Research Article

Tim Diaubalick* and Pedro Guijarro-Fuentes

Consequences of the comparative fallacy for the acquisition of grammatical aspect in Spanish

<https://doi.org/10.1515/iral-2021-0110>

Received June 7, 2021; accepted January 3, 2022; published online February 4, 2022

Abstract: This paper tackles the usefulness of comparing L2 learners against native speakers in empirical SLA studies focusing on grammatical aspect. Adapting the view that interlanguage grammars should be analysed in their own right instead of as a deficient form of the target, we show that expressing perspectivity (fulfilled by grammatical aspect markers) methodologically complicates the analyses of Grammaticality Judgment Tasks in aspect studies. For Spanish past tenses, we show that especially with items constructed as allegedly ungrammatical natives behave heterogeneously. This casts doubt on the question whether these data can be used as a baseline against which learners' data could be compared. By analysing the interlanguage separately (not only in comparison to the controls), our findings among German learners of L2 Spanish suggest the use of the forms depends essentially on temporal markers which can be related to both their L1 lacking grammatical aspect and the pedagogical input. Crucially, though the interlanguage does not match the target (i.e., past tenses do not necessarily correlate with aspectuality), the systems are not chaotic but follow well-defined rules.

Keywords: control group; methodology; second language acquisition; Spanish; tense-aspect

1 Introduction

Empirical testing in SLA research often involves a comparison between the L2 participants and a group of native speakers of the target language. This method

***Corresponding author: Tim Diaubalick**, University of Wuppertal, Wuppertal, Germany,

E-mail: diaubalick@uni-wuppertal.de. <https://orcid.org/0000-0003-0040-6230>

Pedro Guijarro-Fuentes, Universitat de les Illes Balears, Palma de Mallorca, Spain

(often heavily criticized as described below) aims to test for specific differences between non-native and native speakers (or the lack thereof). In this sense, natives are used as a control group. The necessity for such a procedure is especially self-evident for online-studies measuring reaction times or employing eye-tracking methods – here, native speakers can establish precisely measured limits (Hopp 2013). In offline tests, e.g., Grammaticality Judgment Tasks (GJT), the inclusion of a native control group, however, often serves the purpose of validating the methodology and usefulness of constructed items (for example, within the generativist framework, see Rothman and Slabakova 2019).

When native speakers are included in an SLA study the control group may be of a different kind than, e.g., in medical studies. Natives do not just receive a placebo which they could or could not respond to, but they already have the competences learners wish to acquire.¹ Whereas the definition of this conundrum is not new, the issue is still relevant when designing and evaluating research experiments. That is, the use of control groups of native speakers to assess the development of L2 grammars can be problematic.

In the following, we will critically examine this methodology with regards to the study of grammatical aspect focusing on written test formats such as GJT. Since aspectuality often depends on a speaker's viewpoint, a comparison to the native competence is not only logically unnecessary in this realm, but even disadvantageous in some respects, leading quickly to the so-called **comparative fallacy** (see Bley-Vroman 1983; Dekydtspotter et al. 2006), resulting from a simplification of the theory for the sake of a formal approach. On the one hand, the problem could lie in previous definitions and/or delimitations of the theoretical construct of aspectuality, and on the other hand, such a construct itself can raise the possibility of an L1–L2 comparative fallacy. That is, depending on the precision with which aspect is defined and described, an empirical study may lead to different outcomes. In turn, however, observed differences cannot mean that there are two different constructs of aspect (one for native speakers and one for non-native speakers). *A priori*, learners may follow a different development path to learn aspect (e.g., effect of L1 German as the one reported herein), and could end up with a distinct representation of aspect.² A judgment based on a comparison between the two groups is, thus, not useful *per se*. In essence, we support the claim that – especially regarding verbal features – it is time to focus more on the **nature** of the representations L2 learners achieve, rather than to concentrate on revealing the

¹ A noteworthy exception is research on the instruction effect, where the control group consists of non-native speakers who receive a different instruction (e.g. Toth and Guijarro-Fuentes 2013; González 2008) or no instruction at all (e.g. Rothman 2008).

² Formulation thankfully adapted from an anonymous reviewer's suggestions.

causes of the alleged deficits specific to the L2 learning context. This means looking at the non-native grammars in their own right, i.e., to investigate the nature of the interlanguage representation without the necessity to investigate concrete formal features in the narrow sense, recognizing at the same time that it may not always be appropriate to dwell on the question of (in)accessibility to UG as in the research conducted in the eighties and nineties.³

This issue was identified quite early on. Schwartz and Sprouse (1994) and Cook (1997) among many others in the past have reminded us of the dangers of comparing L2 learners to native speakers with respect to UG properties. In his 1983 paper on the comparative fallacy (more on this below), Bley-Vroman alerted us that “work on the linguistic description of learners’ languages can be seriously hindered or side-tracked by a concern with the target language” (p. 2) and that “the learner’s system is worthy of study in its own right, not just as a degenerate form of the target system” (p. 4).

For grammatical aspect in an L2, the issue is even more obvious. Looking at the past tenses in Spanish, which convey the contrast of (im)perfectivity (Zagona 2007), the choice between one or another form is known to depend crucially on the speaker’s perspective (Haßler 2016; Salaberry 2008). Thus, they may not only involve geographic or social variation,⁴ but can have rather different functions in specific idiolects. Nonetheless, these functions are not arbitrary, and (different from German, see below) usually it is not possible to interchange the verb forms without altering the meaning. This notion of perspectivity is difficult to apprehend by formal descriptions as in the generativist framework. We therefore recognize the definition of aspectual knowledge comprises of a much broader range of contextual information than has been accepted in previous works, particularly within formal perspectives.⁵

Our findings from German learners of L2 Spanish exemplarily show that, apart from possible L1 effects that may underly the data, especially the interpretation of constructed items is subject to a high variation. While items assembled as

3 Note, however, that in case of tense-aspect studies, newer studies have indeed not only focussed on UG accessibility. For instance, Slabakova and Montrul (2003) investigate the acquisition of a specific feature (genericity), and Rothman (2008) studies the interference of instruction – both studies are discussed and reviewed in more detail below. In addition to these studies, there is also another body of L2 research on tense-aspect from a functionalist perspective as well as studies examining the aspect hypothesis which have not conducted comparisons to NS data (e.g., research by Bardovi-Harlig (2002); Shirai and Andersen (1995) or Cadierno (2000) for L2 Spanish).

4 As pointed out by one of the reviewers, this also concerns other examples from Latino American varieties (not only Lima’s) and bilingual populations. Interested readers are refer to the work by Schwenter and Torres Cacoullos (2008), González et al. (2019), a.o.

5 This thought was thankfully acknowledged by an anonymous reviewer.

grammatical rarely cause discrepancies in the data, allegedly ungrammatical items can often be interpreted very heterogeneously by the native speakers.⁶ We will argue that when investigating typical hypotheses on L2 acquisition, the problem of the comparative fallacy can be avoided if not only narrow formal features of the phenomenon itself, but also properties inherent to learners are considered (L1, proficiency, etcetera). Coming full circle, one central aim of this paper is to show that the construct of aspect requires a broad scope of contextual information to determine aspectual choices, and that such a proper conceptualization of aspect is applicable to both native and non-native speaker's grammar and, at the same time, it helps us understand its development among L2 learners. Thus, we will focus on the operationalization of aspect, but without leaving aside the nature of the L2 grammars.⁷

The paper is structured as follows: After comparing the tense-aspect systems in Spanish and German (Section 2), we will thoroughly present the comparative fallacy's tenets, connecting the topic L1 variability with the subtleties of the target features (Section 3). Specific hypotheses in relation to acquisition are formulated and then tested in a controlled experiment (Sections 4 and 5). The paper closes with a discussion and conclusion section and recommendations for further studies in relation to how the fallacy can be avoided or diminished.

2 Tense and aspect

2.1 Tense and aspect in Spanish

2.1.1 Variable contexts

Spanish, like other Romance languages (Giorgi and Pianesi 2004), marks grammatical aspect in its tenses. Leaving aside the Present Perfect,⁸ we will concentrate on the notion of perfectivity in the contrast between the Preterit and the Imperfect,

⁶ The heterogeneity often depends on the amount of context given with the item making the design of any elicitation task very challenging.

⁷ By this, we are taking the advice of an anonymous reviewer.

⁸ The reason for excluding the Present Perfect lies in the following observation: Different from perfectivity, this phenomenon is subject to a significantly higher variation both between Romance languages and within Hispanic dialects (Schwenter and Torres Cacoullos 2008). Its use, thus, does not depend mainly on a speaker's subjective perspective, but is heavily influenced by language-external factors such as language contact or dialectal variation. For the main tenet of this paper, this tense is, therefore, less useful.

illustrated in examples (1)–(2). While the Imperfect denotes imperfective aspect, the Preterit marks perfective aspect (Zagona 2007).

- (1) Martín paseaba por el parque.
 Martín walk.IMP for the park.
 ‘Martin was walking inside the park.’
- (2) Martín paseó por el parque.
 Martín walk.PRET for the park.
 ‘Martin walked inside the park.’

Importantly, these examples already show that, on a local level, both forms can be grammatical in the same context. Yet they have quite different functions: while the Preterit marks a clear-cut anteriority to the moment of speech, the Imperfect refers to an unactual layer (e.g., a dream, a movie, a background information in the past, etcetera; see Giorgi and Pianesi 2004 for more details). Crucially, the verb form itself adds the final aspectual interpretation to the sentence.

Adopting a formal approach, Domínguez et al. (2017) suggest that, cross-linguistically, aspectuality can be described by means of four features: [$\pm$ perfective], [$\pm$ habitual], [$\pm$ progressive] and [$\pm$ continuous]. Differences between languages arise as consequence of how these features are mapped onto the corresponding forms. In Spanish, the Preterit corresponds to the perfectivity feature (De Miguel 1992; Fábregas 2015; Leonetti 2004),⁹ whereas the Imperfect, marking unbounded actions that do not need to be presented in their totality, comprises habituality, continuity and progressivity. As Díaz et al. (2008) show, features can also be combined in many ways, resulting for example in a Preterit Progressive in Spanish. By choosing a specific form, a speaker can express a specific perspective, i.e., a certain way of presenting the events (Klein 1994; Salaberry 2008). Within this analysis, thus, these formal features reflect the notion of perfectivity.

We may thus speak of a certain *subjectivity* (i.e., speaker’s perspective), as both forms can appear in the same local context (Salaberry 2008: 192). Crucially, subjectivity does not mean arbitrariness (Haßler 2016: 217), and Preterit and Imperfect are not freely interchangeable. This, consequently, means that there is a certain level of systematicity that the linguistic context will trigger. Consider, for instance, examples (3) and (4), where the Imperfect leaves room for various semantic and pragmatic interpretations:

⁹ Other approaches (e.g. Rojo 1990) explain the differences without the term aspect by assigning different temporal roles to the tenses. Here, we follow Leonetti (2004) who argues that those temporal properties are direct consequences of differences in aspectuality. That is, aspect is the cause, and not an effect.

- (3) *El tren salió a mediodía.*
 The train leave.PRET at midday.
 ‘The train left at noon.’

- (4) *El tren salía a mediodía.*
 The train leave.PRET at midday.
 ‘The train was going to leave at noon.’

While in (3) there is certainty about the train having actually left, (4) could be followed by “*pero se retrasó*” (‘but it was delayed’). Yet, in other instances the semantic and pragmatic difference is very subtle (see examples 5 and 6).

- (5) *Siempre que fuimos a esa tienda, compramos mucha porquería.*
 Always that go.PRET to this store, buy.PRET a lot of junk.
 “Whenever we went to that store, **we ended up** buying a lot of junk.”

- (6) *Siempre que íbamos a esa tienda, comprábamos*
 Always that go.IMP to this store, buy.IMP
mucha porquería.
 a lot of junk.

“Whenever we went to that store, **we bought** a lot of junk.”

(examples and translation from Goodin-Mayeda and Rothman 2007: 136, bold font as original)

Those examples show well that a given context on the local level does not necessarily exclude one form or another (see Goodin-Mayeda and Rothman 2007 for a deeper discussion on that particular example). It is the form itself that determines the definitive aspectual interpretation. The speaker thus has to carry out an agreement function between the features a form is able to express and the perspective they want to adapt. Most grammarians agree that the interplay of contextual features and different perspectives leads to subtle differences which cannot be described by a small narrow set of simple rules (see Llopis-García et al. 2012 for a summary). Morphosyntactic features as well as semantic and pragmatic ones must be considered simultaneously. Additionally, one singular action can be viewed differently so that not only the perspective, but also the grammatical aspect of the event changes. In Spanish, the learning task is further complicated due to the phenomenon’s high variation. Different from variation in other phenomena, however, aspect seems to vary mostly at the individual speaker level. Although

dialectal variation of the perfectivity contrast is rather marginal (Rothman 2008),¹⁰ the selection of one specific verb form depends crucially on how a speaker subjectively evaluates and presents a narrative.

2.1.2 Non-interchangeable contexts

Nonetheless, forms do not vary freely, and context plays a major role. Recall that in the example pairs shown so far ((1) and (2), (3) and (4) and (5) and (6), respectively) all forms were grammatical.¹¹ This is different from (7) where, according to Slabakova and Montrul (2008), only the Imperfect is adequate for the verb *vender* since the Preterit would suggest a completion incompatible with the clause's second half. In such examples, a formal analysis of the underlying features seems appropriate.

- (7) *Los González vendían la casa pero nadie la compró.*
 The González sell.IMP the house but nobody it buy.PRET
 'The González were going to sell the house, but no one bought it.'
 (example and translation taken from Slabakova and Montrul 2008: 463)

Similar differences can be found in modal uses of the Imperfect in which the referenced event takes place in the future and, thus, cannot be replaced with a form that exclusively refers to the past (Salaberry 2008).

- (8) *La fiesta (era/ *fue) mañana, ¿verdad?*
 The party (be.IMP/be.PRET) tomorrow, ¿truth?
 'The party was going to be/was tomorrow, right?'

Yet another incidence is aspectual coercion (Palancar 2005). For example, a stative verb such as *saber* ('know' w.r.t. things) or *conocer* ('know' w.r.t. persons) can be shifted towards an eventive interpretation, depending on the tense.

- (9) *Sabía la noticia.*
 know.IMP the news.
 'She knew about the news.' (stative)
- (10) *Supo la noticia.*
 know.PRET the news.
 'She learned about the news.' (eventive)

¹⁰ Actual dialectal disparity concerns mainly the alternation between the Preterit and the Perfect (see Schwenter and Torres Cacoullos 2008) without impinging on the perfectivity contrast itself. Other variations are restricted to contact varieties (Rojas-Sosa 2008).

¹¹ However, see Section 4 for exertions of that term.

However, Palancar (2005) points out that no direct connection can be established between the form *supe* and the English translation of ‘*I found out*’. The form *supe* may also express a perfective meaning of knowing, as in (11).

- (11) *Siempre supe que un día me dejarías.*
 Always know.PRET that one day me leave.COND
 ‘I always knew that you would leave me [one day].’

(example and translation from Rothman 2008: 84)

In summary, in most cases both Preterit and Imperfect forms are allowed (although with different meanings), which makes tense-aspect features very interesting for investigations of the comparative fallacy and challenging for descriptions, especially when adopting a more formal approach. Thus, it is important that we study how such a contrast is acquired.

As an anonymous reviewer pointed out, the precision with which aspect is theoretically defined and described has an immediate effect on how we can interpret its acquisition. While it is thus mandatory to primarily describe the acquisitional empirical data assuming a broader construct of aspect as the description adopted in the present study, formal definitions of aspect are not broadly conceptualized in terms of contextual information. When contextual, including pragmatic, factors are not properly incorporated into the definitional construct of aspect, we may then end up with unexpected results such as the ones reported below (see Section 3) with native speakers behaving unexpectedly. We believe this is just an artefact of the methodology to collect data whereby previous studies opted for a partial description of the construct of aspect which purposefully put aside some of the contextual information on the grounds that such contextual data belong to “pragmatic knowledge”. More importantly, as previously indicated, only in this way, can the comparative fallacy also be avoided in the first place. Let us now have a look at the German grammar in order to understand other considerations when describing aspectual differences between languages.

2.2 Tense and aspect in German

As previous research on the topic has shown (González and Diaubalick 2020; Diaubalick and Guijarro-Fuentes 2019; González and Quintana Hernández 2018; but see also Díaz and Bekiou 2006 a.o), differences in the expression of tense and aspect can lead to very different acquisition scenarios. German belongs to the so-called non-aspect languages (Schwenk 2012), i.e., it does not possess purely grammatical means to express aspectuality. Although there are different forms available to express the past, these do not represent different aspectual meanings but rather mark different speech styles – as for a formal narrow description based

in semantics, therefore, these differences are practically irrelevant. The most frequent past forms are found in the Present Perfect (*ich habe gefunden* ‘I have found’) and the Preterit (*ich fand* ‘I found’).¹² With most verbs, the Present Perfect is preferred over the Preterit, the latter belonging to more formal contexts or written texts (Heinold 2015: 90). Only for a handful of frequent verbs,¹³ the Preterit is equally common in spoken language. Crucially, interchanging the forms is a mere question of formality (i.e., pragmatic), not of meaning (i.e., semantic). For example, the sentence ‘It was raining, when we landed in Munich’ can be translated in four different ways that at most depend on the formality of the linguistic context, but crucially have the exact same semantic meaning (13–15):

- (12) Es hat geregnet, als wir in München gelandet sind.
It rain-PERF, when we in Munich land-PERF.
- (13) Es regnete, als wir in München landeten.
It rain-PRET, when we in Munich land-PRET.
- (14) Es hat geregnet, als wir in München landeten.
It rain-PERF, when we in Munich land-PRET.
- (15) Es regnete, als wir in München gelandet sind.
It rain-PRET, when we in Munich land-PERF.

Summarizing, we can conclude that in German all grammatical aspect features (especially, perfectivity and imperfectivity) are bundled together. Different from Spanish, the forms do not denote aspectuality and are generally – at least in colloquial context – interchangeable. If semantics require a speaker to mark an aspectual contrast, this has to be done through lexical items (e.g. adverbs/particles) or by reformulating the sentence using another verb or periphrasis. This is, of course, an essential observation that also translates to languages in general: a contribution to the aspectuality by part of the lexicon is always possible, and the semantic value of any such elements as adverbs or temporal-aspectual attributes has always to be taken into account. Crucially, German has no standardized way to express (im) perfectivity analogue to the Spanish grammar. Based on these insights, L1 German learners of Spanish would be receptive to the idea that contextual information (e.g., adverbs) plays a critical role in the selection of aspectual markers. The challenge for a German speaker of L2 Spanish, however, is that (i) some conditions favor conventional, prototypical options that downplay the role of the larger linguistic context,

¹² Apart from that, German also has a pluperfect, and past tense forms in the conjunctive mood. As these convey additional notions beyond pastness, they will not be in the focus of this paper.

¹³ See Heinold (2015: 90) for a full list.

and (ii) the inclination to rely on contextual information is needed but not sufficient to develop the conceptual knowledge of aspect (i.e., semantic contrast of perfective and imperfective relies on notion of distance from the event) (Doiz Bienzobas 2013; Llopis-García et al. 2012).¹⁴ More on this below.

3 The comparative fallacy

3.1 Current approaches to native/non-native comparisons

Turning to Second Language Acquisition, the previously presented comparison between Spanish and German has revealed differences between the two language systems that lead to well-known challenges. However, previous research has shown that even native speakers do not always behave as one would expect. For instance, Slabakova and Montrul (2003) find low accuracy rates for some contexts in a Truth Judgment Task, and García and van Putte (1988) attest unexpected behavior in production tasks. Finally, Salaberry (1999) finds that even in free production natives may differ from expectations. These “unexpected results” whereby native speakers seem to fail to demonstrate knowledge of aspect, lead us to question the adequacy of currently adopted definitional construct of aspect. As shown earlier, every verb in the lexicon can occur with all available tense forms,¹⁵ and speakers choose according to their own perspective and contextual factors. While purely formal approaches adopt a partial description of aspect being an artifact of the methodology employed, in the present study we adopt a much broader description of aspect. Although it is mostly studies within the generativist realm that often strongly rely on comparisons between non-native and natives, in what follows we will try to stay within the formal approaches, but embracing a definition of aspect that is broadly conceptualized in terms of contextual information as the ones described above.¹⁶

In L1 acquisition, it is commonly assumed that the target system is fully acquired: speakers tend to a uniformity. Such claims are not entirely true for L2 learners. Using Bley-Vroman’s (2009) terms, SLA is characterized by unreliability – i.e., a difference between the acquired language (namely, the

¹⁴ We thank one of the anonymous reviewers for advising us to point out this line of thought and adapt their helpfully precise formulation in this paragraph.

¹⁵ There are only a few exceptions to that statement, such as the verb *soler* (‘to use to’) which, given its own lexical content, cannot be combined with other aspectual notions than habituality (see *Diccionario Panhispánico de Dudas*, RAE 2005: 611). Note, also, that we do not refer to periphrases which indeed may be more restrictive (e.g., stative verbs usually do not allow a progressive periphrasis, Comrie 1976).

¹⁶ Also in this paragraph, we have relied on some very precise formulations by the reviewers.

Interlanguage Grammar (ILG)) and the target system – and non-convergence (comparing ILGs in different learners).

In recent UG approaches to SLA (see Rothman and Slabakova 2019), if, in an experiment on a concrete feature, learners behave like native speakers (e.g., in a GJT), they are deemed to have attained that feature. In change, if they differ from natives, then their grammars are assumed as not constrained by UG; hence, UG is not (completely) available/accessible. Nevertheless, the proof of UG access is not equal to the acquisition of a target-like feature configuration (see discussion of the feature reassembly below).¹⁷

Thus, the question remains *how* to describe and evaluate observed contrasts. For instance, Slabakova (2009: 169) argues that differences between L1 and L2 speakers are not qualitative but occur rather gradually depending on the specific input. Still, both L1 and L2 are acquired as natural language for communication purposes (Slabakova 2009: 156). Bley-Vroman (2009) takes a different approach, introducing the concepts of *patches* and *viruses*. Analogously to how these terms are used in informatics, the former ones denote the conscious use of elements to compensate for acquisition failure, whereas the latter characterize sentences on which native speakers disagree regarding their (un)grammaticality. The existence of such phenomena proves that L1 speakers, too, can choose between grammatically defined and shallow representations (cf. Clahsen and Felser 2006). Sometimes, *patches* are even necessary to overcome uncertainties.

While recent approaches focus on persisting learnability problems (see e.g., Tsimpli and Dimitrakopoulou 2007), it has been revealed that the results of a conscious learning strategy may superficially resemble the target system, and thus compensate the absent features (e.g. Hawkins and Hattori 2006). However, the underlying representations are different.

The core system is not working (or not working well), so the patching system is taking the burden. [...] The language faculty has means of dealing with strangeness and uncertainty, and these are heavily employed in foreign language learning. (Bley-Vroman 2009: 193)

Many researchers (see e.g. Judy et al. 2008) have already demonstrated that these observations make it possible to describe the L2 outcome without direct comparisons to native speakers. Apart from circumventing discussions on what counts as “native” at all (see Mesthrie 2010), complicated even further through globalisation and the broad availability of international communication, this proceeding respects proposals on the need to consider ILGs in their own right stated quite early is

¹⁷ For reasons of space, we cannot review more work on that topic herein. Interested readers are referred to Ortega (2013), Rothman and Guijarro-Fuentes (2010), Rothman and Slabakova (2019) among others.

UG/SLA research (e.g. duPlessis et al. 1987; Schwartz & Sprouse 1994 to name just a few) Concentrating only on ILG properties, it can be shown that L2 learners may arrive at stable grammars licenced by UG which indeed account for the L2 input (albeit not in the same way as the grammar of a native speaker). The issue, then, is whether the ILG is a ‘possible’ grammar (i.e., licensed by UG), not whether it is equivalent to the target.

Since language variation is supposed to be located at the feature assembly (see Lardiere 2009 for a discussion), only narrow formal features and their mapping to semantic notions are variable, whereas the “general mechanisms of syntactic and semantic computation” is universal (Slabakova 2009: 165). A similar view is defended in the *Feature Reassembly Hypothesis* (FRH; Lardiere 2008, 2009), according to which SLA consists of several phases (see Hwang and Lardiere 2013): first the L1 feature configuration is transferred, then a subsequent reassembly of features takes place. Allegedly problematic phenomena are explainable through difficulties in mapping the forms to the correspondent features. As pointed out by Rothman (2008), the arduousness of this task may cause insecurities (detectable through an overgeneralization of consciously learned rules that overwrite/blur acquired knowledge). Eventually, the FRH predicts that achieving a target-like feature configuration is possible, while the way to this target is determined by L1 transfer. Of course, this achievement can also be indicted of being a victim of the comparative fallacy, when the target system is understood as the language spoken by native speakers. Important for the matter of this paper, however, the FRH emphasizes the learning process as such (not only the outcome) which is important when contrasting learners on different proficiency levels.

3.2 Redefining the role of native participants in SLA studies

There are studies (see Birdsong and Gertken 2013: 118 for a review) where – maybe unsurprisingly – native speakers showed an even higher variation in their answers than the experimental L2 group. Cook (1997) notes that in the research literature, L1 and L2 learners are treated rather differently in respects of terminology. Concretely, expressions generally avoided in L1 studies such as “false” and “incorrect” are relatively common in the SLA context. This is contradictory (Cook 1997: 43): although the term “interlanguage” suggests a system that can be described in its own rights (see Klein and Purdue’s 1997 *basic variety*), terms such “error” and “failure” are surprisingly frequent.¹⁸ Furthermore, in a later publication, Cook (1999) criticises

¹⁸ Note, however, that the definition of error sometimes is not self-evident, either. It may refer to ungrammaticality, low-marked expressions, less frequent forms, etcetera.

that often the term “native speaker” is equated to “monolingual speaker” (recall Mesthrie 2010 for a redefinition of the concept of ‘native speaker’ in L2 varieties). The only hard criterion to count as native speakers does not concern proficiency but consists of having acquired the language in question from birth, a criterion unfulfillable by L2 learners by definition. Given that bilingual speakers are not only the sum of two monolinguals (Cook 1999: 191; see also Birdsong and Gertken 2013: 111), analyses of differences between L2 speakers and monolingual native speakers *a priori* correspond to a mismatched or unfair comparison. Birdsong and Gertken (2013: 121), therefore, suggest speaking of *high-proficient* instead of *near-native speakers*. As proficiency is an exclusive property of L2 speakers, it allows the consideration of other cognitive variables (such as working memory and attention, but also motivation and emotional factors, see Dörnyei 2006) known to have a general impact on language use.

One might, of course, wonder where a pure analysis of ILGs will lead us. Admittedly, most learners commonly have a certain motivation to improve their proficiency level; i.e., the learners themselves support the idea to compare their performance to native speakers, often supported by external pressures such as possible stigmas attached to foreign accents or similar properties.¹⁹ Admitting that the interlanguage is a system in its own right would not automatically prohibit to carry out comparisons. However, linguistic research must evidently consider both comparative differences and similarities which contribute to understanding the nature of bilingualism (without ignoring null findings, see Birdsong and Gertken 2013).²⁰

In the following, we will show that the richness of tense-aspect phenomena would help us to air the comparative fallacy with new empirical evidence. Thus, possible consequences for deviations from the target L1 grammar must be considered when learners are analysed to avoid the comparative fallacy. We will not, however, exclude the native speakers’ data but use them differently: since in many contexts of the Spanish past tenses a high degree of subjectivity is at play (Salaberry 2008), native speakers do not converge on their judgments and often indicate uncertainties about the grammaticality status. Consequently, we will show that analysing learner data individually can also reveal knowledge about variation and make it possible to conclude general processes for language acquisition instead of contrasting them only to native speakers’ mean values.

19 Although Cook (1997, 1999) is right to state that L2 learners may change their attitude towards wanting to *be like* a monolingual, it would be unreasonable to deny that they want to be able to *speak with* monolingual speakers, i.e., to access another level of communication.

20 We thank to one of the reviewers for alerting us of other approaches including the ones by Truscott and Sharwood Smith (2004) when discussing the fallacy; however, space limitations prevent us from going into too many details.

3.3 Defining research questions

Bringing together the thoughts on the variability of tense-aspect phenomena and the reflections on the comparative fallacy, it seems unsurprising that numerous studies on L2 Spanish have confirmed learning difficulties for past tenses (Comajoan 2014). The L1 and the differences it presents in its tense-aspect system is in some cases decisive: for instance, Díaz et al. (2008)²¹ show that Greek, French or Portuguese-speaking learners whose L1 has a similar system to Spanish have some advantages over Chinese or Japanese learners. As for the reassembly task (see above), the learners must detach the configuration from their L1 and reconfigure the features to achieve the target assembly. Given the role of perspectivity within the realm of tense-aspect, a translation to purely formal features is not easy.²² However, since studies especially within the generativist approach rely on comparisons to control groups (as argued above), we will try to apply the analysis described in 2.1. The learning task could then be described as a mapping of the abstract feature [+perfective] to the Preterit and [+continuous], [+habitual] and of [+progressive] to the Imperfect. Crucially, it is only here where the systematic role of perspectivity comes into play: it is within those abstract aspectual features which are then mapped to the concrete morphophonological forms. Whereas for Anglophone learners, for instance, this represents a standard reassembly task in the sense of the FRH, German learners, in a first step, must comprehend that different past tenses express different forms of aspectuality. That is, they have to understand and define what those abstract aspectual features are in the first place. As pointed out above, important in this context is a focus of the process, and not on comparing the outcome to native speakers. That is, it is also perfectly possible that a fully-functioning system might be developed which, although different from native Spanish, does still work as a language with well-defined grammatical rules. In essence, a difference in the use of forms (deviation from the target system) is not equal to *failure*.

By concentrating on the contrast between the Imperfect and Preterit in L2 Spanish by L1 German learners,²³ we will address three research questions presented here together with our predictions:

21 Here we also would like to thank one of the reviewers for recommending to us an exhaustive list of previous research on the differences between L1 and L2 data, some of which can be found summarized in Comajoan (2014).

22 As pointed out by a reviewer, a cognitive approach (see e.g. Doiz Bienzobas 2013; Salaberry 2018) would provide a more evident description of aspectuality.

23 Although Spanish also has a compound perfect that *a priori* could reinforce a direct transfer from German, this reasoning would be an immediate commitment of the comparative fallacy. In comparison to German, the Spanish Perfect is much more restricted.

1. **Concerning individual variability within the native speakers: do natives behave as a homogeneous group? If not, under which conditions?** Since the expression of perspective is not detachable entirely from a subjective (broad) evaluation a speaker has over the situation, we expect the variation within the native group to be quite high. The categories “ungrammatical” and “grammatical” are, thus, in fact questionable as unequivocal labels for a research experiment.
2. **Concerning individual variability within the non-native speakers: do non-natives behave as a homogeneous group? If not, under which conditions?** Learners shall not be compared to natives in a direct way. However, we expect that the special characteristics which define a learner (proficiency level, access to explicit pedagogical rules) manifest themselves in the data. Since German as L1 lacks grammatical aspect, we expect learners to use the different forms independently of encoding (im)perfectivity. In other words, we prognosticate a certain L1 effect here.
3. **Are different uses of the past tenses as illustrated in Section 2.1 (e.g. completed single events or habits representing frequent usages vs. interrupted actions or coerced verbs representing infrequent or non-standard usages) treated differently?** As shown in the examples above, sometimes it is quite complicated to tease apart ungrammaticality from inappropriateness both from a theoretical as well as from a more practical point of view, i.e., when the interpretation of a syntactically well-built sentence turns illogical due to the verb form used. It does, thus, depend on the participant’s definition of what acceptability means, i.e., whether a seemingly grammatical sentence with an odd interpretation is seen as acceptable or not. We consequently expect high subjectivity in these cases.

4 Methodology

4.1 Participants

Sixty participants with three different language backgrounds were recruited for the study. The native group is formed by 30 native Spanish speakers (median age 23, $\bar{x} = 29.4$, $s = 10.1$), consisting of 22 female and 8 male participants whose variety of Spanish is the European one. For the group of L2 learners, thirty L1 German speakers with a self-reported advanced level of Spanish (i.e., C1-level or higher) were recruited for the study (22 female and 8 male speakers; median age 27, $\bar{x} = 25.2$, $s = 7.9$). Additionally, their proficiency level was confirmed through a 50-items cloze test adapted from the test provided by the Oxford University Language Center. All

participants were living in Spain during the data collection where they had spent at least three months before the test date and reported to use Spanish daily. Eight participants, however, had to be excluded from further analysis, because they fell outside the advanced level according to the proficiency test. Note that the variety of Spanish that German learners are exposed to is European Spanish. Therefore, it seemed logical to us to focus on this variety here, too. Although, due to a stronger representation of the Present Perfect in European Spanish language varieties, this might, to a certain point, induce more of an unfamiliarity with the Preterit;²⁴ the perfectivity contrast between Imperfect and Preterit is still nevertheless valid and present in day-to-day communicative situations.

4.2 Experimental tasks

In addition to an ethnolinguistic questionnaire and the proficiency test, all participants were given the same offline tasks in written form. Data for this study were collected via a Grammaticality Judgment Task (GJT) consisting of 48 items. Whereas 36 items were experimental items with a Preterit or Imperfect form, the other 12 were distracters. The items were constructed in a way that, in accordance with our research questions, a high level of subjectivity (i.e., speaker's perspective) was allowed for. The participants were simultaneously confronted with other tasks belonging to a greater study, in which this GJT was embedded (Diaubalick and Guijarro-Fuentes 2017, 2019). Instructions given to the participants were held to a minimum. Participants were asked to evaluate the items on a Likert-Scale from 1 to 5, and correct ungrammatical items if they wanted to.²⁵ The decision for a Likert-Scale follows the idea that, due to illogical combinations, sentences can turn odd without being univocally ungrammatical. No time limit was given, but in average data collection took about 25–30 min per participant for the GJT, and around the same time for the other tasks of the experiment. Nevertheless, participants were allowed to go back and forth.

24 A future research could be designed to compare the effect of exposure to different varieties of Spanish and its consequences in terms of L2 attainment.

25 As the experiment was designed for a broader group and especially language learners with different proficiency levels in the target language, instructions were kept simple. That is, *Evalúa las siguientes frases indicando una calificación numérica, y corrige aquellas que te parecen incorrectas* ('Evaluate the following sentences by giving a numeric qualification, and correct those which seem incorrect to you'). Crucially, terms such as grammaticality and/or any other terms which could induce learners to guess our research purposes or could confuse participants as we were looking at naïve spontaneous responses if at all possible, were not used within the material. Instructions were clear but neutral, disguising the task main objectives.

Items were constructed in four categories (examples below) which are based on previous studies on tense-aspect-phenomena (e.g. Guijarro-Fuentes 2005; Slabakova and Montrul 2003, 2008): ten items contrast single actions and habitual events and will be called ‘standard context items’ in the following (comprising the fact that these usages are the most frequent and allegedly the least disputable ones, being usually the first ones presented in a textbook of Spanish as foreign language). These items contrast with those of less frequent uses that aim to study the acquisition of all the nuances connected to the phenomenon: ten items are ‘coercion contexts’ (stative verbs with and without an eventive meaning), ten items contain ‘impersonal subjects’ (inspired by Slabakova and Montrul 2008), and further six items represent non-finished telic actions in an imperfective context which, due to their aspectuality, we will call ‘conflicting feature items’. In each category, half of the items were constructed as ungrammatical. Importantly, although we distinguish between more frequent and less frequent cases here, this does not mean that judgments *a priori* must be less homogeneous in the latter. In every item, we have a clear expectation based on previous studies cited above.

To avoid the connection to an alleged ideal L1 speaker (see Cook 1997: 41), however, the terms ‘grammatical’ versus ‘ungrammatical’ must be treated with high caution. As we will demonstrate in the next section, they do not represent unequivocal labels for any participant (neither native nor non-native), since arguably acceptability, grammaticality and appropriateness are individually interpreted very differently. Thus, maintaining the terminological distinction between ‘grammatical’ and ‘ungrammatical items’ can be likewise as misleading as to speak of an error analysis. Nonetheless, Ragheb and Dickinson (2011: 120) affirm that, when error is used as a neutral term devoid of content – i.e., as a method and not as a judgment – then its analysis might eventually lead to a result. In the same sense, we will use especially the term *ungrammatical* to express a deviation from our expectations, not necessarily meaning inaccuracy. Crucially, it is not intended to express any normative notion of falseness but rather a label. Nevertheless, some methodological consequences remain: whereas so-called grammatical items are constructed as expectedly acceptable, the so-called ungrammatical items contain an expectedly unacceptable form. This determines the use of statistics as we will translate our expectations to numerical judgments in order to calculate deviations. Considering what has been presented above, of course, any such statistical deviations would be little surprising.

Examples from our test battery are given in (16)–(19).²⁶

²⁶ Although one of the reviewers would have liked to see the complete test battery or to include more examples of the items used, space limitations prevent us from repeating them here. However, in Diaubalick (2019), the test battery in its entirety (investigated with a different focus) can be consulted.

- (16) Normalmente iba andando a la universidad.
 Normally go.IMP walked to the university.
 ‘Usually, I walked to the university.’ (habitual context, Imperfect)
- (17) Carlos y Andrea se conocieron el 3 de octubre de 1991.
 Carlos y Andrea self know.PRET the 3 of October of 1991.
 ‘Carlos and Andrea got to know each other on the 3rd of October, 1991’
 (stative verb coerced into an eventive reading, Preterit)
- (18) Creo que allí se vivía con más preocupaciones.
 I-believe that there self live.IMP with more worries.
 ‘I think, over there people live with more worries’ (generic subject, Imperfect)
- (19) Ayer me comía dos platos de paella, pero el
 Yesterday me eat.IMP two plates of paella, but the
 segundo no me lo pude terminar.
 second not me it could.PRET finish.
 ‘Yesterday, I was going to have two plates of paella, but I couldn’t finish the second one’ (telic verb phrase in imperfective context, i.e. contrasting levels of aspectuality)

All test items are constructed with the intention that the tense forms used cannot be replaced with an alternative one. That is, in designing the task we were very careful in creating items with no room to choose both forms without changing the meaning of the item.

5 Data analysis and results

Our data analysis follows a similar idea to the one defended by Ragheb and Dickinsons (2011), i.e., using errors and deviations as unprejudiced labels and not as attesting deficits. Different from their approach, however, we are rather interested in the interpretation of the verbal forms than in their production. Participants rated each item on a 5-point Likert scale, where 1 means acceptable and 5 means rejection.²⁷ In this paper, native and non-native speakers are analysed separately. For each individual and each category (grammatical and ungrammatical), we

²⁷ Note that this seemingly unconventional direction of the scale stems from the German school grade systems, where 1 is the highest and 4 the lowest passing grade. The German participants are thus familiar with the number 5 meaning failure.

calculated means, resulting in eight values per participant. Based on these, we conducted several (multivariate) ANOVAs. Only the relevant results for the purposes of the present paper are presented below.²⁸

5.1 Native data – group results

The data from the native group show, items constructed as grammatical and ungrammatical are judged in a very different way. While there is generally a high consistency in accepting grammatical items, the ungrammatical ones are not rejected with the same strength. This confirms our safeguard to use the term ungrammatical as a non-strict label. Table 1 (organised according to the categories shown in examples 16–19) shows means and standard deviation for the two groups of native speakers.

Two main observations arise: first, the only case where we find standard deviation values greater than 1 is in the ungrammatical items of contrasting levels. Second, the means found here were not greater than 3. This means that for all the grammatical items, most speakers agree on their status. This is important because by using a five-point Likert scale, the value of 3 is the one expected to be selected if a speaker is uncertain about an item; other values, on the contrary, indicate

Table 1: Native data in the GJT – means and standard deviation.

	Grammatical items			Ungrammatical items		
	Spain	Hispano-America	Total	Spain	Hispano-America	Total
Standard contexts	<i>M</i> = 1.52 (<i>s</i> = 0.67)	<i>M</i> = 1.49 (<i>s</i> = 0.53)	<i>M</i> = 1.51 (<i>s</i> = 0.59)	<i>M</i> = 4.28 (<i>s</i> = 0.54)	<i>M</i> = 4.05 (<i>s</i> = 0.47)	<i>M</i> = 4.17 (<i>s</i> = 0.51)
Coercion contexts	<i>M</i> = 1.69 (<i>s</i> = 0.62)	<i>M</i> = 1.72 (<i>s</i> = 0.60)	<i>M</i> = 1.71 (<i>s</i> = 0.61)	<i>M</i> = 3.71 (<i>s</i> = 0.61)	<i>M</i> = 3.41 (<i>s</i> = 0.66)	<i>M</i> = 3.56 (<i>s</i> = 0.64)
Impersonal subjects	<i>M</i> = 2.77 (<i>s</i> = 0.74)	<i>M</i> = 2.49 (<i>s</i> = 0.73)	<i>M</i> = 2.63 (<i>s</i> = 0.74)	<i>M</i> = 3.97 (<i>s</i> = 0.69)	<i>M</i> = 3.69 (<i>s</i> = 0.62)	<i>M</i> = 3.83 (<i>s</i> = 0.66)
Contrasting levels	<i>M</i> = 2.15 (<i>s</i> = 0.82)	<i>M</i> = 1.87 (<i>s</i> = 0.72)	<i>M</i> = 2.01 (<i>s</i> = 0.78)	<i>M</i> = 2.69 (<i>s</i> = 1.45)	<i>M</i> = 2.47 (<i>s</i> = 1.32)	<i>M</i> = 2.58 (<i>s</i> = 1.37)

²⁸ For a more detailed comparison with occasionally somewhat loaded burdening results that indeed reflect a comparative fallacy, the interested reader is referred to our previous work (Diabulick and Guijarro-Fuentes 2019) with partial results. Here, however, we present all data collected together with all results as part of this research endeavor.

tendencies or certainties. Newly, the difference between ungrammaticality and inappropriateness can have an influence on the certainty about the status of an item. Note that especially the judgments within the categories “impersonal subjects” and “contrasting levels” often deviate from expectations (see below).

Observing the native group, Figure 1 shows convincingly that not all categories are treated homogeneously. It seems that, only in standard contexts (i.e., habitual/episodic events) speakers sharply distinguish between grammatical and ungrammatical items, thus showing a greater unanimity than in all other categories. Conversely, our participants do not show clear tendencies in the judgments of impersonal subjects, contrasting levels and coercion contexts.

Whereas in other research on this topic (including Diaubalick and Guijarro-Fuentes 2019), standard ANOVAs have been performed for finding differences between groups, in the following, we will take a different approach. Instead of asking if groups behave differently, we want to know if items are treated differently according to the context in which they are given and contrast those for each participant group separately. Consequently, we will contrast items for each separate category, asking whether speakers rate these items differently.

In a first step, however, we need to know if those items constructed as grammatical are indeed treated as such; the same later goes for the ungrammatical ones. One way to do this is by performing one-tailed t-tests (assuming normal distribution within the Likert scale). In order to possibly translate the threshold limits required by the test, in the following we only use natural numbers. Table 2 shows the results of these tests that strengthen the impression that only in some cases, items are accepted or rejected with a high level of certainty. Especially in those cases in which the mean values are higher than 2, but lower than 4, we can assume either a high proportion of variability or a general uncertainty – this

Figure 1: Native data in graphical form, where the mean is represented as a cross, the quantiles by vertical lines, and outliers as dots.

Table 2: One-tailed t-tests over the grammatical items.

Category	Null hypothesis	t-Value	p-Value
Gram. standard contexts (mean was 1.51)	$H_0: \mu \geq 2$	Right-tailed: $t(29) = -4.534$	$p < 0.001$ (***)
Gram. coercion contexts (mean was 1.71)	$H_0: \mu \geq 2$	Right-tailed: $t(29) = -2.655$	$p = 0.007$ (**)
Gram. impersonal subjects (mean was 2.63)	$H_0: \mu \leq 2$	Left-tailed: $t(29) = 4.705$	$p < 0.001$ (***)
Gram. contrasting levels (mean was 2.01)	$H_0: \mu \leq 2$	Left-tailed: $t(29) = 0.078$	$p = 0.469$

concerns the grammatical items with an impersonal subject and those with contrasting aspectual levels. Therefore, for these two categories another right-tailed t-test is not applicable. Instead, a left-tailed t-test with 2 as threshold value was conducted.

For standard and coercion contexts, the obtained values are significantly lower than 2, such that we can assume the items are indeed acceptable. Also for the contrasting levels, although the obtained value is slightly higher, there is not enough evidence to reject the idea that the speakers regard the items as grammatical – certainly, they are not unacceptable (with a mean lower than 4). Thus, only in the case of the impersonal subject contexts, the significant result indicates that the items are not consistently judged as grammatical. Nonetheless, they are clearly not judged to be ungrammatical, either.

Turning to the ungrammatical items, the numbers indicate a stronger tendency towards unforeseen judgments. Only regarding the standard contexts, the mean value is higher than 4, whereas in all other contexts it is lower. Thus, the standard contexts can be analysed by means of a left-tailed t-test, whereas in all other cases we must conduct right-tailed t-tests with the value 4 as threshold (Table 3).

The statistical analyses show that only standard contexts constructed as ungrammatical are consistently rejected. As for coercion contexts and contrasting levels, we cannot say that they are absolutely rejected (means significantly lower than 4). In fact, in the case of the contradicting levels, the examples even show to be generally acceptable.²⁹ Likewise, the non-significant value in the impersonal subject items means we cannot dismiss the hypothesis that the participants are

²⁹ In fact, an additional t-test with 2 as threshold (with $H_0: \mu \leq 2$, left-tailed: $t(29) = 2.310$, $p = 0.28$) shows that we cannot dismiss the assumption the items are judged as grammatical.

Table 3: One-tailed t-tests over the ungrammatical items.

Category	Null hypothesis	t-Value	p-Value
Ungram. standard contexts (mean was 4.17)	$H_0: \mu \leq 4$	Left-tailed: $t(29) = 1.792$	$p = 0.042$
Ungram. coercion contexts (mean was 3.56)	$H_0: \mu \geq 4$	Right-tailed: $t(29) = -3.741$	$p < 0.001$ (***)
Ungram. impersonal subjects (mean was 3.83)	$H_0: \mu \geq 4$	Right-tailed: $t(29) = -1.387$	$p = 0.088$
Ungram. contrasting levels (mean was 2.58)	$H_0: \mu \geq 4$	Right-tailed: $t(29) = -5.686$	$p < 0.001$ (***)

rejecting the items. In summary, the results concerning the ungrammatical items suggest that there is clearly less certainty and less consistency among the native speakers in rejecting ungrammatical items than in accepting grammatical ones.

What seems essential is to distinguish sharply between frequent and less frequent usages. Consequently, two additional ANOVAs contrasting the different test categories (Table 4) show that neither in the grammatical nor in the ungrammatical categories, items are rated homogeneously across test conditions.

To track these differences down, we applied a post-hoc test using Tukey’s HSD deviding the categories in homogeneous subgroups (Tables 5 and 6). In sum, in non-standard items there is a high effect of subjectivity. The next logical step must thus be to look closer at the individual data.

Table 4: ANOVA results for variation across categories (natives).

Test conditions	ANOVA result	p-Value
Grammatical categories	$F(3, 116) = 14.170$	$p < 0.001$
Ungrammatical categories	$F(3, 116) = 19.208$	$p < 0.001$

Table 5: Homogeneous grouping of grammatical categories (Tukey’s HSD).

Grammatical categories		
Group 1 (rated lowest)	Group 2	Group 3 (rated highest)
Standard, coercion	Coercions, contrasting levels	Impersonal subjects

Table 6: Homogeneous grouping of ungrammatical categories (Tukey’s HSD).

Ungrammatical categories		
Group 1 (rated lowest)	Group 2	Group 3 (rated highest)
Contrasting levels	Coercions, impers. subj.	Impers. subj., standard contexts

5.2 Native data – individual results

Again, a clear unanimity among native speakers cannot be expected. Table 7 summarizes the number of individuals behaving as intended. Different from the analyses applied to the group results, the means presented here are calculated for every participant individually. Again, we see that the labels “grammatical”/“ungrammatical” are not categorical. These observations complicate designs involving comparisons between native speakers and L2 learners.

Most observations from the group results are confirmed. However, two values deserve a further comment here. Similar to Slabakova and Montrul’s (2003) findings, we can attest no systematicity; quite an interesting result when trying to consider the FRH (see below). Second, the Preterit in those items of conflicting features triggers a contradiction to the context. However, it is not entirely clear if this contradiction deserves the status of *ungrammatical*. The illogic produced by the combination of the Preterit with an event that, according to the context, is not completed clearly affects the acceptability, but speakers do not seem to equate this with ungrammaticality. As for standard and coercion context, the results from the group analyses are confirmed.

Taking together the results, and providing an answer to our first research question, that is, *do natives behave as a homogeneous group? If not, under which conditions?*, the data presented here have shown that none of the natives acted

Table 7: Individual rates in the native group.

	Number of individuals with a mean judgment of 2 or below in grammatical items	Number of individuals with a mean judgment of 4 or above in ungrammatical items
Standard contexts	$\#(\bar{x} \leq 2) = 25$ [83%]	$\#(\bar{x} \geq 4) = 19$ [63%]
Coercion contexts	$\#(\bar{x} \leq 2) = 22$ [73%]	$\#(\bar{x} \geq 4) = 12$ [40%]
Impersonal subjects	$\#(\bar{x} \leq 2) = 7$ [23%]	$\#(\bar{x} \geq 4) = 13$ [43%]
Contrasting levels	$\#(\bar{x} \leq 2) = 19$ [63%]	$\#(\bar{x} \geq 4) = 7$ [23%]

homogeneously. Note that this observation mainly proves that an individually stable system in each speaker does not equal a systematicity on the group level. We find a certain degree of variation regarding acceptance or rejection of items that feature tense-aspect phenomena; this, thus, confirms a problem that had been identified in many previous empirical studies. A prejudiced comparison between non-native and native speakers is thus, *a priori*, a mistaken one. In the following, we will therefore describe the non-native data in its own right by using the same means as with the native group, i.e., without speaking of failure or errors, but of variation.

5.3 Non-native data – group results

Table 8 summarizes the mean judgments of the 22 highly advanced speakers of Spanish.

We detect various mean judgments close to the value of 3. Again, this may express a general uncertainty. But is it that different from what happened in the native group? Let us now concentrate on the distinction of grammatical versus ungrammatical items. Figure 2 in which Table 8 is graphically visualized makes the contrast visible: In non-standard contexts, the judgments do not clearly diverge. The figure illustrates why we opted for a contrast between the different contexts, first for the grammatical, then for the ungrammatical items, instead of contrasting grammatical with ungrammatical items in each category: it definitely seems to be the context that defines how judgments are made.

To confirm the impression of a high uncertainty with most non-standard items, the data will be analysed using the same statistical test as in the case of native speakers. Since no mean judgment is passing the threshold (under 2 in the

Table 8: Results of the grammaticality judgment task, German learners.

	Grammatical items	Ungrammatical items
Standard contexts	$M = 2.14$ ($s = 0.55$)	$M = 3.81$ ($s = 0.88$)
Coercion contexts	$M = 2.56$ ($s = 0.63$)	$M = 3.07$ ($s = 0.62$)
Impersonal subjects	$M = 2.60$ ($s = 0.72$)	$M = 3.45$ ($s = 0.94$)
Contrasting levels	$M = 2.38$ ($s = 0.87$)	$M = 2.40$ ($s = 1.02$)

Figure 2: Graphic representation of mean judgments in the German group.

Table 9: One-tailed t-tests over the grammatical items (German group).

Category	Null hypothesis	t-Value	p-Value
Gram. standard contexts (mean was 2.14)	$H_0: \mu \leq 2$	Left-tailed: $t(21) = 1.156$	$p = 0.131$
Gram. coercion contexts (mean was 2.56)	$H_0: \mu \leq 2$	Left-tailed: $t(21) = 4.187$	$p < 0.001$ (***)
Gram. impersonal subjects (mean was 2.60)	$H_0: \mu \leq 2$	Left-tailed: $t(21) = 3.924$	$p = 0.001$ (**)
Gram. contrasting levels (mean was 2.38)	$H_0: \mu \leq 2$	Left-tailed: $t(21) = 2.031$	$p = 0.028$ (*)

Table 10: One-tailed t-tests over the ungrammatical items (German group).

Category	Null hypothesis	t-Value	p-Value
Ungram. standard contexts (mean was 3.81)	$H_0: \mu \geq 4$	Right-tailed: $t(21) = -0.989$	$p = 0.167$
Ungram. coercion contexts (mean was 3.07)	$H_0: \mu \geq 4$	Right-tailed: $t(21) = -6.993$	$p < 0.001$ (***)
Ungram. impersonal subjects (mean was 3.45)	$H_0: \mu \geq 4$	Right-tailed: $t(21) = -2.726$	$p = 0.007$ (**)
Ungram. contrasting levels (mean was 2.40)	$H_0: \mu \geq 4$	Right-tailed: $t(21) = -7.330$	$p \leq 0.001$ (***)

grammatical, over 4 in the ungrammatical items), all t-tests are conducted under the null hypotheses that overall learners still behaved as expected, i.e., t-tests for grammatical items are left-tailed, t-tests for ungrammatical items are right-tailed. A significant result is, therefore, a deviation from expectation. According to Tables 9 and 10, our first impressions are entirely confirmed.

For all non-standard contexts, grammatical items are not strictly accepted, and ungrammatical items are not rigorously rejected. Though recalling the importance of non-significant findings (Section 3.1), affirmations on the standard contexts need more data to avoid possible type II errors. Conducting ANOVAs for the several categories, there is no indication that grammatical items are indeed treated heterogeneously. In the case of ungrammatical items, conversely, learners do seem to differentiate between categories (Table 11).

Table 11: ANOVA results for variation across categories (non-natives).

Test conditions	ANOVA result	<i>p</i> -Value
Grammatical categories	$F(3, 84) = 1.998$	$p = 0.120$
Ungrammatical categories	$F(3, 84) = 10.293$	$p < 0.001$

However, a post-hoc test, using Tukey’s HSD, shows there are no disjoint subgroups (Table 12). It is thus safe to say that most grammaticality judgments among our learners indicate a high variation and possible uncertainty.

Table 12: Homogeneous grouping of ungrammatical categories (Tukey’s HSD).

Ungrammatical categories		
Group 1 (rated lowest)	Group 2	Group 3 (rated highest)
Contrasting levels, coercions	Coercions, impers. subj.	Impers. subj., standard contexts

5.4 Non-native data – individual analyses

The analyses on an individual level confirm the group data shown above (Table 13). In non-standard contexts, less than half of the participants behaved as expected. Also in the standard contexts, although to a lesser extent, there is high variation.

In the same way as we argued above that individual stability does not imply group homogeneity, here, we can argue the opposite: the lack of homogeneity on a group level does not mean that each individual could not have a stable inter-language system. Overall, it seems impossible to expect participants to sharply distinguish allegedly grammatical from allegedly ungrammatical items. Since this applies both to native and non-native speakers, there is not objective reason to take

Table 13: Individual results of the German participants heaving as expected.

	Number of individuals with a mean judgment of 2 or below in grammatical items	Number of individuals with a mean judgment of 4 or above in ungrammatical items
Standard contexts	$\#(\bar{x} \leq 2) = 14$ [64%]	$\#(\bar{x} \geq 4) = 12$ [55%]
Coercion contexts	$\#(\bar{x} \leq 2) = 5$ [23%]	$\#(\bar{x} \geq 4) = 2$ [9%]
Impersonal subjects	$\#(\bar{x} \leq 2) = 7$ [32%]	$\#(\bar{x} \geq 4) = 7$ [32%]
Contrasting levels	$\#(\bar{x} \leq 2) = 9$ [41%]	$\#(\bar{x} \geq 4) = 2$ [9%]

differences between the mean values of the two groups as indication of access problems. Instead, we should focus on revealing possible effects a category can have on the judgment. Thus, answering our second research question, that is, *do non-natives behave as a homogeneous group? If not, under which conditions?*, one needs to consider possible (explicit) learning strategies besides the information that our statistical analyses may provide. For L2 learners whose L1 lacks the perfectivity contrast, strategies known from previous literature on the topic concern the consideration of the lexical aspect (e.g., stativity, telicity, see De Miguel 1992; Salaberry 2008) or the focus of temporal adverbs (Diaubalick and Guijarro-Fuentes 2019; Rothman 2008). This observation is directly connected to our third research question; namely, *are different uses of the past tenses treated differently?* Meant as guideline, the presence of adverbs supposedly facilitates the choice of the adequate verb form. For instance, *ayer* (‘yesterday’) denotes a completed time interval that often co-appears with the perfective aspect and, thus, seems to favour the Preterit. Adverbs as *frecuentemente* (‘frequently’), in contrast, are more often combined the Imperfect. Again, it is crucial at this point to remind ourselves that, of course, adverbials generally do contribute to the aspectual readings of the sentence (as well as lexical aspect does). However, an alleged correlation between some selected adverbs and (im)perfectivity are only rules-of-thumb which do not invariably lead to a target-like sentence; other elements contribute to aspectuality, too. Since in German aspectual differences are not expressed via verbal morphology but can only be made explicit via participles, spontaneous/non-grammaticalized periphrases and adverbs, it is plausible to assume that German-speaking learners base their strategies on similar lexical items. As revealed repeatedly in usage-based studies (see e.g. Wulff and Ellis 2018), the focus on adverbials can have a blocking effect for further learning, since learners know lexical cues often to be a reliable tool in any language. Therefore, the next logical step of analysis is to focus on those items in which the appearing marker is misleading; for instance, an imperfective context co-appearing with the adverb *ayer* (‘yesterday’). If in such a case the item contains a form of the Imperfect,

it is grammatical, but may be rejected due to too rigorous a learning strategy. For an illustration, recall our third research question.

In our experiment, such misleading markers were present in all four categories (see examples 16–19), and therefore reflect the learners’ answers in the standard- and non-standard contexts simultaneously. Our data seem to confirm that indeed in such cases, grammatical and ungrammatical items are not sharply distinguished anymore (Table 14).

A further t-test on these values confirms the deviation from expectation to be a significant one (Table 15). Since learners base their decisions on those misleading markers, the item is not recognized as unequivocally grammatical or ungrammatical.

These results effectively stem from the application of learning strategies, as revealed by the individual performance (Table 16). Whereas only five learners treat grammatical items with a misleading temporal marker as grammatical, there is only one learner who rejects the ungrammatical items.

Table 14: Judgments for items with misleading temporal markers (German group).

	Grammatical items	Ungrammatical items
Misleading temporal marker	$M = 2.56$ ($s = 0.57$)	$M = 2.83$ ($s = 0.63$)

Table 15: One-tailed t-tests over the items with misleading temporal marker.

Category	Null hypothesis	t-Value	p-Value
Misleading temporal marker in grammatical items	$H_0: \mu \leq 2$	Left-tailed: 4.610	<0.001 (***)
Misleading temporal marker in ungrammatical items	$H_0: \mu \geq 4$	Right-tailed: -8.722	<0.001 (***)

Table 16: Individual analyses in items with misleading marker.

	Number of individuals with a mean judgment of 2 or below in grammatical items	Number of individuals with a mean judgment of 4 or above in ungrammatical items
Misleading temporal marker	$\#(\bar{x} \leq 2) = 5$ [23%]	$\#(\bar{x} \geq 4) = 1$ [5%]

In light of the answers of our three main research questions, the data are congruent with observations of previous studies: in empirical experiments specifically constructed to reveal the learners' accuracy with tense-aspect forms, native speakers included as a control group do often not behave as expected. We next discuss our findings in more detail.

6 General discussion and conclusion

By questioning the role of a native control group in SLA studies of tense-aspect phenomena, the main goal of this paper was to show evidence that it is more fruitful to describe the learners' interlanguage in its own rights instead of contrasting it to the target system. Taking the Spanish past tenses as a concrete example, deviations from expectations, thus, do not serve as counterarguments against a successful acquisition. As argued, due to the dependence on one's perspective in many cases, various verb forms would fit into the linguistic contexts, but express different points of view. While depending on which perspective the speaker adopts, which would reaffirm the description of the broader construct of aspect in the present study, the context can be interpreted as [+perfective] or [–perfective], and each of these features would transcend to another morphophonological forms. Taking inspiration from Bley-Vroman's (2009) terminology, this property qualifies certain items as occasions for a virus, since their grammaticality status is unclear and depends on subjective factors (such as perfectivity). Concretely, we cannot know if an apparently target-deviant verb form results from a mismatch between the contextual feature and the form, or if it stems from an actually expectation-deviant interpretation of the context, assigning a different abstract aspectuality feature.

According to the results reported herein, accuracy is in fact a value not analysable in experimental settings – especially when considering less frequent uses of the past tense. The differences found in the data could not be statistically attributed to language variation but in fact seem to mainly depend on subjective and individual factors such as the expression of one's personal viewpoint. Crucially, and following what has been exposed in the introducing sections, if this is true for the native speakers, it must be also true for second language learners.

The (implicit or explicit) expectation that L2 speakers do not differ significantly from native speakers with respect to performance on some syntactic features is one

of the clearest manifestations of the Comparative Fallacy.³⁰ As a logical consequence, the learners' language must be analysed in its own rights, describing *how* and *why* a certain verb form is selected. The key, thus, does not lie in a contrast of several speaker groups, but in focusing on how the speakers treat different categories. Although the data could arguably support the *Feature Reassembly Hypothesis* (Lardiere 2009), as the learners arrive at a stable system in which the verbal features systematically correspond to other features present in the context (i.e., lexical elements such as temporal adverbs), we found no complete acquisition which would support the last step predicted by the FRH. One could argue, hence, the formal definitions of aspect that do not rely on a broad range of linguistic information are not relevant for the assessment of the construct of aspect because they provide a narrow window into the construct. Following this line of thought, the FRH may not be relevant for the definition of the aspect of our L2 learners are developing, and for that matter, for the interpretation of our results, because it does not encompass the actual scope of the construct of aspect.³¹

However, the systematic (or for that matter, the narrow and delimited definitions of aspect) focus on adverbials proves that the learners indeed follow organized rules, i.e., their representation of the Spanish grammar is not chaotic. This observation strengthens our proposal that it helps us to conceptualize that the L2 learners' grammar as not defective (when compared to the native standard grammar) leading us to describe it as systemic and coherent on its own. Nonetheless, simultaneously stable yet (apparently) target-deviant interlanguages follow from learning difficulties. Recurring once more to Bley Vroman's (2009) terminology, shallow representations may serve as patches for processing complex linguistic phenomena, and, according to our data, the association of verbal features with adverbs can be a compensating measure constructed to process the Spanish past tenses. As adverbials generally play a role both for native as well for non-native speakers, the orientation towards them cannot be taken as a failure after all.³² The only issue raised here is that, possibly, learners weigh the different aspectual factors in another way than the native speakers do. The processes of

30 Ragheb and Dickinson (2011) state that another case would be to explain all observed "errors" by comparing the structures to the L1. They suggest describing only what a learner uses instead of conjecturing what they intended to use.

31 This thought comes from an anonymous reviewer whose arguments we integrated into this section.

32 Quoting an anonymous reviewer, "the construct of aspect that BOTH native speakers and L2 learners are attentive to when assessing the examples depicted in this study is NOT the one that leaves out of the picture the adverbial and lexical elements that seem to determine the appropriate selection in some cases."

making a decision based on elements from the (local or global) context as such, however, is generally the same.

Having found accuracy as a factor difficulty applicable to tense-aspect phenomena (recall the reflection on grammaticality vs. acceptability, Section 4.2), this is a very important result for further applications of the findings for follow-up studies with a didactic focus: in a guided learning environment, learners should be made aware that tense-aspect morphology is not just a mere reflection of other elements that appear in the sentence but contribute themselves to the interpretation of aspectuality. That is, different from grammatical features that depend on agreement functions such as gender and number morphemes in adjectives, by adopting a broader description of aspect many contexts permit several verb forms. These are not interchangeable but influence meaning differently. Currently, there are still often tasks used in class that do not reflect this property accurately. For example, fill-in-the-blank tasks employed repeatedly suggest to the students that the meaning of a text is already clear, even if verbs are only given in their infinitive form in brackets. Such instruction may lead to stable representations as shown above, yet they do not correspond to the target system completely. This is why the variation observed in both groups of native and non-natives has very different causes and could not be compared without taking into consideration innerlinguistic factors. A shift towards the explanation of perspectivity and aspectuality seems therefore mandatory for future teaching approaches (see e.g. González 2008; Quintana Hernández 2010 for proposals). This may be the one of the most important conclusions of this paper. Adapting explanations within a more cognitive framework (e.g. Doiz Bienzobas 2013; Jansen 2013), new teaching methods could benefit from what has been found in linguistic research.

Summing up, we propose to describe the patterns by means of the properties of the learners only, i.e., by considering their L1 and their proficiency. Since the L1 of our learners was German, and the German language is known to mark contrasts rather lexically than morphologically, the L1 properties can influence how a compensating strategy is constructed. Instead of causing its construction, however, we sustain that it merely strengthens it. The ability to form rule-based learning mechanisms is inherent to any second language learner, but the concrete way in which this happens may be determined by individual variables such as the L1.

A comparison of the learners' data to a native control group is unnecessary. For future studies we therefore recommend not to compare learners with natives, but to contrast different uses and different phenomena in relation to different learner variables (Diaubalick and Guijarro-Fuentes 2019). At the same time, we want to make clear that this does not mean, as has been shown in the analyses carried out above, that a native control group has to be discarded completely – analyses of native data can still validate the methodology used. When analysed in

isolation (i.e., not as a background check for the evaluation of the learners' data), native data can deliver a good insight into the functionality of features themselves which is especially relevant for heavily context-dependent phenomena such as tense aspect systems. As stated, the mere fact that a comparison has been carried out is not automatically equivalent to a comparative fallacy. Hence, we strongly recommend carrying out such analyses with a larger data base, also considering some shortcomings of our study such as the lack of spontaneous production data.

All in all, the findings of the present paper are useful in precluding any description of the L2 learners' grammars as unsystematic and unprincipled. The analysis of our data indicates that such a conclusion is valid insofar as one provides an accurate description of the construct that is relevant for both native and non-native speakers. The latter outcome no doubt helps us to dispel the myth about the categorical choices of aspect attributed to native speakers (as exemplified in narrow definitions of aspect that eschew the important contribution of a broad range of contextual factors as described here).³³ It has been clearly shown that the equation of native-like performance and normative grammar is – especially in what regards tense-aspect phenomena – a clear fallacy as it is linked to participants' viewpoint and other factors that are discourse-related. Future studies can prevent this issue from affecting the participants' behaviour in a task by designing a task in which the items are inserted in a context which biases one of the two interpretations as it is not generally the case that the two options can appear in the same context. That said, only the contrast between various learner groups and between different kinds of contexts has revealed the rules upon which the interlanguage is constructed. This procedure has revealed important insights into the learners' representations of tense-aspect features and allows for further investigations both from a linguistic as well as from a didactic point of view.

References

- Bardovi-Harlig, Kathleen. 2002. Analyzing aspect. In M. Rafael Salaberry & Yasuhiro Shirai (eds.), *The L1 acquisition of tense-aspect morphology*, 129–154. Amsterdam: Benjamins.
- Birdsong, David & Libby M. Gertken. 2013. In faint praise of folly: A critical review of native/non-native speaker comparisons, with examples from native and bilingual processing of French complex syntax. *Language, Interaction and Acquisition* 4(2). 107–133.
- Bley-Vroman, Robert. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1). 1–17.

³³ By this formulation, we adapt a precise recommendation coming from one of the anonymous reviewers, who thankfully pointed out the main conclusion drawn from our data.

- Bley-Vroman, Robert. 2009. The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition* 31(2). 175–198.
- Cadierno, Teresa. 2000. La enseñanza gramatical y el aprendizaje de la gramática: el caso del aspecto en español. *Revista española de lingüística aplicada* 14. 53–73.
- Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27. 3–42.
- Comajoan, Llorenç. 2014. Tense and aspect in second language Spanish. In Kimberly L. Geeslin (ed.), *The handbook of Spanish second language acquisition*, 7235–7252. Chichester, West Sussex: Wiley.
- Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Cook, Vivian. 1997. Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingeleses* 34. 35–49.
- Cook, Vivian. 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly* 33(2). 185–209.
- De Miguel, Elena. 1992. *El aspecto en la sintaxis del español: perfectividad e imperfectividad*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- Dekydspotter, Laurent, Bonnie D. Schwartz & Rex A. Sprouse. 2006. The comparative fallacy in L2 processing research. In Mary Grantham O'Brien, Christine Shea & John Archibald (eds.), *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006). The Banff conference*, 33–40. Somerville, MA: Cascadia Press.
- Diaubalick, Tim. 2019. *La adquisición del sistema verbal español por aprendices alemanes. Una comparación entre los tiempos de pasado y los tiempos de futuro*. Tübingen: Narr.
- Diaubalick, Tim & Pedro Guijarro-Fuentes. 2017. L1 effects as manifestation of individual differences in the L2 acquisition of the Spanish tense-aspect system. In Kate Bellamy, Michael W. Child, Paz González, Antje Muntendam & M. Carmen Parafita Couto (eds.), *Multidisciplinary approaches to bilingualism in the hispanic and lusophone world*, 9–40. Amsterdam/Philadelphia: John Benjamins.
- Diaubalick, Tim & Pedro Guijarro-Fuentes. 2019. The strength of L1 effects on tense and aspect: How German learners of L2 Spanish deal with acquisitional problems. *Language Acquisition* 26. 282–301.
- Díaz, Lourdes, Aurora Bel & Konstatina Bekiou. 2008. Interpretable and uninterpretable features in the acquisition of Spanish past tenses. In Juana Muñoz Licerias, Helmut Zobl & Helen Goodluck (eds.), *The role of formal features in second language acquisition*, 484–512. New York: Lawrence Erlbaum.
- Díaz, Lourdes & Konstatina Bekiou. 2006. Lo que las reformulaciones y repeticiones (halladas en los relatos orales) de los aprendices de español L2 pueden decirnos acerca de la adquisición del aspecto verbal del español. In Milka Villayandre Llamazares (ed.), *Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística*, 400–418. León: Universidad de León (Dpto. de Fil. Hispánica y Clásica).
- Doiz Biezobas, Aintzane. 2013. The Spanish preterite and imperfect from a cognitive point of view. In M. Rafael Salaberry & Llorenç Comajoan (eds.), *Research design and methodology in studies on L2 tense and aspect*, 57–88. Boston: De Gruyter.
- Domínguez, Laura, María J. Arche & Florence Myles. 2017. Spanish imperfect revisited: Exploring L1 influence in the reassembly of imperfective features onto new L2 forms. *Second Language Research* 33(4). 431–457.

- Dörnyei, Zoltán. 2006. Individual differences in second language acquisition. *Association Internationale de Linguistique Appliquée Review* 19(1). 42–68.
- duPlessis, Jean, Doreen Solin, Lisa Travis & Lydia White. 1987. UG or not UG, that is the question: A reply to Clahsen and Muysken. *Second Language Research* 3. 56–75.
- Fábregas, Antonio. 2015. 'Imperfecto' and 'indefinido' in Spanish. What, where and how. *Borealis – An International Journal of Hispanic Linguistics* 4(2). 1–70.
- García, Erica C. & F. C. M. van Putte. 1988. The value of contrast: Contrasting the value of strategies. *International Review of Applied Linguistics* 26(4). 263–281.
- Giorgi, Alessandra & Fabio Pianesi. 2004. On the speaker's and the subject's temporal representation: The case of the Italian imperfect. In Jacqueline Guéron & Jacqueline Lecarme (eds.), *The syntax of time*, 259–298. Cambridge, Mass: MIT Press.
- González, Paz. 2008. Towards effective instruction on aspect in L2 Spanish. *International Review of Applied Linguistics* 46(2). 91–112.
- González, Paz & Lucía Quintana Hernandez. 2018. Inherent aspect and L1 transfer in the L2 acquisition of Spanish grammatical aspect. *The Modern Language Journal* 102(3). 611–625.
- González, Paz, Margarita J. Yupanqui & Carmen Kleinherenbrink. 2019. The microvariation of the Spanish perfect in three varieties. *Isogloss A Journal on Variation of Romance and Iberian Languages* 4(1). 113–133.
- González, Paz, & Tim Diaubalick. 2020. Task and L1 effects: Dutch students acquiring the Spanish past tenses. *Dutch Journal of Applied Linguistics (DujAL)* 8(1). 24–40.
- Goodin-Mayeda, C. Elisabeth & Jason Rothman. 2007. The acquisition of aspect in L2 Portuguese & Spanish: Exploring native/non-native performance differences. In Sergio Baauw, Frank Drijkoningen & Manuela Pinto (eds.), *Romance languages and linguistic theory 2005. Selected papers from 'Going Romance', Utrecht, 8–10 December 2005*, 131–148. Amsterdam/Philadelphia: John Benjamins.
- Guijarro-Fuentes, Pedro. 2005. Tiempo y Aspecto: estudio comparativo entre nativos y no-nativos de nivel avanzado de español. *Hispanic Research Journal* 6(2). 99–115.
- Haßler, Gerda. 2016. *Temporalität, Aspektualität und Modalität in romanischen Sprachen*. Berlin, Boston: De Gruyter.
- Hawkins, Roger & Hajime Hattori. 2006. Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research* 22(3). 269–301.
- Heinold, Simone. 2015. *Tempus, Modus und Aspekt im Deutschen: Ein Studienbuch*. Narr Studienbücher. Tübingen: Narr Verlag.
- Hopp, Holger. 2013. Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research* 29(1). 33–56.
- Hwang, Sun Hee & Donna Lardiere. 2013. The acquisition of Korean plural marking by native English speakers. *Second Language Research* 29(1). 57–86.
- Jansen, Silke. 2013. Tempus und Aspekt als linguistisches und sprachdidaktisches Problem: Perspektiven der Kognitiven Linguistik. *Zeitschrift für romanische Sprachen und ihre Didaktik* 7(1). 105–128.
- Judy, Tiffany, Pedro Guijarro-Fuentes & Jason Rothman. 2008. Adult accessibility to L2 representational features: Evidence from the Spanish DP. In Melissa A. Bowles, Rebecca Foote, Silvia Perpiñán & Rakesh Bhatt (eds.), *Selected proceedings of the 2007 second language research forum*, 1–21. Somerville, MA: Cascadilla Press.
- Klein, Wolfgang. 1994. *Time in language. Germanic linguistics*. London: Routledge.

- Klein, Wolfgang & Clive Perdue. 1997. The basic variety, or: Couldn't language be much simpler? *Second Language Research* 13. 301–47.
- Lardiere, Donna. 2008. Feature assembly in second language acquisition. In Juana Muñoz Liceras, Helmut Zobl & Helen Goodluck (eds.), *The role of formal features in second language acquisition*, 106–140. New York: Lawrence Erlbaum.
- Lardiere, Donna. 2009. Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2). 173–227.
- Leonetti, Manuel. 2004. Por qué el imperfecto es anafórico. In Luis García Fernández & Bruno Camus Bergareche (eds.), *El pretérito imperfecto*, 481–508. Madrid: Gredos.
- Llopis-García, Reyes, Juan Manuel Real Espinosa & José Plácido Ruiz Campillo. 2012. *Qué gramática enseñar, qué gramática aprender*. Madrid: Edinumen.
- Mesthrie, Rajend. 2010. New Englishes and the native speaker debate. *Language Sciences* 32. 594–601.
- Ortega, Lourdes. 2013. SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning* 63(1 Suppl). 1–24.
- Palancar, Enrique L. 2005. Los verbos 'saber' y 'conocer' en español: aspecto léxico y morfodinámica. In Margaret Lubbers Quesada & Ricardo Maldonado (eds.), *Dimensiones del aspecto en español*, 17–54. México: Universidad Autónoma de México.
- Quintana Hernández, Lucía. 2010. El aprendizaje del contraste indefinido/imperfecto a través del concepto de aspecto. *Sintagma: Revista de Lingüística* 22. 101–113.
- Ragheb, Marwa & Markus Dickinson. 2011. Avoiding the comparative fallacy in the annotation of learner corpora. In Gisela Granena, Joel Koeth, Sunyoung Lee-Ellis, Anna Lukyanenko, Goretti Prieto Botana & Elizabeth Rhoades (eds.), *Selected proceedings of the 2010 second language research forum*, 114–124. Somerville, MA: Cascadilla Proceedings Project.
- Rojas-Sosa, Denyanira. 2008. The use of past tenses in the Spanish of Lima: Variation in a situation of contact. In Joyce Bruhn de Garavito & Elena Valenzuela (eds.), *Selected proceedings of the 10th hispanic linguistics symposium*, 265–275. Somerville, Mass: Cascadilla Press.
- Rojo, Guillermo. 1990. Relaciones entre temporalidad y aspecto en el verbo español. In Ignacio Bosque (ed.), *Tiempo y aspecto en español*, 17–43. Madrid: Cátedra.
- Rothman, Jason. 2008. Aspect selection in adult L2 Spanish and the competing systems hypothesis. When pedagogical and linguistic rules conflict. *Languages in Contrast* 8(1). 74–106.
- Rothman, Jason & Roumyana Slabakova. 2019. The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition* 40. 417–442.
- Rothman, Jason & Pedro Guijarro-Fuentes. 2010. Input quality matters: Some comments on input type and age-effects in adult SLA. *Applied Linguistics* 31(2). 301–306.
- Salaberry, M. Rafael. 1999. The development of past tense verbal morphology in classroom L2 Spanish. *Applied Linguistics* 20(2). 151–178.
- Salaberry, M. Rafael. 2008. *Marking past tense in second language acquisition: A theoretical model*. New York/London: Continuum.
- Salaberry, M. Rafael. 2018. Advanced conceptualizations of tense and aspect in L2 acquisition. In Paul A. Malovrh & Alessandro G. Benati (eds.), *The handbook of advanced proficiency in second language acquisition*, 361–380. Oxford: Wiley Blackwell.
- Schwartz, Bonnie D. & Rex A. Sprouse. 1994. Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In

- T. Hoekstra & Bonnie D. Schwartz (eds.), *Language acquisition studies in generative grammar*, 317–368. Amsterdam: John Benjamins.
- Schwenk, Hans-Jörg. 2012. Die Vergangenheitstempora im Deutschen und ihr semantisches Potential. *Lublin Studies in Modern Languages and Literature* 36. 35–49.
- Schwenter, Scott A. & Rena Torres Cacoullos. 2008. Defaults and indeterminacy in temporal grammaticalization: The ‘perfect’ road to perfective. *Language Variation and Change* 20. 1–39.
- Shirai, Yasuhiro & Roger W. Anderson. 1995. The acquisition of tense-aspect morphology: A prototype account. *Language* 7(4). 743–762.
- Slabakova, Roumyana. 2009. L2 fundamentals. *Studies in Second Language Acquisition* 31(2). 1–19.
- Slabakova, Roumyana & Silvina A. Montrul. 2003. Genericity and aspect in L2 acquisition. *Language Acquisition* 11(3). 165–196.
- Slabakova, Roumyana & Silvina A. Montrul. 2008. Aspectual shifts: Grammatical and pragmatic knowledge in L2 acquisition. In Juana Muñoz Licerias, Helmut Zobl & Helen Goodluck (eds.), *The role of formal features in second language acquisition*, 106–140. New York: Lawrence Erlbaum.
- Toth, Paul D. & Pedro Guijarro-Fuentes. 2013. The impact of instruction on second-language implicit knowledge: Evidence against encapsulation. *Applied Psycholinguistics* 34(6). 1163–1193.
- Truscott, John & Mike Sharwood-Smith. 2008. Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition* 7(1). 1–20.
- Tsimpli, Ianthi M. & Maria Dimitrakopoulou. 2007. The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research* 23(2). 215–242.
- Wulff, Stefanie & Nick C. Ellis. 2018. Usage-based approaches to second language acquisition. In David Miller, Fatih Bayram, Jason Rothman & Ludovica Serratrice (eds.), *Bilingual cognition and language: The state of the science across its subfields*, 37–56. Amsterdam: John Benjamins.
- Zagona, Karen. 2007. Some effects of aspect on tense construal. *Lingua* 117(2). 464–502.