LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

A Generative Approach to the Instructed Second Language Acquisition of Spanish *se*

Alan Juffs and Shaohua Fang
University of Pittsburgh

Abstract: This article focuses on the role of crosslinguistic patterns with verbs in the mapping of noun phrases/semantic roles to positions in morphosyntax, with a particular focus on second language (L2) development of Spanish *se*. The data set derives from high school learners of Spanish in the United States under broadly deductive and inductive learning treatments leading to explicit awareness. Using linear mixed effects modeling (LME) and binomial logistic regression, an analysis of high school learners from three schools (total $n = 138$) showed that learners based their acceptability judgments of aurally presented sentences and written production on verb classes proposed in formal linguistic theory. However, effects of the instructional intervention were limited to production data. No advantage for either deductive or inductive instruction was identified. The data show a clear role for formal linguistic categories in explaining patterns in the data. Implications for fine-tuning instructional intervention and testing of verb classes are discussed.

Keywords Spanish *se*; argument structure; generative SLA; instructed SLA

Introduction

The second language acquisition (SLA) of Spanish morphosyntax has been studied from many perspectives, including input processing (VanPatten, 1996, 2002), variationist approaches (Kanwit, 2017; Kanwit & Geeslin, 2014), and

Greatest thanks must go to Paul D. Toth for his creativity and incredible hard work in designing the instruction and collecting these data. Thanks are due to him and Judit Kormos for their painstaking editing as well as the anonymous *Language Learning* reviewers whose comments led to a much clearer and focused article. However, we alone are responsible for any misrepresentations of data and for all errors.

Correspondence concerning this article should be addressed to Alan Juffs, University of Pittsburgh, Department of Linguistics, University of Pittsburgh, Pittsburgh, PA 15260. Email: juffs@pitt.edu

The handling editor for this article was Paul D. Toth.

DOI: 10.1111/lang.12483

generative approaches (Montrul, 1999, 2000, 2001; Toth, 2000, 2008). The editor of this special volume invited researchers to provide their perspective on a large set of preexisting materials and data that had resulted from an instructional intervention to help high school learners acquire aspects of the morpheme *se* in Spanish (Toth, 2021c). Toth and Guijarro-Fuentes (2013) had already reported on a subset of these data. This article extends their findings by (a) comparing the impact of not one but two types of instructional treatment; (b) considering previously unanalyzed sentence types in the data, and (c) deploying different statistical techniques. A more recent theoretical formulation of Spanish *se* (MacDonald, 2017) also supports the linguistic analysis and discussion. Thus, the article demonstrates the value of reanalyzing data with more sophisticated techniques that was emphasized by Marsden and Torgerson (2012).

The article is organized as follows. First, we introduce theoretical constructs in generative linguistics, the evidence for their operation across languages and in SLA, and then we focus specifically on previous research on the acquisition of Spanish *se* by second language (L2) learners. Next, we summarize the methodology of this study and present results using linear mixed effects (LME) and binomial logistic regression (Levshina, 2015). We then discuss the dataset in light of generative linguistic theory, the items in the tests, and issues of teachability and learnability. We conclude by offering some suggestions for future classroom-based research based on the analysis presented in this article.

Background Literature

Generative Theory

Generative theories of language provide a precise account of linguistic competence and the limited hypothesis space within which learners can identify properties of the target language (White, 2003; 2012). The limited hypothesis space means that learners' internal cognitive resources provide sensitivity to key aspects of the input and prevent them from considering impossible grammars that might fit that input. The cognitive structures that comprise this hypothesis space are known as Universal Grammar (UG). Widely accepted theoretical and empirical arguments support the role of UG in first language (L1) acquisition (e.g., Chomsky, 1986; Crain & Thornton, 2006; and many articles in journals such as *Language Acquisition*, e.g., Orita, Ono, Feldman, & Lidz, 2021). Just as for L1 acquisition, research has demonstrated that L2 learners know more than can be explained by the input alone. This is the familiar "logical problem" argument that learners know more than the linguistic input could have

provided (White, 2003, p. 3). The evidence for such knowledge derives from experiments with learners who judge and interpret sentences that they may never have encountered, correctly rejecting sentence forms and meanings that are communicatively plausible, but ungrammatical. If learners had no knowledge of UG, these judgments would be impossible. Moreover, L2 learners do not make many plausible overgeneralizations in production (e.g., White & Juffs, 1998; studies discussed in Slabakova, 2016), although they do exhibit developmental errors, particularly in morphology of over- and undergeneralization in production that do not match the target grammar (Slabakova, 2016, Chapter 7). Generative linguistics views language as a system of interacting modules, each of which has its own internal structure. One module includes the principles of clause structure and constraints on relations among elements of a clause such as those discussed in Chomsky (1986) and Carnie (2012). These constraints are largely internal to the operation of UG. However, the UG components known as *parameters*, which account for limited structural variation among human languages, must be learned by comparing the parameter options with evidence from the input. Finally, some aspects of language are obviously acquired from input via frequent implicit and explicit exposures. For example, vocabulary must be learned by matching arbitrary forms with meanings (e.g., *cabbage* = *repollo* in Spanish and *báicài* in Chinese).

Although vocabulary learning depends on input, some elements of word knowledge rely on universal structures at the interface of cognition and linguistic representation (Pinker, 1989). Verb meanings in particular exhibit limited variation that affects clause structure (Pinker, 1989; Talmy, 1985; Wasow, 1977), which has also been the focus of extensive L2 studies (e.g., Juffs, 2000; 2009). We now focus on theories of verb meaning and morphosyntax that are relevant to the data from the study.

Verb Argument Structure

As part of its lexical entry, a verb's argument structure dictates how many and which types of noun phrases and prepositional phrases are required by a verb's meaning. For example, a verb such as *destroy* might be represented in the lexicon as [*destroy*, v.t. [AGENT, THEME]], meaning that the transitive verb *destroy* assigns two theta [θ](semantic) roles: the thing that destroys [AGENT] and the entity that undergoes destruction [THEME]. This argument structure requires two noun phrases in an active voice clause to satisfy its requirements (see 1a). In addition, semantic Agents are usually grammatical *subjects*, while semantic Themes are usually grammatical *objects* in nominative-accusative languages (Baker, 1988). When this alignment of theta role to grammatical

function is not met, languages often mark such deviation with morphology (Baker, Johnson, & Roberts, 1989) as illustrated in (1e) with the passive form of the verb.[1] Thus, the theta role, which is the semantic part a noun phrase plays in an event (Agent, Theme), is separate from the grammatical function, which is the syntactic part a noun plays in a sentence (Subject, Object). As a result, differences in verb meaning (theta roles) and morphology mean a noun phrase can appear as subject or object in a clause but still have the same semantic role. In Spanish, the morpheme *se* is involved in indicating changes in grammatical functions and as a result was the target of instruction in this study.

The Current Study

This article focuses on four main verb types in the data set relevant to *se*. First, transitive verbs in (1) that require both an overt subject and object in the active voice; next, transitive verbs in (2) that may participate in a transitive-intransitive alternation with and without morphology (depending on the language); and, last, intransitive verbs of two types, namely *unergative* and *unaccusative* verbs in (3) and (4). Note that an asterisk before a sentence means that the sentence is ungrammatical and "$_i$" and "t_i" indicate a syntactic link between a noun phrase and an "empty" position in the verb phrase.

1. Transitive: e.g., *destroy*, v.t. [AGENT, THEME].
   (a) The machine destroyed the house. AGENT verb THEME
   (b) *The machine destroyed. AGENT verb. No Theme Object.
   (c) *The house destroyed. THEME verb. No Agent, Theme in Subject position without morphology.
   (d) [$_{IP}$ The house $_i$ [was [$_{VP}$ destroyed t_i]]]. THEME Passive verb/ morphology.
   (e) The house was destroyed by the machine. THEME Passive verb with AGENT.
2. Alternator: e.g., *open*, v.t.; v.i., [(AGENT); THEME]
   (a) The driver opened the door.
   (b) *The driver opened.
   (c) [$_{IP}$ The door $_i$ [VP opened t $_i$]]
   (d) [$_{IP}$ The door $_i$ [was [VP opened t $_i$]]] (by the driver).
3. Unergative: e.g., *laugh*, v.i, [AGENT]
   (a) [$_{IP}$ A child [$_{VP}$ laughed]].
   (b) *The clown laughed a child. (= *made the child laugh)
   (c) *There laughed a child.
   (d) *A child was laughed.

4. Unaccusative: e.g., *arrive*, v.i., [THEME]
   (a) [$_{IP}$ A train $_{i}$ [$_{VP}$ arrived t$_{i}$]] (at the station)
   (b) *The driver arrived a train (at the station).
   (c) There arrived a/?the train (at the station).
   (d) *A train was arrived.

Example (1) shows that transitive verbs require both the Agent and the Theme to appear in the clause in English. The example in (2) shows that verbs that describe a change of state (e.g., open/close; melt/freeze) have an optional Agent. If the Agent is absent, the Theme appears as the grammatical subject in English without additional morphology as in (2c). For transitive verbs in (1), grammatical function-changing morphology in the form of the passive construction is required in (1d) and (1e) to permit the Theme to appear in subject position, which is also possible for alternators (2d).

Intransitive verbs in (3) and (4) belong to two distinct types (Perlmutter, 1978). Unergative verbs, which often describe internally caused events such as *cough* and *laugh* semantically require an Agent. Unaccusative verbs usually describe an entity moving, such as *arrive*, or changing state, such as *die* and require a Theme. Because Themes are prototypical grammatical objects, some unaccusatives exhibit syntactic properties that reflect this, especially when the noun phrase (NP) is indefinite. In English, the single argument is compatible with the *there* construction in (4c) for unaccusatives, but not unergatives (3c). The generative account is that the Theme subject is linked to a position inside the verb phrase indicated by "t $_{i}$" in (4a) passive in (1d) and (2d), and inchoative in (2c), which is the normal position for the object noun phrase.[2] Neither intransitive class permits the passive voice in English.

Other languages demonstrate different morphosyntactic reflexes of these verb classes. To accommodate universals and variation, some theories of verb meaning propose that a verb root's meaning can be analyzed into components such as CAUSE, MANNER, and PATH (Pinker, 1989; Talmy, 1985) and that these components are actually the source of the theta roles assigned by the verb. Furthermore, crosslinguistic variation in how meaning components are combined in verb roots that superficially "mean the same" in different languages is well-documented (Talmy, 1985). An example is that in English one can say "The movie disappointed me," while in Chinese the same idea must be expressed by a periphrastic causative "The movie made me disappointed," indicating that English allows CAUSE to be part of the root meaning of a verb of emotion, while Chinese prefers not to include such a meaning component in the verb root (Juffs, 1996). Research has shown that L2 learners can develop

knowledge of abstract regularities in the mapping of meaning components (CAUSE, PATH, MANNER, etc.) in verb roots and the impact these regularities have on clause structure (Juffs, 2000). However, learners face challenges in acquiring the morphology associated with such mapping, especially when the morphology affects meaning components associated with CAUSE as is the case with Spanish *se* (Toth, 2000), and as we document in the next section.

Although verb root meaning and morphosyntax links vary crosslinguistically, universal regularities certainly exist. For example, the *split intransitive hypothesis* illustrated in (3) and (4) has been validated in at least three separate language families and explains developmental patterns in the acquisition of the associated morphosyntax of those languages. Examples include Indo-European languages such as English, French, Italian, Spanish, and Greek (Lozano, 2006; Sorace, 1993; Zobl, 1989), Sino-Tibetan (Yuan, 1999), and Japanese (Hirakawa, 2001). Where alternating change-of-state verbs are concerned (e.g., *freeze/melt* and *open/close*), many languages require additional morphology to indicate an unambiguously caused change of state (Haspelmath, 1993; Tai, 1984). In contrast, many Indo-European languages favor semantically caused, syntactically transitive constructions as the unmarked root verb and require morphology to indicate "spontaneous," inchoative events with a Theme in subject position. This morphology may overlap in form with passive voice.

The conclusion from this research is that constrained crosslinguistic variation exists that results in patterns of theta roles, morphological changes, and mapping to grammatical functions across the world's languages (Baker, 1988). Although there are "verb classes," they do not operate in isolation as unrelated structures but form a unified system of subtle and abstract meaning-to-morphosyntax links. From a generative perspective, therefore, one cannot simply examine one verb type alone because to do so would fail to appreciate the interlocking nature of the system. Spanish is a language that requires morphology to change the mapping of theta roles to grammatical functions and we now turn to the focus of the original study.

Spanish *se*

Zyzik (2006, p. 454) lists many functions of *se*: anticausative, passive, reflexive, impersonal, intransitive, aspectual (telic), and inherent. In addition to these complex functions, English speakers must learn that Spanish is a subject *pro-drop* language and that transitive, causative constructions can make reference to Agent arguments through subject-verb agreement suffixes rather than overt noun phrases. Thus, language development requires properly attending

to and interpreting the presence or absence of overt subjects along with the subject-verb agreement suffixes, the presence or absence of *se*, and the semantics of the verb.

This section illustrates only the various forms of *se* which are relevant to the data set. MacDonald's (2017) article constitutes a recent generative analysis of *se*, using both syntactic and semantic interpretation tests to support his analysis. He demonstrates that different meanings of *se* have separate syntactic sources, even though the superficial clausal syntax may appear the same. The first *se* is the reflexive illustrated in (5).[3] This form was not the focus of the study but learners had received explicit instruction on reflexive *se* in the previous year and lesson transcripts revealed knowledge of this use.

5. *Juan se vio (a sí mismo) en el espejo.*
   John SE_{REFL} saw (self-same) in the mirror.
   "John saw himself in the mirror."

MacDonald (2017) further distinguishes among three non-reflexive *se* meanings. First, the SE_{IMPERS} in (6), which could be also paraphrased in English as impersonal *one*. The SE_{IMPERS} in (6) is the only *se* that allows the direct object marker (DOM) *a*. The passive SE_{PASS} is illustrated in (7a) with the clause *para acabar* indicating that some Agent deliberately burned the forest. The anticausative $SE_{ANTICAUS}$ is in (7b) with the phrase *por si solo* indicating that no Agent is implied.

6. *Se conoce a María como buena madre.*
   Se_{IMPERS} know DOM Maria as good mother.
   "Mary is known as a good mother." or "One knows Mary as a good mother."

7.
   a. *Se quemó el bosque para acabar con la plaga.*
      SE_{PASS} burned the forest for finish with the blight.
      "The forest was burned in order to end the blight."
   b. *Se quemó el bosque por si solo anoche.*
      $SE_{ANTICAUS}$ burned the forest by self alone last night.
      "The forest burned by itself last night."

Generative theory provides a precise notation to reflect the subtle facts in (6) and (7). Within this theory, the verb phrase is governed by inflectional functional categories (Carnie, 2012). Functional categories account for tense, number, aspect, agreement, and other closed class grammatical phenomena. For *se*, MacDonald proposes the functional category *Voice Phrase* [VoiceP], which

indicates whether the clause is in passive or active voice. In his account of the *se* sentences in (6) and (7), MacDonald used a variety of syntactic tests with purpose clauses (e.g., [7b], with adverbs) to demonstrate that *se* impersonal (6) and *se* passive (7a) constructions have an implicit, unexpressed Agent theta role, whereas *se* anticausative (7b) does not. Therefore, VoiceP is "thematic" in SE_{IMPERS}/SE_{PASS} (8a) and (8b) as indicated by [θ], which means that some external agent or experiencer must be present for correct interpretation. No such external role is implied in (8c) $SE_{ANTICAUS}$ (MacDonald, 2017, p. 367).

8. Functional Features associated with *se* in Spanish
   a. [VoiceP *pro* $Voice_{[\theta]\ se}$] → SE_{IMPERS}/SE_{PASS}
   b. [VoiceP $Voice_{[\theta]}$] → Periphrastic Passive
   c. [VoiceP $Voice_{se}$] → $SE_{ANTICAUS}$

The Spanish passive *se* (8a) is equivalent in meaning to the periphrastic passive (8b) in which the agent optionally appears in a "by" phrase, for example, *El bosque fue quemado por el granjero*, (The forest was burned by the farmer). It is these parameterized properties of [VoiceP] in (8), requiring the morphological expression by *se* of impersonal passive and anticausative in Spanish (cf. Alexiadou, 2014), that English-speaking learners must acquire. It also predicts a difference between the passive/impersonal *se* and anticausative *se* in acquisition.

In addition, English-speaking learners of Spanish must learn that in the right context, pro-drop is allowed with transitive verbs, making them appear with only one overt NP in contrast to English. Thus, the Spanish sentence, *(pro) Quemó el bosque*, is equivalent to English "He (or she or it) burned the forest." Learners also should not use *se* with active transitive sentences when an Agent is mentioned in the clause or prominent in the discourse. Intransitives likewise do not require *se*, particularly unaccusative verbs. Acquisition will involve knowledge that *se* is required for alternators with a Theme in subject position while English requires no morphology as in (2c). In contrast, English-speaking learners of Spanish may overgeneralize *se* morphology to mark Themes in Subject position with unaccusatives (e.g., Toth, 2000), which is comparable to learners of English L2 who overgeneralize passive morphology with unaccusatives (Oshita, 2001; Zobl, 1989). Thus, although simply hearing Spanish in class and elsewhere might expose the learners to all the functions listed by Zyzik (2006), theoretically informed instruction can provide targeted, clear input for the various uses that may assist learners in untangling the diverse form to meaning mappings. Teachers who understand that the uses of *se* are constrained by the transitivity of verbs could also be

more effective at providing explicit feedback than those who do not know why learners may be overgeneralizing.

Generative Linguistics and Instructed SLA

The goal of instructed SLA is to improve learning outcomes, preempt "plateaus in development" (i.e., *fossilization*), and expedite developmental processes (Larsen-Freeman & Long, 1991, p. 361). The concern for GenSLA and grammar instruction is whether explicit metalinguistic teaching can achieve such results (Ellis, 1993; Han & Finneran, 2014; Long, 2007, pp. 121-123). There are various ways instruction might affect linguistic competence in communicatively oriented, task-based curricula. The first option is that declarative knowledge about the grammar can directly translate into unconscious linguistic competence. This is a version of the *strong interface* position (Ellis, 1993). The GenSLA point of view is that the strong interface is not a viable position. The theoretical reason is that declarative knowledge of metalinguistic rules is not of the same cognitive type as linguistic competence (Schwartz, 1999; Truscott & Sharwood Smith, 2011, VanPatten, 2016). For example, a rule for Spanish verb classes in (1) – (4) would not capture the subtleties in (8). In contrast, the *weak interface* position is that declarative instruction aimed at creating implicit knowledge might aid learners in noticing certain phenomena, but that linguistic competence cannot incorporate this declarative knowledge directly. Instruction that manipulates the primary data to be more salient to the abstract module for easier processing will thus be encouraged (cf., VanPatten, 2002). Output practice and corrective feedback can reinforce such processing (Ranta & Lyster, 2017). The third position is that declarative knowledge is not useful in changing linguistic competence even for the purpose of noticing. For example, Truscott & Sharwood-Smith (2011, p. 524) specifically state that "The language module deals with linguistic information automatically; awareness of input is sufficient for its development and noticing and understanding are not directly relevant."[4]

Despite this controversy, Whong, Gil, and Marsden (2013, p. 4) have emphasized the importance of instruction informed by GenSLA: "GenSLA is offered as a principled basis of language pedagogy, with contributions in terms of the understanding of properties of language as well as how knowledge of these properties develops as a result of exposure to language input." A nuanced GenSLA view is that theory can aid instructional design to align the abstract form-meaning categories with the pedagogical intervention, thereby providing the learner with input and output practice in the optimum format for acquisition. This position reflects evolving thinking on the integration of learner

internal contributions, processing, cognition, and pedagogy (Slabakova, Leal, Dudley, & Stack, 2020).

The intervention in this study deployed two modes of instruction discussed in the procedures section. The first was deductive instruction and the second was PACE instruction, which was a kind of guided inductive instruction based on Adair-Hauck and Donato (2016). Both instructional groups were encouraged to develop an explicit understanding of the target structures, albeit via different paths. Cognitivists such as Leow (2015) suggest that PACE-like guided instruction may lead to deeper processing and thus better learning outcomes. However, GenSLA would predict no difference between the two because both offer opportunities for noticing the input through explicit rules that cannot directly interface with linguistic competence described in MacDonald (2017). The advantage of instruction over simple exposure is the provision of clear input for learning.

This article principally considers the study from the vantage point of linguistic theory and so it concentrates on the test responses to four specific verb types and the grammatical/ungrammatical use of *se* including regular transitive verbs, alternator verbs, unaccusative, and unergative verbs.

Research Questions

Based on the data provided in Toth (2021c, 2021d) and publicly available on IRIS (https://www.iris-database.org), our analysis includes four main effects and their interactions: Transitivity with four levels: Accusative, Alternator, Unaccusative, and Unergative; TestAdmin with three levels: Pretest, Posttest, Delayed Posttest; Treatment, with three levels: Control, Deductive, and PACE, and Morphology, *SE* or NULL. Factors that were not part of the planned experimental design were School, Participant, and Test Item; however, these factors were included in the statistical analysis and are discussed in the results.

The following research questions were addressed, with the interaction between knowledge of verb class and morphology, change over time, and instruction being the most important consideration.

1. Do the learners distinguish null morphology (i.e., the omission of *se*) or *se* by verb class/function? In other words, do they show knowledge of the three kinds of *se* in (8), but not extend the use of *se* to inappropriate verb classes and clause types, such as accusative verb sentences with both arguments in the clause, or unaccusatives?
   (a) In acceptability judgment data
   (b) In production data

2. Is there a change over time in the learners' knowledge of the uses of *se*?
   (a) In acceptability judgment data
   (b) In production data
3. Is there a role for instruction type: no specific instruction (control), deductive, and PACE (inductive) on participants' use of null vs. *se* morphology?
   (a) In acceptability judgment data
   (b) In production data

Method

Participants

The 138 L2 learners were third year high school students from three Midwestern U.S. institutions with the pseudonyms: Walton ($n = 69$), Grey ($n = 43$), and LaForge ($n = 23$). On a demographic questionnaire, none of the participants reported ever having lived in a Spanish-speaking country or using Spanish on a regular basis outside of the classroom. Heritage language learners were not included in the analysis and any student who had reported regular use of Spanish outside the classroom was excluded. In the available data set, 37 native speakers (NS) provided a validation of the test materials in the study. NS judgments are not reported in this article because they took the test only once, received no instruction, and are not part of the statistical analysis of the learner data.

Materials

All of the materials had been piloted and the three different versions of both tasks used in data collection had been shown to be equivalent based on Cronbach alpha statistics (Toth & Guijarro-Fuentes, 2013). The complete instructions, lesson plans and materials, as well as the test instruments (Toth, 2021a; Toth, 2021b) are publicly available on IRIS (https://www.iris-database.org). The guided production task was administered first, so that the acceptability judgment task (AJT) would not influence production task responses for the pretest data collection. Learners were given a pretest, an immediate posttest after instruction, and a delayed posttest about six weeks after instruction.

Acceptability Judgment Task (AJT)

The complete AJT contained 88 items (Toth, 2021c) in which the participants listened to a sentence and assessed how well the sentence described the picture that they saw. Accusatives and alternators each had eight clause types that varied word order, use of *se*, and pro-drop. This diversity makes analysis challenging because of the number of variables that were not strictly controlled

(e.g., word order). Therefore, we chose only four clause types that matched with either null morphology or *se* that were clearly marked in the materials in Toth (2021c) as either grammatical or ungrammatical. We avoided sentence types marked as pragmatically questionable in the key to the grammaticality of each sentence type in Toth (2021c) because under certain circumstances such sentence types were marginally acceptable.

Table 1 presents the 12 sentence types from the AJT with null morphology and with *se* which were used for the statistical analysis. First, four clause types with accusative (Acc) verbs were included: The first omitted the overt subject in favor of *pro*, including only the Theme (3 items). The second represents an ungrammatical overuse of *se* because both arguments are present (3 items); the third is also ungrammatical because the Theme is in subject position with *se* absent (3 items); finally, there is the grammatical passive/impersonal *se* (3 items). The total items for accusatives per test were 12. The same sentence types were used with alternator verbs, with the difference being that the interpretation of alternators with only the Theme and *se* is inchoative, rather than passive with theme in subject position. Alternators were represented by 12 items on each test. Intransitives were represented by eight items each, four with *se* (ungrammatical) and four without *se* (grammatical). Therefore, we analyzed 40 out of 88 items on each version of the test. The three versions of each task with varied lexical items were given to the learners randomly in as a pretest, and then individuals cycled through the remaining versions in the immediate and delayed posttests. A Likert scale of 1-6, with 1 being totally unacceptable and 6 being totally acceptable, was used. A 4-second gap occurred between each item to force more implicit judgments, that might tap abstract knowledge rather than metalinguistic knowledge (Ellis, 2009; Toth & Guijarro-Fuentes, 2013). The complexity of the pictures also varied, with some pictures having only inanimate objects with no action (e.g., Test A, #1, books with titles) and others having multiple people, buildings, and objects with action (e.g., Test A, #5, with a machine and buildings). Example pictures are provided in Figure 1.

The principal focus of the instructional intervention were the two types in bold in Table 1, the agentless passive ($SE_{IMPERS, PASS}$) and anticausative uses of *se* ($SE_{ANTICAUS}$), which are both grammatical. However, it is vital to determine if the learners also know what is not possible: That is, do they know the constraints on the morpheme's distribution by verb class beyond the instructed clause types? Recall that GenSLA does not treat verb classes as structures, but seeks a deeper understanding of the entire semantics, syntax, and argument changing morphology system.

Table 1 Sentence picture matches with and without *se* and grammaticality

Sentence/picture match type and one example. (Abbreviations in parentheses for key to Figures and Appendices)	*Null/Se*	*Grammaticality: √ or * (English comment)*
Accusative-causative-transitive: Destruyó la casa. ([*pro*] Destroyed the house.) (Acc-caus-tr)	Null	√ (*in English with pro-drop subject)
Accusative-causative-transitive-SE: La máquina se destruyó la casa. (The machine SE destroyed the house.) (*Acc-caus-trans SE)	*Se*	* (?The machine destroyed the house [by] itself)
Accusative-intransitive: La casa destruyó. (The house destroyed.) (Acc-tr)	Null	*
Accusative-intransitive-SE: **Se destruyó la casa. (SE destroyed the house.)** (Acc-intr SE)	***Se***	**√ (=Agentless passive in English; "The house was destroyed.)**
Alternator-causative-transitive: Abrío la ventana. ([*pro*] Opened the window.) (Alt-caus-tr)	Null	√ (* in English with pro-drop subject)
Alternator-causative-transitive-SE: Gabriel se abrío la ventana. (Gabriel SE opened the window.) (*Alt-caus-trans SE)	*Se*	* (* in English)
Alternator-Intransitive: La puerta abrío. (The door opened.) (Alt-Intr)	Null	* (√ in English)

(*Continued*)

Table 1 (Continued)

Sentence/picture match type and one example. (Abbreviations in parentheses for key to Figures and Appendices)	*Null/Se*	*Grammaticality: √ or * (English comment)*
Alternator-Intransitive-SE: **La puerta se abrío. (The door SE opened.)** (Alt-intr SE)	***se***	√ **(The door opened)**
Unaccusative-intransitive: El avión desciende. (The plane descends.) (Unacc-intr)	Null	√ The plane descends.
Unaccusative-intransitive-SE: El avión se desciende. (The plane SE descends.) (*Unacc-intr SE)	*se*	Compare: *The plane is arrived/?descended.
Unergative-intransitive: El bebé llora. (The baby cries.) (Unerg-intr)	Null	√ The baby cries.
Unergative-Intransitive-SE: El bebé se llora. (The baby SE cries.) (*Unerg-intr-se)	*se*	*The baby is cried. Note: "The baby cries to itself,"could be plausible.

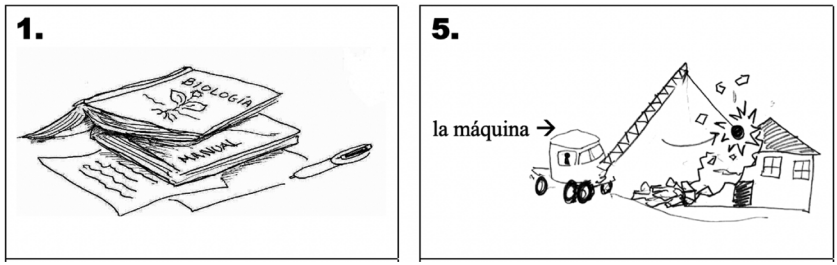

Figure 1 Example pictures.

Guided Production Task

The guided production task was a written production task based on pictures. Participants were instructed to use the pictures to create well-formed sentences in Spanish. Each picture prompt included only two vocabulary items: a nonfinite verb, followed by a logical direct object of the verb if accusative (e.g., Test A, #8, "paint, the wall" or a subject if intransitive (e.g., Test A, #4 "cry, Cristina"). Three items each targeted accusative, alternator, unaccusative, unergative verbs, and *optional theme* verbs such as *write* (a card), *eat* (a pizza), and *drink* (lemonade), which are inherently accusative but allow optional omission of the theme argument.

Instructional Materials and Procedure

In a quasi-experimental design, students were assigned in intact classes to one of three groups: deductive instruction, PACE instruction (based on Adair-Hauck & Donato, 2016), and a control group that received regular classroom teaching (with no explicit grammar focus) in tasks writing to Argentine pen pals. The assumptions underlying the deductive approach in some theories are that explicit/deductive instruction provides an explicit jump start for later procedural, automatized learning (DeKeyser & Juffs, 2005, pp. 440-442). Learners received explicit rules followed by whole class and small group practice in reading, writing, and speaking. The PACE instruction constituted teacher-guided, inductive instruction with students developing explicit rules and reaching a consensus through teacher-led *instructional conversation* (Tharp & Gallimore, 1988) which directed students to an acceptable rule. Once the explicit rule was derived, the class engaged in the same whole-class and small-group communicative activities as the deductive group. This instruction reflects a range of focus-on-form practices that involve noticing of

content and form, which is thought to promote acquisition in instructed SLA (Ranta & Lyster, 2017).

There was one class in each of the three schools assigned to the PACE treatment ($n = 49$), one in each school assigned to the deductive treatment ($n = 49$), and two in the Walton school for controls ($n = 40$). Both instructed groups received input on the same structures over three consecutive 90-minute lessons during class meetings that occurred every other day. The first lesson involved the impersonal *se*; Lesson 2 concerned the agentless passive with accusative verbs (SE_{IMPERS}/SE_{PASS}); Lesson 3 added the use of *se* with alternator verbs ($SE_{ANTICAUS}$), building on Lessons 1 and 2. Thus, instruction was informed by generative grammar and divided to maximize the effect for noticing under assumption in instructed SLA that noticing promotes acquisition.

Results

Recall that Research Question 1 concerned the appropriate and inappropriate use of *se* with different verb classes. Specifically, learners should not use *se* with accusative verbs with both arguments present, nor should they extend the use of *se* with intransitive verbs (unaccusatives and unergatives). Moreover, research questions two and three targeted the possible effects of instructional treatment over time, in the pre-, post-, and delayed posttests. An acceptability judgment task (AJT) and a production task were used to evaluate these research questions.

Statistical Approach

Mixed-effects regression models can incorporate the random influence of participants, items, and school to permit a focus on the main factors that are the focus of the study, which are the interaction of verb class, morphology, instructional treatment, and time of test. Another advantage is that these models allow group comparisons with missing data points due to a participant who missed a test, which is not possible with ANOVA. (See Bates, Maechler, Bolker, & Walker, 2015, for a review and Moranski & Zalbidea, 2022). All statistical analyses were conducted in the R environment (version 3.6.2; R Core Team, 2019).

Acceptability Judgment Task

AJT: Descriptive Statistics

Raw score means and standard errors are illustrated by box plots in Figure 2 for the learners by treatment group and test administration. The data visualization in Figure 2 suggests possible differences among the groups over

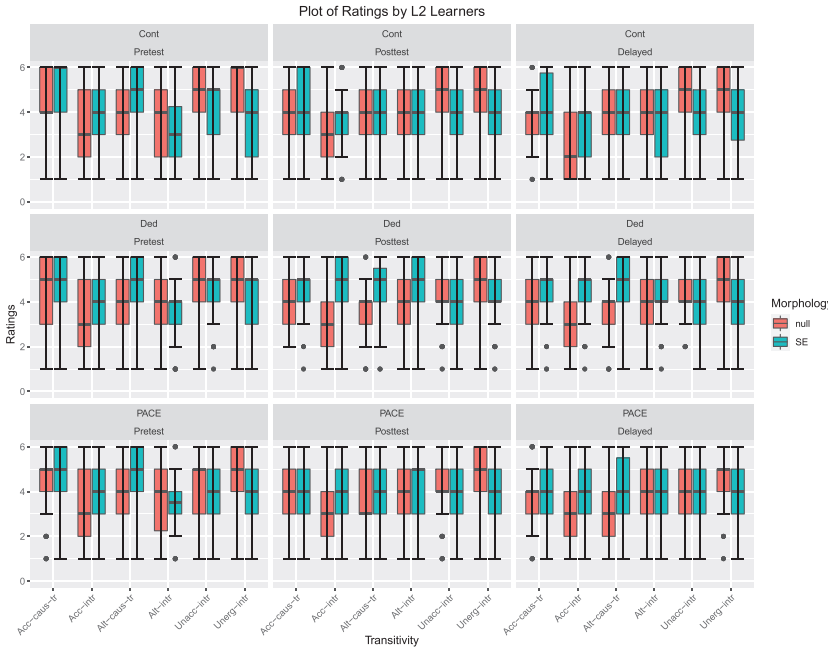

Figure 2 Boxplots of raw scores on 12 clause type-picture matches for learners by treatment group and time. [Color figure can be viewed at wileyonlinelibrary.com]

time. For example, for the mean of the alternator-intransitive sentences (e.g., in English *The door opened*, which requires *se* in Spanish), the students in the deductive group increased their approval of sentences with *se* from the pretest to the posttest, with the median score increasing from 4 to 5, and 50% of the scores ranging from 5 to 6.

Because every student might use the Likert scale differently, raw score data were z-score transformed to mitigate the scale biases among participants, following Schütze and Sprouse (2013). All participants' scores appear in Table 2. The dependent variable for statistical analysis was thus a set of z-scores (mean = 0; standard deviation = 1) derived from ratings on the Likert scale for each sentence type, and for all the learner groups, by treatment group and time of testing. Each transformed score represents how many standard deviations the raw score is away from that specific participant's average score of 40 items. Thus, the z-score transformation was not simply a transformation of all the raw scores to a z score, but designed to account for each participant's use of the scale. The formula used is:

Table 2 Descriptive results by normalized scores by verb class, treatment, and time (z:mean = 0, *SD* = 1)

	Contrl			PACE			Ded		
Verb class/ sentence type	Pre	Post	Del	Pre	Post	Del	Pre	Post	Del
Acc-caus-trans: Destruyó la casa. ([*pro*] Destroyed the house.)	0.25 (1.02)	−0.16 (0.86)	−0.34 (0.91)	0.21 (1.01)	−0.02 (0.87)	−0.26 (0.95)	0.14 (1.09)	−0.27 (0.85)	−0.40 (0.99)
*Acc-caus-trans-SE: La máquina se destruyó la casa. (The machine SE destroyed the house.)	0.56 (0.98)	0.12 (0.93)	0.01 (0.93)	0.43 (0.95)	0.02 (0.94)	0.12 (1.0)	0.31 (0.92)	0.21 (0.86)	0.27 (0.84)
*Acc-intr: La casa destruyó. (The house destroyed.)	−0.64 (1.10)	−0.65 (1.06)	−1.01 (0.91)	−0.46 (1.20)	−0.76 (1.04)	−0.79 (0.99)	−0.53 (1.31)	−0.78 (1.01)	−0.93 (1.0)
Acc-intr-SE: Se destruyó la casa. (SE destroyed the house.)	**−0.20 (0.87)**	**−0.27 (0.84)**	**−0.38 (0.82)**	**−0.19 (0.91)**	**0.14 (0.90)**	**0.03 (0.86)**	**−0.10 (1.05)**	**0.41 (0.83)**	**0.17 (0.93)**
Alt-caus-tr: Abrío la ventana. ([*pro*] Opened the window.)	−0.05 (0.99)	−0.23 (0.87)	−0.22 (0.87)	−0.17 (1.09)	−0.34 (0.97)	−0.45 (0.92)	−0.22 (1.03)	−0.47 (0.87)	−0.50 (0.83)
*Alt-caus-tr-SE: Gb. se abrío la ventana. (Gb. se opened the window.)	0.48 (0.95)	0.06 (0.89)	0.16 (0.86)	0.53 (0.88)	0.13 (1.02)	0.21 (0.97)	0.36 (0.92)	0.23 (0.91)	0.26 (0.96)

(*Continued*)

Table 2 (Continued)

	Contrl			PACE			Ded		
Verb class/ sentence type	Pre	Post	Del	Pre	Post	Del	Pre	Post	Del
*Alt-intr: La puerta abrío. (The door opened.)	0.19 (1.05)	0.09 (0.86)	−0.20 (0.99)	−0.20 (1.12)	−0.18 (1.02)	−0.20 (1.02)	−0.32 (1.08)	−0.23 (0.91)	−0.2 (1.04)
Alt-intr-SE: La puerta se abrío. (The door SE opened.)	**−0.38 (0.89)**	**−0.12 (0.92)**	**−0.39 (0.84)**	**−0.38 (0.95)**	**0.19 (1.10)**	**0.09 (0.94)**	**−0.49 (1.01)**	**0.42 (0.88)**	**0.12 (0.89)**
Unacc-intr: El avión desciende. (The plane descends)	0.45 (1.01)	0.48 (0.85)	0.39 (0.83)	0.24 (0.93)	0.22 (0.93)	0.06 (0.84)	0.35 (0.98)	0.16 (0.79)	0.11 (0.81)
*Unacc-intr-SE: El avión se desciende. (The plane SE descends)	0.16 (0.96)	0.04 (0.82)	−0.14 (0.95)	0.03 (0.92)	−0.14 (0.91)	0.02 (0.86)	0.12 (0.93)	0.05 (0.85)	0.01 (0.83)
Unerg-intr: El bebé llora. (The baby cries)	0.63 (1.02)	0.58 (0.91)	0.62 (0.81)	0.63 (0.87)	0.47 (0.88)	0.38 (0.88)	0.50 (1.04)	0.29 (0.88)	0.20 (0.98)
*Unerg-intr-SE: El bebé se llora. (The baby se cries)	−0.12 (0.99)	−0.17 (0.94)	−0.17 (0.87)	0.08 (1.0)	−0.23 (0.97)	−0.09 (0.88)	0.12 (1.00)	0.07 (0.82)	−0.09 (0.88)

$z = (\gamma - \mu)/\sigma$
γ: raw rating from each observation
μ: mean of all ratings from a given participant
σ: standard deviation of these ratings

AJT: Statistical Analysis

To check for assumptions required by the statistical tests, plots of model residuals against fitted values and Q-Q plots based on z-scores revealed no obvious deviations from normality and homoscedasticity. These assumptions being met, linear mixed-effects models (Baayen, Davidson, & Bates, 2008) were fitted using the lme4 package (version 1.1.23; Bates et al., 2015) and graphs were generated using the ggplot2 package (Wickham, 2016). The coding by R default procedure is explained for each test. Following a stepwise model selection suggested in Baayen (2008), the simplest model with a random intercept for items fitted was used, and independent variables and their interactions were incrementally added to the same model to estimate which predictors significantly improve model fit by means of likelihood ratio test (LRT). LRTs estimate the amount of variability in the data explained by an effect, and if an effect does not contribute to explaining variance in the model, it is then dropped from the regression analysis.

Following the focus of the research questions regarding use of *se* with verb classes, type of instruction and change over time, possible main effects included Transitivity (verb class), Test Administration (TestAdmin), Instructional Treatment, and Morphology (use or omission of *se*). Possible random effects included School, Test Item, and Participant. The extent to which each predictor and their interactions contributed to the variance in the normalized grammaticality judgments was assessed by model comparison with likelihood ratio tests (LRT). As a result, by means of LRT, the optimal model was finalized comprising only Transitivity, TestAdmin as main effects and the interactions of Transitivity* Morphology; Transitivity* Morphology* TestAdmin.

Importantly, Treatment (Control, Deductive, and PACE) did *not* contribute to an explanation of the variability in the acceptability judgment data and therefore did not appear in the final model, which means that for Research Question 3 there was no effect for treatment in the AJT data. A by-participant random intercept was used to model the variation resulting from the fact that individual participants may judge the same items differently as they are assumed to perceive the scale along 1-6 differently.[5]

In LME models, sources of variance beyond the factors manipulated in the design (i.e., Treatment, Transitivity, Morphology, TestAdmin) are controlled

for. These include responses by individuals and the items themselves, which used to be handled in ANOVAs by fixed effects with participants and item analysis (Barr, Levy, Scheepers, & Tily, 2013, p. 260). Such effects are *random*, and only Item and School were treated as random effects in the model. Following Barr et al. (2013), we maintained a maximal random effects structure justified by the experimental design, including random intercepts for Item and School, with by-Item random slopes for Treatment and by-School random slopes for every within-subject factor. We eventually removed the by-School random intercept and by-Item random slopes for Treatment because they did not contribute to explanation of variance in the model, thus only keeping random intercepts for Item. The variance of School is rather small ($SD < 0.0001$), and therefore excluding its effect from our model would not make any reliable difference. What this means in lay person's terms is that we tested whether school or items explained variability in the AJT scores and concluded that they did not. Eleven percent (11%; Marginal R^2) of the variance was explained by the fixed effects and 13% (Conditional R^2) of the variance was explained by the both the fixed and random effects. The final model appears in Appendix S1, where the *baseline* score to which all the other scores are compared is calculated and is labeled (Intercept) in the first row of the table. Thus, the (Intercept) in Appendix S1 represents the scores for Accusative-causative-transitive, Null, Pretest, which was selected as it was the first alphabetically, but it also makes sense as a base line comparison point because it represents scores for pretest sentences, without *se*, for transitive verbs.

The focus of the analysis is the potential interaction of Transitivity, Morphology, Treatment, and TestAdmin, so we report those first. Reliable interactions among Transitivity, Morphology, and TestAdmin were found. That is, changes in rating occur over time for the use of *se* among the verb classes, even though the effect for instructional group was not significant. To clarify this result, we concentrate on the delayed posttest interactions as these results would most likely reflect competence grammar rather than short term instructional effects. Recall that instruction focused on impersonal *se* in Lesson 1, passive *se* in Lesson 2, and inchoative *se* in Lesson 3. Ratings by the learners increased for accusative with passive SE_{IMPERS}/SE_{PASS} (Transitivity [Acc-intr] × Morphology [SE] × TestAdmin [Delayed], $\beta = 0.267$, $SE = 0.029$, $t = 3.068$, $p < 0.01$) from Pretest to Delayed test. That is, accusative verbs with one argument and *se* were accepted more over time (e.g., *Se destruyó la casa*). The focus of instruction was also on $SE_{ANTICAUS}$ and again ratings increased reliably from Pretest to Delayed posttest (e.g., *La puerta se abrió*; Transitivity [Alt-intr] × Morphology [SE] × TestAdmin [Delayed],

$\beta = 0.743$, $SE = 0.089$, $t = 8.358$, $p < 0.001$). However, ratings of alternators without $SE_{ANTICAUS}$ did not change reliably, although they increased slightly, away from the target grammar, from Pretest to Delayed posttest (e.g., **La puerta abrió*). This interaction approached significance (Transitivity [Alt-intr] × Morphology [null] × TestAdmin [Delayed], $\beta = 0.162$, $SE = 0.088$, $t = 1.832$, $p = 0.07$).

Two other reliable changes involve transitive verbs without *se* (Transitivity [Acc-caus-tr] × Morphology [null] × TestAdmin [Delayed], $\beta = -0.379$, $SE = 0.087$, $t = -4.344$, $p < 0.001$). The ratings of grammatical sentences with pro-drop in a causative sentence without *se* (e.g., [*pro*] *Destruyó la casa*) decreased reliably (see Table 2), again moving learners away from the target grammar. Moreover, ungrammatical sentences without *se* (e.g., **La casa destruyó*) also decreased in ratings (Transitivity [Acc-intr] × Morphology [null] × TestAdmin [Delayed], $\beta = -0.227$, $SE = 0.087$, $t = -2.598$, $p < 0.01$). No other interactions were reliable.

The results also revealed an interaction between Transitivity and Morphology ($\chi^2 = 50.58$, $df = 5$, $p < 0.001$). These reliable interactions are indicated in Appendix S1 with asterisks. Crucially, both unaccusatives and unergatives with the *se* morpheme are correctly rated lower than without *se* and this does not change over time, even though the acceptance of ungrammatical use of *se* with intransitives did increase slightly but not significantly. Grammatical accusative intransitives with *se* are rated higher than without *se* (e.g., *Se destruyó la casa* vs. **La casa destruyó*).

Finally, the results from a likelihood ratio test (LTR) showed a main effect of Transitivity ($\chi^2 = 15.87$, $df = 5$, $p < 0.01$), indicating that ratings for other verb and clause types differed from the reference level (Intercept) by verb class; TestAdmin ($\chi^2 = 71.51$, $df = 2$, $p < 0.001$) was also reliable, indicating that overall, the average normalized ratings decreased from the pretest session to the delayed posttest session in a significant way based on all 40 items. This result means that learners rated sentences lower overall in the experiment from pretest to delayed posttest, but we note the interactions previously discussed.

In sum, acceptance of grammatical sentences with *se* for both passive and alternators increased reliably (see Research Question 1), but the instructional treatment did not explain this increase statistically (see Research Question 2). Acceptance of alternators without *se* did not reliably decrease as intended by the instruction. The incorrect rejection of pro-drop in fact increased reliably in transitive sentences over test administration, contributing to the lower experiment-wise ratings for these items from pretest to delayed posttest.

Figure 3 Learner production by test administration and treatment group. [Color figure can be viewed at wileyonlinelibrary.com]

Guided Production Task

Guided Production Data: Descriptive Statistics

We again analyzed four Transitivity verb classes: Accusative, Alternator, Unaccusative, and Unergative. Each verb class was represented by three items in each test form. Recall that each prompt contained a nonfinite verb with one noun phrase that could be an argument of the verb. Previous researchers had already coded production using nominal categories, depending on the "type of sentence" that the learner had produced to complete the prompt (Toth, 2000; Toth & Guijarro-Fuentes, 2013). Figure 3 shows the responses of the learners by treatment group and test administration.

The learner data showed variation across morphology and verb class and also over time. As can be observed from Figure 3, in the pretest, learners appropriately used transitive sentences for transitive verbs. However, for alternators, pretest responses were the ungrammatical intransitive construction without *se* approximately 25% of the time. Unaccusatives and unergatives were mostly intransitive with no inappropriate *se*. In the immediate posttest, learners in

the instructed groups began using the grammatical $SE_{ANTICAUS}$ 50% of the time for alternators; the control group also used some $SE_{ANTICAUS}$. In addition, ungrammatical intransitives without *se* decreased for alternators. However, the instructed groups, but not the control group, expanded the use of *se* to unaccusative and unergative verbs, more so for the unaccusatives, as expected based on previous research that shows a crosslinguistic tendency to morphologically mark a Theme in subject position (Toth, 2000; Zobl, 1989). In the delayed posttest, the pattern of results is similar to the posttest, with fewer uses of *se* with intransitives and a slight increase in use of ungrammatical intransitive with *se*. Ungrammatical *se* persisted most of all in unaccusatives. Impersonal *se* was rare in the learners' production, as was the syntactic passive.

Guided Production Data: Statistical Analysis

The original data set contained six nominal categories of response. A multinomial regression model could be used in such cases. However, based on the transitivity properties of the verbs and the experimental design, it was logical to collapse these six types into a binary categorical outcome. In R, we recoded the six response types into a binomial Null or SE response: A response with *transitive*, *intransitive*, *passive*, and *other* without *se* were coded as Null; *intransitive se and transitive se* were coded as SE. This coding made the data suitable for mixed effects binomial logistic regression. The distribution of the recoded responses is illustrated in Figure 4 providing insight into the use of *se* by verb class over test administrations of whether they are grammatical or not. Note that for the alternators, as for the accusative verbs, a sentence with two arguments, for example, *Gabriel abrió la ventana* (Gabriel opened the window), is also grammatical. The figures represent sentence types produced by the learners.

We fit a mixed-effects binomial logistic regression on L2 learners' data ($n = 138$). We started with a null model without any predictors and proceeded stepwise by including predictors and their interaction one by one. Any predictor or interaction which failed to significantly improve model fit was removed (Baayen, 2008). Following Barr et al. (2013), the random effects structure was maximal with by-Item random slopes for Treatment. By-Participant and by-School random slopes for Transitivity*TestAdmin were also included, but the model only converged by dropping random slopes. Therefore, the best fitting model that accounts for the patterns in the data contained Transitivity, TestAdmin and Treatment as main effects, and the interaction of Transitivity*TestAdmin and TestAdmin*Treatment. The results

Figure 4 Learner production by test administration and treatment group: Null/SE. [Color figure can be viewed at wileyonlinelibrary.com]

are summarized in Appendix S2. As with the previous analysis, the model began with a baseline of predictors, called the (Intercept) in Appendix S2, against which the scores in all the other conditions are compared. In this case, the (Intercept) is the responses of the control group to accusative prompts. This reference level is coded as Transitivity = Acc and Treatment = Cont. As for TestAdmin (Pretest, Posttest, and Delayed), the reference level Posttest was selected manually. The reference level of the dependent variable is Null, namely, when learners did not use *se*. The algorithm compares the second level (*se*) to the reference level (null).

As with the AJT, the focus of the experiment is the potential interaction of Transitivity, Treatment, and TestAdmin. Transitivity and TestAdmin significantly interacted ($\chi^2 = 30.41$, $df = 6$, $p < 0.001$). For guidance on interpretation of the output of logistic regression see Levshina (2015, pp. 261-262). Thus, the comparisons in Appendix S2 reflect the proportions illustrated in Figure 4 in that for the posttest, L2 learners were significantly more likely to produce *se* when Transitivity is Alt as opposed to Acc (logit coefficient: +1.81, $SE = 0.228$, $z = 7.93$, $p < 0.001$). The probability of producing *se* was also

higher for Alt than for Unacc (logit coefficient: +2.24, $SE = 0.259$, $z = 8.66$, $p < 0.001$) or Unerg (logit coefficient: +3, $SE = 0.277$, $z = 10.83$, $p < 0.001$). Thus, $SE_{ANTICAUS}$ appeared appropriately more often with alternating verbs in the post-test. In addition, for $SE_{IMPERS/PASS}$, the probability of producing *se* was higher for Acc than for Unerg (logit coefficient: +1.42, $SE = 0.271$, $z = 5.23$, $p < 0.001$) or Unacc (logit coefficient: +0.65, $SE = 0.254$, z = 2.57, $p < 0.05$), and the probability of producing *se* was higher for Unacc than for Unerg (logit coefficient: +0.76, $SE = 0.276$, $z = 2.77$, $p < 0.05$). Taken together, the results mean that learners appropriately used *se* in transitive and alternator verbs but resisted using *se* in intransitives. However, when overgeneralization occurred, the unaccusative class was more often the target than unergatives.

There were also reliable changes in production of *se* over time conditioned by verb class. The overall probability of producing *se* is higher for Posttest (logit coefficient: +2.71, $SE = 0.265$, $z = 10.28$, $p < 0.001$) and Delayed (logit coefficient: −2.47, $SE = 0.25$, $z = -9.86$, $p < 0.001$) than for Pretest. However, this result is modulated by verb class and treatment group. For the delayed test, the probability of producing *se* remained higher for Alt than for Acc (logit coefficient: −0.82, $SE = 0.246$, $z = -3.34$, $p < 0.05$), and higher for Acc than for Unerg (logit coefficient: +1.09, $SE = 0.286$, $z = 3.82$, $p < 0.001$). Meanwhile, the probability of producing *se* is higher for Alt than for Unacc (logit coefficient: +1.18, $SE = 0.254$, $z = 4.65$, $p < 0.001$), and higher for Alt than for Unerg (logit coefficient: +1,91, $SE = 0.279$, $z = 6.87$, $p < 0.001$).

In addition, there is a reliable interaction between TestAdmin and Treatment ($\chi^2 = 56.69$, $df = 4$, $p < 0.001$). For the posttest, the probability of producing *se* is higher for Deductive than for Control (logit coefficient: −3.968, $SE = 0.54$, $z = -7.35$, $p < 0.001$), and higher for PACE than for Control (logit coefficient: −3.527, $SE = 0.545$, $z = -6.48$, $p < 0.001$). Likewise, during the delayed posttest, the probability of producing *se* is higher for Deductive than for Control (logit coefficient: −1.579, $SE = 0.479$, $z = -3.29$, $p < 0.01$), and higher for PACE than for Control (logit coefficient: −1.256, $SE = 0.487$, $z = -2.58$, $p < 0.05$). In addition, both Deductive (logit coefficient: 1.072, $SE = 0.165$, z = 6.49, $p < .001$) and PACE (logit coefficient: 0.953, $SE = 0.172$, $z = 5.55$, $p < .001$) decreased their use of *se* in the delayed posttest compared to that in the posttest phase. This result indicates that the two treatment groups reliably decreased their use of *se* in the delayed posttest, did not reliably differ from each other, and remained different from the control group.

Finally, all main effects were significant. The LRT indicated a reliable effect for Transitivity ($\chi^2 = 38.73$, $df = 3$, $p < 0.001$), TestAdmin ($\chi^2 = 724.09$, $df = 2$, $p < 0.001$) and Treatment ($\chi^2 = 32.09$, $df = 2$, $p < 0.001$). For Transitivity, the probability of producing *se* is higher for Alt than Acc (logit coefficient: +1.34, $SE = 0.15$, $z = 8.88$, $p < 0.001$), higher for Acc than Unacc (logit coefficient: −0.43, $SE = 0.16$, $z = -2.67$, $p < 0.01$), higher for Acc than Unerg (logit coefficient: −1.17, $SE = 0.18$, $z = -6.53$, $p < 0.001$); higher for Alt than Unacc (logit coefficient: −1.78, $SE = 0.16$, $z = -11.1$, $p < 0.001$), higher for Alt than Unerg (logit coefficient: −2.51, $SE = 0.18$, $z = -13.95$, $p<0.001$), and higher for Unacc than Unerg (logit coefficient: −0.73, $SE = 0.18$, $z = -4.02$, $p < 0.001$). Thus, the verb class where *se* was most likely to appear was Alt, followed by Acc, Unacc, and Unerg.

The treatment effect revealed the following results: The probability of producing *se* is higher for Deductive than for Control (logit coefficient: +3.26, $SE = 0.59$, $z = 5.5$, $p < 0.001$) and higher for PACE than for Control (logit coefficient: +2.93, $SE = 0.6$, $z = 4.85$, $p < 0.001$). Thus, the two instructional groups differed from the control group, but not from each other.

The TestAdmin effect showed that the probability of producing *se* by learners is higher on the Posttest than the Pretest (logit coefficient: −5.33, $SE = 0.35$, $z = -15.13$, $p < 0.001$), and the Delayed Posttest (logit coefficient: −0.99, $SE = 0.12$, $z = -8.27$, $p < 0.001$) and that the probability of producing *se* is higher on the Delayed Posttest than the Pretest (logit coefficient: 4.34, $SE = 0.35$, $z = 12.55$, $p < 0.001$). This means that learners produced more *se* at the immediate posttest, and although use declined at the delayed posttest it did not retreat to the level of the pretest.

In sum, the results show a reliable increase of the production of *se* in appropriate verb classes over time, particularly at the posttest, though learners used *se* less at the delayed posttest. Only the instructional treatment groups increased use of *se*; however, the deductive group and the PACE group did not differ from each other. The use of *se* varied by verb class; *se* was used reliably more with alternators than with accusatives. Although *se* was used inappropriately with both unaccusatives and unergatives, it was reliably used more with unaccusatives.

Discussion

Acceptability Judgment Data

We first discuss the research questions and then elaborate on patterns emerging from the data. The research questions addressed: (1) whether the learners

distinguish the forms of *se* by verb class/function as predicted by generative theory; (2) whether a change occurred over time; and (3) whether instructional treatment played a role.

The answer to the first question is clearly affirmative. First, unergative and unaccusative verbs were rated lower with *se* than without *se* throughout the experiment and, although the learners became a little less certain over time, this change was not reliable. Thus, the learners demonstrated knowledge of constraints on the use of grammatical function-changing morphology, even with unaccusative verbs, in the acceptability judgment task. This is an important result because it shows that the targeted instructional intervention did not cause learners to inappropriately overgeneralize on this task.

Where alternating verbs are concerned, an overall effect of time is reliable in changing acceptability judgments to $SE_{ANTICAUS}$ sentences. Initially, learners reject $SE_{ANTICAUS}$ with alternating verbs; however, over time, learners (even somewhat in the control group), increase their acceptance. Thus, the learners added $SE_{ANTICAUS}$ to their grammar of Spanish, which should "drive out" the ungrammatical bare forms based on theories of blocking (Rainer, 2016). However, $SE_{ANTICAUS}$ has not preempted (Rutherford, 1989) acceptance of alternators with null morphology. This result suggests continued L1 influence that may require additional feedback in instruction.

With transitive verbs an interesting pattern emerges with *se*. Looking at the pretest scores, learners accept ungrammatical *se* with sentences such as **La máquina se destruyó la casa* (The machine SE destroyed the house), a response that does not change significantly over time and is not reliably different from the grammatical causative (*pro*) *Destruyó la casa.* It is possible, although undocumented, that learners believe that the *se* being used is telic or even emphatic, as in English, *The machine itself destroyed the house*. It is also possible that they simply did not know what it meant in this context.

The statistical model for the AJT found no main effect for instructional treatment (Control, Deductive, PACE) and no interaction between test time and instruction, even though there were changes over time in how students rated the sentences. Thus, no effect of instruction was found which was the focus of Research Question 3. This result is most likely due to the large variability in the scores. Reviewing Figure 2, it is clear that the raw scores for the deductive group do change from pre- to posttest with accusative-intransitive with $SE_{IMPERS/PASS}$ and alternator-intransitive clauses with $SE_{ANTICAUS,}$ both increasing in acceptability based on changes in the interquartile range. However, the variation in the scores remains very high in all treatment groups, which underlies the failure to find a significant treatment effect. For example,

although the median of the deductive group for alternator-intransitive clauses with $SE_{ANTICAUS}$ increased from 4 to 5 in the posttest, with 50% of scores above 5 in the posttest, the minimum and maximum without outliers still ranged from 1 to 6, just as for the control group and the PACE group. In the delayed posttest, all groups' 50th percentile score for this sentence type reverted to 4. Once normalized through z-scores, any group differences remained statistically insignificant. The failure to detect instructional effects could also be due to the large variability in the interpretations of pro-drop and sentences with and without *se* with transitive verbs. Thus, some learners may have interpreted the testing instruments as intended and showed benefits for the instructional treatment, while others did not, such that consistent improved results within and among groups were not evident.

An important point about the test instrument itself is in order at this juncture. The complexity and labeling of the pictures could have affected the lack of reliable effects for instruction. For example, in Test Version A, the picture with a machine knocking down a house shown in Figure 1 was used four times with different sentences, some grammatical and some not. For some of these, a label next to a picture featuring an external cause (the machine) was intended to create a discourse context for a null subject pronominal reference. Thus, the machine was labeled for Items 5 and 66, with stimulus sentences *(pro) Destruyó la casa* (It destroyed the house) and **La casa destruyó* (The house destroyed) respectively. However, for Item 32, **La máquina se destruyó la casa*, and Item 48, *Se destruyó la casa* no label was used. Although Item 5 is grammatical, it is not clear that the pragmatic conditions for a null subject are met in this context: In ordinary discourse, one might expect the agent labeled in the picture to nonetheless be mentioned in a sentence, not referred to simply through a null subject. Normally, pro-drop is permitted when an entity has been mentioned in the discourse and the referent is pragmatically shared by both speaker and hearer. In the AJT task, no discourse had been established but attention was directed to the Agent, "the machine," through the label on the picture. For sentence (48), the machine is in the picture, but not labeled, and hearers might expect the machine to be mentioned despite the passive *se*. The problem is not just with these sentences, however. Throughout the AJT, various items, times of day, and other aspects of the pictures are labeled, which might have distracted learners in different ways. Thus, the picture labels were not consistently manipulated as a variable and therefore it is impossible to know what effects these labels may have had.

In sum, failure to statistically detect reliable instructional effects in the AJT could be due to the large variability in the ratings the AJT. The sources

of such variation could reside in the varied complexity of the visual stimuli and uncertainty caused by the 4-second time pressure in matching the number of noun phrases labeled in the pictures with those mentioned or not in the sentences that participants heard during the task. For this reason, the tendencies shown in the boxplots in Figure 2 may not provide evidence for instructional effects. The results from the production task may be a better indication of such effects.

Guided Production Task

In the learner production data, a significant effect of instruction was evident, which was the main focus of Research Question 3. Both the deductive group and the PACE group produced more *se* with accusative verbs and with alternators over time, peaking in the use of *se* in the immediate posttest, yet reliably declining in the delayed posttest. Thus, these data confirm findings in Zyzik's study (2006, p. 475) that alternators cause persistent problems even at the more advanced third-year college Spanish level (c.f., also Montrul, 1999; 2000; Toth, 2000). They did overgeneralize *se* to unaccusative intransitive verbs more than unergative transitive verbs over time, and when overgeneralization to inherent intransitives occurred, unaccusatives were preferred more than unergatives. This result is predicted by the split intransitive hypothesis (Perlmutter, 1978). They did not use *se* with intransitives more than correct use with transitive and alternating verbs at any stage of their learning.

General Discussion

From the point of view of generative linguistics, our review of data from this study of high school Spanish L2 revealed several important findings. First, we note that generative linguistics guided the original design and sequence of the instructional materials and test instruments. This approach made Research Question 1 possible as it focused on the role of verb meaning and morphosyntax, specifically the correct use of *se* with different verb classes. Despite inconsistent prompts with some AJT items, learners' knowledge of verb classes constrained their judgments. The basis of this claim is the reliable interaction of Transitivity with Morphology in the AJT: learners did not overgeneralize *se* to unaccusatives and unergatives. This result suggests that their linguistic systems recognized verb classes (accusative, alternator, unaccusative, unergative), albeit possibly based on subconscious L1 knowledge. They were then able to recognize in the AJT the appropriateness of *se* or null morphology depending on the other arguments present in the clause. In production, they did not extend use of *se* to unaccusatives or unergatives to a

large degree. Without generative linguistic categories, it would not be possible to predict or discuss these effects because other theories treat separate clause types as *constructions* that are unrelated to any deeper system of theta roles, argument structure, and grammatical function-changing morphology.

Reliable interactions were found among Transitivity, Time, and Morphology both for the AJT and production data, suggesting that changes in the learners' grammars depended on verb class. However, only in the production data did instructional treatment play a role in explaining the interaction with Transitivity, TestAdmin, and Morphology among the groups. No statistical difference between the deductive and PACE groups was found in the production data. Clearly, the instruction helped the learners produce more accurate Spanish sentences for both the instructed groups, but it did not matter whether the instruction was declarative with deductive rules or inductive with teacher lead co-construction as in the PACE treatment. Both groups reached an explicit metalinguistic formulation of a rule for *se* and participated in output practice with feedback. We assume that these activities resulted in making the input available to subconscious grammatical processing (Sharwood Smith & Truscott, 2014). By this we mean that compared to the control group, the learners' exposure to focused input through instruction led to their internal grammar "noticing" that the morpheme *se* occurred with some verb meaning/contexts but not others: for example, with alternators such as *cook* with a theme subject, but not intransitives. The addition of the morpheme *se* in the L2 where no morphology is required in the L1 for alternators with a theme subject triggered the development over time for the alternator class.

The study also shows the importance of understanding subtle differences of what may superficially appear to be the same form. Conceptually and theoretically, $SE_{ANTICAUS}$ and $SE_{PASS/IMPERS}$ are not "the same." Thus, although the instruction designers were aware of the meaning difference, they did not have the advantage of MacDonald's (2017) more recent syntactic analysis. It is possible that teaching these forms back-to-back over three days resulted in confusion among the learners and certainly did not help create separate linguistic categories based on MacDonald's (2017) analysis. Importantly, [VoiceP] in $SE_{ANTICAUS}$ does not contain an implicit Agent, whereas $SE_{PASS/IMPERS}$ does. An analogy from ESL might be that when teaching vocabulary for fruit, one would not teach *pear*, *peach*, and *plum* in the same lesson if one wants learners to effectively distinguish each form and meaning. Thus, a recommendation would be to space out the different uses of *se* across time. Moreover, the inclusion of pro-drop sentences in the AJT challenged the learners. Learners are exposed to pro-drop early in learning Spanish, so a focus on pro-drop as part

of the explanations for when arguments can be omitted with *se* could be added to instruction on verb classes. Finally, the presence of the implicit Agent in $SE_{PASS/IMPERS}$ may also have affected the interpretation of the varied labeling in the AJT, but that is a new research question that was not part of the design and cannot be answered without further research.

Another important contribution of the analysis in this article is that it has deployed different techniques of statistical analysis than in Toth and Guijarro-Fuentes (2013). A comparison of ANOVA techniques (using group means) with LME and binomial logistic regression with judgment and production data falls beyond the scope of this article. However, due to these techniques, the effect of deductive instruction found in similar data analyzed by Toth and Guijarro-Fuentes (2013) was not replicated in the AJT data vs. the control group. Toth and Guijarro-Fuentes (2013) did not compare deductive instruction with PACE, but learners in Toth and Guijarro-Fuentes (2013) did incorrectly extend *se* to unaccusatives, which was a finding not replicated in this paper. However, reliable, sustained improvement in these data for instructed groups in delayed posttests in production is consistent with Toth and Guijarro-Fuentes (2013). It is possible that z-score normalization and including more items and participants as part of the model increased the detection of variability among groups. Increased variability from three schools rather than one and three groups rather than just two as in Toth and Guijarro-Fuentes (2013) reduces the probability of finding reliable differences among groups. This observation is especially true when one considers the whole linguistic system's interactions with *se* across verb classes, and not just with the target(s) of instruction. Between-school variation arising from various teachers can be well-accounted for by mixed-effects models, but none were found in the current analysis. This result might suggest that learner-internal resources constrain overgeneralizations no matter the context.

Limitations and Future Directions

The analysis presented in this article is one view of a subset of all the data collected during the original study. For example, we did not consider the pragmatically questionable sentence types in the AJT. In addition, while the AJT data revealed no reliable effects of instruction, learners did change over time based on verb class. It should therefore be noted that some individual learners may have benefitted more than others. Our analysis in this article focused on an estimation of a population and is not an article on potential individual differences, which could show that for specific individuals one form of instruction may be more suitable than another. Thus, future studies could examine whether

individual difference factors such as aptitude and working memory might interact with instructional effects and as the numbers of participants in the study were relatively small larger studies with greater statistical power could be conducted.

In other future directions, researchers could investigate instructional effects with other forms that may superficially look the same, but where the underlying abstract linguistic system may be different. In this article, MacDonald's (2017) analysis showed that the representation of passive *se* contains an implicit Agent, whereas anticausative *se* does not. Another example of a challenging similarity for English-speaking learners would be the well-known complexities of German word order (e.g., Jackson & Ruf, 2017). Supported by such detailed knowledge, language professionals could make better pedagogical choices in instruction and targeted assessment that, together with internal learner resources, lead to measurable gains in proficiency. Another future direction would be investigating the effect of spacing out and recycling lessons on this challenging morpheme based on well-known effects of spacing of practice exercises in vocabulary instruction Pavlik and Anderson (2005) and work by Suzuki and Sunada (2020) on the effects of different types and patterns of practice.

Conclusion

The role of generative linguistics in SLA and linguistics in general has been challenged recently, albeit based on misconceptions in many cases (Rothman & Slabakova, 2018). In this article, generative approaches to the links between verb argument structure and morphosyntax were able to shed light on L2 developmental patterns in subtle properties of Spanish morphosyntax. First, generative accounts of *se* guided the development and sequencing of the materials. The generative categories of verbs showed reliable differences among verb types and changes over time suggesting that learners are constrained by abstract, internal resources. The data from production tasks showed that either both types of form-focused instruction (whether deductive or PACE) was superior to incidental classroom exposure. This finding is important, as free production is thought to provide better insights into underlying proficiency in general (e.g., Crossley & Skalicky, 2019, p. 386). While a lack of difference in explicit instructional delivery is predicted by GenSLA, the results provide support for form-focused teaching. Finally, for teacher training programs, the study reinforces the view in Whong et al. (2013) that instructional design and assessment need to be informed by fine-grained linguistic analysis. It further underscores the need for language teacher preparation programs to provide

enhanced training in the linguistics of the target language and theories of language representation and acquisition.

Final revised version accepted 22 November 2021

Notes

1 We note that pragmatic and other reasons may trigger the use of passive. A speaker may wish to avoid mentioning the cause of an event to avoid assigning responsibility to the agent ("The window was broken when the ball hit it" vs. "Maria broke the window when she angrily threw the ball") or the speaker may wish to foreground the Theme in discourse, for example, "The window was broken when the ball hit it, not when the wind slammed it shut."

2 [IP] stands for *inflectional phrase*, often just "S" in introductory textbooks in linguistics. The theoretical status of this term has evolved considerably, with IP splitting into [TP] (Tense Phrase) and other functional categories. In addition, theories concerning the origin of subjects in a double VP shell have been introduced. Nevertheless, based on editorial feedback, we retain this notation here as it would obscure some of the main points of this section that relate to verb argument structure. The original experiment was also not designed to test the VP internal subject hypothesis. See Carnie (2012) for further discussion of these developments.

3 We do not discuss further other uses, for example, the telic *se*:
Juan se leyó el libro.
Juan SE read the book
"*John read the book*." (= he read all of it, read it to the end/completely).
This telic *se* is used primarily with transitive verbs of consumption, either of food, or drink, or knowledge/information (read, learn, know, eat, drink, swallow, etc.) It can also be used with certain common intransitives, mostly unaccusatives, to direct attention to the boundedness of a state that that might otherwise be interpreted as prolonged (changing from a Vendler, 1957, *activity* to an *accomplishment* or *achievement*). These verbs include the Spanish equivalents of *die*, *come*, *go*, *sleep*, *arrive*, *escape*, and *laugh* (among others). The testing and teaching materials were very carefully designed to avoid contexts where these intransitives with telic *se* would be appropriate or required.

4 The exception may be for the most basic form-meaning vocabulary instruction, but even then frequency and spacing of practice may need to be carefully considered (Pavlik & Anderson, 2005).

5 To handle this issue, recall that we had transformed the data points into z-scores within individual participants; thus, all participants have a mean of 0. The intercept is the base line score to which all the other scores in the model are compared to assess whether any differences are caused by those factors or interactions.

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

References

Adair-Hauck, B., & Donato, R. (2016). PACE: A story-based approach for dialogic inquiry about form and meaning. In J. L. Shrum & E. W. Glisan (Eds.), *Teacher's handbook: Contextualized language instruction* (5th ed, pp. 206–230). Boston, MA: Cengage Learning.

Alexiadou, A. (2014). Active, middle, and passive: The morpho-syntax of Voice. *Catalan Journal of Linguistics*, *13*, 19–40. https://doi.org/10.5565/rev/catjl.153

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. Chicago, IL: University of Chicago Press.

Baker, M. C., Johnson, K., & Roberts, I. (1989). Passive arguments raised. *Linguistic Inquiry*, *20*(2), 219–252. https://doi.org/10.4324/9781315310572-16

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carnie, A. (2012). *Syntax: A generative approach* (3rd ed.). Oxford, UK: Wiley-Blackwell.

Chomsky, N. (1986). *Knowledge of language*. New York, NY: Praeger.

Crain, S., & Thornton, R. (2006). Acquisition of syntax and semantics. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 1073–1010). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-012369374-7/50029-8

Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara (2011). *Language Teaching*, *52*(3), 385–405. https://doi.org/10.1017/S0261444817000362

DeKeyser, R. M., & Juffs, A. (2005). Cognitive considerations in L2 learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 461–478). New York, NY: Routledge. https://doi.org/10.4324/9781410612700-36

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly*, *27*(1), 91–113. https://doi.org/10.2307/3586953

Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 31–64). Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767-004

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (Eds.), *Causatives and transitivity* (pp. 87–120). Philadelphia, PA: Benjamins. https://doi.org/10.1075/slcs.23.05has

Han, Z., & Finneran, R. (2014). Re-engaging the interface debate: Strong, weak, none, or all? *International Journal of Applied Linguistics*, *24*, 370–389. https://doi.org/10.1111/ijal.12034

Hirakawa, M. (2001). L2 acquisition of Japanese unaccusative verbs. *Studies in Second Language Acquisition*, *23*, 221–246. https://doi.org/10.1017/S0272263101002054

Jackson, C., & Ruf, H. T. (2017). The priming of word order in second language German. *Applied Psycholinguistics*, *38*, 315–345. https://doi.org/10.1017/S0142716416000205

Juffs, A. (1996). *Learnability and the lexicon: Theories and second language acquisition research*. Philadelphia, PA: Benjamins. https://doi.org/10.1075/lald.12

Juffs, A. (2000). An overview of the second language acquisition of the links between verb semantics and morpho-syntax. In J. Archibald (Ed.), *Second Language Acquisition and Linguistic Theory* (pp. 187–227). Oxford, UK: Blackwells.

Juffs, A. (2009). Second language acquisition of the lexicon. In W. Ritchie & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 181–205). Bingley, UK: Emerald.

Kanwit, M. (2017). What we gain by combining variationist and concept-oriented approaches: The case of acquiring Spanish future-time expression. *Language Learning*, *67*, 461–498. https://doi.org/10.1111/lang.12234

Kanwit, M., & Geeslin, K. L. (2014). The interpretation of Spanish subjunctive and indicative forms in adverbial clauses. *Studies in Second Language Acquisition*, *36*, 487–533. https://doi.org/10.1017/S0272263114000126

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London, UK: Longman.

Levshina, N. (2015). *How to do linguistics with R*. Philadelphia, PA: Benjamins. https://doi.org/10.1075/z.195

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York, NY: Routledge. https://doi.org/10.4324/9781315887074

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.

Lozano, C. (2006). Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish. *Second Language Research*, *22*(2), 145–187. https://doi.org/10.1191/0267658306sr264oa

MacDonald, J. E. (2017). An implicit projected argument in Spanish impersonal- and passive-*se* constructions. *Syntax*, *20*, 353–383. https://doi.org/10.1111/synt.12146

Marsden, H., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*, 583–616. https://doi.org/10.1080/03054985.2012.731208

Montrul, S. (1999). Causative errors with unaccusative verbs in L2 Spanish. *Second Language Research*, *15*(2), 191–219. https://doi.org/10.1191/026765899669832752

Montrul, S. (2000). Transitivity alternations in L1 acquisition: Toward a modular view of transfer. *Studies in Second Language Acquisition*, *22*, 229–273. https://doi.org/10.1017/S0272263100002047

Montrul, S. (2001). First language constrained variability in argument structure changing morphology with causative verbs. *Second Language Research*, *17*(2), 144–194. https://doi.org/10.1177/026765830101700202

Moranski, K., & Zalbidea, J. (2022). Context and generalizability in multisite L2 classroom research: The impact of deductive versus guided inductive instruction. *Language Learning*, *72*(S1), 41–82. https://doi.org/10.1111/lang.12487

Orita, N., Ono, H., Feldman, N. H., & Lidz, J. (2021). Japanese children's knowledge of the locality of zibun and kare. *Language Acquisition*, *28*, 327–343. https://doi.org/10.1080/10489223.2021.1899181

Oshita, H. (2001). The unaccusative trap in second language acquisition. *Studies in Second Language Acquisition*, *23*, 279–304. https://doi.org/10.1017/S0272263101002078

Pavlik, P. I. J., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Sciences*, *29*, 559–586. https://doi.org/10.1207/s15516709cog0000_14

Perlmutter, D. M. (1978). Impersonal passives and the Unaccusative Hypothesis. In J. Jaeger (Ed.), *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society* pp. 157–189). Berkeley, CA: Berkeley Linguistics Society.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rainer, F. (2016). Blocking. In M. Aronoff, *Oxford research encyclopedia of linguistics*. New York, NY: Oxford University Press. https://doi.org/10.1093/OBO/9780199772810-0185

Ranta, L., & Lyster, R. (2017). Form-focused instruction. In P. Garrett & J. M. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 40–56). New York, NY: Routledge.

Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, *40*, 417–442. https://doi.org/10.1017/S0272263117000134

Rutherford, W. E. (1989). Preemption and the learning of L2 grammars. *Studies in Second Language Acquisition*, *11*, 441–458. https://www.jstor.org/stable/44488284

Schwartz, B. D. (1999). Let's make up your mind: "Special nativist" perspectives on language, modularity of mind, and nonnative language acquisition. *Studies in Second Language Acquisition*, *21*, 635–655. https://doi.org/10.1017/S0272263199004052

Schütze, C. T., & Sprouse, J. (2013). Judgment data. In R J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge, UK: Cambridge University Press.

Sharwood Smith, M., & Truscott, J. (2014). Explaining input enhancement: A MOGUL perspective. *International Review of Applied Linguistics in Language Teaching*, *52*(3), 253–281. https://www.degruyter.com/document/doi/10.1515/iral-2014-0012/html

Slabakova, R. (2016). *Second Language Acquisition*. Oxford, UK: Oxford University Press.

Slabakova, R., Leal, T., Dudley, A., & Stack, M. (2020). *Generative second language acquisition*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108762380

Sorace, A. (1993). Incomplete versus divergent representations of unaccusativity in non-native grammars of Italian. *Second Language Research*, *9*(1), 22–48. https://doi.org/10.1177/026765839300900102

Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language: The potential and limits of skill transfer and practice schedule. *Studies in Second Language Acquisition*, *42*(1), 169–197. https://doi.org/10.1017/S0272263119000470

Tai, H.-Y. J. (1984). Verbs and Times in Chinese: Vendler's Four Categories. In D. Testen, V. Mishra, & J. Drogo (Eds.), *Papers from the parasession on lexical semantics* (pp. 289–296). Chicago, IL: Chicago Linguistics Society.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical patterns. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. *III*, pp. 57–149). Cambridge, UK: Grammatical Categories and the Lexicon.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139173698

Toth, P. D. (2000). The interaction of instruction and learner internal factors in the acquisition of L2 morphosyntax. *Studies in Second Language Acquisition*, *22*(2), 169–208. https://doi.org/10.1017/S0272263100002023

Toth, P. D. (2008). Teacher-and learner-led discourse in task-based grammar instruction: Providing procedural assistance for L2 morphosyntactic development. *Language Learning*, *58*, 237–283. https://doi.org/10.1111/j.1467-9922.2008.00441.x

Toth, P. D. (2021a). *Teaching materials for deductive lessons. Materials from "What do the data show? Multiple perspectives on classroom L2 learning from a single data set*" [Instructional /Intervention / Teaching / Training materials]. IRIS Database, University of York, UK. https://doi.org/10.48316/5w1g-ky22

Toth, P. D. (2021b). *Teaching materials for PACE lessons. Materials from "What do the data show? Multiple perspectives on classroom L2 learning from a single data set*" [Instructional /Intervention / Teaching / Training materials]. IRIS Database, University of York, UK. https://doi.org/10.48316/5xcb-m335

Toth, P. D. (2021c). *Test scores from the Grammaticality judgment task. Materials from "What do the data show? Multiple perspectives on classroom L2 learning from a single data set*" [Data: Scores on measures /tests]. IRIS Database, University of York, UK. https://doi.org/10.48316/jm90-7t67

Toth, P. D. (2021d). *Test scores from the picture description task. Materials from "What do the data show? Multiple perspectives on classroom L2 learning from a single data set*" [Data: Scores on measures /tests]. IRIS Database, University of York, UK. https://doi.org/10.48316/ms1w-fe12

Toth, P. D., & Guijarro-Fuentes, P. (2013). The impact of instruction on second-language implicit knowledge: Evidence against encapsulation. *Applied Psycholinguistics*, *34*, 1163–1193. https://doi.org/10.1017/S0142716412000197

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness. The Quest for a Theoretical Foundation. *Studies in Second Language Acquisition*, *33*, 497–528. https://doi.org/10.1017/S0272263111000295

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Santa Barbara, CA: Greenwood Publishing Group.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*, 755–803. https://doi.org/10.1111/1467-9922.00203

VanPatten, B. (2016). Why explicit knowledge cannot become implicit knowledge. *Foreign Language Annals*, *49*, 650–657. https://doi.org/10.1111/flan.12226

Vendler, Z. (1957). Verbs and Times. *Philosophical Review*, *66*(2), 143–160. https://doi.org/10.2307/2182371

Wasow, T. (1977). Transformations and the Lexicon. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 327–360). Cambridge, MA: Academic.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511815065
White, L. (2012). Universal grammar, crosslinguistic variation, and second language acquisition. *Language Teaching*, *45*, 309–328. https://doi.org/10.1017/S0261444812000146
White, L., & Juffs, A. (1998). Constraints on Wh-movement in two different contexts of non-native language acquisition: Competence and processing. In S. Flynn, G. Martohardjono, & W. O'Neill (Eds.), *The generative study of second language acquisition* (pp. 111–130). Mahwah, NJ: Erlbaum.
Whong, M., Gil, K.-H., & Marsden, H.((2013). *Universal Grammar and the Second Language Classroom*. New York, NY: Springer. https://doi.org/10.1007/978-94-007-6362-3
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
Yuan, B. (1999). Acquiring the unaccusative/unergative distinction in a second language: Evidence from English-speaking learners of L2 Chinese. *Linguistics*, *37*, 275–296. https://doi.org/10.1515/ling.37.2.275
Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 203–221). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524544.015
Zyzik, E. (2006). Transitivity alternations and sequence learning: Insights from L2 Spanish production data. *Studies in Second Language Acquisition*, *28*, 449–486. https://doi.org/10.1017/S0272263106060190

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Supplementary information

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Instructional Effects in the Learning of Complex Morpho syntax in Spanish as a Second Language

What This Research Was About and Why It Is Important

Generative linguists are interested in how abstract properties of language guide learning and whether explicit teaching of rules in various ways can affect unconscious knowledge. This article explained how an abstract theory of verb meaning and knowledge of the Spanish morpheme *se* may differ from

pedagogical grammar, showed that abstract knowledge may guide learners' development, and provided evidence that explicit instruction helps learners develop greater accuracy with complex grammar in production.

What the Researchers Did

- At three U.S. high schools, teachers used different techniques of grammar instruction with Spanish se: in three days of instruction: deductive rule-based instruction and inductive "discovery" instruction. The sentence types included impersonal, passive, and inchoative (i.e., a "spontaneous" event without external cause). Other classes did not receive any grammar instruction with *se*.
- The researchers administered an aural acceptability judgment task (AJT) and a picture description (production) task at three points: before instruction, immediately after instruction, and six weeks later.
- The researchers analyzed changes in test scores over time on the two tasks using different verb types, type of instruction, and time of test as the main factors.

What the Researchers Found

- In the AJT, the researchers found that learners increased their acceptance of *se* in sentences such as *La puerta se abrío* (The door SE opened). However, learners still accepted sentences without *se*, for example, **La puerta abrío* (The door opened) at the delayed posttest. This is acceptable in English grammar, but not in Spanish.
- In addition, the learners did not incorrectly extend the use of *se* to intransitive verbs such as in **El avión se desciende* (The plane SE descends), where *se* is not permitted.
- In the production task, participants who had received either kind of instruction fared better than the group that had received no instruction. Learners showed increased use of *se* with verbs such as open in Spanish, but again they did not incorrectly overgeneralize to verbs such as *cry* or *descend*.

Things To Consider

- The same form, for example, *se*, may have different abstract representations depending on the meaning that is intended. Therefore, different meanings of the "same" structure might be better taught spaced out rather than in a block.
- The underlying (abstract) system of the grammar of a language should be taken into account in language teaching.

- Teachers need fine-grained linguistic knowledge to understand developmental patterns in learning.

Materials, data, open access article: Materials and/or data are publicly available on IRIS (https://www.iris-database.org).

How to cite this summary: Juffs, A., & Fang, S. (2022). Instructional effects in the learning of complex morphosyntax in Spanish as a second language. *OASIS Summary* of Juffs & Fang (2022) in *Language Learning*. https://oasis-database.org