AUTHOR ACCEPTED MANUSCRIPT

Preterite and Imperfect in Spanish Instructor Oral Input and Spanish Language Corpora

Danielle Daidone

Indiana University

Abstract

This study compares Spanish instructors' use of preterite and imperfect in the foreign language classroom to the distribution of these forms in large-scale corpora, which represent the input learners would potentially receive in a naturalistic learning context. Twenty-four 50-minute class sessions were recorded, and all tokens of preterite and imperfect spoken by the instructors were transcribed. Each token was coded for the verb, its morphological ending, the instructor's native language, the class activity during the token's utterance, and the class level. For all verbs that were analyzed in the instructor input, their preterite and imperfect token frequencies in the oral data of two Spanish corpora were determined. Results reveal low token and type frequencies for imperfect morphology in instructor input. The imperfect accounted for only 18% of past tense forms, with class activity as a significant predictor of its use. In contrast, around 60% of past tense tokens of these same verbs in large-scale corpora were imperfect, and fewer verbs were biased toward use in the preterite. In sum, the distribution of past tense forms in instructor input differed substantially from that of the corpora.

1. Introduction

All theories of second language acquisition agree that input is an essential component. Input has an especially dominant role in usage-based theories of acquisition, which hypothesize

that learners' experience with language shapes their grammars (Bybee 2008; Ellis 2002, 2013). According to a usage-based framework, cognitive representations emerge based on speakers' experiences of the frequency of linguistic units and their contexts of use. These tenets of usage-based grammar apply not only to first language (L1) acquisition, but second language (L2) acquisition as well. Given the importance of experience with language in shaping a learner's grammar, it follows that studying patterns in the input that learners receive would yield insights into how and why their grammars are organized as they are. Nevertheless, as a field we have little research on the input to which learners are exposed (Ellis & Collins 2009).

Many L2 learners receive the vast majority of their target language input in the foreign language classroom, making the classroom an important target of investigation. Those acquiring the language in a foreign language classroom are likely to be exposed to input that is "restricted" or "distorted" in some way (Lightbown 1985). For example, classroom learners are less likely to get native-like exposure to low frequency structures and conventionalized sequences (Bybee 2008). Furthermore, if classroom input is different enough from that present in the target language environment, instructed learners' interlanguage may differ from naturalistic learners' interlanguage, at least transiently, in the types of errors they produce or in the order in which they acquire structures (Lightbown 1985; Rothman & Guijarro-Fuentes 2010). Despite these possibilities, exactly how the use of certain forms differs between the foreign language classroom and the target language environment is largely unknown.

The current study aims to provide some insight into potential input differences by focusing on the use of two particular verb forms: the Spanish preterite and imperfect. The preterite, which encodes perfective aspect, presents an event as a single unit, encompassing the beginning, middle, and end of the action. Although an action may have distinct parts, perfective

aspect indicates a view of the event as a unified whole (Comrie 1976). In contrast, the imperfect indicates a view of the event without reference to temporal boundaries. In other words, an imperfective action is either typical of the period of time that includes the reference point, i.e. an habitual action, or is ongoing at that particular reference point (Bybee et al. 1994). Because languages like English do not mark this aspectual distinction on past tense verbs, many learners of Spanish are faced with the challenge of understanding and producing this unfamiliar morphological distinction in their L2.

Given its difficulty, the L2 acquisition of the preterite and imperfect has been extensively studied, and research has shown that the preterite generally emerges earlier in learners' interlanguage and is overextended into imperfect contexts even by advanced learners (see Comajoan 2014 for a review). The aspect hypothesis (e.g., Andersen 1991; Andersen & Shirai 1994), which emphasizes the importance of the inherent lexical aspect of the predicate, and the discourse hypothesis (e.g. Bardovi-Harlig 1995), which emphasizes the role of foregrounding and backgrounding in discourse, have helped explain learners' acquisition of these forms. Yet as Salaberry (2000) points out, both of these theories were originally created based on data from learners living in an L2 environment, a situation in which it is assumed that learners are receiving input that is comparable to the input native speakers receive. Classroom learners of Spanish in a foreign language environment may or may not receive input that exhibits similar patterns of this verbal morphology, and differences could affect their acquisition. In fact, Salaberry (2002) argues that there is possibly a distributional bias toward the perfective form in the input given to learners. Based on the principles of usage-based language acquisition, this potential bias in the input could be one factor leading to the earlier emergence and the overgeneralization of the preterite in learner' interlanguage. Thus, the current study examines

whether there is indeed a bias towards the preterite in instructor oral input in the foreign

language classroom, and if so, whether characteristics of the class session and instructor

background affect the use of the past tense in this input. Additionally, the use of preterite and

imperfect forms in classroom input are compared to the distribution of these same verbs in oral

data of Spanish language corpora in order to determine to what extent the patterns of use differ

between the classroom and the target language environment.

2. Effects of characteristics of the input

Under a usage-based approach, learners' grammars are hypothesized to consist of

constructions, which are form-function mappings between semantic or discourse functions and

linguistic units. These may be as specific as morphemes, words, and idioms, or as abstract as

syntactic patterns. Learners acquire these constructions by generalizing from their usage in

linguistic input. This is possible because speakers are sensitive to the frequency of linguistic

elements in the input and are continuously tracking the contexts in which they occur (Bybee

2008; Ellis 2002, 2013).

According to Bybee (2008), high token frequency, or the frequency of a particular item,

is necessary in learners' input because a specific form needs to be repeated often enough to have

a solid mental representation before other forms in its paradigm can be learned as derivatives of

that form. For example, she cites the $1^{st}/3^{rd}$ person singular as being the most frequent form of

the imperfect in Spanish, such as *cantaba*. This form serves as an anchor from which other forms

can be derived, such as the 1st person plural *cantábamos*. As for type frequency, or the frequency

of a pattern, a construction must appear with a variety of verbs in order for it to be productive in

a learner's grammar. If learners hear the suffix *-aba* only with the verb *cantaba* for instance, they are unlikely to realize that *-aba* is used to mark imperfective aspect on other verbs as well.

In general, learners need to be exposed to numerous prototypical tokens of a construction in order to learn its core meaning, and also to more peripheral cases so that they can acquire the breadth of possibilities in that construction (Ellis 2009). To take imperfect morphology as an example construction, learners would ideally be presented with many instances of imperfect used with verbs that denote a state, like *ser*, so that they could learn how imperfect is typically used, and then learners would be exposed to the use of imperfect morphology with verbs it is less commonly associated with, such as *llegar*. Exposure to a low variance sample aids in category formation, while higher type frequency aids in the productivity of a construction with different items. Evidence for the importance of exposure to numerous tokens of a prototypical verb for a construction can be found in the results of Ellis and Ferreira-Junior (2009). Using corpus data, they discovered that the verbs learned first in a construction were those that were the most prototypical, frequent, and generic, e.g. *to give* in the English ditransitive construction. Furthermore, the necessity of exposure to a range of types has been corroborated by McDonough and Kim (2009), who manipulated the type frequency of verbs occurring with *wh*-questions in input to learners. They found that although high token frequency and low type frequency may have been sufficient for learners to perceive patterns in the input, those learners exposed to many different types in their prompts were more accurate at forming questions.

The tendency of acquisition to mirror patterns in the input has not only been found in experimental studies or correlated with native speech in corpora, but also has been shown with teacher talk in classrooms. Collins, Trofimovich, White, Cardoso, and Horst (2009) examined the instructor input given to ESL students in Quebec, analyzing specifically the distribution and

saliency of progressive aspect, simple past, and third-person possessive determiners. They observed differences in type frequency, perceptual saliency, and semantic scope between the progressive, which is acquired earlier, and the simple past and possessive determiners, which are acquired later. Similarly, one part of Hamayan and Tucker's (1980) investigation on teacher talk examined the token frequency of nine syntactic structures in the speech of 3rd and 5th grade French immersion teachers and the oral production of these structures by 3rd and 5th grade students. The frequency of structures in teacher input was significantly correlated with students' correct production of these structures. Thus, distributions of forms in instructor input appear to relate to both the order of acquisition of structures as well as to students' accuracy with forms.

3. Characteristics of native speaker discourse and classroom input

For native speakers outside of the classroom, their production of perfective and imperfective morphology is said to demonstrate a distributional bias according to the inherent lexical aspect of the verb (Andersen & Shirai 1994). Verbs have traditionally been divided into four classes according to their inherent lexical aspect: states, activities, accomplishments, and achievements (Vendler 1967). States do not entail a beginning or endpoint in time and continue without change, such as the verbs *be*, *know*, and *have*, while activities also do not imply temporal boundaries but are dynamic actions, e.g. *run*, *sleep.* Unlike states and activities, accomplishments and achievements have a natural endpoint. Accomplishment verbs imply a duration of time before the endpoint is reached, e.g. *melt*, *dry*, while for achievements the endpoint of the event denoted is almost instantaneous, e.g. *pop*, *arrive* (Andersen 1991). According to Andersen and Shirai (1994), native speakers tend to use more past perfective morphology with achievements and accomplishments, which have a natural endpoint, and more imperfective morphology with

states and activities, which do not imply temporal boundaries. This tendency has been termed the Distributional Bias Hypothesis and has been supported in a corpus study by Tracy-Ventura (2007) (as cited in Tracy-Ventura 2008). She investigated the distribution of different verb types in a Spanish native speaker corpus with over 9.5 million words from conversations and literature. She found that 1) verbal predicates that had a higher token count in the imperfect tended to be activities and states, 2) verbal predicates that had a higher token count in the preterite tended to be accomplishments and achievements, and 3) verbal predicates that were distributed evenly between preterite and imperfect tended to be verbs that could be classified into multiple inherent lexical classes. Thus, it seems that the language that Spanish native speakers use among themselves exhibits a bias according to the lexical aspect of the verb.

However, speaking to students in a foreign language classroom is a different communicative act than talking to fellow native speakers, and may not demonstrate the same distributions. Andersen and Shirai (1994) suggested the possibility that "language directed to learners conforms much more closely to the aspect hypothesis" (138), but Salaberry (2002) argued that perfective forms are likely to be more frequent in instructor input due to the type of discourse that is used in the classroom.

One reason to suspect that classroom input may differ from native speaker discourse in terms of the distribution of perfective and imperfective forms is that classroom learners have displayed different acquisitional paths compared to untutored learners living in the L2 environment. Classroom learners often overgeneralize perfective morphology into contexts for the imperfective. For example, Camps (2002) studied the acquisition of L2 Spanish by first-year university students and observed that the preterite accounted for 39% of errors when a different form was used instead of the required imperfect, while the imperfect made up only 4.5% of such

errors in the preterite context. Swain (1991) reported similar findings for the acquisition of French by middle school students. In her investigation, learners frequently overextended the *passé composé* (the perfective form) into imperfective contexts. This overgeneralization of the perfective form is so widespread among instructed learners that some researchers have argued that the perfective form initially functions as default marker of past tense and the imperfective emerges later in certain contexts (Bonilla 2013; Comajoan 2005; Lubbers Quesada 2006; Salaberry 2002; Tracy-Ventura 2008). Although the overextension of the preterite diminishes as proficiency increases, even learners in 3rd semester university courses have been shown to display this tendency (Salaberry 2002).

Untutored learners living in the L2 environment have not exhibited this overgeneralization of the perfective (Salaberry 2000). While they have often been found to use little verbal morphology (e.g. Sato 1990), the morphology they do produce, whether it be target-like or not, has generally been reported to be divided along the lines of inherent lexical aspect (e.g. Andersen 1991, 1994; Clements 2003; Lopez-Ortega 2000). For example, Andersen (1994) found that the use of the preterite by two English-speaking learners of Spanish living in Puerto Rico was almost entirely restricted to accomplishments and achievements and the use of the imperfect was restricted to states and activities.

Apart from data on tutored versus untutored learners, the few studies that have analyzed past tense forms in teacher input have found evidence for a higher use of perfective forms compared to imperfective. Swain (1991) analyzed both learner production and the teacher talk of ten middle school French immersion classes. She found that, of the past tense verbs produced by instructors, about two-thirds were perfective forms while one-third were imperfective forms. A similar result was reported by Kaplan (1987), who analyzed a small sample of instructor oral

input as part of a study on the acquisition of the perfective and imperfective past tense by university students in French language classes. In the teacher talk, the perfective accounted for 84% of all past tense forms.

4. The current study

Previous research has demonstrated that properties of the input influence learners' acquisition, yet the input that students receive in the language classroom – which is the main source of input for most foreign language learners – has not been well described. Differences between classroom and naturalistic learners' output and research on the past tense in French teacher talk point to the possibility that the perfective may be overrepresented in the foreign language classroom, making a comparison between typical native speaker discourse and instructor input especially warranted. As such, the current study investigates the use of past tense forms in Spanish instructor oral input and contrasts this pattern of use with that found in corpora of native speech. These datasets are compared in order to determine how much the distribution of past tense forms differs between their use across a variety of communicative situations in the target language environment and the limited discourse contexts found in teacher talk. Since instructor input has not been extensively documented, three factors that may influence instructors' oral production are explored in this study: class level, the (non)native-speaker status of the instructor, and the class activity when the form was produced. The following research questions guide this investigation:

1. What is the distribution of preterite and imperfect morphology by token and type frequency in the oral input provided by Spanish language instructors?

2. Are there any differences in the proportion of preterite and imperfect forms due to extralinguistic variables (class level, (non)native status of instructor, and activity)?

3. Does the distribution of the preterite and imperfect for the verbs that appear in classroom oral input mirror their distribution in the preterite and the imperfect in corpora of native speaker oral discourse?

5. Method

5.1 Instructor input data

The corpus analyzed for this study was the instructor oral input present in 24 videos of 50-minute class sessions at Indiana University. These sessions were evenly divided between 3rd, 4th, and 5th semester Spanish language classes, resulting in 8 videos per level. The 3rd and 4th semester classes were part of the basic language program, and the 5th semester class was a grammar in context course for Spanish majors and minors. All classes followed synthetic syllabi arranged by grammatical concepts, and none of the classes were scheduled to cover the preterite or imperfect on the day of taping. In the basic language program, students study the regular forms of the preterite in 1st semester, and both regular and irregular preterite and imperfect forms are taught in 2nd semester. Both of these courses meet 4 days a week for 50 minutes. The preterite and imperfect are reviewed in 3rd and 5th semester courses, which meet for 2.5 hours a week. At the time of taping, the preterite and imperfect review had not yet occurred in the 3rd semester course and had already happened in the 5th semester course.

Four instructors per level were recorded: two native Spanish speakers (1 male, 1 female) and two non-native Spanish speakers whose L1 was English (1 male, 1 female). Native Spanish speakers were from a variety of regions: Cuba, Mexico, Puerto Rico, and Spain. Although there

is variation in the use of past tense forms by dialect, such as the expansion of the present perfect into preterite contexts in Peninsular Spanish (e.g. Schwenter & Torres Cacoullos 2008), this mix of native speakers from different dialects and non-native instructors was chosen in order to be representative of the diversity of instructor backgrounds present in the university setting. Each instructor was recorded teaching two class sessions at the same level.

All tokens of preterite and imperfect forms spoken by the instructors in the videos were transcribed and coded according to their verbal morphology. The suffixes following the verb stem were coded as one unit, without further division into thematic vowel, aspect marker, and/or person and number suffixes. For example, *decía* was coded as containing -/ˈia/, which marks imperfect in the 1st and 3rd person singular for *-er/-ir* verbs. Variations in spelling were ignored for the purposes of coding as long as they represented phonologically regular patterns. For example, both *vi* and *rompí* were coded for the presence of -/i/, the 1st person singular preterite marker for *-er/-ir* verbs. Only verb forms with regular suffixes and no changes to the stem were included in the analysis of the type frequency of morphological patterns. For the type frequency of different verbs, reflexive and non-reflexive uses were counted as the same verb type; for example, both *se tiraban* and *tiró* were considered instances of the verb *tirar*.

In terms of extralinguistic variables, each token was coded for the class level (3rd, 4th, or 5th semester), the native language of the instructor (native or non-native Spanish speaker), and the class activity when the preterite or imperfect form was used (readings, vocabulary, grammar, or other). The category "readings" refers to the discussion of cultural or literary texts; "vocabulary" signifies the learning or review of vocabulary from the textbook; "grammar" refers to the teaching of grammatical concepts, excluding preterite or imperfect; and "other" consists of

comments before class, recapping what they had learned that day at the end of class, or other instances of speech that were not during the course of the lesson.

5.2 Spanish language corpora data

In order to determine the differences between the distribution of past tense forms found in the instructor input and the distribution produced by native speakers in a wide variety of discourse contexts, the oral data of the *Corpus de Referencia del Español Actual* (CREA) (Real Academia Española) was searched. These data, totaling nearly 9 million words, consist of mainly radio and television recordings from various Spanish-speaking countries from 1975-2004, although other types of data are also represented such as telephone conversations, political speeches, answering machine messages, and sociolinguistic interviews. These data were chosen as a comparison for the instructor input because they are potentially similar to the oral input that students may be exposed to in a study abroad or naturalistic learning context. For each verb that appeared in the classroom input in the past tense, token frequency was calculated for that verb in CREA for preterite and imperfect forms. Because the oral data of this corpus are not yet lemmatized or tagged for part of speech, the searches were done by individual word forms, e.g. *hablé*, *hablaste*, *habló*, etc. All 1st person plural forms of regular *-ar* and *-ir* verb classes were excluded from this analysis due to identical morphology in the present tense and preterite. Any other forms that had polysemous verbal morphology were also excluded, such as *viste*, which may be either the 2nd person singular preterite form of *ver* or the 3rd person singular present form of *vestir*. If a verb was excluded in one aspect, it was also excluded in the other; for example, because the preterite form *viste* was excluded, the imperfect form in the same grammatical person and number, *veías*, was also excluded. In addition, the verb *haber* was excluded from the

corpus analysis because it was not possible to separate presentational uses of *haber* in the imperfect from its use in the imperfect as an auxiliary verb to form the pluperfect. The verbs *ser* and *ir* were combined in the analysis due to their identical forms in the preterite.

Given the high number of exclusions and the necessity of combining *ser* and *ir* in the analysis of verbs in CREA, the author also chose to analyze preterite and imperfect verb forms from the oral data of the *Corpus del español* (CDE) (Davies 2002a). The data in this corpus primarily come from sociolinguistic interviews, although some transcriptions of conferences and newspaper interviews are also included (Davies 2002b). These data are less representative of what a student would typically be exposed to in a naturalistic learning environment as compared to the data from CREA; nevertheless, this corpus presents an advantage in that the data are lemmatized and verb tokens are tagged for their tense, mood, and aspect. Because of this, it was not necessary to exclude verb forms based on polysemous morphology. However, in order to exclude instances of *haber* as an auxiliary verb (which are not tagged separately from other instances of the imperfect in the corpus), all occurrences before a past participle were excluded from the search. This was done for *haber* in both the imperfect and preterite for the sake of consistency. Additionally, three verbs – *cepillar*, *explorar*, and *maquillar* – did not occur in the oral data in the *Corpus del español* and thus could not be included in the analysis.

6. Results

6.1 Distribution of forms in instructor input

In two of the class sessions, the preterite was discussed and practiced although it was not on the syllabus for that day. This did not occur in any class sessions for the imperfect. In a 4th

semester class, students had to conjugate a verb in the preterite as a quick exercise at the start of class for "Conjugation Wednesday," while it is unclear why the instructor in a 3rd semester class chose to include the preterite in her lesson plan on that day. The 17 tokens from the 3rd semester class and 6 tokens from the 4th semester class that were produced by the instructors during discussion of the preterite were excluded from the analyses. By excluding these tokens, the classroom sample exclusively represents the use of the past tense in the classroom when it is not the focus of instruction.

After excluding these tokens, a binomial test revealed that forms in the preterite, at 81.8% of past tense tokens, were significantly more frequent than forms in the imperfect ($p < .001$). A total of 562 tokens of the preterite and 125 tokens of the imperfect occurred across the 24 class sessions. An average of 23 tokens of the preterite (range: 0-124, SD: 25.89) and 5 tokens of the imperfect (range: 0-31, SD: 8.47) were spoken by an instructor in a single class session. In terms of type frequency, a total of 113 different verbs occurred in the preterite and 30 different verbs appeared in the imperfect, with an average of 11 verb types in the preterite (range: 0-37, SD: 8.85) and 2 verb types in the imperfect (range: 0-11, SD: 3.45) per class session.

Table 1 shows how many instances of regular preterite morphology occurred, as well as how many different verbs were used with each morphological marker. Table 2 displays the same information for regular imperfect morphology. No tokens of the preterite or imperfect occurred with *vosotros* morphology in the instructor oral input included in the analyses. For all other grammatical persons, instances of preterite morphology were more frequent than imperfect morphology.

Table 1. Type and Token Frequency of Regular Preterite Morphology

	Grammatical Person		Nb of Diff. Verbs	Nb of Tokens	Grammatical Person		Nb of Diff. Verbs	Nb of Tokens
-ar verbs	1st sg.	*-é*	31	63	1st pl.	*-amos*	8	16
	2nd sg.	*-aste*	8	13	2nd pl.	*-asteis*	0	0
	3rd sg.	*-ó*	35	107	2nd/3rd pl.	*-aron*	15	45
-er/ir verbs	1st sg.	*-í*	10	17	1st pl.	*-imos*	4	25
	2nd sg.	*-iste*	8	10	2nd pl.	*-isteis*	0	0
	3rd sg.	*-ió*	17	57	2nd/3rd pl.	*-ieron*	14	42

Table 2. Type and Token Frequency of Regular Imperfect Morphology

	Grammatical Person		Nb of Diff. Verbs	Nb of Tokens	Grammatical Person		Nb of Diff. Verbs	Nb of Tokens
-ar verbs	1st/3rd sg.	*-aba*	11	26	1st pl.	*-ábamos*	2	2
	2nd sg.	*-abas*	4	6	2nd pl.	*-abais*	0	0
	--	--	--	--	2nd/3rd pl.	*-aban*	7	16
-er/ir verbs	1st/3rd sg.	*-ía*	9	25	1st pl.	*-íamos*	3	6
	2nd sg.	*-ías*	3	8	2nd pl.	*-íais*	0	0
	--	--	--	--	2nd/3rd pl.	*-ían*	2	6

6.2 Factors affecting proportion of preterite and imperfect in instructor input

A mixed effects logistic regression was conducted in Rbrul 2.3.2 (Johnson 2009) to determine which factors affected the proportion of preterite and imperfect in the instructor input. Four instructors, two from 3rd semester (a native and non-native speaker) and two from 4th semester (native speakers), were excluded from the analysis because they categorically produced preterite tokens. After exclusions, a total of 577 tokens of past tense forms at 78.3% preterite were included in the analysis. An examination of the data revealed that not all activities occurred in every class level, resulting in empty cells. Because of the correlation of activity with class level and the exclusion of both native speakers from the 4th semester data, the author chose to include class activity instead of class level in the final analysis. The final analysis was run with past tense form as the dependent variable, instructor as a random effect, and activity and instructor (non)native status as fixed effects. Results are displayed in Table 3.

Table 3. Factors Conditioning the Use of Imperfect

Factor	Logistic Coefficient	Nb of Tokens	% Imperfect	Factor Weight
Activity ($p < .001$)				
Readings	0.787	219	28.8%	0.68
Other	0.187	104	27.9%	0.55
Vocabulary	0.012	65	18.5%	0.50
Grammar	-0.946	189	11.2%	0.28
(Non)native status (n. s.)				
Native	--	371	16.2%	--
Non-native	--	206	31.6%	--

Factor weights above 0.5 indicate that this factor favors the use of the imperfect, while factors with weights below 0.5 disfavor the use of the imperfect. Only activity was a significant predictor, with the discussion of readings favoring the use of the imperfect, instructor comments unrelated to the lesson slightly favoring the use of the imperfect, and grammar instruction disfavoring the use of the imperfect. Reviewing vocabulary neither favored nor disfavored the imperfect.

6.3 Comparison of instructor and corpora data

After the token frequency of all verbs that occurred in the instructor input was computed for CREA and CDE, a binomial test was performed on the data from each of these corpora in order to determine whether preterite or imperfect forms were more frequent. A binomial test of the token frequency of preterite and imperfect forms for the relevant verbs in CREA revealed that unlike in the instructor data, the use of preterite forms was significantly less frequent than the use of imperfect forms ($p < .001$). Preterite tokens constituted 37.2% of the data. A binomial test performed on the data from CDE also showed that the preterite, accounting for 45.8% of forms, was significantly less frequent than the imperfect ($p < .001$).

In order to examine whether the higher token frequency of imperfect forms in the large-scale corpora was due only to a few verbs, each verb was coded as occurring more often in the preterite or the imperfect. In the instructor oral input, 86% of verb types that occurred in the past tense had a higher token frequency in the preterite. When the verb types that occurred in the past tense in the instructor corpus were searched in the large-scale Spanish corpora, 67% of these verb types in CREA and 72% of these verb types in CDE were found to have a higher token frequency in the preterite. Thus, around 15% of verbs that occurred more often in imperfect

contexts in the corpora were more frequent in preterite contexts in the classroom. For example, *entender* was more frequent in the imperfect in the large-scale corpora, but was used exclusively in the preterite in the instructor corpora, e.g. "¿Han escuchado lo que dijo ella? ¿Entendieron?" (non-native speaker, 5th semester).

In order to determine which verbs constituted the majority of tokens for each construction, the conditional probability of a verb occurring in either the preterite or imperfect was determined. The number of tokens of the verb in a certain construction was divided by all the tokens of that construction. For example, *querer* occurred in the imperfect in the instructor input 10 times, and there were 125 occurrences of the imperfect in total; thus, *querer* accounted for 8% of imperfect tokens. Table 4 displays the verbs with the five highest conditional probabilities in each construction. The results are similar across the instructor input and large-scale corpora for the top two or three verbs.

Table 4. Verbs with the 5 Highest Conditional Probabilities in the Preterite and Imperfect

	Preterite			Imperfect		
	Instructor	CREA	CDE	Instructor	CREA	CDE
1.	ser 8.5%	ser/ir 25.1%	ser 17.3%	ser 23.2%	ser/ir 35.6%	ser 27.6%
2.	decir 4.8%	decir 8.8%	decir 9.0%	estar 12.8%	estar 13.5%	estar 12.2%
3.	nacer 4.4%	hacer 8.0%	hacer 7.7%	querer, tener 8.0%	tener 12.3%	tener 11.6%
4.	ver 3.6%	tener 5.9%	tener 5.5%	haber, pensar, saber 5.6%	decir 5.4%	ir 6.2%
5.	terminar 3.4%	estar 4.0%	estar 5.1%	hablar 4.0%	hacer 3.8%	decir 5.4%

Wulff, Ellis, Römer, Bardovi-Harlig, and LeBlanc (2009) have argued that learners are likely to be sensitive to the bias of verbs toward one construction or another rather than the token

frequency of verbs within a particular construction. Thus, a distinctive collexeme analysis (DCA) (Gries & Stefanowitsch 2004) was also computed on the data using Gries' R-script coll.analysis 3.2a in order to determine which verbs were particularly associated with either the preterite or imperfect. DCA uses an exact binomial test in order to determine the association strength between a word and two different constructions, in this case a verb with either preterite or imperfect. The absolute value of the log-transformed p value for each verb is used to determine if the verb is significantly associated with either construction. An association strength of 0 means that a verb is not distinctively associated with either construction, while a value above 1.3 indicates that the verb is significantly associated with one or the other of the constructions and is equivalent to $p < .05$.

In the instructor input results, 104 verbs were associated with the preterite (5 significantly so at $p < .05$, and 99 trended in that direction), while 24 verbs were associated with the imperfect (11 significantly so at $p < .05$, and 13 trended in that direction). In the results from CREA, 92 verbs were associated with the preterite (74 at $p < .05$), and 34 verbs were associated with the imperfect (22 at $p < .05$). Similarly, in CDE, 95 verbs were associated with the preterite (69 at $p < .05$), 30 verbs were associated the imperfect (21 at $p < .05$), and 3 verbs were not distinctively associated with the preterite or the imperfect due to the absence of tokens in either form. Therefore, within the same set of verbs, more were associated with the preterite in the instructor input data than in the corpora, although it is important to note that the majority of these verbs merely trend in that direction. When the 10 verbs with the highest association strength for each construction are examined, they are quite similar across datasets; 5 verbs in the preterite and 7 verbs in the imperfect are in the top 10 in both the instructor input and the data for at least one of the corpora (see Table 5).

Table 5. Top 10 Verbs Distinctively Associated with the Preterite and Imperfect

	Preterite			Imperfect		
	Instructor	CREA	CDE	Instructor	CREA	CDE
1.	nacer	empezar	hacer	querer	estar	estar
2.	ver	dar	empezar	estar	ser/ir	ser
3.	terminar	hacer	nacer	saber	tener	tener
4.	decir	nacer	dar	ser	querer	haber
5.	hacer	morir	ver	haber	saber	saber
6.	aprender	pasar	decir	tener	existir	querer
7.	perder	decir	pasar	pensar	poder	existir
8.	quemar	casar	morir	dibujar	andar	poder
9.	romper	decidir	terminar	esperar	deber	ir
10.	olvidar	comenzar	decidir	gustar	esperar	deber

7. Discussion

The results of the present study suggest that classroom learners are exposed to fewer imperfect tokens than preterit tokens in instructor input. In exactly half of the class sessions, no tokens of the imperfect were spoken at all by the instructors. In contrast, preterite forms occurred in all of the class sessions, although in one class they only occurred during a review activity of preterite conjugations. It is interesting to note that in two class sessions by different instructors the preterite was reviewed although it was not on the syllabus for that day, but the same never happened for the imperfect, likely due to the more complicated nature of the conjugations for the preterite.

According to principles of usage-based linguistics, exposure to such a low number of imperfect tokens would not be conducive to the acquisition of this construction. Even if students notice every one of the few imperfect forms spoken by their instructors, which is unlikely, these forms contain a limited number of verb types and few tokens per morphological ending.

Nevertheless, students presumably were exposed to preterite and imperfect tokens in written materials and also heard more of these forms on the days in which the past tense was the focus of the lesson, apart from any input outside the classroom. This additional input would likely be needed for robust acquisition, since the present data suggests that typical classroom oral input alone is insufficient.

Additionally, there was a clear bias toward the use of preterite forms, which totaled over 80% of past tense tokens in the instructor input, as compared to approximately 40% of past tense tokens of those same verbs in CREA and CDE. This is consistent with Salaberry's (2002) hypothesis as to the distribution of forms, as well as with previous findings on the French foreign language classroom (Kaplan 1987; Swain 1991). The analysis of factors affecting this distribution showed that activity significantly conditioned the use of past tense forms, such that the discussion of readings or comments unrelated to the teaching of the lesson resulted in more use of the imperfect. A closer examination of the data showed that this was principally due to instructors reviewing literature or conversing with students at the beginning of class. For example, one instructor began class by explaining why the camera was there, saying "Algunos de Uds. que estaban aquí antes vieron a mi colega que armó ese…esa cámara" (non-native speaker, 5th semester). Another instructor, while discussing a story the students had read, explained that "Margarita se sentía que ella era la que estaba pescando mucho" (native speaker, 5th semester). These activities, in which narrative discourse often occurred, encouraged the presence of background clauses and therefore the use of the imperfect (see Bardovi-Harlig 1995). In contrast, the teaching of grammar disfavored the use of the imperfect. The majority of the observed classes focused on grammatical concepts in which theoretically the preterite or imperfect could have been incorporated, yet these cases still showcased a bias toward the use of preterite forms.

For example, in an activity to practice reflexive verbs, the instructor asked the students to describe what they did that morning and explicitly told them to use preterite: "¿Qué hiciste hoy? Y obviamente pretérito, ¿verdad?" (native speaker, 3rd semester). Furthermore, instructors often gave isolated example sentences to explain grammatical concepts, and thus the prototypical context for imperfect, background clauses in narrative, was not present. Although it might be hypothesized that non-native instructors, as L2 learners of Spanish, could differ from native speakers in the proportion of preterite and imperfect in their speech, the non(native) status of the instructor did not have a significant impact on their production.

Analyses by verb revealed that the preterite bias was not due to only one or two verbs behaving differently across the datasets. The DCA for each construction showed that approximately ten verbs that were associated with preterite in the instructor input were instead associated with imperfect or neither construction in the large-scale corpora. However, which types of verbs were associated with which aspect was quite similar across the instructor oral input and Spanish language corpora. Verbs that were most strongly associated with the preterite are principally accomplishments and achievements, while verbs most strongly associated with the imperfect are primarily states. This suggests that there is indeed a distributional bias according to inherent lexical aspect even in instructor data. This bias is not as clear, however, when we look at token frequency, particularly for the verb *ser*.

In the corpora data *ser* had a much higher frequency in the imperfect, e.g. 11733 tokens of imperfect to 6206 tokens of preterite in CDE. While *ser* was the most frequent verb in both the preterite and imperfect constructions, it was more common in the imperfect. In the instructor data, however, most instances of imperfect were tokens of *ser*, but most tokens of *ser* were not in the imperfect. Some examples of instructors' use of *ser* in the preterite include discussing

famous soccer players as an extension of a textbook reading: "Maradona fue un gran jugador de fútbol, ¿no? Entonces, casi siempre se dice...unos dicen que Maradona es el mejor, o fue el mejor" (native speaker, 3rd semester), and asking about the homework at the beginning of class: "¿La tarea fue fácil o difícil?" (native speaker, 4th semester). While there were 29 tokens of *ser* in the imperfect and 48 tokens of *ser* in the preterite, this verb was still classified as significantly associated with the imperfect because it differed from the expected distribution of only 14 imperfect tokens and 63 preterite tokens of *ser*, based on the overall ratio of imperfect to preterite forms in the dataset. Nevertheless, it is unclear how this distribution would affect acquisition. Would learners be sensitive to the expected frequency of *ser* in preterite and imperfect based on other verbs and conclude that *ser* is more prototypical of the imperfect construction? Or would the much higher use of *ser* in the preterite falsely lead learners to believe this verb is prototypical of the preterite? If this pattern holds true in lower level courses as well, and *ser*, the most frequent and basic stative verb, has a higher token frequency in the preterite, then this could perhaps help explain the findings of previous research in which learners initially interpreted the preterite as a general marker of past tense and overextended the preterite into imperfect contexts even after imperfect use had emerged.

8. Conclusion

The results of this study support a bias in instructor oral input toward the preterite, which was not the case for the same verbs in large-scale corpora. Although the imperfect and preterite tended to be used according to verbs' inherent lexical aspect in the instructor input, this was not always reflected in raw token frequency, particularly for *ser*. Additional research is needed in order to determine if this type of distribution is misleading for acquisition. Furthermore, an

investigation of past tense forms in other sources of input to which learners are exposed in the classroom, such as the textbook, could clarify whether this preterite bias is unique to oral input.

In order to present more balanced input, instructors could incorporate more activities that encourage the use of narrative, such as the discussion of literature. However, given that academic discourse is different from most other communicative situations outside the classroom, it is unlikely that classroom input could ever mirror the patterns of use found in native speaker corpora. For more comprehensive experience with past tense forms in spoken Spanish, it appears necessary for learners to seek input beyond the classroom.

WORKS CITED

Andersen, Roger W. (1991). "Developmental Sequences: The Emergence of Aspect Marking in Second Language Acquisition." *Crosscurrents in Second Language Acquisition and Linguistic Theories.* Ed. Thom Huebner and Charles A. Ferguson. Amsterdam: John Benjamins. 305-24. Print.

---. (1994). "The Insider's Advantage." *Italiano lingua seconda/lingua straniera*. Ed. Anna Giacalone Ramat and Massimo Vedovelli. Rome: Bulzoni. 1-26. Print.

Andersen, Roger W., and Yasuhiro Shirai. (1994). "Discourse Motivations for Some Cognitive Acquisition Principles." *Studies in Second Language Acquisition* 16.2: 133-56. Print.

Bardovi-Harlig, Kathleen. (1995). "A Narrative Perspective on the Development of the Tense/Aspect System in Second Language Acquisition." *Studies in Second Language Acquisition* 17.2: 263-91. Print.

Bonilla, Carrie L. (2013). "Tense or Aspect?: A Review of Initial Past Tense Marking and Task

    Conditions for Beginning Classroom Learners of Spanish." *Hispania* 96.4: 624-39. Web. 22

    Jan. 2014.

Bybee, Joan. (2008). "Usage-Based Grammar and Second Language Acquisition." *Handbook of*

    Cognitive Linguistics and Second Language Acquisition. Ed. Peter Robinson & Nick C.

    Ellis. New York: Routledge. 216-36. Print.

Bybee, Joan, Revere Perkins, and William Pagliuca. (1994). *The Evolution of Grammar: Tense,*

    Aspect, and Modality in the Languages of the World. Chicago: University of Chicago Press.

    Print.

Camps, Joaquim. (2002). "Aspectual Distinctions in Spanish as a Foreign Language: The Early

    Stages of Oral Production." *International Review of Applied Linguistics in Language*

    Teaching (IRAL) 40.3: 179-210. Web. 21 Jan. 2014.

Clements, J. Clancy. (2003). "The Tense-Aspect System in Pidgins and Naturalistically Learned

    L2." *Studies in Second Language Acquisition* 25.2: 245-81. Web. 11 Apr. 2013.

Collins, Laura, Pavel Trofimovich, Joanna White, Walcir Cardoso, and Marlise Horst. (2009).

    "Some input on the easy/difficult grammar question: An empirical study." *The Modern*

    Language Journal 93.3: 336-53. Web. 17 Jan. 2014

Comajoan, Llorenç. (2005). "The Early L2 Acquisition of Past Morphology: Perfective

    Morphology as an Aspectual Marker or Default Tense Marker?" *Selected Proceedings of*

    the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second

    Languages. Ed. David Eddington. Somerville: Cascadilla Proceedings Project. 31-43. Web.

18 Jan. 2014.

---. (2014). "Tense and Aspect in Second Language Spanish." *The Handbook of Spanish Second Language Acquisition*. Ed. Kimberly L. Geeslin. West Sussex: Wiley- Blackwell. 235-52. Web. 17 Jan. 2014.

Comrie, Bernard. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. New York: Cambridge University Press. Print.

Davies, Mark. (2002a). *Corpus del Español: 100 Million Words, 1200s-1900s*. Web. 21 Jul. 2016.

---. (2002b). "Un corpus anotado de 100.000.000 palabras del español histórico y moderno." *Procesamiento Del Lenguaje Natural* 29: 21-7. Web. 7 Jul. 2016.

Ellis, Nick C. (2002). "Frequency Effects in Language Processing." *Studies in Second Language Acquisition* 24.2: 143-88. Web. 28 Nov. 2012.

---. (2009). "Optimizing the Input: Frequency and Sampling in Usage-Based and Form-Focused Learning." *The Handbook of Language Teaching*. Ed. Michael H. Long and Catherine J. Doughty. West Sussex: Wiley-Blackwell. 139-58. Web. 11 Jul. 2016.

---. (2013). "Frequency-Based Grammar and the Acquisition of Tense and Aspect in L2 Learning." *Research Design and Methodology in Studies on L2 Tense and Aspect.* Ed. M. Rafael Salaberry and Llorenç Comajoan. Boston: De Gruyter Mouton. 89-117. Web. 19 Nov. 2013.

Ellis, Nick C., and Laura Collins. (2009). "Input and Second Language Acquisition: The Roles of Frequency, Form, and Function." *The Modern Language Journal* 93.3: 329-35. Web. 18

Dec. 2016.

Ellis, Nick C., and Fernando Ferreira–Junior. (2009). "Construction Learning as a Function of Frequency, Frequency Distribution, and Function." *The Modern Language Journal* 93.3: 370-85. Web. 17 Jun. 2016.

Gries, Stefan T., and Anatol Stefanowitsch. (2004). "Extending Collostructional Analysis: A Corpus-Based Perspective on 'Alternations.'" *International Journal of Corpus Linguistics* 9.1: 97–129. Web. 18 Jun. 2016.

Hamayan, Else V. and G. Richard Tucker. (1980). "Language Input in the Bilingual Classroom and Its Relationship to Second Language Achievement." *TESOL Quarterly* 14.4: 453–68. Web. 21 Aug. 2014.

Johnson, Daniel E. (2009). "Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis." *Language and Linguistics Compass* 3.1: 359-83. Web. 20 May 2017.

Kaplan, Marsha A. (1987). "Developmental Patterns of Past Tense Acquisition Among Foreign Language Learners of French." *Foreign Language Learning: A Research Perspective*. Ed. Bill VanPatten, Trisha R. Dvorak, and James F. Lee. New York: Newbury House. 52-60. Print.

Lightbown, Patsy M. (1985). "Input and Acquisition for Second-Language Learners In and Out of Classrooms." *Applied Linguistics* 6.3: 263-73. Web. 19 Jun. 2015.

López-Ortega, Nuria R. (2000). "Tense, Aspect, and Narrative Structure in Spanish as a Second Language." *Hispania* 83.3: 488-502. Web. 1 Oct. 2014.

Lubbers Quesada, Margaret. (2006). "La percepción de las propiedades semánticas y la adquisición de la morfología verbal en el español como L2." *Estudios de Lingüística Aplicada* 24.44: 11-35. Web. 20 Jan. 2014.

McDonough, Kim, and Youjin Kim. (2009). "Syntactic Priming, Type Frequency, and EFL Learners' Production of Wh-Questions." *The Modern Language Journal* 93.3: 386-98. Web. 21 Jan. 2014.

Real Academia Española. *Corpus de referencia del español actual (CREA)*. Web. 20 Oct. 2013.

Rothman, Jason, and Pedro Guijarro-Fuentes. (2010). "Input Quality Matters: Some Comments on Input Type and Age-Effects in Adult SLA." *Applied Linguistics* 31.2: 301-6. Web. 18 May 2017.

Salaberry, M. Rafael. (2000). *The Development of Past Tense Morphology in L2 Spanish*. Philadelphia: John Benjamins. Print.

---. (2002). "Tense and Aspect in the Selection of Spanish Past Tense Verbal Morphology." *The L2 Acquisition of Tense-Aspect Morphology*. Ed. M. Rafael Salaberry and Yasuhiro Shirai. Philadelphia: John Benjamins. 397-416. Print.

Swain, Merrill. (1991). "Manipulating and Complementing Content Teaching to Maximise Learning." *Foreign/Second Language Pedagogy Research.* Ed. Robert Phillipson, Eric Kellerman, Larry Selinker, Mike Sharwood Smith, and Merrill Swain. Clevedon: Multilingual Matters. 234-50. Print.

Schwenter, Scott A., and Rena Torres Cacoullos. (2008). "Defaults and Indeterminacy in Temporal Grammaticalization: The 'Perfect' Road to Perfective." *Language Variation and*

Change 20: 1–39. Web. 13 Oct. 2017.

Tracy-Ventura, Nicole. (2008). *Developmental Readiness and Tense/Aspect: An Instructional Study of L2 Preterite and Imperfect Acquisition in Spanish.* Thesis. Northern Arizona University. Web. 10 Jan. 2014.

Vendler, Zeno. (1967). *Linguistics in Philosophy*. Ithaca: Cornell University Press. Print.

Wulff, Stephanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig, and Chelsea J. LeBlanc. (2009). The Acquisition of Tense-Aspect: Converging Evidence From Corpora and Telicity Ratings. *The Modern Language Journal* 93.3: 354–369. Web. 18 Jun. 2016.