

The importance of task variability in the design of learner corpora for SLA research

Nicole Tracy-Ventura & Florence Myles

University of South Florida & University of Essex

Key words: oral learner corpora, task variability, tense and aspect, L2 Spanish, Aspect

Hypothesis

Abstract

This cross-sectional study investigates task variability focusing on the use of Spanish past tense morphology in a spoken learner corpus. Sixty L2 learners of Spanish (English L1) from three different proficiency levels (20 per group) and fifteen native speakers completed three communicative tasks (a guided interview, a picture-based narrative, and a historical figures description) and an experimental task, all designed to investigate the acquisition of tense and aspect in L2 Spanish. Data were transcribed in CHAT, and analysed and coded using a specially created interactive coding program that works in combination with the CLAN program (MacWhinney 2000). Results demonstrate significant differences in the emergence and accurate use of past tense morphology across tasks. An additional analysis showed that the less controlled tasks encouraged few instances of more advanced features, suggesting that not all task types are equally successful at eliciting the range of tense-aspect morphological contrasts theoretically relevant for SLA research on tense and aspect.

1. Introduction¹

As the field of learner corpus research (LCR) continues to grow, the number of studies which utilize corpus tools and are informed by theories of second language acquisition (SLA) has also been increasing. This change has come about as both LCR and SLA continue to mature, and as their needs evolve, several researchers have urged for a

¹ This research was funded by Economic and Social Research Council (UK), award number RES-062-23-1075. We are grateful to our colleagues on the SPLLOC project (Laura Dominguez, Rosamond Mitchell, Maria Arche, and Tim Boardman) for their contribution to this research, and we thank the many students who participated in the SPLLOC project. Lastly, we are grateful to two anonymous reviewers and the editors of the IJLCR for their detailed feedback on our initial manuscript; any remaining errors are our own.

rapprochement of both fields (Granger 2009, Hasko 2013, Myles 2008). For example, LCR now includes more hypothesis-testing studies (Bonilla 2014, Lozano & Mendikoetxea 2010, Rankin 2009), following an initial phase where much of the research consisted of descriptions of learner language. Additionally, SLA research has begun to recognise both the potential of large datasets to improve generalizability of findings and the usefulness of corpus tools.

As both disciplines begin to appreciate the strengths that each contributes toward the shared goal of understanding second language (L2) development, issues remain that are worthy of discussion. One important issue focuses on what constitutes a learner corpus, as LCR and SLA seem to adopt different definitions. The field of corpus linguistics (CL) developed alongside computers, when it became possible to investigate actual language use in large collections of electronic texts. Thus, when the field of LCR emerged, it naturally aligned with corpus linguistics methods and theoretical frameworks, and adopted their definition of what constitutes a (learner) corpus, that is a principled collection of naturally occurring spoken or written language. In particular, the emphasis is on authentic, continuous, open-ended, and spontaneous language. In CL the aim is to “investigate how speakers and writers exploit the resources of their language” (Biber et al. 1998: 1), and this aim has been adopted in LCR as well. The field of SLA seeks to understand the underlying L2 knowledge system of learner language, its development, and what impacts upon both. Therefore, SLA research is not only about how learners use a second/foreign language (L2) but also what accounts for changes in language use over time. For this reason most SLA research is hypothesis-testing and analyses both experimental and corpus data; compared to most learner corpora, datasets used in SLA research have tended to be small and typically analysed by hand.

Furthermore, SLA research agendas usually require the occurrence of the linguistic feature under investigation in different contexts of use, hence the use of carefully designed data elicitation tasks where learners are asked to produce spoken and/or written language. Although the data collected through these various means would count as a corpus in a wide sense (i.e. a collection of data), many corpus linguists would object to the use of this term because of the experimental nature of the tasks, and therefore the lack of authenticity and spontaneity (Gilquin & Gries 2009). In fact, some corpus linguists have argued that corpora should not be designed based on internal criteria (i.e., to elicit particular linguistic features) but rather external criteria based on the communicative function of texts (Lozano & Mendikoetxea 2013b, Sinclair 2005). This view implies that corpora should be for general use and not be biased towards certain grammatical or lexical features over others. In relation to LCR, Granger (2008: 261) argues “it is best to restrict the term ‘learner corpus’ to the most open-ended types of tasks, viz. those tasks that allow learners to choose their own wording rather than being requested to produce a particular word or structure”.

While general-purpose learner corpora (i.e. corpora constructed without a specific research agenda in mind) are certainly necessary, the issue becomes more complex depending on the research agenda. If a specific rare structure is the object of a study, then a general-purpose learner corpus or a specialized learner corpus representing a specific text type or register will probably not suffice. For example, the use of quantifiers has attracted much theoretical interest in SLA recently, and they are rather rare in spontaneously occurring speech (Dekydtspotter 2001, Gil & Marsden 2013, Marsden 2009). In order to address such SLA research agendas, it is imperative to ensure the corpus contains multiple examples of the feature(s) under investigation. For this reason, collecting more open-ended samples of learner language can be a

gamble, potentially leading to somewhat limited productions, with learners ‘playing safe’ in order to avoid making errors and not fully demonstrating how much they know. When investigating the development of a specific L2 property, the issue of construct underrepresentation (Norris & Ortega, 2003) can become problematic: how do we know if a structure is not present in the learner’s interlanguage because they have not acquired it yet, or simply because the speech or writing task used does not require its production? In the field of SLA the issues of construct underrepresentation and over-representation (i.e., the use of default forms or the use of unanalysed chunks, see e.g. Myles 2004) are of serious concern and one reason why many SLA studies include experimental methods like cloze-tests and grammaticality judgement tasks that have been designed to test a particular linguistic feature (Mackey & Gass 2005). Such tasks typically also go through vigorous piloting, and validity and reliability testing. One known advantage of experimental data over corpus data is that “they allow the study of phenomena that are too infrequent in corpora” (Gilquin & Gries 2009: 9), which is one reason why several researchers have suggested the combined use of corpora and experimental methods (e.g., Lozano & Mendikoetxea 2013a, Meunier & Litte 2013). For hypothesis-testing LCR, complementing corpora with experimental methods can be particularly important.

On the other hand, corpora have many advantages over experimental methods when it comes to linguistic data analysis. One advantage noted by Gilquin & Gries (2009: 8) is that “the data are from natural contexts; thus, they make it possible to study register/genre questions that are difficult to study experimentally and come with a higher degree of external validity than some experimental designs”. Corpora are thus thought to be more representative of ‘real’ or ‘authentic’ language than decontextualized samples, in the sense that speakers choose how they express

themselves to achieve a communicative goal. This characterisation of corpora, however, can be a little problematic in the context of L2 learners, whose language productions are often far from ‘authentic’ and tend to serve an instrumental goal such as passing an exam or getting a good mark, rather than a purely communicative one. In this regard, we agree with Gilquin & Gries (2009: 6) who claim that “there is actually no strict corpora-experiments dichotomy. Rather, just as linguistic data in general form a continuum of naturalness of production/collection, so do corpora”. They discuss learner corpora as examples of less prototypical corpora, compared to the *British National Corpus* (BNC) for example. What is considered a natural communicative setting can differ for native speakers and L2 learners, and also between L2 learners who are instructed versus those learning naturalistically. Therefore, the content of a learner corpus will, more often than not, be exactly those activities which are natural in the context of a second language classroom (Gilquin & Gries 2009: 7, Granger, 2002: 8). In other words, instructed learner data, unlike the native-speaker data used in CL, will likely be the result of some kind of classroom task. Instructed learners engage in a wide range of communicative activities in the classroom: role plays, speaking and reading activities, writing etc., and learner corpora should reflect this, especially as it is well known that different communicative contexts require different language use for native and non-native speakers alike. For example, corpus linguists have long been interested in documenting the difference between written and spoken language generally, and between various written and spoken registers (see Biber 1988, Biber & Conrad 2009). Learner corpora also need to reflect a variety of classroom language use in order to ensure that the different registers and types of language use are represented.

Having varied data is important from an SLA perspective as well, as it is well known that different tasks elicit different kinds of data (Ellis & Barkhuizen 2005,

Foster & Tavakoli 2009, Tavakoli & Foster 2011). This issue of task variability is also discussed from a language testing perspective where multiple tasks are used to gain a more generalizable picture of learners' language abilities (see Norris et al. 2002). The fact that LCR usually uses large samples of learner language goes some way to addressing the issue of construct underrepresentation, yet it still cannot always guarantee a sufficient number of examples of some linguistic feature(s) of current theoretical interest in SLA, as in the case of quantifiers mentioned above. For that reason, the combined use of rich corpora containing varied data and experimental methods is likely the optimal solution.

In sum, if learner corpora are to help tackle some of the SLA issues of current theoretical interest, the definition of what counts as a learner corpus needs to be expanded to take account of the large body of research in SLA related to avoidance, construct underrepresentation, and task variability. To conduct SLA studies based on learner corpora, we must design corpora that are representative of the different types of language used by the learners and that naturally provide contexts for the use of the grammatical or lexical features under investigation. In this paper, we provide an example of a corpus designed with these considerations in mind, focusing on a much debated SLA topic: the acquisition of tense/aspect morphology. We compare the performance of the same learners across three communicative tasks (interview, narrative retell, and historical figure description) which all lead to (semi)spontaneous continuous discourse. These particular tasks were chosen because they are known for their natural high use of past tense, thus allowing us to investigate the theoretically important lexical and grammatical aspectual combinations relevant for testing the Aspect Hypothesis (Andersen & Shirai 1994, 1996). An additional experimental task was administered, which consisted of a controlled narrative with verb phrase prompts

provided. Having four samples of oral data from each learner enables us to demonstrate the importance of taking into account task variability when designing learner corpora in order to provide sufficiently rich and representative data to analyse learners' interlanguage.

2. Background

2.1. Grammatical and lexical aspect

For English-speaking learners of Spanish, the distinction between the Preterit and the Imperfect² is one of the biggest challenges to overcome. This difficulty is due to the differences in how aspectual meaning is represented in the two languages as will be explained next.

Differences exist across languages between situations presented with a specific endpoint (bounded) and those which are on-going (unbounded). Bounded situations are perfective and reflect an external perspective, whereas unbounded situations are imperfective and reflect an internal perspective. The difference between perfective and imperfective is a grammatical aspect contrast and it is realized across languages by different means. For instance English uses both periphrasis and inflectional forms to encode perfective and imperfective aspect:

- (1) Yesterday it **rained**. - *perfective*
- (2) It **was raining** when I left. - *imperfective, progressive*

² Capital letters are used to differentiate between the grammatical markers (e.g., Preterit and Imperfect) and the semantic categories (e.g., perfective and imperfective) following Comrie (1976).

- (3) It **used to rain/would rain/rained** all the time when I lived in Seattle. – *imperfective, habitual*
- (4) He **was** happy when I saw him. – *imperfective, stative*

In contrast, Spanish primarily uses inflectional forms to mark the perfective/imperfective distinction. Examples (5) – (8) demonstrate how examples (1) – (4) can be translated into Spanish:

- (5) Ayer **llovió**_{pret} - *perfective*
- (6) Ayer **llovía**_{impf} cuando salí - *imperfective, progressive*
- (7) **Llovía**_{impf} todo el tiempo cuando vivía en Seattle – *imperfective, habitual*
- (8) **Estaba**_{impf} contento cuando lo vi. – *imperfective, stative*

In Spanish the Preterit (e.g. example 5) is used for situations in the past that are viewed as being complete, with clear beginning and ending points (perfective). In English, this meaning would typically be expressed by the simple past. The Spanish Imperfect is used to express situations in the past that are still in progress (6), as well as those which are habitual (7) or stative (8). In English each of these situations are expressed using different means. For example, progressivity is expressed by the past progressive. Habituality could be expressed by either the simple past, *used to*, or the conditional *would*. The stative meaning is expressed in English with simple past morphology. In sum, because Spanish grammaticalizes the perfective-imperfective distinction, English speakers must learn to remap their representational knowledge of aspectual concepts onto verbal morphology. As shown above, this is a complex task and it takes time to master.

Another way in which aspect is expressed in languages is by the intrinsic semantic qualities of the predicate, what is known as lexical aspect. Terminology used to describe lexical aspect tends to vary although it shares similar semantic distinctions. For example, Vendler (1967) identified four categories of verbs according to their inherent aspect properties: states, activities, accomplishments, and achievements. Achievements are punctual and instantaneous events (e.g., *At five o'clock we stopped for dinner*). Accomplishments are different in that they have duration (e.g., *Paul built our house in 6 months*) but they are similar to achievements because both are telic and have inherent endpoints. Activities, in contrast, are atelic and lack inherent endpoints but they have duration (e.g., *I swam in the sea*). Finally, states are situations that continue to exist indefinitely until something is done to change them (e.g., *We loved that big oak tree in the park*).

Although all verbs in Spanish are able to take both Preterit and Imperfect morphology, some pairings of lexical and grammatical aspect are prototypical, whereas others are non-prototypical (see Table 1). For example, a telic predicate (achievements and accomplishments) with Preterit morphology is prototypical because both share the feature of boundedness. In contrast, a telic predicate with Imperfect morphology is non-prototypical because achievements and accomplishments are both bounded and have specific beginning and ending points, but the use of the Imperfect forces an unbounded interpretation (e.g., *Paul was building our house*). Activity verbs are atelic and lack inherent endpoints, yet when paired with the Preterit a bounded interpretation is forced (e.g., *Yesterday I swam in the sea*).

Insert Table 1 Here

2.2 L2 acquisition: The Aspect Hypothesis

The Aspect Hypothesis (AH) predicts that learners' emerging use of verbal morphology will reflect lexical aspect differences before grammatical aspect differences (for theoretical discussions see Andersen & Shirai 1994, 1996; Bardovi-Harlig 2000; Salaberry 2008). Although there is still some debate about the specific order of acquisition in L2 Spanish (e.g., Dominguez et al. 2013, Salaberry 2008), in general research has demonstrated that the Preterit emerges first with telic predicates (achievements and accomplishments), whereas the Imperfect emerges first with states and later activities. In other words, when learners begin to use past tense morphology, the prototypical combinations of lexical and grammatical aspect will emerge before the non-prototypical pairings. In fact, most of the production data has provided strong evidence for these early prototypical combinations; however, at what stage learners begin to use non-prototypical combinations has not been well documented.

Major design differences exist across the studies testing the AH, including notable task differences (see Bardovi-Harlig 2013 for a description of the various task types used), and these differences are a potential source of the conflicting results. In fact, some authors (e.g., Bonilla 2013, Shirai 2004) found that support for the AH was linked to the tasks used. In Bonilla's (2013) survey of studies on L2 Spanish, she found that open-ended tasks better supported the AH, whereas Shirai's (2004) survey of L2 English studies demonstrated that studies which used paper and pencil tests, such as cloze-tests or fill-in-the-blank, supported the AH more consistently. The following section will look at this issue in more detail.

2.3. Task variability

Although several researchers (e.g., Bonilla 2013, Shirai 2004, Sugaya & Shirai 2007) have made reference to the role of task type in explaining differential outcomes in research testing the AH, few studies have empirically tested this claim. In a study comparing accuracy rates across tasks in L2 Spanish, Salaberry & Lopez-Ortega (1998) found that only the lower proficiency group's use of past tense morphology varied between a written narrative and a grammar test (either fill-in-the-blank or multiple choice depending on the group), with more mistakes occurring in the grammar test; the more advanced group scored consistently across tasks. The findings suggest that while attention to form was one source of task variability, it was not the only one.

Communicative control of the L2 grammar or “the learner’s ability to manage and utilize their linguistic resources in the TL” (Salaberry & Lopez-Ortega 1998: 518), also influenced learners’ performance. In this study, the written narrative was found to provide learners with more communicative control than the more controlled grammar test because learners were able to use those linguistic resources they were most comfortable with and avoid problem areas, thus increasing accuracy of past tense morphology use.

Contrary to Salaberry & Lopez-Ortega’s (1998) finding that learners were least accurate with a multiple choice grammar test, Bardovi-Harlig & Reynolds (1995) found that learners were more accurate with this task type than in written narratives. Additionally, Bardovi-Harlig (1998) and Camps (2002) showed that oral narratives tend to elicit the least accurate use of past tense morphology. Based on these results, it is difficult to conclude whether task mode (written vs. spoken) may have been the cause of learners’ variable performance or whether these tasks differed in the kinds of lexical-grammatical aspect combinations they elicited. For example, Lafford (1996)

found no examples of the Imperfect with achievements or accomplishments (non-prototypical pairings) in her study using a retelling of “The Sorcerer’s Apprentice” from the film *Fantasia*. As Bardovi-Harlig (2000, 2005) points out, film retellings tend to elicit a much higher percentage of achievement verbs than the three other lexical aspect classes. These results highlight the fact that certain types of discourse are much more likely to lead to the use of certain types of verbs, e.g., in storytelling, telic predicates typically appear more often in the foreground, whereas atelic predicates appear in the background.

The variety of lexical-grammatical aspect combinations can also be influenced by other variables such as task topic, prompts, and type of narrative. Duff (1993) compared participants’ verb production across three different tasks and found that both images and topic influenced the range of verb types used. The widest range of verb types were used in a discussion task compared to a picture-description task and a folk story narrative task. Similarly, research on lexical diversity more broadly has also demonstrated wide variation exhibited by the same learners on different tasks. For example, story retellings elicit less diverse vocabulary than more open-ended tasks (David 2008), and conversations on familiar topics elicit a much larger proportion of (less diverse) formulaic language than story retelling tasks (Cordier 2013). Additionally, Ayoun (2004) found that task topic influenced learners’ use of the French ‘*Imparfait*’. The pre-test topic about a birthday party elicited more habitual meanings than the post-test topic about a vacation which elicited more progressive meanings.

Related to task prompts, many researchers have noted the difficulties in eliciting the past tense with prompts such as “What happened in the story?” as both learners and native speakers often use the historical present in response (e.g., Salaberry 1999).

This kind of prompt has also inadvertently caused learners and native speakers to focus on main events (i.e., foreground) rather than on background descriptions (Bardovi-Harlig 2013). What this means for past tense morphology is a higher use of the Preterit (usually found in the foreground with achievements and accomplishments) when compared to the Imperfect (usually found with activities and states in the background). The type of narrative, whether impersonal or personal, also appears to impact the amount of foreground or background since impersonal narratives tend to provide more examples of foreground than background (Liskin-Gasparro 2000), whereas personal narratives often include more background (Camps 2002, Salaberry 2003).

To summarize, when studies have investigated learners' use of past tense morphology in more than one task, the results have shown that their use varies across tasks. However, most studies have tested students on tasks of different modes such as a fill-in-the-blank task and a written narrative, or a written narrative and a spoken narrative. This is problematic because these tasks tap different kinds of knowledge (explicit or implicit) and differ in planning time and processing demands. One notable exception is Pienemann (1998) who investigated the performance of six learners and native speakers across a range of six oral tasks, three with an interviewer and three in pairs. The three tasks with the interviewer were 1) describing the daily activities of specific people (habitual actions), 2) a story completion task based on pictures, and 3) an informal interview. The pair tasks included 1) a jigsaw picture sequence, where each partner had different images and they needed to put a story together, 2) a spot the difference task, and 3) a "Meet your partner" activity where they had to find out information about each other to introduce the other person to the researcher. The consistency of learners' interlanguage production was investigated for selected

syntactic and morphological structures using measures of both emergence and accuracy. Important to this analysis was whether in some tasks learners overproduced or underproduced structures above or below the developmental level apparent in other tasks.

The results for both syntax and morphology demonstrated that some tasks failed to provide any linguistic contexts for a specific stage, or too few to adequately judge whether a learner has acquired a rule or not. Additionally, lexical choice, as determined by the task, affected accuracy rates both positively and negatively. That is, if a learner used the same verb with correct/incorrect inflection several times within one task, then the accuracy rate increased/declined. Based on these results, Pienemann argues for emergence as a more stable acquisition criterion because it is less sensitive to frequency, as it is categorical: a form has emerged or it has not.

Pienemann's study, although not specifically about the acquisition of tense-aspect morphology, has several relevant implications. For example, studies of tense-aspect need not only count how often a certain form is supplied but also how many obligatory contexts were available in order to know whether the learner had an opportunity to demonstrate productive use of the form in all appropriate contexts. Unfortunately, very few studies have provided information about the number of obligatory contexts (but see Camps 2002). Furthermore, a full investigation of the AH should include all possible combinations of lexical and grammatical aspect of theoretical interest (e.g., eight in Spanish if using Vendler's four-way classification system: four lexical aspect classes x two grammatical aspect forms) in order to investigate whether learners are overproducing prototypical forms in non-prototypical obligatory contexts.

The mixed results of previous research could be due at least in part to the unlikelihood of open-ended tasks to provide all the contexts necessary to draw an accurate picture of learners' use of aspectual contrasts. One way to get around this issue is to use carefully designed tasks that, as much as possible, naturally create contexts for the forms under investigation. In fact, it may be the case that the best solution is a range of communicative activities completed by the same learners in combination with more controlled experimental tasks. To test this claim, the current study investigates task variability, both intra- and inter-learner, and focuses on the following research questions:

1. Does learners' use of past tense morphology vary across oral tasks (both in terms of emergence and accuracy)?
2. Do certain task types provide learners with fewer/more opportunities to demonstrate both prototypical and non-prototypical combinations of lexical and grammatical aspect in oral production?

Our first hypothesis for question one is that when acquisition is measured using an emergence criterion, then learners' performance will vary depending on whether the task provides contexts for the more advanced stages of acquisition, i.e. non-prototypical pairings which are the most difficult to elicit/acquire. If these contexts are absent, learners' underlying competence could be underrepresented in the data. Similarly, we hypothesize that if accuracy is used as the criterion for acquisition, then learners' performance will also vary depending on the task. The more open-ended tasks will allow learners to use frequent examples of the forms and verb types they know best and avoid more difficult ones, leading to more accurate use overall (Salaberry & Lopez-Ortega 1998, Pienemann 1998).

Our hypothesis for question two is that task type matters. Specifically, more open-ended tasks will encourage the use of more prototypical pairings, or early acquired forms, than nonprototypical pairings because prototypical pairings occur more frequently in learner and native speaker production (Bardovi-Harlig 2000, 2005; Lafford 1996). In order to ensure that learners have the opportunity to produce nonprototypical pairings, it might be necessary to resort to experimental tasks which provide the required contexts for such pairings.

3. Methodology

3.1 Participants

Participants included sixty English L1 learners of Spanish representing three distinct proficiency levels (twenty learners each) based on their length of Spanish instruction and fifteen age-matched native speakers from Spain. Although we acknowledge that an independent measure of proficiency would have been a more objective measure than institutional level (see e.g. Callies et al. 2014, Leclercq & Edmonds 2014), we felt that the large differences between the groups in amount of instruction would ensure very distinct proficiency levels. We will return to the issue of proficiency again when we describe the learner corpus and how we used it to validate these proficiency differences.

The beginner group included secondary school learners who were in year 10 of the British school system (14-15 years old) and in their second year of studying Spanish. The intermediate group included learners who were in year 13, the final year

of a sixth form college (17-18 years old). They had been studying Spanish for five years. The advanced group included learners who were in their final year of a Spanish bachelor's degree at a UK university (21-23 years old), and had spent a year abroad in a Spanish-speaking country. The native speaker group included five people from each of the age groups represented by the learners and were all from Spain (fifteen in total).

3.2 Procedure

Participants met individually with a member of the research team (the first author and two experienced research assistants) at their school/university. All members of the research team participated in training sessions to limit the potential effect that the interviewer could have on data collection. All tasks were audio-recorded. The order of tasks was randomized for each participant except for the “Famous People” task, described below, which was used as a warm-up for the guided Interview:³ All instructions were given in English for the learners and in Spanish for the native speakers. Vocabulary help was provided as needed, but the interviewers were instructed not to provide any past tense forms. For example, if a learner asked how to say “he lived”, the interviewer would reply giving the infinitive form of the verb, saying something like “to live is *vivir*”. If learners spoke in English, they were encouraged to change to Spanish. Time on task was not controlled as completion of the task was the priority. All tasks were piloted to ensure that all students, including the youngest group, could manage the task and that vocabulary was appropriate and instructions clear. Tasks were revised according to the results of piloting.

³ All the data used in this study are freely available through the SPLLOC website: <http://www.splloc.soton.ac.uk> and on <http://www.talkbank.org>. In the recordings and transcripts available online, the “Famous People” task is not a separate task but rather part of the interview. For the purpose of this study, however, we decided to separate them into different tasks.

3.3 The corpus

In order to investigate the L2 development of tense and aspect, effective oral tasks need to 1) naturally provide contexts for the past tense as authentically as possible, 2) elicit a variety of lexical aspect classes in both prototypical and non-prototypical contexts, and 3) be rich in background and foreground (if a narrative). A review of the literature showed that no task was found to meet our design criteria; therefore, three open-ended communicative tasks were designed specifically for the project, in addition to an experimental task that included a large number of items focusing on non-prototypical pairings. A summary of each task is provided in Table 2 and word counts are provided in Table 3. All tasks are described in more detail below.

Insert Table 2 Here

Insert Table 3 Here

Although a separate measure of proficiency was not administered, to further demonstrate proficiency differences between the groups, we calculated a measure of lexical diversity known as *D* (Malvern & Richards 2002) which has been shown to correlate positively with general language proficiency (e.g., Tracy-Ventura et al. 2014, Yu 2010). A one-way analysis of variance (ANOVA) confirms the difference between groups on average *D*-scores is significant [$F(3, 71)=86.82, p<.001$] and post-hoc comparisons using the Tukey HSD test indicated the average *D*-scores were significant between all groups at $p<.01$.

The interview task, based on participants describing memorable events from their childhood, can be considered the most open-ended and is shown in Figure 1.

Insert Figure 1 Here

Learners were first shown a timeline with five boxes to represent different stages in their lives. For example, the first box read *mi primer recuerdo* ('my first memory') and the last box *el fin de semana pasado* ('last weekend'). Participants were asked to describe what they remembered about these times in their lives in hope that they would use short personal narratives to describe these events. The interviewer would ask several follow-up questions to try to elicit more extended language samples. Learners were given approximately two minutes to consider the events they would talk about before starting. The following excerpt is taken from the interview of an advanced student talking about what she did during her year abroad in Spain⁴

(9) *INT: y qué hacías_{simpf} en España ? [What were you doing in Spain?]

*I88: Yo trabajé_{pret} como auxiliar en un instituto con jóvenes que tienen doce hasta dieciocho años y yo ayudé_{pret} <en las> [/] en las clases con su inglés y bueno. Tenía_{impf} que preparar ejercicios y actividades para que practiquen. Más importante para que hablan porque en el instituto hacían mucho trabajo escrito en inglés.

[I worked as an assistant in an institute with kids who were 12-18 years old and I helped <in the> [/] in the classes with their English and yea. I

⁴ As described later, the transcription conventions used in this study follow CHAT (MacWhinney 2000). < > is used to indicate language which is repeated and is followed by [/] if it is purely a repetition or [//] for retracing with a correction.

had to prepare exercises and activities so that they'd practice. It was more important that they talked because in the institute they did a lot of written work in English.]

In the "Famous People" picture-description task participants were shown pictures of famous historical figures and asked to explain why these people were famous, what they had accomplished in their lives, and what they were known for. All pictures were of deceased people so that the use of the past tense would be natural. Some examples were Princess Diana, Queen Victoria, Winston Churchill, John Lennon, and Albert Einstein. The learners were asked to pick four people to describe and were given one minute to plan what they would say before starting. The interviewer also asked questions when the learners had difficulty and provided vocabulary help as needed. It was assumed that this task would encourage the use of more non-prototypical Preterit pairings including activities and states, which would be natural in this context as the learners would be describing someone's life as complete and bounded. The following excerpt is from a participant in the beginner group who struggles to produce appropriate vocabulary to describe John Lennon. He also overwhelmingly relies on the stative verb *ser* ('be') which he uses in the Preterit only, sometimes correctly.

- (10) *I03: John Lennon <fue_{pret} un> [//] (.) fue_{pret} una mús(xx) [/]
musician@s:d [//] cantiamos@n⁵ [//] cantía@n [//] cantista@n y fue_{pret}
en [/] en un grupo que [/] que [/] que se llama el the@s:det Beatles .

⁵ The @n symbolizes an invented form. Because this student uses many to come up with a word for the concept "musician", the English translations are not exact.

<Fue_{pret} mató_{pret}> [//] (.) <fue_{pret} un> [//] fue_{pret} mató_{pret} en [/] en América .

[John Lennon <was a> was a mus(xx) [/] musician [//] sing [//] sing [//] singist and was in [/] in a group that [/] that [/] that is called the Beatles. <He was he killed> [//] <he was a> [//] he was he killed in [/] in America.].

Two picture-based impersonal narratives were also included, one more open-ended and the other more experimental. The first narrative, “Nati y Pancho”, was adapted from the children’s book *Missing* (Langley 2000). The storyline was changed slightly to allow for the inclusion of several habitual actions. In order to naturally create contexts for this meaning, two discourse prompts were included to raise awareness of habitual and one-time events. For example, on the first page of the picture-book was the sentence *Todas las mañanas eran iguales* (‘Every morning was the same’) (see Figure 2 for an example). A few pages later, the phrase *hasta que un día...* (‘until one day...’) appeared to signal that the habitual background information was over and the main plot of the story was about to begin. Several of the illustrations depicted telic events (e.g., reading a book, painting a picture) which appeared in habitual (imperfective) contexts, exactly those examples of non-prototypical pairings that are characteristic of more advanced stages but are less frequent in natural discourse.

Insert Figure 2 Here

All participants were given time to preview the story before starting the oral retelling to a member of the research team. During the retelling participants followed along with the book, turning the pages as they went. The following excerpt is from a participant in the intermediate group:

- (11) *C57: Todas las mañanas eran_{impf} iguales. Natalia su día <empezó_{pret} o> [//] siempre empezaba_{impf} cuando Natalia se despertó_{pret} [: despertó_{pret}] [*] y leía_{impf} una [//] un libro a sus juguet(es) [/] juguetes y ella pintaba_{impf} una [/] una pintura y construía_{impf} un [//] una casa de cartón (.) y quizás salió_{pret} con su [/] su amigo en bicicleta y jugaba_{impf} en [/] en la culum(xx) [//] +/.

*INT: columpio .

[*C57: Every morning was the same. Natalia, her day <started or> [//] always started when Natalia woke up and would read a [//] a book to her toy [/] toys and she would paint a [/] a picture and she would build a [//] a cardboard house and perhaps she went out with her [/] her friend on her bike and she would play on [/] on the play(xx).

INT: playground]

The other picture-based narrative, “Las Hermanas”, was the controlled experimental task. In contrast to “Nati y Pancho”, it included illustrations with accompanying infinitival verb phrases (e.g., *leer un libro* ‘read a book’) as shown in Figure 3.

Insert Figure 3 Here

Participants were told to use these verb phrases when retelling the story but that they could also add more information if they wanted. An artist was commissioned to illustrate the story, which was written by the research team. This story specifically targeted non-prototypical preterit pairings and non-prototypical imperfect pairings. In fact, twenty of the twenty-five verb phrases provided were non-prototypical pairings (e.g., read a book, paint a picture, write a story, wake up early, and finish her homework early in imperfective contexts, see Figure 3). Occasional discourse prompts were also added to clarify some parts of the storyline (e.g., *Gwen de niña...cada fin de semana*–‘When Gwen was a child...every weekend’). As in the other tasks, participants were given time to preview the story before beginning the oral retell. Although this task is more controlled than the others, the benefit is that the researcher can be more confident of what the learner is trying to express and how, in order to assess their ability to produce non-prototypical pairings in particular. (12) is an excerpt from a native speaker retelling this narrative.

- (12) *H94: Gwen de niña cada fin de semana leía_{impf} un libro, escribía_{impf} un cuento, pintaba_{impf} un cuadro y durante la semana se despertaba_{impf} temprano. Terminaba_{impf} sus deberes temprano. Mientras que Sarah de niña los fines de semana jugaba_{impf} al fútbol y veía_{impf} una película. Durante la semana iba_{impf} al colegio en bicicleta y llegaba_{impf} tarde a clase.

[When Gwen was a child, every weekend she would read a book, write a story, paint a picture, and during the week she would wake up early. She would finish her homework early. Meanwhile Sarah as child, every

weekend she would play football and watch a film. During the week she would go to school by bike and she would arrive late to class.]

3.4 Data coding and analysis

The advantage of large electronic corpora is that the data can be coded systematically and (semi-)automatically, making subsequent analyses much easier and more powerful. We chose to transcribe all audio recordings according to the Codes for Human Analysis of Transcripts (CHAT) conventions for later use with the Computerised Language Analysis (CLAN) program (MacWhinney 2000) which is freely available as part of the Child Language Data Exchange System (CHILDES)⁶. Transcriptions were then checked, anonymised, and tagged using the morphosyntactic analysis program (MOR), also available as part of CLAN. An interactive coding program, called VCX ('verb in context') was developed by the research team to add an extra layer of tagging in all transcripts, exactly those variables that were theoretically relevant to investigating our research questions (see also Housen 2002, who created similar analytic codes in CLAN). These additional tags allowed for later automatic analysis of various aspectual and discursive features (e.g., lexical aspect class, obligatory context, foreground/background, correct suppliance etc). The coded features which are pertinent to the current study include lexical aspect class, obligatory context, and correct suppliance. An example of the coding is shown in (13).

- (13) *C70: todas las mañanas a Pancho le gustaba dormir hasta muy tarde.

⁶ <http://talkbank.org/>

```

%mor: det:indef|todo-FEM-PL=all det:art|el&FEM-PL=the
      n|mañana-PL&FEM=morning prep|a=to n:prop|Pancho
      pro:per|le=him vpas|gusta-13S&PAS=like vinf|dormi-
      INF=sleep prep|hasta=until adv|muy=very adv|tarde=late.
%vcx: verb_STA|gusta-13S&PAS=like IMPF CORR BACK
%vcx: verb_ignore|dormi-INF=sleep INFIN CORR IGNORE

```

In this example, the participant's utterance is shown in the line beginning with *C70. The %mor line includes a morphosyntactic tag for each word occurring in the utterance. Below that is the first %vcx line, corresponding to the first verb in the utterance: *gustaba* ('he liked'), which was tagged as vpas|*gusta*-13S&PAS. V is the tag for verb, PAS for Imperfect, 1 for first person, and 3 for third person (*gustaba* is the same form for first and third person). Because this verb occurs in an obligatory past context, it was coded for aspectual and discursive features. In this example, it was a stative (verb_STA), imperfective was the obligatory context (IMPF), it was correctly inflected (CORR), and it occurred in the background of the narrative (BACK). In comparison, the second %vcx line corresponds to the verb *dormir* ('sleep') which is an infinitive form and, therefore, need not be coded for aspectual and discursive features because it is correctly non-finite. In that case, the coding option is 'IGNORE'. When infinitive forms were incorrectly produced in cases when an inflected form should have been used, the %vcx lines were coded normally for all the aspectual and discursive features. An example of this is shown in (14) with the verb *volver* ('return'). The Imperfect *volvían* ('they returned') was required in this context, but this participant from the beginner group produced the infinitive:

(14) *C12: volver to@s:prep la casa juntos .
 %mor: vinf|volve-INF=return L1|to det:art|el&FEM&SG=the
 n|casa&FEM=house adj|junto-MASC-PL=together .
 %vcx: verb_ACH|volve-INF=turn IMPF INCR BACK

Additional programs were written by the research team to automatically count the frequency of specific coding features in the %vcx line that are relevant to the research questions investigated in the current study (e.g., the number of obligatory perfective and imperfective contexts by lexical aspect class).

Two raters coded all verb phrases occurring in obligatory past contexts for these features. A total of 8,743 verb phrases were coded by each rater. Inter-rater reliability was 90%, and all disagreements were resolved through discussion. The lexical aspect class coding scheme followed is supplied in Appendix 1 and is based on Vendler's (1967) four categories of state, activity, accomplishment, and achievement and was adapted from Camps (2002) and Arche (2006). Obligatory context was coded based on the form required (PRET, IMPF, PRES, INFIN, etc) and correct suppliance was coded as CORR or INCR depending on the form produced, either Preterit or Imperfect. Correct subject-verb agreement did not influence the coding. For example, if the context required a Preterit form, any Preterit form was considered correct. When coding for accurate suppliance the surrounding discourse was considered, not just the sentence the verb phrase occurred in.

For research question one we chose to consider that a form had emerged when two non-formulaic verb types were supplied in each category (i.e. excluding very common fixed routines used in classrooms such as *me llamo* 'I'm called' which learners produce as a single unit, without being able to productively use the verb

llamar in other contexts - see Pienemann et al. 1988). This decision was made to avoid counting a learner as having acquired the forms on the basis of just one example which might have been rote-learned with a common verb prior to the productive use of that structure. We also used the same criterion for obligatory contexts (OC). That is, in order for a learner to be counted as having an OC, they needed to have at least two OCs for that lexical aspect class.

Accuracy was operationalized as suppliance in obligatory context (Pica 1983). Separate mixed between-within subjects ANOVAs were conducted to investigate whether accuracy scores varied across tasks (research question one). The two independent variables were group (between-subject) and task (within-subject), and the dependent variable was either the Preterit or Imperfect score.

For research question two we only analysed the obligatory contexts for each lexical aspect class in both past forms. In other words, we did not analyse the actual forms produced, only the contexts for the Preterit and the Imperfect that the participants had created in each task. Various chi-square tests of independence were conducted to test whether frequency differences across tasks were statistically significant.

4. Results

4.1 Use of aspectual morphology across tasks

Research question one aimed to investigate whether learners' use of past tense morphology varies across tasks using both emergence and accuracy as

operationalizations of acquisition. The first set of results presented focus on prototypical and non-prototypical Imperfect combinations. For reasons of space, we chose to focus solely on the Imperfect here (rather than both the Preterit and Imperfect), because previous research has shown that much less is known about the Imperfect. This is likely due to the inability of the tasks used in previous studies to create sufficient contexts for the Imperfect compared to the Preterit (Bardovi-Harlig 2005, 2013).

Table 4 displays results for the number of participants, by task and group, who a) created opportunities to produce all four lexical aspect classes in the Imperfect, and b) demonstrated emergence of each class in the Imperfect. By opportunities, we mean that there were obligatory contexts (OC) in the task requiring them to use the Imperfect with these lexical aspect classes. As a reminder, the prototypical Imperfect pairings are states and activities, and the non-prototypical pairings are achievements and accomplishments.

Insert Table 4 Here

As evident from Table 4, fewer participants created contexts for non-prototypical pairings (achievements and/or accomplishments in the Imperfect) in the interview and the “Famous People” tasks compared to the narratives where those contexts were specifically created in the tasks through the use of pictures and discourse prompts (e.g., “Every morning was the same” to start the “Nati y Pancho narrative”). For example, in the beginner group data there were no participants who created obligatory imperfective contexts with achievements in the interview or the “Famous People” task. In contrast, all participants created obligatory contexts in the “Hermanas” story. The difference across tasks between opportunities to produce the Imperfect with achievements and accomplishments did not affect the number of learners who demonstrated emergence in the beginner group because they had not yet begun to use the form. This finding differs from the intermediate and advanced groups where large differences are found in the number of participants demonstrating emergence across the tasks. Using the advanced data as an example, no learners showed evidence of emergence on the “Famous People” task, only one learner on the interview task, eight learners on the “Nati y Pancho” narrative, and eleven learners on the “Hermanas” story. Therefore, if the narrative tasks had not been used, there would be very little evidence to suggest that the advanced learners were at this late stage of acquiring non-prototypical pairings.

Another result here worthy of discussion is the increased number of imperfective obligatory contexts with achievements and accomplishments evident in the native-speaker group in the interview compared to the three learner groups. For example, eight participants in the NS group created opportunities to use achievements in the Imperfect compared to only one student in the advanced group. Also in the interview, seven participants from the advanced group and ten from the NS group created opportunities to produce accomplishments in the Imperfect. The difference between

groups is likely due to overall proficiency and a more diverse range of vocabulary as demonstrated by their *D*-scores (see Table 3). Note example (14) of a native speaker from the interview. Here she talks about her first memory and describes many habitual activities she did as a child which requires use of the Imperfect. This example also includes an achievement verb, *llegar* ('arrive') at the end.

(15) *I95: Mi primer recuerdo probablemente fuera ... con mi abuela y con mis primos porque de pequeña yo pasaba_{impf} los veranos con ella en y con todos mis primos en una casa cerca de la playa .

[My first memory probably was ... with my grandma and with my cousins because when I was little I spent the summers with her and with all my cousins in a house close to the beach].

*INT: y qué cosas hacían ? [And what things would you all do?]

*I95: Solíamos_{impf} ir a la playa . Salíamos_{impf} siempre con la bicicleta a jugar con los amigos. Siempre llegábamos_{impf} heridos a casa .

[We used to go to the beach. We'd always go out on our bikes to play with friends. We'd always arrive home hurt.]

In contrast, participants from the intermediate and beginner groups did not provide as much description about their first memory and tended to use the same frequent stative verbs (e.g., *ser/estar* 'be', *tener* 'have'). In example (15), a participant from the intermediate group also incorrectly produces the Preterit in several contexts where the Imperfect was required.

- (16) *I66: Mi primer recuerdo...mi primer recuerdo tuve_{pret} tres años y fui_{pret} mi cumpleaños y tuve_{pret} un oso de peli(xx) peluche que quiero [//] queré_{pret} [: quise_{pret}] [*] mucho y fue muy fa(xx) muy feliz.
[My first memory...my first memory I was three years old and it was my birthday and I had a bear of ... stuffed that I like liked [*] a lot and it was very happy.]

Table 4 also includes examples of prototypical pairings of states and activities in the Imperfect. In the interview, all groups created more opportunities to produce these prototypical pairings than non-prototypical pairings. In the “Famous People” task, few students created contexts for activities in the Imperfect, yet when states and activities are taken together, learners had several more opportunities to produce the Imperfect with these lexical aspect classes than achievements and accomplishments. Participants in all groups created several contexts for both activities and states in the Imperfect in the “Nati y Pancho” narrative, although this was not the case for the experimental “Hermanas” task where learners created few opportunities overall to produce activities in the Imperfect. This is due to the design of the task which focused more on the non-prototypical activity-Preterit pairings.

Results of learners’ accuracy scores across tasks are displayed in Table 5. Native speakers are not included because their accuracy scores were 100%. As shown, all groups were more accurate using the Preterit than the Imperfect in all tasks except for the beginner group who scored slightly higher on the Imperfect in the “Famous People” task (30.89 on the Imperfect compared to 26.39 on the Preterit). This higher score was due to the large number of stative verbs produced in the Imperfect (e.g., *era*,

estaba – ‘was’). All groups were more accurate on the interview and “Famous People” tasks compared to both narratives.

Insert Table 5 Here

To test whether these differences were statistically significant, separate mixed between-within subjects ANOVAs were conducted with group (between subject) and task (within-subject) as independent variables and either Preterit or Imperfect as dependent variables. These analyses test whether there are main effects for group and task and whether there is an interaction effect between these two variables.

Specifically, they will tell whether learners’ Preterit and Imperfect scores change across tasks. They also test whether the group averages differ, as well as whether the change in scores by task is different for the three groups. The results of the Preterit scores are presented first.

A significant interaction effect was found [$F(6,110) = 2.16, p = .05$, partial eta squared = .105], however the effect size was low, suggesting a weak interaction between task and group. Additional tests were conducted to examine differences across tasks within each group to explain the interaction effect. Because multiple paired sample t-tests were conducted (six for each group), we adjusted the p -value to reflect this using .0083 to determine statistical significance ($.05/6$). Significant differences were found for the beginner and intermediate groups between certain tasks only, although the advanced group showed no differences. In other words, task type did not affect the advanced group’s performance on the Preterit. The beginner group performed significantly better on the interview compared to the “Nati y Pancho” narrative ($p = .004$), and on the interview compared to “Las Hermanas” ($p = .002$). The

only difference for the intermediate group was that they scored significantly higher on the interview compared to “Las Hermanas” ($p=.005$).

The same tests were conducted on the Imperfect scores. Significant main effects were found for task [$F(3, 52)=5.03, p=.004$, partial eta squared=.225] and group [$F(2, 54)=27.58, p<.0005$, partial eta squared=.505] only. The interaction effect did not reach statistical significance [$F(6, 104)=.96, p=.454$]. Post-hoc tests revealed significant differences between all groups on their Imperfect scores. Post-hoc tests comparing each groups’ scores across tasks only revealed significant differences for the intermediate group. They performed significantly better on the “Famous People” task compared to the “Nati y Pancho” narrative ($p=.005$) and on the interview compared to “Nati y Pancho” ($p=.007$).

In sum, the results for research question one demonstrate that learners’ use of past tense morphology varies across oral tasks, irrespective of whether emergence or accuracy criteria are used. Using emergence as the criterion for acquisition, we see that when the task does not naturally provide learners with opportunities to demonstrate advanced stages of acquisition (e.g., telic events in the Imperfect), learners do not create the necessary contexts, and as a result we would have underestimated their knowledge of these target forms. The lack of opportunities to produce more advanced forms affected the intermediate and advanced learners more than the beginner group because many of the intermediate and advanced learners were actually able to produce the more advanced forms when given the opportunity (see Table 4). Using accuracy as the criterion for acquisition, the most variability was evident for the intermediate group. They scored nearly 80% on the Preterit in the interview and “Famous People” tasks but only 64% and 54% in the “Nati y Pancho” and “Las Hermanas” narratives, respectively. Their Imperfect scores were also highest on the interview and “Famous

People” tasks (49% and 54% respectively) compared to the “Nati y Pancho” and “Las Hermanas” narratives (32% and 47% respectively). The beginner group only varied across tasks on the Preterit, scoring 42% on the interview but 26% or below on the other three tasks. The advanced group’s performance on the Preterit and Imperfect was consistent across tasks. In conclusion, tasks varied widely in the extent to which they provide opportunities for beginner and intermediate learners to demonstrate their acquisition of target aspectual contrasts. We will return to this finding later.

4.2 Variability across tasks in eliciting necessary contrasts

Research question two aimed to investigate whether certain tasks provided learners with more/fewer opportunities to produce both prototypical and non-prototypical pairings of lexical and grammatical aspect in oral production. The purpose of this analysis was to investigate how effective the tasks were in eliciting the various combinations of lexical and grammatical aspect.

Table 6 displays the number of obligatory perfective and imperfective contexts in the four tasks based on the participants’ actual production (tokens, not types, across all participants).

Insert Table 6 Here

Taken together, roughly equal amounts of obligatory perfective and imperfective contexts were created, 49% and 51% respectively, with some variation in number of tokens produced across tasks. When the tasks are considered on their own and the frequency of obligatory perfective and imperfective contexts are separated by lexical aspect classes, a different picture emerges. We will now consider each task in turn. For this analysis we combined the counts for the learners and the native speakers because

as shown in Table 6 they produced similar proportions of perfective and imperfective obligatory contexts in all tasks.

Figure 4 displays the results of the interview. Achievements (ACH) and accomplishments (ACC) overwhelmingly occurred in perfective contexts (88% and 84% respectively), whereas activities (ACT) are fairly evenly split between perfective and imperfective contexts (48% vs 52%) and states (STA) more often occurred in imperfective contexts (73%). It is also of interest to note that states were by far the most frequent lexical aspect class produced in this task, with 50% (1287) of the total verbs produced. Accomplishments made up 20% (511), followed by activities with 17% (441), and achievements with 14% (360).

Insert Figure 4 Here

The purpose of the “Famous People” task was to create contexts for more non-prototypical Preterit pairings (activities and states). Compared to the interview, the percentage of perfective obligatory contexts increased across all lexical aspect classes, although most notably with activities and states, both nonprototypical pairings (see Figure 5). To test whether these differences were statistically significant, a chi-square test of independence was performed. The result was significant, $\chi^2 = (3, N=2217) = 114.01, p < .001$, and the effect size, Cramer’s V, was moderate $= .23$ (Cohen, 1988). 76% of activities occurred in perfective contexts (compared to 48% in the interview) and 39% of states (compared to 27% in the interview). States were the most frequent lexical aspect class produced, making up 61% (956) of all obligatory contexts, followed by achievements with 17% (267), activities with 13% (210), and accomplishments with only 9% (145).

Insert Figure 5 Here

Figure 6 displays results for the “Nati y Pancho” story. Here we see that the percentages of perfective and imperfective obligatory contexts differ from the two tasks described so far, especially with regards to the number of imperfective contexts with accomplishments and achievements. A chi-square test of independence confirms that the difference across tasks for accomplishments and achievements is statistically significant, $\chi^2 = (2, N=598) = 8.86, p=.01, \text{Cramer's } V=.12$. As a reminder, this task began with a series of habitual actions, all requiring the Imperfect, and several of the pictures were chosen because they depicted telic events (e.g., painting a picture, reading a book, meeting at the corner, leaving somewhere) thus all non-prototypical pairings. The percentage of imperfective contexts with accomplishments was 53% (236) and 29% (215) with achievements. In contrast to the two tasks described thus far, the number of obligatory contexts across the four lexical aspect classes was more balanced in the Nati y Pancho story. Achievements had the most contexts overall making up 31% (748), followed by states with 27% (654), activities with 24% (586), and accomplishments with 18% (447). In both the interview and the “Famous People” task, states represented 49% and 61% (respectively) of all the past obligatory contexts.

Insert Figure 6 Here

Lastly, results from the experimental task, “Las Hermanas”, show a very different picture (see Figure 7). This task was designed specifically to elicit non-prototypical pairings of both the Preterit and the Imperfect and infinitival verb phrases were

provided underneath each picture (unlike in “Nati y Pancho”). The highest number of non-prototypical imperfective obligatory contexts were elicited in this task compared to the other three: 78% of accomplishments and 49% of achievements. A chi-square test of independence confirms that these differences are significant: $\chi^2 = (3, N=1278) = 24.84, p < .001$, Cramer’s $V = .14$. The number of non-prototypical perfective obligatory contexts is nearly 78% with activities but only 34% with states. The result for states is not surprising, as stative verbs more frequently depict unbounded background information, e.g. *There was a large group of people*, rather than the corresponding bounded interpretation more often found in the foreground, e.g. *There was a loud noise*.

Insert Figure 7 Here

In sum, our analysis of the obligatory contexts elicited in the different tasks demonstrates that task variability is an important factor to consider when investigating the acquisition of tense-aspect morphology. The interview and “Famous People” tasks were not as effective as the two narrative tasks in creating contexts for a range of prototypical and nonprototypical pairings.

5. Discussion

The purpose of this study was to investigate to what extent learners’ use of Spanish past tense morphology varies across a range of oral tasks. In particular, we investigated how differences in the range of verb types produced by learners influenced their performance when measured by emergence and/or accuracy criteria.

Additionally, we were interested in examining how well different task types provided learners with opportunities to demonstrate use of past tense morphology across all combinations of lexical and grammatical aspect, which is crucial in order to understand learners' underlying interlanguage grammar and to test the predictions of the Aspect Hypothesis.

Research question one aimed to investigate whether learners' use of past tense morphology varied across the four tasks using both emergence and accuracy as measures of acquisition. The results highlight the importance of using more than one task type to ensure a representative picture of the learner's ability to use the target structures in all their relevant contexts. As demonstrated in this study, if either or both the interview or the "Famous People" task had been the only tasks used, we might have concluded that several of the intermediate and advanced learners had not yet begun to acquire the Imperfect with telic predicates. Because these same learners had completed all four tasks in this study, we have evidence showing that several of them have in fact reached those advanced stages. This is important because the production of the correct aspectual form in non-prototypical contexts shows us that the underpinning interlanguage grammar correctly selects this form on the basis of grammatical aspect rather than lexical aspect. Prototypical forms do not allow us to conclude whether the target grammar has been acquired, as both lexical and grammatical aspect coincide in these contexts. The interview in particular, being the most open-ended context of use, encourages learners to remain in their comfort zone and play safe, and did not provide opportunities for them to demonstrate this knowledge. The "Famous People" and "Nati y Pancho" narratives provided visual input which necessitated the use of relevant contrasts, but learners were still free to choose their own wording, thus producing authentic discourse with a real

communicative goal. The results were quite different between these two tasks because of the different discourse requirements. The “Nati y Pancho” narrative was better at creating contexts for the Imperfect with non-prototypical pairings, whereas the “Famous People” task was better at creating contexts for the Preterit with non-prototypical pairings.

When acquisition was operationalised by accuracy, the results demonstrate that learners’ accurate use of Preterit and Imperfect morphology varies across tasks but only for learners at the intermediate and beginning proficiency levels. This finding relates to the variable of communicative control discussed in Salaberry & Lopez-Ortega (1998). When learners have more communicative control and can choose their own wording, such as in the interview and the “Famous People” task, they tend to produce more contexts for prototypical pairings, the earlier acquired stages, which in turn inflates their accuracy scores. The tasks with more non-prototypical pairings were those tasks where the accuracy scores were the lowest. This finding is similar to that of Pienemann (1998) suggesting that the lexical demands of the task influence accuracy scores; however, our results suggest this only happens at lower proficiency levels and once the form has emerged. The beginner group performed significantly differently across tasks on their Preterit score but not their Imperfect score. We believe this is because the Preterit has been shown to emerge before the Imperfect, so their accuracy scores were more influenced by task type on the Preterit because they have already begun to acquire it. The intermediate group performed significantly differently across tasks on both the Preterit and Imperfect because both forms have emerged in their interlanguage, but they are not as accurate when forced to produce more advanced forms, hence their significantly lower scores on those tasks with more advanced obligatory contexts.

Research question two focused specifically on whether certain task types provide learners with fewer/more opportunities to demonstrate both prototypical and non-prototypical combinations of lexical and grammatical aspect in oral production. The results demonstrated that in the more open-ended tasks, the interview especially but also the “Famous People” task, learners created fewer contexts for telic predicates (achievements and accomplishments) in the Imperfect. In comparison, the two narrative tasks were designed specifically to target non-prototypical contexts, and although they differed in design (“Las Hermanas” being more experimental in nature with infinitival phrases provided underneath each picture), by including pictures of telic events and creating habitual contexts with prompts, more obligatory imperfective contexts with achievements and accomplishments were created in these tasks. Therefore, it seems that in order to ensure learners have ample opportunities to demonstrate knowledge of all theoretically relevant contexts, tasks need to be carefully designed with this in mind. Other more prototypical combinations of lexical and grammatical aspect which are characteristic of early stages of acquisition (telic predicates in the Preterit and atelic in the Imperfect) occur frequently without carefully designed tasks.

Non-prototypical Preterit pairings, i.e. atelic activities and states, present a different picture. Whereas contexts for activities in the Preterit occurred quite frequently in all tasks, the opposite was found for states in the Preterit. Even in the “Las Hermanas” task where we tried to include several obligatory contexts for the Preterit, many native speakers still produced the Imperfect, demonstrating the difficulty in eliciting a stative verb in a perfective context.

Comparing the results of the “Las Hermanas” task to the other three, we see that it was the most successful at eliciting nonprototypical imperfective contexts, in large

part due to its controlled experimental design. This finding highlights the importance of complementing corpora with experimental methods. Because non-prototypical pairings are the true test cases for the AH, it is imperative that learners have opportunities to demonstrate their ability to produce the past tense with these combinations of lexical and grammatical aspect. As shown in this study, many of the non-prototypical pairings rarely occurred in the corpus data. Therefore, when designing learner corpora to investigate the acquisition of specific linguistic structures, it is important to triangulate results from multiple tasks (more open-ended and more controlled) to ensure that all contexts relevant to the understanding of the acquisition of a particular structure can be analysed.

Another issue highlighted in this study (and others) is how the number of predicates from each lexical aspect class varies across task types. In this study, states made up 50% and 61% of all lexical aspect classes in the interview and “Famous People” task respectively. In contrast, the two narrative tasks showed a more balanced range of lexical aspect classes. This bias is obviously a concern if a single task is used. It also demonstrates how the frequent and accurate use of the same verbs can impact on learners’ accuracy scores (see also David et al. 2009). As discussed in the results section, the beginner group scored higher on the Imperfect in the “Famous People” task because of their frequent use of the verb *ser/estar* (‘be’).

The results of this study demonstrate that communicative tasks which some may consider to be more ‘authentic’ such as the interview, do not provide sufficient opportunities for learners to demonstrate their abilities to use past tense morphology. Only using the interview task to investigate learners’ use of past tense morphology would have misled the analysis. We are mindful of the necessity for learner corpora to be as authentic as possible and to provide continuous and contextualised samples of

language use. Yet, we also believe that the results of this study demonstrate that it is possible to design learner corpora so that they include a range of authentic communicative activities that naturally provide multiple contexts for the linguistic feature(s) under investigation. Furthermore, by complementing oral corpus data with experimental methods we were able to gain a better understanding of our learners' current state of interlanguage development. It would be worthwhile to extend this type of analysis to investigate task variability between oral and written tasks, as well as among written tasks.

6. **Conclusion**

The results of this study demonstrate that different task types lead to very different kinds of language produced by learners. This in turn leads to different conclusions being reached about what stage of development individual learners have reached, in terms of both emergence and accuracy, depending on which task is analysed. This result is important because, in the context of the L2 acquisition of the specific structure examined in this paper, some of the tasks do not enable us to tell whether the correct underpinning rule has been acquired or not, as the relevant contexts are not present. The demands of the tasks not only impact the stage of acquisition learners are placed into using an emergence criterion but also their level of accuracy. If they are required to use more advanced forms, and they have not reached advanced proficiency, then their accuracy will be lower. However, when they have more communicative control, like in the interview, they can use the verbs they are most comfortable with, therefore increasing their accuracy (see also Salaberry & Lopez-Ortega 1998, Pienemann 1998).

When designing learner corpora for the purposes of SLA research, it is crucial to be aware of what contexts need to be present in order to understand the interlanguage grammar underpinning learner productions at a given stage of development. These needs have to be reconciled with the necessity for learner corpora to be as authentic as possible. This means designing a corpus that includes a range of open-ended tasks representative of the language used by the learners and complementing the corpus data with experimental data, to ensure drawing the most accurate and generalizable picture of learners' developmental stage.

References

- Andersen, R. 1991. "Developmental sequences: the emergence of aspect marking in second language acquisition". In T. Huebner & C. Ferguson (Eds.), *Crosscurrents in SLA and linguistic theories*. Amsterdam: John Benjamins, 305-324.
- Andersen, R. & Shirai, Y. 1994. "Discourse motivations for some cognitive acquisition principles". *Studies in Second Language Acquisition*, 16, 133-156
- Andersen, R. W., & Shirai, Y. 1996. "Primacy of aspect in first and second language acquisition: The pidgin/Creole connection". In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Second Language Acquisition*. San Diego, CA: Academic Press, 527-570.
- Arche, M. J. 2006. *Individuals in Time: Tense, Aspect and the Individual/Stage Distinction*. Amsterdam & Philadelphia, PA: John Benjamins.
- Ayoun, D. 2004. "The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study". *Modern Language Journal*, 88, 31-55.
- Bardovi-Harlig, K. 1998. "Narrative structure and lexical aspect: Conspiring factors in second language acquisition of tense-aspect morphology". *Studies in Second Language Acquisition*, 20, 471-508.
- Bardovi-Harlig, K. 2000. *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Malden, MA: Blackwell Publishers, Inc.
- Bardovi-Harlig, K. 2005. "Tracking the elusive Imperfect in adult second language acquisition: Refining the hunt". In P. Kempchinsky & R. Slabakova (Eds.), *Aspectual Inquiries*. Dordrecht: Kluwer, 397-419.

- Bardovi-Harlig, K. 2013. "Research design: From text to task". In M. R. Salabery & L. Comajoan (Eds.), *Research Design and Methodology in Studies on L2 Tense and Aspect*. Berlin: Walter de Gruyter, 219-269.
- Bardovi-Harlig, K., & Reynolds, D. W. 1995. "The role of lexical aspect in the acquisition of tense and aspect". *TESOL Quarterly*, 29, 107-131.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. 1998. *Corpus Linguistics: Investigating Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D. & Conrad, S. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Bonilla, C. 2013. "Tense or Aspect?: A Review of Initial Past Tense Marking and Task Conditions for Beginning Classroom Learners of Spanish". *Hispania*, 96(4), 624-639.
- Bonilla, C. 2014. From number agreement to the subjunctive: evidence for Processability Theory in L2 Spanish. *Second Language Research*, Prepublished July 14 2014.
- Callies, M., Díez-Bedmar, M.B., & Zaytseva, E. 2014. "Using Learner Corpora for Testing and Assessing L2 Proficiency" In P. Leclercq, A. Edmonds, & H. Hilton (Eds), *Measuring L2 Proficiency: Perspectives from SLA*. Bristol: Multilingual Matters, 71-90.
- Camps, J. 2002. "Aspectual distinctions in Spanish as a foreign language: The early stages of oral production". *IRAL*, 40, 179-210.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Comrie, B. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Cordier, C. 2013. *The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study*. Unpublished PhD thesis, Newcastle University, UK.
- David, A. 2008. "Vocabulary breadth in French L2 learners". *Language Learning Journal*, 36, 167-180.
- David, A., Myles, F., Rogers, V., & Rule, S. 2009. "Lexical development in instructed L2 learners of French: Is there a relationship with morphosyntactic development?" In H. Daller, Malvern, D., Meara, P., Milton, J., Richards B., Treffers-Daller, J. (Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application*. Basingstoke: Palgrave, 147-163.
- Dekydtspotter, L., & Sprouse, R. 2001. "Mental design and (second) language epistemology: adjectival restrictions of *wh*-quantifiers and tense in English-French interlanguage". *Second Language Research*, 17(1), 1-35.
- Dominguez, L., Tracy-Ventura, N., Arche, M., Mitchell, R., & Myles, F. 2013. "The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology". *Bilingualism: Language and Cognition*, 16, 558-577.
- Duff, P. 1993. "Tasks and interlanguage performance: An SLA perspective". In G. Crookes & S. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice*. Clevedon, Avon: Multilingual Matters
- Ellis, R. & Barkhuizen, G. 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

- Foster, P., & Tavakoli, P. 2009. "Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity". *Language Learning*, 59(4), 866-896.
- Gil, K-H., & Marsden, H. 2013. "Existential quantifiers in second language acquisition: a feature reassembly account". *Linguistic Approaches to Bilingualism*, 3(2), 117-149.
- Gilquin, G. & Gries, S. Th. 2009. "Corpora and experimental methods: A state-of-the-art review". *Corpus Linguistics and Linguistic Theory* 5(1), 1-26.
- Granger, S. 2002. "A bird's-eye view of learner corpus research". In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.
- Granger, S. 2008. "Learner corpora". In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter, 259-275.
- Granger, S. 2009. "The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation". In K. Aijmer (Ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, 13-32.
- Hasko, V. 2013. "Capturing the Dynamics of Second Language Development via Learner Corpus Research: A Very Long Engagement". *Modern Language Journal*, 97 (Suppl 1): 1-10.
- Housen, A. 2002. "A corpus-based study of the L2-acquisition of the English verb System". In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 77-116.
- Lafford, B. 1996. "The development of tense/aspect relations in L2 Spanish

- narratives: evidence to test competing theories”. Paper read at SLRF 96, at Tucson, AZ.
- Langley, J. 2000. *Missing!* London: Francis Lincoln Limited.
- Leclercq, P. & Edmonds, A. 2014. “How to Assess L2 Proficiency? An Overview of Proficiency Assessment Research” In P. Leclercq, A. Edmonds, & H. Hilton (Eds), *Measuring L2 Proficiency: Perspectives from SLA*. Bristol: Multilingual Matters, 3-24.
- Liskin-Gasparro, J. 2000. “The use of tense-aspect morphology in Spanish oral narratives: Exploring the perceptions of advanced learners”. *Hispania*, 83, 830-844.
- Lozano, C., & Mendikoetxea, A. 2010. Interface conditions on postverbal subjects: a corpus study of L2 English. *Bilingualism: Language and Cognition*, 13(4), 475-497.
- Lozano, C. & Mendikoetxea, A. 2013a. “Corpus and experimental data: subjects in second language research”. In S. Granger, G. Gilquin & F. Meunier (eds). *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead*. Louvain-la-Neuve: Presses universitaires de Louvain, 313-323.
- Lozano, C. & Mendikoetxea, A. 2013b. “Learner corpora and second language acquisition: The design and collection of CEDEL2.” In A. Diaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins, 65-100.
- Mackey, A. & Gass, S. 2005. *Second Language Research: Methodology and Design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. 2000. *The CHILDES Project: Tools for analyzing talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

- Malvern, D. & Richards, B. 2002. "Investigating accommodation in language proficiency interviews using a new measure of lexical diversity." *Language Testing* 19 (1), 85-104.
- Marsden, H. 2009. "Distributive Quantifier Scope in English-Japanese and Korean-Japanese Interlanguage". *Language Acquisition*, 16, 135-177.
- Meunier, F. & Littré, D. 2013. "Tracking learners' progress: Adopting a dual 'corpus cum experimental data' approach". *The Modern Language Journal* 97 (Suppl. 1): 61-76.
- Myles, F. 2004. "From data to theory: the over-representation of linguistic knowledge in SLA". In R. Towell & R. Hawkins (Eds.), *Empirical Evidence and Theories of Representation in Current Research in Second Language Acquisition* (Vol. 102): Transactions of the Philological Society., 139-168
- Myles, F. 2008. "Investigating learner language development with electronic longitudinal corpora: Theoretical and methodological issues". In L. Ortega & H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities*. Hillsdale: Lawrence Erlbaum, 58-72.
- Norris, J.M., Brown, J.D., Hudson, T.D., & Bonk, W. 2002. "Examinee abilities and task difficulty in task-based second language performance assessment." *Language Testing* 19 (4), 395-418.
- Norris, J. & Ortega, L. 2003. "Defining and measuring SLA". In C. Doughty, & M. H. Long, (Eds.), *Handbook of Second Language Acquisition*. London: Blackwell, 716-761.
- Pica, T. 1983. "Methods of morpheme quantification: their effect on the interpretation of second language data". *Studies in Second Language Acquisition*, 6, 69-78.
- Pienemann, M. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.
- Pienemann, M. & Johnston M., & Brindley, G. 1988. "Constructing an acquisition-

- based procedure for second language assessment.” *Studies in Second Language Acquisition*, 10, 217-224.
- Rankin, T. 2009. Verb-second in advanced L2 English: A learner corpus study. In M. Bowles, T. Ionin, S. Montrul & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference* (pp. 46-59). Somerville, MA: Cascadilla Proceedings Project.
- Salaberry, M. R. 1999. “The development of past tense verbal morphology in classroom L2 Spanish”. *Applied Linguistics*, 20, 151-178.
- Salaberry, M. R. 2000. *The Development of Past Tense Morphology in L2 Spanish*. Amsterdam: John Benjamins.
- Salaberry, M. R. 2003. “Tense aspect in verbal morphology”. *Hispania*, 86, 559-573.
- Salaberry, M. R. 2008. *Marking Past Tense in Second Language Acquisition: A Theoretical Model*. London: Continuum.
- Salaberry, M.R. & Lopez-Ortega, N. 1998. “Accurate L2 production across language tasks: Focus on form, focus on meaning, and communicative control”. *The Modern Language Journal*, 82(4), 514-532.
- Shirai, Y. 2004. “A multiple-factor account to form-meaning connection in the acquisition of tense-aspect morphology”. In VanPatten, B. et al. (Ed.) *Form-Meaning Connections in Second Language Acquisition* (91-112). Mahwah, NJ: L. Erlbaum Associates, 91-112.
- Sinclair, J. 2005. “How to build a corpus”. In M. Wynne (Ed), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow books, 79–83.
- Sugaya, N., & Shirai, Y. 2007. “The acquisition of progressive and resultative

- meanings of the imperfective aspect marker by L2 learners of Japanese: Transfer, universals, or multiple factors”. *Studies in Second Language Acquisition*, 29, 1-38.
- Tavakoli, P., & Foster, P. 2011. Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 61, 37-72.
- Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. 2014. “ ‘Repeat as much as you can’: Elicited imitation as a measure of oral proficiency in L2French”. In P. Leclercq, A. Edmonds, & H. Hilton (Eds), *Measuring L2 Proficiency: Perspectives from SLA*. Bristol: Multilingual Matters, 143-166.
- Yu, G. 2010. “Lexical diversity in writing and speaking task performances”. *Applied Linguistics* 31 (2), 236-259.
- Vendler, Z. 1967. *Linguistics in philosophy*. Ithaca, New York: Cornell University Press.

Appendix 1: Tests of lexical aspect

Common tests for STATES vs. NONSTATES:

A. Can the verb be used in the progressive form and it sounds natural?

- (1) *Juan está siendo alto. (states)
Juan está paseando. (activity)
- (2) Juan está trazando un círculo. (accomplishment)
- (3) Juan está dándose cuenta de que su madre tiene razón. (achievement)

If the answer is NO, then code as a state.

If the answer is YES, then not a state.

B. Can the verb have a habitual interpretation in the present tense?

- Ex. *Juan sabe matemáticas. (state)
Juan pasea. (activity)
Juan traza un círculo. (accomplishment)
Juan se da cuenta de que su madre tiene razón. (achieve)

If the answer is NO, then code as a state. If the answer is YES, then not a state.

Common tests for ACTIVITIES vs. TELICS (accomplishments & achievements):

A. Can you add the phrase *in x time*?

- (5) *Pablo viajó en tres semanas. (activity)
- (6) Pablo construyó una casa en tres semanas. (accomplishment)
- (7) Pablo encontró la aguja en tres horas. (achievement)

If the answer is NO, then code as activity. If the answer is YES, then check accomplishment test.

Other, further tests are:

A. If you stop in the middle of the action, does that entail that you did the action?

- (8) Juan estaba andando (activity) → Juan ha andado
- (9) Juan estaba construyendo una casa → Juan has **not** built a house

If the answer is YES, then code as an activity. If the answer is no, then not an activity. Check accomplishment test.

B. Does the event become ambiguous when the adverb 'almost' is used?

- (10) Juan casi corrió
→ Not ambiguous: only meaning: he did not start running. (Activity).
- (11) Juan casi construyó una casa
→ Ambiguous:
 - a) He did not start building the house
 - b) He did not accomplish building the house. (Accomplishment).

If the answer is NO, then code as an activity. If the answer is YES, then check accomplishment test.

Common tests for ACCOMPLISHMENTS vs. the rest:

A. Can you say that X has done (verb) *hasta la mitad*? (positive result only with accomplishments)

- (12) *Juan ha odiado a su primo hasta la mitad (state)
J has hated his cousin half way
- (13) *Juan vagó por las calles hasta la mitad (activity)
J wandered on the streets half way
- (14) Ha escrito la tesis más o menos hasta la mitad. (accomplishment)
J has written his thesis half way
- (15) *La bomba ha explotado hasta la mitad (achievement)
The bomb exploded half way

If yes, then code as an accomplishment. If no, then check other complementary tests to carry on diagnosing.

B. Can the verb be a complement of *finish*?

- (16) Pablo ha terminado de construir una casa. (accomplishment)
Pablo has finished building a house
- (17) *Pablo ha terminado de darse cuenta de que su madre tenía razón.(ach)
Pablo has finished realizing his mum is right
- (18) * Pablo has finished wandering around the streets (activity)

If yes, then code as an accomplishment. If no, then double check the achievement test.

Tests for ACHIEVEMENTS (vs. accomplishments):

A. Can you ask “At what time did you (verb)?” and see if it can be answered with a specific time?

- 19. ¿A qué hora llegaste a la fiesta? (achievement)
*¿A qué hora construiste la casa? (accomplishment- in this case it would have the meaning of starting to do the action)

If yes, then it is an achievement. If no, then it is an accomplishment.