assign them both telic and atelic interpretations; (iii) L2 speakers will start by interpreting telicity according to the parameters available in their L1, and will require sufficient lexical knowledge and contextual cues to understand the aspectual entailments of telicity markers that are not present in their native language.

3. The present study

The aim of the present study is to analyse the interpretation of telicity entailments by different types of adult speakers of German: late L2 learners, early bilinguals and monolingual speakers. Throughout the paper, the L2ers will be described as non-native speakers, while the early bilinguals and monolinguals will be called native speakers of German[3]. First, we will look at the speakers' interpretation of strong telicity markers in contrast to bounded Path DPs. Then, we will see whether speakers make any sort of distinction between particle and adjectival markers. We will also analyse whether there are differences between the groups of speakers, by means of a between-subject analysis. Based on the literature, I put forward the following research questions and the corresponding predictions:

i) Are speakers more lenient to reject cancellation implicatures with strong telicity markers (i.e., resultative particles and adjectives) than with bounded Path DPs? I expect the patterns found in child L1 acquisition (van Hout, 1998, 2000; Schulz & Penner, 2002) to hold for adult speakers as well, i.e., being weak telicity markers, bounded DP objects are expected to be more easily accepted as felicitous with cancellation implicatures than resultative particles and adjectives.

ii) Regarding strong telicity markers, are speakers more lenient to reject cancellation implicatures with resultative adjectives than with resultative particles? Based on the Transparency Principle (van Hout, 1998) and the assumption that lexical saliency plays a prominent role in language processing and acquisition (Ellis, 2017; Lempert, 1990; Rodina & Westergaard, 2017), I assume that late L2 learners of German will have less difficulty assigning telicity to adjectival resultative constructions than to resultative particle constructions, not because of a better acquisition of the former, but because resultative adjectives are more salient than particles and lexicalise the event's final state.

iii) Are there differences between native and non-native speakers in their interpretations of telicity markers? As stated above, L2 speakers are expected to have more difficulty interpreting the telicity entailments of resultative verb particles. They will most likely be driven by lexical transparency to assign completion to adjectival resultatives, being potentially less certain about particle markers, while bilingual and monolingual speakers are expected to treat them both similarly.

Experimental data was collected by means of a Sentence Conjunction Task (SCT, based on Slabakova, 2001). Sociolinguistic information was retrieved using a background questionnaire (LEAP-Q; Kaushanskaya, Blumenfeld & Marian, 2020). The experiment was conducted online via the testing platform Gorilla (gorilla.sc) and comprised an informed consent, followed by the language background questionnaire and the SCT. Statistical analysis was conducted in the RStudio Build 492 software using Bayesian hierarchical ordinal regression models with the *brms* package (Bürkner, 2017).

---

[2]As a reviewer pointed out, some theories like the Interface Hypothesis (Sorace, 2011; Tsimpli & Sorace, 2006) assume that these operations are particularly painstaking for late second language learners. Although we acknowledge the significance of these approaches, the goal of the present study was not to present evidence in favour of any particular theory of acquisition and we will, therefore, refrain from discussing more general hypotheses related to linguistic interfaces.

[3]I am aware of the contentious terminological discussion on the "native vs. non-native" contrast (Dewaele, 2018; Putnam, Kupisch & Pascual y Cabo, 2018). This labelling will be used to simplify data presentation, but the four experimental groups will be considered on their own throughout the analysis.

3.1. Participants

A total of 129 adult speakers were tested ($M_{Age} = 33.95$; $SD_{Age} = 10.6$), who were divided into non-native and native speakers of German, depending on the age of onset of bilingualism and the context of acquisition. Our main experimental group corresponds to the (non-native) L2 speakers, while the native speakers were subdivided into three groups: Portuguese–German bilinguals who were born and/or grew up in Portugal with German as a minority language (i.e., heritage language) and Portuguese as their majority language (MIN); Portuguese–German bilinguals who were born and/or grew up in Germany with German as a majority language and Portuguese as their minority language (MAJ) and German monolinguals (MON). What distinguishes the late L2 learners from these early acquirers is the age of onset of acquisition (after age 12 in the L2 group) and the learning context (the L2 speakers started acquisition in a formal classroom setting)[4]. The proficiency of the L2 speakers was measured using an 11-point self-assessment task incorporated into the background questionnaire. Table 1 reports groupwise descriptive data about the participants.

Before participation, additional screening measures were taken to make sure that the participants fulfilled the desired criteria for inclusion in the study, while still accounting for diverse acquisitional backgrounds. This was obtained via informed consent, which included a detailed description of the desired speaker profile, that had to be virtually signed before participants could start the experiment. The L2 group had to be made up of Portuguese late learners of L2 German with a minimum of 3 years of acquisition, but no restriction was imposed as to whether they had lived or were living in a German-speaking country at the time of testing or whether they only learned German in school and never lived abroad. The Portuguese–German bilinguals were required to be descendants of native speakers of German or Portuguese and to have been born in or moved to Portugal or Germany, respectively, as a child, having, thus, acquired both Portuguese and German either as simultaneous or as early successive bilinguals[5]. German monolinguals were required to have grown up in Germany with exposure to German. Any participants who turned out not to match the desired profiles were removed from the sample.

---

[4]We understand that the decision to not include a standard proficiency test is a significant setback of the present study. Some screening measures (e.g., self-assessment, years of acquisition and language exposure) were implemented to guarantee that the selected L2 speakers would not be elementary users of German. The self-assessment task served as an additional qualifying criterion for this purpose, but it will not be used as a predictor in the analysis. Several studies have confirmed the validity of self-assessment measures to determine speaker proficiency in language acquisition research (Edele, Seuring, Kristen, & Stanat, 2015; Flores, Zhao, & Eira, 2022; Ma & Winke 2019), although these measures should be merely regarded as complementary instruments to traditional proficiency assessment tasks. In a forthcoming paper (Author, Author & Author, in preparation), I hope to focus on the individual differences between speakers and these could reflect a potential effect of proficiency on their overall performance.

[5]As pointed out by a reviewer, there are several criteria to define simultaneous and early sequential bilinguals, and these should be indicated. For this sample, I understand simultaneous bilinguals as bilingual speakers who started acquiring both languages simultaneously, and early sequential bilinguals as bilingual speakers who started learning their majority language shortly after their heritage language. During the recruitment process, both types were conflated in the same groups, because the main purpose for selecting such a heterogeneous speaker sample was to account for this diversity in the German-speaking world and not necessarily to further the discussion on different types of bilinguals, which is not of interest for the present study.

Table 1. Descriptive statistics about the participants' age and self-assessment of the L2 group

	N	Age				Self-assessment			
		Mean	SD	Min	Max	Mean	SD	Min	Max
L2	45	34.2	11.7	20	66	7.5	1.7	3.8	10.3
MIN	29	35.7	12.5	18	58	-	-	-	-
MAJ	34	35.4	8.5	22	51	-	-	-	-
MON	21	28.7	6.6	20	44	-	-	-	-

L2 speakers were recruited with the help of ASPPA (*Association of Portuguese Postgraduates in Germany*) and from the researcher's own social and academic circle. MIN speakers were recruited from the *alumni* communities of the German bilingual schools in Portugal and via divulgation of the study on the professional social network LinkedIn. The MAJ and MON groups were also recruited via LinkedIn. The experimental design and recruitment policy received the formal assent of the Ethics Commission for Research in the Social Sciences and Humanities (CEICSH) of the University of Minho. Participation in the study was completely voluntary and included no monetary compensation.

3.2. Sentence Conjunction Task

The Sentence Conjunction Task aimed at assessing how sensitive speakers are to the aspectual entailments associated with telicity markers. The participants were presented with complex sentences whose felicity they had to judge using a 4-point Likert-type scale (from 0 to 3). The sentences were composed of a main clause and a conjoined coordinated clause introduced by either an adversative (*aber* 'but') or a copulative conjunction (*und* 'and'), which specified that the event described in the main clause was not completed or interrupted. The stimuli were divided into three target conditions: (i) resultative particles (ResP) and (ii) adjectival resultative (ResA), which are infelicitous and (iii) bounded Path DPs (bDP), which are felicitous. A fourth felicitous condition with partitive PPs (PPP) was included as a control, to balance out the number of infelicitous stimuli. In contrast to the other conditions, partitive PP constructions are atelic and the assumption of non-attainment of an endpoint is a natural entailment of these predicates (Filip, 1989; Krifka, 1992).

Recall that resultative particles (ResP) and adjectival resultatives (ResA) overtly identify a culmination point, with resultative adjectives further specifying a lexically salient resulting state. Speakers are, therefore, expected to reject conditions with these structures, since they are incompatible with conversational implicatures of endpoint cancellation. Bounded Path DPs (bDP) also identify a culmination point, but they are weak telicity markers because the result subevent is not overtly specified, which makes them acceptable with cancellation implicatures (for a detailed discussion on event structure, see Ramchand, 2008). Partitive PPs (PPP) do not specify an endpoint, but rather "an attempt at the action" (Broccias, 2001; Frense & Bennett, 1996; Perek & Lemmens, 2010) or the durative property of an atelic event[6], and the subsequent negation

---

[6]In our analysis, partitive PPs encompass prepositional objects specifying an attempt at the action or a subpart of the matrix event denoted by the verb. We use the term "partitive" as a generic designation for these constructions, although they can be subdivided into further groups (Broccias, 2001; Perek & Lemmens, 2010).

of completion is a pleonastic elaboration of the derived predicate (see [8]).

8. Peter   hat   **vom**   **Kaffee**   getrunken,   aber   die
    Peter   has   from-the   coffee   drunk   but   the
    Kaffeekanne ist   noch   fast   voll.
    coffeepot   is   still   almost   full
    'Peter took a few sips of coffee, but the coffeepot is still almost full.'

The task comprised a total of 40 items: 24 experimental items (six per condition) and 16 fillers. The verbs selected for the stimuli were not always consistent between conditions, given that not all verbs restrictively allow constructions of the different paradigms. The selection of the target verbs relied on the criteria of frequency adopted in previous studies (Slabakova, 2001). Table 2 gives examples of the verbal predicates used in the SCT and their expected acceptability ratings (for a complete list of the items, see Table S1 in the Supplementary Materials).

The SCT was administered online in the Gorilla platform after the language background questionnaire. On the opening screen, the participants were told that sentences would show up one at a time and they had to rate them using a 4-point differential scale from 0 to 3, in which 0 meant "does not make sense" and 3 meant "makes a lot of sense". The scale was presented as a response slider; as the participants moved the slider along the scale, the tooltip showed the numerical value about each position. Each sentence had a time limit of 30 seconds to be judged, after which the screen would skip to the next item and the response would count as missing. Both the instructions and the labels for the scale were given in German; the order of the sentences was randomized for each participant.

As a first step, the items of the SCT were coded as ordered factors with four levels (0, 1, 2 and 3) and missing values, which were excluded from the analysis. In the descriptive analysis, the response values were counted per condition and group. Thereafter, the

Table 2. Examples of the verbal predicates used in the SCT with expected acceptability ratings

Condition	Verbal predicate	Meaning	Expected ratings
Resultative particle (ResP)	*das Wasser austrinken* *den Fisch aufessen* *die Zeitschrift durchlesen*	'to drink up the water' 'to eat up the fish' 'to read the whole magazine'	0 = does not make sense or 1 = makes little sense
Resultative adjective (ResA)	*das Glas leer trinken* *den Teller leer essen* *den Tisch trocken wischen*	'to drink the glass empty' 'to eat the plate empty' 'to wipe the table dry'	
Bounded DP (bDP)	*den Kaffee trinken* *die Pizza essen* *den Boden fegen*	'to drink the coffee' 'to eat the pizza' 'to sweep the floor'	2 = makes sense or 3 = makes a lot of sense
Partitive PP (PPP)	*vom Kaffee trinken* *von dem Fisch essen* *aus dem Album hören*	'to take a sip of the coffee' 'to eat *from* the fish' 'to hear (songs) from the album'	

proportion of each response type was calculated to ease visualization of the data distribution. For the inferential statistical analysis, the dataset was transformed into long format, in which each row described a single observation, and a column was created for the response variable, i.e., the corresponding acceptability ratings on the ordinal scale.

4. Results

4.1. Descriptive analysis

This section reports the descriptive statistics of the task per group of speakers. The SCT is essentially a felicity judgment task with an ordinal rating scale. Each rating on the scale represents an abstract level of acceptability (0 = does not make sense, 1 = makes little sense, 2 = makes sense, 4 = makes a lot of sense). Data tidying and visualization were performed in the RStudio software, using the *tidyverse* (Wickham et al., 2019) and *ggplot2* (Wickham, Chang & Wickham, 2016) packages. Figure 3 reports the proportions of the acceptability ratings per condition for each experimental group.

Figure 3 confirms that the speakers' intuitions about the conditions are congruent with their expected acceptability ratings (see Section 3.2). Conditions ResP (resultative particles) and ResA (resultative adjectives) were assigned lower values on the scale, relative to conditions bDP (bounded DPs) and PPP (partitive PPs). The generally similar acceptability rates of bDP and PPP are indicative that the latter worked as a control condition the way it was expected to, showing that the participants understood the task. Speakers also appear to be less accepting of cancellation implicatures with adjectival resultatives than with resultative particles, as seen by the higher density of 0-responses (L2: 74% [ResA] vs. 51% [ResP]; MIN: 80% [ResA] vs. 60% [ResP]; MAJ: 80% [ResA] vs. 66% [ResP]; MON: 76% [ResA] vs. 67% [ResP]). The differences between these two conditions, however, do not seem substantial. Both non-native and native speakers of German are also evidently more accepting of bounded Path DPs than both particles and adjectival markers, showing a much higher proportion of three-responses in this category (L2: 70%; MIN: 76%; MAJ: 66%; MON: 71%).

4.2. Bayesian analysis

Our statistical analysis will focus on Bayesian inference, rather than on the more traditional frequentist approach. One point of distinction is that Bayesian modelling assigns probabilities to hypotheses and reports a posterior distribution, which is a compromise between prior knowledge about the parameters and the data at hand (for a more elaborate discussion on the Bayesian-frequentist debate, see Bayarri & Berger, 2004). Some of the advantages of Bayesian statistics are: (i) its natural expression of uncertainty (through the estimation of a distribution with credible intervals (or highest-density intervals, HDIs) instead of a single significance value); (ii) its ability to integrate prior information and (iii) its modelling flexibility (Kruschke, 2021; McElreath, 2020; Schad, Betancourt & Vasishth, 2021). Therefore, the main goals of this analysis were to quantify effect sizes and to determine whether there is convincing evidence for the existence of such effects.

Given the ordinal nature of our response variable, the 4-point rating scale was treated as an ordered factor with four distinct categories and flexible thresholds. An *ordered probit model* was applied to describe the acceptability ratings of the SCT. This model assumes the existence of a normally distributed population and a
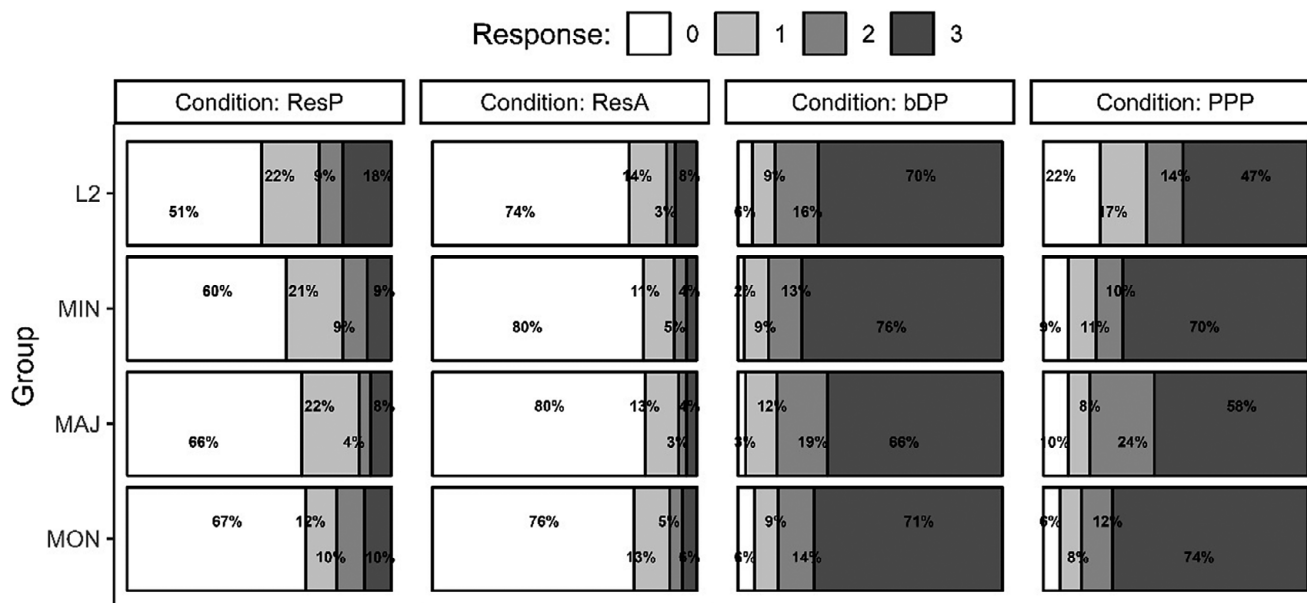
Figure 3. Acceptability ratings (%) in the SCT per condition and group.

latent continuous variable, that underlies the ordinal scale (for a more elaborate explanation of ordered probit models, see Liddell & Kruschke, 2018; Veríssimo, 2021).

I fitted a mixed-effects regression model with an interaction between Condition and Group, as well as a "maximal" random-effects structure, which included by-participant random slopes for Condition and by-item random slopes for the interaction between Condition and Group[7]. To ease interpretation, our categorical predictors were sum-coded (e.g., −0.25, 0.75, −0.25 and −0.25; see Schad, Vasishth, Hohenstein & Kliegl, 2020), and ResP and L2 were set as the reference levels for Condition and Group, respectively. Model convergence was checked using the R-hat, Bulk_ESS and Tail_ESS diagnostics (Gelman, Hill & Vehtari, 2020) and plotted using the *mcmc_trace()* function of the *bayesplot* package (Gabry et al., 2019). Posterior predictive checks confirmed that the simulated data generated by the model suitably mimicked observed data (see Supplementary Materials for the full summary statistics and posterior predictive checks).

For prior specification, I chose to use mildly informative priors that would not overwhelm our posterior distributions and coerce the model into calculating unrealistic effects (for a discussion on the specification of priors, see Gelman, Jakulin, Grazia Pittau & Su, 2008; Nicenboim, Schad & Vasishth, 2022). Our priors followed a normal distribution of $N(1, 2)$ for the model intercept and of $N(0, 1)$ for the slopes. The priors for the random effects followed a normal distribution of $N(0, 0.5)$ for the random intercepts and of $N(0, 0.25)$

for the random slopes. For the correlation matrices, I used the standard regularizing prior LKJ(2).

To determine whether there is evidence for or against the estimated effects, I computed Bayes factors (BF) using the Savage-Dickey density ratio method, in which the ratio of the heights of the posterior and the prior distributions were calculated at the point hypothesis (i.e., zero) for each target comparison. Interpretation of Bayes factors relied on Lee and Wagenmakers' (2013) scale, in which a BF_{10} greater than 1 indicates evidence for the alternative hypothesis (H_1), i.e., values between 1 and 3 indicate anecdotal evidence in favour of an effect, between 3 and 10 moderate evidence, between 10 and 30 strong evidence, between 30 and 100 very strong evidence and over 100 extreme evidence. In opposition, a BF_{10} below 1 indicates evidence against H_1, i.e., values between 1 and 0.3 are considered anecdotal evidence, between 0.3 and 0.1 moderate evidence and below 0.1 strong evidence. Given that the Bayes factor is sensitive to prior information, I also performed sensitivity analyses using a range of different prior standard deviations to determine how different priors could influence the results.

For parsimony, I will only report the posterior distributions of our target conditions, i.e., resultative particles, resultative adjectives and bounded DPs (the estimates for partitive PPs can be retrieved from Table S2 in the Supplementary Materials).

4.2.1. Main effects of condition and group
Figure 4 plots the posterior distributions of the target main effects as estimated by our model[8]. The main effects are estimated across the levels of the other variable (for example, the effects of Group are

---

[7]Including by-participant random slopes for Condition will allow the effects of this predictor to vary for each speaker. In the same way, by-item random slopes of Condition and Group (and their interaction) allows the explanatory variables to have different effects in each item. Sometimes, we can run "simpler" models if the random slopes do not lead to an improvement (Ciaccio & Veríssimo, 2022), however, in this study, we will make use of the more "conventional" approach by fitting a maximal model (Barr, Levy, Scheepers & Tily, 2013).

[8]Note that not all of the posterior distributions reported in the figures were available in the model output. The model only prints the contrasts specified for each predictor. To avoid refitting it with different contrasts, I computed the missing estimates manually by extracting the model's posterior draws and performing the appropriate algebraic operations between the parameters (see R code).
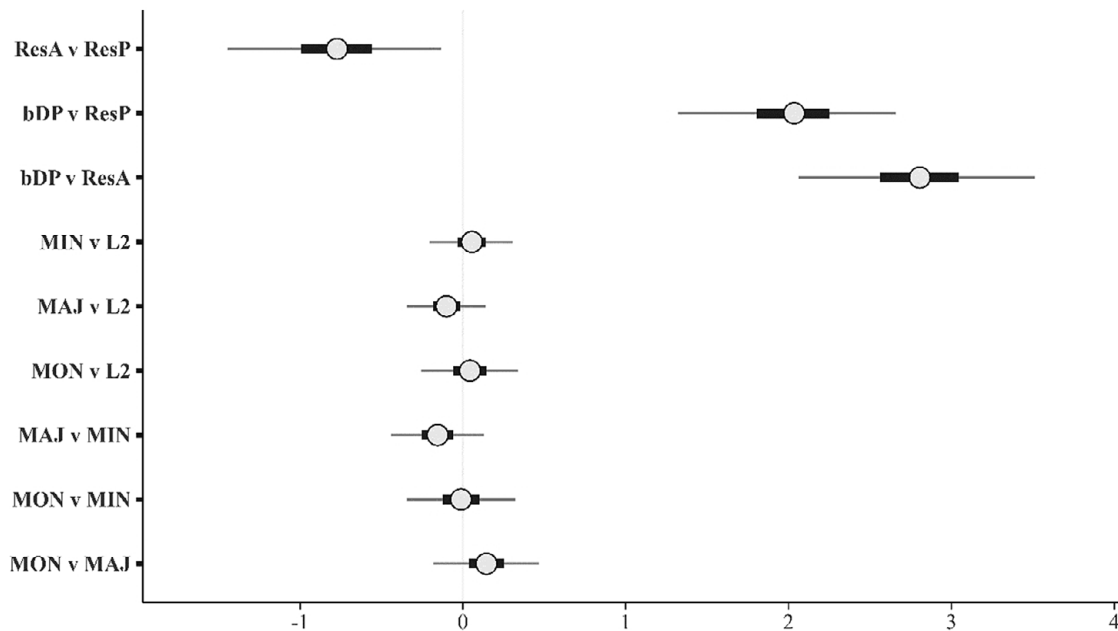
Figure 4. Posterior distributions of the target main effects of Condition and Group, plotted in log space. The circle in the middle represents the mean estimate of the distribution and the lines illustrate the credible intervals (CrIs). The thick line represents the 50% CrI and the thin line represents the 95% CrI.

estimated across the four conditions taken together). Positive values specify an increase in log-odds of the estimated parameter (relative to the reference level), while negative values indicate a decrease[9].

Regarding the first research question, the posterior estimate of the difference between bounded DPs and particles was 2.02 [1.32, 2.66], with the entirety of the probability mass covering positive values. A Bayes factor of 1×10^{17} indicated extreme evidence for a difference between bDP and ResP. As for the difference between DPs and adjectival resultatives, the posterior estimate was 2.80 [2.06, 3.51]. A Bayes factor of 5×10^{19} indicated extreme evidence in favour of a difference between bDP and ResA. A sensitivity analysis showed that the evidence in favour of a difference between bounded DPs and strong telicity markers did not decrease substantially with wider prior standard deviations (see Figure S2 in the Supplementary Materials). With respect to the second research question, the posterior mean of the difference between resultative adjectives and particles was −0.78 log-odds with its 95% CrI ranging from −1.44 to −0.13. Although the credible interval was relatively wide, it also contained values consistent with a predictive effect. A Bayes factor of 5.66 indicated moderate evidence in favour of a difference between these two conditions. A sensitivity analysis using different prior standard deviations showed that this Bayes factor was sensitive to prior information, with both smaller and larger priors decreasing the support for a difference between resultative adjectives and particles (see Figure S2 in the Supplementary Materials).

As for the effects of Group, our posterior distributions are clustered together around zero with very narrow credible intervals, suggesting that all groups performed quite similarly across conditions. The posterior estimates were 0.05 [−0.20, 0.31] for the difference between MIN and L2 ($BF_{10} = 0.14$), −0.10 [−0.34,

0.14] for the difference between MAJ and L2 ($BF_{10} = 0.17$), and 0.04 [−0.26, 0.34] for the difference between MON and L2 ($BF_{10} = 0.16$). The reported Bayes factors indicated moderate evidence against a difference between each group of native speakers and the L2 group. As for the comparisons between groups of native speakers, the posterior estimates were − 0.15 [−0.44, 0.13] for the difference between MAJ and MIN ($BF_{10} = 0.18$), −0.01 [−0.35, 0.32] for the difference between MON and MIN ($BF_{10} = 0.12$), and 0.14 [−0.18, 0.47] for the difference between MON and MAJ ($BF_{10} = 0.17$). The Bayes factors indicated moderate evidence against differences between native speakers. A sensitivity analysis using wider prior standard deviations increased the support against H_1.

4.2.2. Interactions between condition and group
To determine whether the differences between conditions are substantially larger or smaller in one group relative to the others, I looked more closely at the interaction parameters of our model (see Figure S3 in the Supplementary Materials for the posterior distributions of the interactions between groups and the target conditions). At first glance, there seems to be no indication of potentially relevant interactions. The posterior distributions have very wide intervals, covering both negative and positive values. For example, the difference between ResA and ResP in the MIN versus L2 comparison yielded a mean posterior estimate of 0.03 with the 95% CrI ranging from −0.42 to 0.46, suggesting that the difference between adjectives and particles is similar in both groups.

A series of Bayes factors showed that most of the interaction parameters yielded no evidence in favour of a predictive effect, indicating either anecdotal or moderate evidence against H_1 (see Table S3 in Supplementary Materials). However, the posterior estimate for the interaction between bDP vs. ResP and MIN vs. L2 was 0.54 [0.05, 1.04]. A Bayes factor of 2.80 indicated anecdotal evidence in favour of an effect. Similarly, the interaction between bDP vs. ResP and MAJ vs. L2 rendered a posterior estimate of 0.48 [0.01, 0.96], with a Bayes factor of 1.80 also suggesting

---

[9]For reference, the zero in our distribution represents chance (50%). Values above or below zero indicate that a potential predictive effect is either above or below chance. This will come in handy for the interpretation of our figures, where zero is marked with a vertical line.
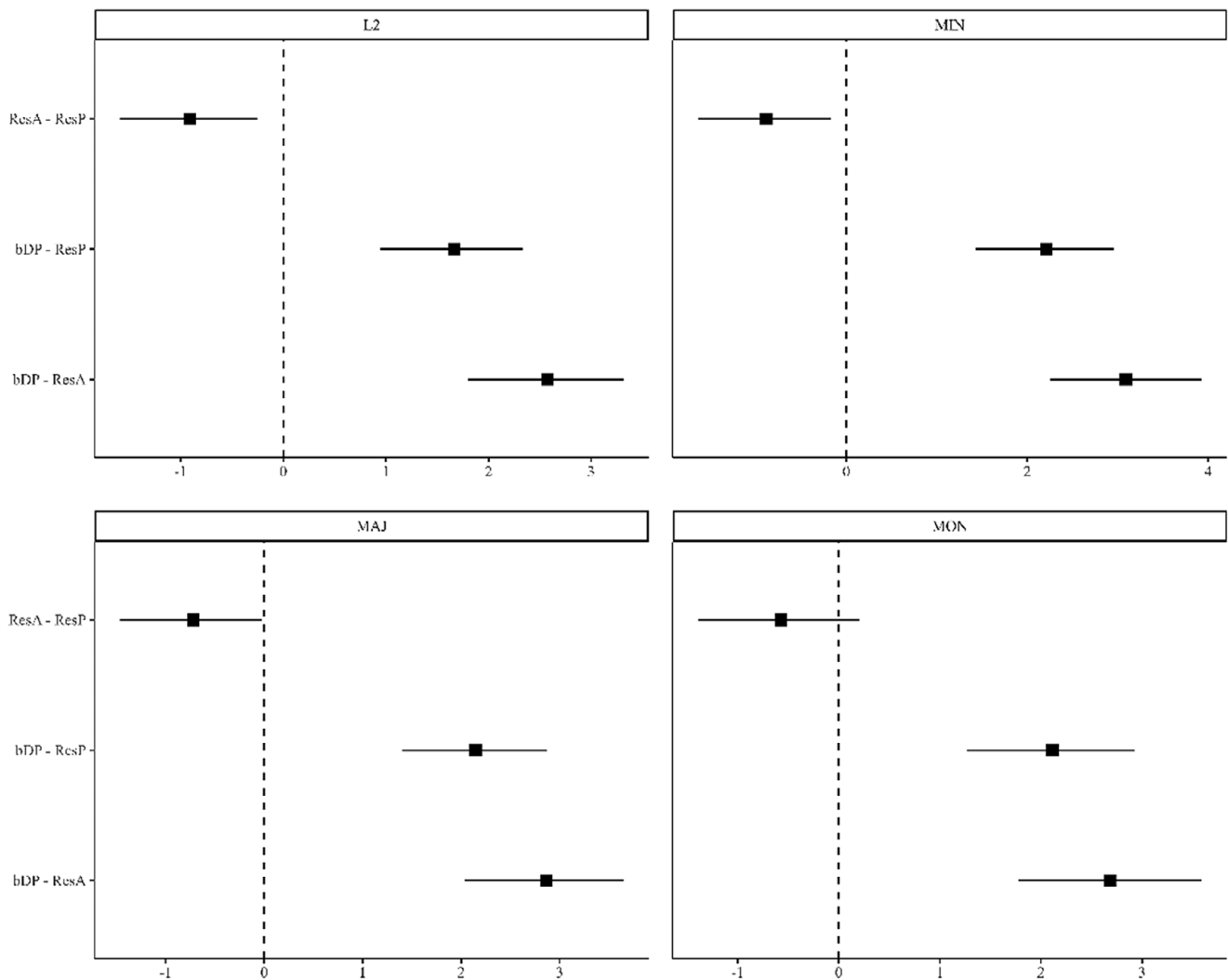
Figure 5. Forest plots reporting the posterior distributions of the nested effects of Condition.

anecdotal evidence for an effect. Only the interaction between bDP vs. ResP and MON vs. L2 yielded anecdotal evidence against an effect (0.45, [−0.14, 1.05], BF10 = 0.96). A sensitivity analysis using different prior standard deviations confirms that the evidence for these interactions decreases with wider less informative priors (see Figure S4 in Supplementary Materials). To further investigate these potential differences and attempt to answer the third research question concerning the differences between native and non-native speakers in all target conditions, we will now estimate the differences between conditions nested by Group[10].

4.2.3. Nested effects of condition by group

The posterior estimates for the nested effects of Condition were calculated using the *emmeans* R package (Lenth, 2022) and the Bayes factor analyses were conducted with *bayestestR* (Makowski, Ben-Shachar & Lüdecke, 2019). Figure 5 plots the posterior distributions of the target parameters.

---

[10]We will only be focusing on the effects of Condition nested by Group, and not on the effects of Group nested by Condition. The reason for that is that, frequently, between-group comparisons can be related to more general individual differences (e.g., biases, response trends, etc.).

Overall, the differences between bDP and both ResP and ResA have very similar effect sizes in all groups. The centre of the distribution deviates from zero and the credible intervals cover only positive values, which is indicative of an increase in acceptability. The Bayes factor analyses indicated extreme evidence in favour of these differences in all groups (see Table S4 in Supplementary Materials for the posterior distributions and respective Bayes factors). The subsequent sensitivity analysis confirmed that the evidence for a difference between bounded DPs and both particles and adjectives does not change substantially with larger prior standard deviations (see Figure S5 in Supplementary Materials).

On the other hand, the posterior estimates for the difference between ResA and ResP were quite different across groups. In general, most of the probability mass covered negative values, consistent with a decrease in acceptability. However, the evidence decreases as we go from L2 to MON. The posterior estimate for L2 was −0.92 [−1.59, −0.25], with a Bayes factor of 11.94 indicating strong evidence in support of a predictive effect. As for MIN, the posterior estimate was −0.89 [−1.64, −0.17], and a Bayes factor of 5.72 indicated moderate evidence in favour of a difference. In the MAJ group, the posterior estimate was −0.72 [−1.47, −0.03] and the evidence decreased once again, with a Bayes factor of 2.12

suggesting moderate evidence in favour of H_1. Finally, the MON group rendered a posterior estimate of -0.58 $[-1.39, 0.20]$, with a Bayes factor of 0.87 indicating anecdotal evidence against an effect. Our sensitivity analysis showed that, across all groups, the evidence in support of H_1 decreased substantially with weakly informative priors, particularly in the groups of native speakers (see Figure S5 in Supplementary Materials).

5. Discussion

The present study sought to address how speakers of German from different acquisitional backgrounds understand the telicity entailments of specific telicity markers. The participants were required to judge the acceptability of complex sentences in which an overtly defined endpoint was cancelled by means of pragmatic implicatures. We have seen that bounded DPs are weak telicity markers, in that they allow for endpoint annulment, while strong telicity markers (i.e., resultative particles and adjectives) do not. By analysing the participants' acceptability judgments using Bayesian techniques, we were able to predict their response patterns on the basis of differences between conditions and groups. In this section, I will attempt to answer the research questions in accordance with our results.

Our first research question concerned itself with potential differences between strong and weak telicity markers. Based on the existing research, we postulated that speakers, in general, would be less accepting of resultative adjectives and particles with endpoint cancellation than of bounded direct objects (van Hout, 1998; Schulz & Ose, 2008; Schulz & Penner, 2002). Our main effects analysis confirmed that speakers disapprove endpoint cancellation with resultative particles and adjectives, while they have no problems accepting bounded object events whose culmination was halted or interrupted. This shows that, regardless of more focused differences, speakers tend to understand that resultative particle and adjectival constructions are intrinsically linked with an aspectually determined boundary, which is not reversible by any pragmatic means. By inspecting these differences further, our nested-effects analysis showed that this pattern applies to every group of speakers individually, i.e., both non-native and native speakers show higher acceptability of bounded DPs relative to both types of strong telicity markers. To no surprise, these findings corroborate the theoretical proposals put forward by previous research (Putstejovsky, 1995; Ramchand, 2008), i.e., it is quite likely that the aspectual (im)permeability of such events rests on their internal structure, rather than on mere discourse-pragmatic factors.

Regarding our second research question, we wanted to determine whether lexical transparency would play a role in the speakers' interpretation of telicity entailments with strong aspectual markers. We wondered about whether resultative adjectives – being that they overtly specify a consequent state – would be less acceptable for speakers than resultative particles, whose resultativity is not as salient. Our main effects analysis showed that there is a general tendency for speakers to reject endpoint cancellation with resultative adjectives relative to resultative particles, with a considerable amount of evidence in favour of this difference. This suggests that speakers are overall less tolerant about the aspectual flexibility of these adjectival constructions, which must be due to the overt specification of their resulting state. While a cancellation implicature of a telic event marked by a resultative particle may yield some reprocessing of the state-of-affairs, it is unlikely that such a proposition with a resultative adjective would require much revaluation, since it mainly constitutes a very noticeable semantic contradiction

(e.g., *Samuel hat die Hausarbeit **fertig** geschrieben, und morgen schreibt er sie **weiter***, 'Samuel has *finished* writing the school paper and he will *continue* writing it tomorrow'). It is quite apparent that these observations are not categorical and allow for various degrees of variation even among native speakers, which shows how complex these structures are and how far we probably are from reaching a comprehensive understanding of the phenomena under study. Theoretically, both types of strong telicity markers should be interpreted in the same manner, but speakers show a clear leniency towards accepting resultative particle constructions with a cancelled endpoint, a disposition that is not extended to their adjectival counterparts.

Our third and final research question concerned itself with potential differences between groups regarding the speakers' interpretations of telicity markers. As expected, the results showed that non-native and native speakers are not substantially different regarding their interpretations of bounded DPs, i.e., they tend to be much more accepting of bounded objects with a cancelled endpoint than of resultative particles and adjectives in the same contexts. This shows that all speakers understand these objects as weak telicity markers, allowing for aspectual reinterpretations of the main event when the appropriate contextual framing is specified.

However, when comparing the participants' acceptability of resultative adjectives *versus* resultative particles, non-native speakers are sensibly different from the groups of native speakers. Our nested effects analysis shows that L2 speakers are more rejecting of adjectival constructions than of resultative particle structures in contexts of endpoint cancellation. As mentioned before, this trend may be justified by van Hout's (1998) Transparency Principle, in which the overt specification of a consequent state may move the speakers away from accepting these sentences, not necessarily because of their sensitivity to aspectual mismatches, but rather due to the expression of a conspicuous contradiction between clauses. It could also be the case that L2 speakers have not yet fully acquired the aspectual intricacies of resultative particle verbs, which leads them to interpret their telicity more flexibly. Another explanation could rest on the speakers' individual perceptions of the target items, i.e., if the verb used allows for a more pliable interpretation of the end state, they may be more lenient to evaluate the state-of-affairs in a "non-target-like" way, as in the case of, e.g., *abzeichnen* 'to draw' versus *aufessen* 'to eat up' (a future study targeting the speakers' individual differences should provide clearer answers to these questions; see Author, Author & Author, in preparation). Still, the evidence in favour of this trend decreases as we move from non-native to native speakers. In other words, while L2 speakers are more opposed to cancellation with resultative adjectives than with particles, this distinction is not as salient for bilinguals and monolinguals.

If we were to assume that this decrease in evidence, as we go from L2 to native speakers, is scientifically motivated, could we be observing an "acquisitional continuum" (L2 → MIN → MAJ → MON), in which a more target-like acquisition of German results in a stricter interpretation of particle markers? As a reviewer pointed out, telicity entailments seem to be acquired very early and should not be dependent on extensive exposure to the language, however, given the semantic-conceptual nature of this phenomenon, one should not expect it to be completely impermeable to the influence of the second L1's conceptual system. If that is indeed the case, a more fine-grained research methodology is required to concretely identify what potential (external) factors (e.g., language exposure, type and quantity of input) may be playing a role here. Given the limitations of our current study, we hope that these open questions

motivate researchers to target similar properties in future studies, especially given that linguistic research in the last decades has tended to deviate from meaning-oriented approaches, which has taken a toll on the resources we have available to investigate semantic-pragmatic phenomena across different languages.

6. Conclusions

A sentence conjunction task investigated the acceptability patterns of speakers of German from different acquisitional settings regarding the telicity entailments of three types of markers. We found that both native and non-native speakers follow the theoretical assumptions of previous research, in that they are more lenient to accept endpoint cancellation with bounded direct objects, while they find such implicatures with adjectival and particle markers infelicitous. Our analysis also suggested that the telicity values associated with adjectives and particles may not be intrinsically similar, at least from the perspective of the speakers' intuitions. There is substantial evidence indicating that non-native speakers have more difficulty understanding the telicity entailments of resultative particle constructions than those of adjectival resultatives. I argue that this difference is not due to a lack of familiarity of one structure relative to the other, but rather to the lexical transparency of adjectival markers as identifiers of a resulting state. There also appears to be some variation regarding this difference in the groups of bilingual speakers, but the evidence is not sufficiently robust for us to assume that their understanding of strong telicity markers deviates from that of monolinguals. Ideally, the present study can serve as theoretical input for future research targeting similar properties, especially if the authors seek to implement Bayesian inference in their analyses, for which prior information can be directly derived from our results.

References

Andersen, R. W., & Shirai, Y. (1996). The primacy of aspect in first and second language acquisition: The pidgin-creole connection. In W. C. Ritchie & T. J. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 527–570). Academic Press.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Blackwell.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278. 10.1016/j.jml.2012.11.001.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, **19**(1), 58–80.

Behrens, H. (1998). How difficult are complex verbs? Evidence from German, Dutch and English, *Linguistics*, **36**(4), 679–712.

Borer, H. (2005). *Structuring sense: Volume 2: The normal course of events*. Oxford University Press.

Broccias, C (2001). Allative and ablative at-constructions. CLS, 37(1), 67-82. In M. Adronis, C. Ball, E. Heide, & S. Neuvel (Eds.), *CLS 37: The main session. Papers from the 37th meeting of the Chicago Linguistic Society* (pp. 67–82). Chicago Linguistic Society.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, **80**, 1–28.

Ciaccio, L. A., & Veríssimo, J. (2022). Investigating variability in morphological processing with Bayesian distributional models. *Psychonomic Bulletin & Review*, **29**(6), 2264–2274. 10.3758/s13423-022-02109-w.

Davies, A. (2003). *The native speaker: Myth and reality*. Multilingual Matters. 10.21832/9781853596247.

Dewaele, J. M. (2018). Why the dichotomy "L1 versus LX user" is better than "native versus non-native speaker". *Applied Linguistics*, **39**, 236–240. 10.1093/applin/amw055.

Dimroth, C. (2009). Stepping stones and stumbling blocks. Why negation accelerates and additive particles delay the acquisition of finiteness in German. In C. Dimroth, & P. Jordens (Eds.), *Functional categories in learner language* (pp. 137–170). Mouton de Gruyter.

Dowty, D. R. (1979). *Word meaning and montague grammar*. Reidel.

Duarte, I., & Oliveira, F. (2010). Particípios resultativos. In A. M. Brito, M. F. H. Silva, J. Veloso, & A. Fiéis (Eds.), *Textos Seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística* (pp. 397–408). APL.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, **52**, 99–123.

Ellis, N. C. (2017). Salience in usage-based SLA. In S. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 21–40). Routledge.

Filip, H. (1989). Aspectual properties of the an-construction in German. In W. Abrahamand & T. Janssen (Eds.), *Tempus – Aspekt – Modus. Die lexikalischen und grammatischen Formen in den germanischen Sprachen* (pp. 259–292). Niemeyer.

Filip, H. (2014). Weak and strong telicity via maximalization. In Paper presented at workshop on identification and delimitation of events, Paris.

Flores, C., Zhao, C., & Eira, C. (2022). "I no longer count in German." On dominance shift in returnee heritage speakers. *Applied Psycholinguistics*, 1–25. 10.1017/S0142716422000261.

Frense, J., & Bennett, P. (1996). Verb alternations and semantic classes in English and German. *Language Sciences*, **18**(1–2), 305–317.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society*, **182**(2), 389–402. 10.1111/rssa.12378.

Gelman, A., Jakulin, A., Grazia Pittau, M., & Su, Y. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, **2**(4), 1360–1383.

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

Jeschull, L. (2007). The pragmatics of telicity and what children make of it. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd conference on generative approaches to language acquisition North America* (pp. 180–187). Cascadilla Proceedings Project.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The language experience and proficiency questionnaire (leap-q): Ten years later. *Bilingualism: Language and Cognition*, **23**(5), 945–950.

Kratzer, A. (2005). Building resultatives. In C. Maienborn & A. Wöllstein (Eds.), *Event arguments: Foundations and applications* (pp. 177–212). Niemeyer.

Krifka, M. (1989). Nominal reference, temporal constitution, and quantification in event semantics. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expression* (pp. 75–115). Foris.

Krifka, M. (1992). In I. A. Sag & A. Szabolcsi (Eds.), *Thematic relations and links between nominal reference and temporal constitution, lexical matters* (pp. 29–53). CSLI Publications.

Krifka, M. (1998). The origins of telicity. In S. Rothstein (Ed.), *Events and grammar* (pp. 197–235). Kluwer. 10.1007/978-94-011-3969-4_9.

Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour*, **5**(10), 1282–1291.

Lee, M., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lempert, H. (1990). Acquisition of passives: The role of patient animacy, salience, and lexical accessibility. *Journal of Child Language*, **17**(3), 677–696.

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means.* R package version 1.7.3. https://CRAN.R-project.org/package=emmeans

Liddell, T. M., & **Kruschke, J. K.** (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, **79**, 328–348.

Ma, W., & **Winke, P.** (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals*, **52**(1), 66–86.

Makowski, D., **Ben-Shachar, M. S.**, & **Lüdecke, D.** (2019). Bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, **4**(40), 1541. 10.21105/joss.01541.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Chapman and Hall/CRC. 10.1201/9780429029608.

Müller, S. (2002). *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German.* CSLI Publications.

Nicenboim, B, **Schad, D.**, & **Vasishth, S.** (2022). *An introduction to Bayesian data analysis for cognitive science.* Published online via bookdown on September 12, 2022. https://vasishth.github.io/bayescogsci/book/

Perek, F., & **Lemmens, M** (2010). Getting at the meaning of the English at-construction: The case of a constructional split. *CogniTextes*, **5**. 10.4000/cognitextes.331. Published online by Open Edition Journals on December 20, 2010.

Pustejovsky, J. (1995). *The generative lexicon.* The MIT Press.

Putnam, M. T., **Kupisch, T.**, **Pascual, Y.**, & **Cabo, D.** (2018). Different situations, similar outcomes: heritage grammars across the lifespan. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 251–280). John Benjamins.

Ramchand, G. (2008). *Verb meaning and the lexicon: A first-phase syntax.* Cambridge University Press.

Richter, M., & **van Hout, R.** (2013). Interpreting resultative sentences in German: Stages in L1 acquisition. *Linguistics*, **51**(1), 117–144.

Rodina, Y., & **Westergaard, M.** (2017). Grammatical gender in bilingual Norwegian-Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition*, **20**(1), 197–214.

Rothman, J., & **Treffers-Daller, J.** (2014). A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics*, **35**, 93–98. 10.1093/applin/amt049.

Rothman, J, **Bayram, F**, **DeLuca, V**, **Di Pisa, G**, **Duñabeitia, JA**, **Gharibi, K**, **Hao, J**, **Kolb, N**, **Kubota, M**, **Kupisch, T**, **Laméris, T**, **Luque, A**, **van Osch, B**, **Soares, SMP**, **Prystauka, Y**, **Tat, D**, **Tomić, A**, **Voits, T.**, & **Wulff, S** (2022). Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *Applied Psycholinguistics*, 1–14. 10.1017/S0142716422000315. Published online by Cambridge University Press on November 11, 2022.

Schad, D. J., **Vasishth, S.**, **Hohenstein, S.**, & **Kliegl, R.** (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, **110**, 104038. 10.1016/j.jml.2019.104038.

Schad, D. J., **Betancourt, M.**, & **Vasishth, S.** (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, **26**(1), 103–126. 10.1037/met0000275.

Schulz, P. (2018). Telicity in typical and impaired acquisition. In K. Syrett & S. Arunachalam (Eds.), *Semantics in language acquisition* (Vol. **24**, pp. 123–150). John Benjamins.

Schulz, P., & **Ose, J.** (2008). Semantics and pragmatics in the acquisition of telicity. In *Paper presented at 29. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS)*, Bamberg.

Schulz, P., & **Penner, Z.** (2002). How you can eat the apple and have it too: Evidence from the acquisition of telicity in German. In J. Costa & M. J. Freitas (Eds.), *Proceedings of the GALA 2001 conference on language acquisition* (pp. 239–246). APL.

Schulz, P., & **Wittek, A.** (2003). Opening doors and sweeping floors: What children with specific language impairment know about telic and atelic verbs. In B. Beachley, A. Brown, & F. Colin (Eds.), *Proceedings of the 27th annual Boston University conference on language development* (Vol. **2**, pp. 727–738). Cascadilla Press.

Schulz, P., **Wymann, K.**, & **Penner, Z.** (2001). The early acquisition of verb meaning in German by normally developing and language impaired children. *Brain and Language*, **77**, 407–418. 10.1006/brln.2000.2411.

Slabakova, R. (2001). *Telicity in the second language.* John Benjamins.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, **1**(1), 1–33. 10.1075/lab.1.1.01sor.

Tenny, C. (1994). *Aspectual roles and the syntax-semantics interface.* Kluwer.

Travis, L. D. (2010). *Inner aspect.* Springer.

Tsimpli, L., & **Sorace, A.** (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th annual Boston-University conference on language development* (pp. 653–664). Cascadilla Press.

van Hout, A. (1998). On the role of direct objects and particles in learning telicity in Dutch and English. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd annual Boston University conference on language development* (pp. 397–408). Cascadilla Proceedings Project.

van Hout, A. (2000). Event semantics in the lexicon-syntax interface. In C. Tenny & J. Pustejovsky (Eds.), *Events as grammatical objects* (pp. 239–282). CSLI Publications.

van Hout, A. (2008). Acquiring telicity cross-linguistically: On the acquisition of telicity entailments associated with transitivity. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 255–278). Routledge.

Veríssimo, J. (2021). Analysis of rating scales: A pervasive problem in bilingualism research and a solution with Bayesian ordinal models. *Bilingualism: Language and Cognition*, **24**(5), 842–848.

Verkuyl, H. (1972). *On the compositional nature of the aspects.* Reidel.

Verkuyl, H. (1993). *A theory of aspectuality.* Cambridge: Cambridge University Press.

Weist, R. M. (2002). The first language acquisition of tense and aspect: A review. In R. Salaberry & Y. Shirai (Eds.), *Tense–aspect morphology in L2 acquisition* (pp. 21–78). John Benjamins.

Wickham, H., **Chang, W.**, & **Wickham, M. H.** (2016). Package 'ggplot2'. *Create Elegant Data Visualisations Using the Grammar of Graphics*. *Version*, **2**, 1–189.

Wickham, H., **Averick, M.**, **Bryan, J.**, **Chang, W.**, **McGowan, L. D. A.**, **François, R.**, **Grolemund, G.**, **Hayes, A.**, **Henry, L.**, **Hester, J.**, **Kuhn, M.**, **Pedersen, T. L.**, **Miller, E.**, **Bache, S. M.**, **Müller, K.**, **Ooms, J.**, **Robinson, D.**, **Seidel, D. P.**, **Spinu, V.**, … **Yutani, H.** (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, **4**(43), 1686.

Wiese, H., **Alexiadou, A.**, **Allen, S.**, **Bunk, O.**, **Gagarina, N.**, **Iefremenko, K.**, **Martynova, M.**, **Pashkova, T.**, **Rizou, V.**, **Schroeder, C.**, **Shadrova, A.**, **Szucsich, L.**, **Tracy, R.**, **Tsehaye, W.**, **Zerbian, S.**, & **Zuban, Y.** (2022). Heritage speakers as part of the native language continuum. *Frontiers in Psychology*, **5982**. 10.3389/fpsyg.2021.717973.