

Third-Person Subjects in Native Speakers' and L2 Learners' Narratives: Testing (and Revising) the Givenness Hierarchy for Spanish

Sarah E. Blackwell and Margaret Lubbers Quesada
The University of Georgia

1. Introduction

Two recurrent questions posed in studies on reference and anaphora have been what allows speakers to felicitously use a particular referential form at a given point in discourse and what, in turn, enables the hearer to successfully interpret that form. The present study examines native Spanish speakers' and Spanish as a second language (L2) learners' use of third-person expressions to refer to grammatical subjects in oral narratives in light of Gundel, Hedberg and Zacharski's (1993) proposal that "the form of a referring expression...depends on the assumed cognitive status of the referent" (1993:275). As Gundel, Bassene, Gordon, Humnick and Khalfaoui (2010) explain, the basic premise of the theoretical framework proposed by Gundel et al. (1993) "is that some determiners and pronouns encode information about the assumed cognitive (memory and attention) status of the intended referent for the addressee" (2010:1770). In other words, cognitive statuses "correspond to memory and attention states" (Gundel, Hegarthy and Borthen 2003:283). Thus, speakers use different referring forms (e.g. pronouns, NPs) to encode information about the status of a discourse referent in memory, and in so doing, restrict "possible interpretations from among those that fit the conceptual content encoded by the form" (Gundel et al. 2010:1770). Because the cognitive status of a referent depends on the speaker's assumptions about the memory and attention states of the addressee, the notion of cognitive statuses is, as Hedberg and Zacharski note, "inherently pragmatic" (2007:4). However, without objective access to the short-term memories and attention states of speakers and hearers, determining the status of a referent based on the analysis of discourse is inferential at best.¹

In this study we compare the subject choices made by L2 learners at three proficiency levels (beginning, intermediate and advanced) with those of native Spanish speakers by assessing the discourse status of the referent in order to infer, via discourse analysis, the cognitive status of the intended referents of third-person subject expressions in oral, film-retell narratives.² The objectives of our analysis are: 1) to gain a better understanding of the extent to which the cognitive status of the referent influences native speakers' and L2 learners' choice of third-person referring forms; 2) to show how cognitive status (as inferred from the discourse) interacts with other constraints on the choice of third-person forms (e.g. the recoverability of antecedents based on discourse context and lexical semantics); and 3) to characterize the distribution of third-person subject expressions used by native speakers and learners of Spanish in the film narratives. More generally, the theoretical approach we adopt enables us to test the viability and explanatory adequacy of the six cognitive statuses proposed by Gundel et al. (1993) as constraints on the form and distribution of referring expressions in Spanish

* Data collection was funded by UCMEXUS and CONACYT (Mexico) (Grant No. CN0255). We express our gratitude to research assistant, Adam Lassila of UGA, to Kim Love-Myers and Andrea Moncayo González of the Statistical Consulting Center of UGA, and to the three anonymous reviewers for their insightful comments. Any remaining errors are ours.

¹ Thanks to one anonymous reviewer for pointing this out.

² Because it is impossible to directly assess the "cognitive" status of referents in the minds of speakers and hearers through discourse analysis, we use the term "discourse" status here, as suggested by the same reviewer, but hereafter we refer to this concept as "cognitive" status, following the terminology used by Gundel et al. (1993).

as well as other languages, and to propose modifications in the cognitive statuses which, we believe, can enable us to better account for reference in Spanish narrative discourse.

2. Previous studies on subject reference in Spanish

The study of reference in discourse, including reference to subjects, has been approached from various perspectives. Several studies have examined subject expression versus omission in Spanish within a variationist framework, making use of Silva-Corvalán's (1982) notions of "switch reference" and "same reference"³ in order to determine the conditioning factors influencing the probability of use of overt subject pronouns in contexts where alternation with a null subject is possible—i.e. contexts within the "envelope of variation" (e.g. Silva-Corvalán 1982, on Mexican-American Spanish; Bentivoglio 1987, on Caracas Spanish; Ranson 1991, on Andalusian Spanish; Cameron, e.g., 1992, 1995, on San Juan and Madrid Spanish; and Cameron and Flores-Ferrán 2004, on San Juan, Madrid, and New York City Spanish, to name a few). The variationist approach predicts the use of an overt subject in contexts involving switch reference. However, as we show later in this paper, there are other factors that account for the use of overt subjects in contexts where switch reference occurs.

Studies in second language acquisition have focused on English-speaking learners' difficulty in "resetting" the null pro-drop parameter and related properties, such as subject-verb inversion and *that*-trace (Licerias 1989; Licerias and Diaz 1998, 1999; Al-Kasey and Pérez-Leroux 1998; Isabelli 2007). Others have explored the effect of the Overt Pronoun Constraint (Montalbetti 1984) and to what extent learners master the semantic contrasts that guide null pronoun use in Spanish (Pérez-Leroux and Glass 1997, 1999; Lozano 2002).

More recent studies have used pragmatic or sociolinguistic principles to examine the problem. Quesada and Blackwell (2009) proposed a set of pragmatic rules for first-person subject pronoun use and found that learners gradually incorporate these rules into their learner grammar; learners produce null subject pronouns for all contexts but prefer overt subject pronouns for contrastive focus contexts (the predicted form for native speakers); they still have difficulty producing overt subjects to express switch focus and few have learned the rule of using overt first-person subject pronouns for topicalizing or adding "pragmatic weight" to utterances (Davidson 1996). Rothman (2009), in several experimental studies, examined the L2 acquisition of the syntax-pragmatics interface of subject pronouns and found that although learners acquire the syntax of null subjects early on, they have difficulty with the discourse distribution of null and overt subject pronouns. Geeslin and Gudmestad (2008) use sociolinguistic methodology and theory to study the variable use of subject pronouns in native and advanced learners' speech and found that person and number of the verb and specificity of the referent interact to determine subject expression. Theirs is the first large-scale study in SLA to look at all subject expressions, including lexical noun phrases and overt, null, demonstrative, interrogative and indefinite pronouns, in a variety of specificity contexts, including specific, non-specific and group referents. They found both similarities and differences between the learner and NS groups' use of subject expressions. The range of forms was similar for learners and NSs, and both groups used a wider range of forms for third-person subjects, but the frequency with which each form was produced was significantly different. Most importantly, learners were more likely to use null subjects in all except the first-person singular and third-person plural contexts where NSs used more nulls. In addition, the learners used a greater number of subject pronouns for group referents and fewer subject pronouns in non-specific contexts.

Although each of these studies has made important contributions to the study of the acquisition of subject expression, by and large, research in SLA has ignored the discourse (cognitive) constraints that determine subject use in narrative structure.

Other studies on subject expression in Spanish have focused on the interaction of syntactic, semantic, pragmatic and/or discourse-interactional motivations for the use of overt versus null

³ According to Silva-Corvalán, "[s]witch-reference is established whenever the subject referent of the preceding finite verb, regardless of whether the subject is expressed or deleted, is different from the referent of the expressed subject of the sentence in question... Same reference is established when the subject of the sentence in question is the same as that of the preceding sentence, regardless of whether it is expressed or not expressed" (1982:104).

subjects, oftentimes looking only at the alternation between null and overt subject pronouns. For example, Luján (1985, 1986, 1999) looks at personal pronoun use (both subjects and objects) in terms of neutrality (absence of emphasis) versus contrastiveness (emphasis) on pronominal expressions, the syntactic constraints on null and overt alternates, and the meanings that these alternates convey. She shows that the distinction between stressed and unstressed pronouns in non-pro-drop languages such as English functions like the overt/null pronominal distinction in pro-drop languages such as Spanish, producing similar systematic contrasts in anaphoric interpretation. For instance, in the following examples, the overt/null contrast in Spanish produces the same effect as the stressed/unstressed contrast in English (from Luján 1986:255):

- (1) a. Ana ama a Elsa y Ø/ella lo sabe. (Ø=Ana; ella=Elsa)
 b. 'Ana loves Elsa and she/SHE knows it.' (she=Ana; SHE=Elsa)

Here, the null subject selects the preceding subject as its antecedent, whereas the overt subject pronoun picks out the object antecedent.

Mayol (2010) also investigates the contrastive function of pronouns in null-subject Romance languages, but adopts a largely semantic perspective and applies the mathematical framework of Game Theory to account for pronoun use. She argues that non-focal, contrastive overt subject pronouns in Catalan, Spanish and Italian function as contrastive topic markers that convey an uncertainty contrast, which may involve a "double contrast", an "implicit contrast" or a "weak contrast". Double contrasts involve two-clause discourses in which two referents occupy the subject position and their verb phrases predicate two different actions or states, as in (2) (Mayol 2010:2500, citing Cameron 1992):

- (2) **Ellos** fueron pero **yo** no fui.

Implicit contrasts involve a contrast between the antecedent of the pronoun and another implicitly understood entity that is "highly salient in the discourse" (Mayol 2010:251):

- (3) La ranita se pone a llorar porque Ø se ha hecho daño y además **ella** quería que las dos fueran amigas.

Here, overt *ella* (where a null subject would refer to the same antecedent, *la ranita*) conveys that "while the small frog (the referent of the pronoun) wanted to be friends with the big one, the opposite was not true" (Mayol 2010:2501).⁴ Finally, in the following example, involving a weak contrast, the speaker uses overt *yo* to contrast what is true where s/he (the speaker) is concerned, regardless of what may be the case for others (Mayol 2010:5502):

- (4) Entonces cuando por la mañana sabes que se convoca una manifestación de estudiantes o, vamos, una cosa similar, pues te informas un poco del tema. Vamos **yo** por lo menos pues miro si ha pasado en días anteriores.

Although Mayol's study focuses on the contrastiveness of subject pronouns, other overt nominal expressions (e.g. NPs, demonstrative pronouns) could also convey the kinds of contrast she identifies. For instance, the demonstrative *esta* in place of *ella* in (3) could produce the same implicit contrast and refer anaphorically to *la ranita*. The observations cited above regarding the contrastive capacity of overt subject pronouns illustrate the ability of a referential form (e.g. a pronoun) to signal information to the hearer regarding the intended referent, which another expression (e.g. a null subject) cannot signal. In cases such as (1) (from Luján 1986), the overt pronoun tells the hearer to select the structurally and cognitively less salient potential antecedent, the grammatical object.⁵ However, in

⁴ Martín Rojo and Meeuwis (1993:95-96) also cite examples of implicit contrast requiring an overt subject, calling them cases of "contraste intrínseco".

⁵ See, for example, Huang (1994), Blackwell (2003), and Taboada and Wiesemann (2010) and earlier studies cited therein on the greater saliency (and thus foregrounded cognitive status) of topics and subjects over objects.

examples (2)-(4) (from Mayol), where subject expression makes no difference in referential interpretation, we see how overt subjects convey an additional message or “implicature” to the hearer indicating a contrast in actions or attitudes on the part of the referents mentioned.

Bentivoglio (1983), working within the functional-typological framework championed by Givón (e.g. 1983), analyzes third-person references in conversational narratives from Mexico City and Caracas in order to identify the most common linguistic forms used to maintain and interrupt topic continuity in Spanish. According to Huang (2000a), the basic premise of the topic-continuity model of anaphora is that choice of NP expression in discourse is determined by topic continuity versus discontinuity, where “topic” may be defined as what the discourse is *about* at a given point.⁶ Topic continuity is determined mainly by (1) linear distance (i.e. the number of clauses/sentences between two mentions of a referent); (2) interference, or the number of interfering referents in the discourse context at a given point; and (3) thematic information, such as maintenance or change of the protagonist (Huang 2000a:153). The theory predicts that the shorter the distance between two mentions of a referent, the fewer the interfering or competing referents; and, the more stable the thematic status of the protagonist, the greater the likelihood that it will be referred to by a reduced anaphoric expression. Thus, distance, interference and thematic status correlate with the use of more minimal anaphoric expressions, while greater distance between two mentions of an entity and multiple interfering NPs, along with non-protagonist status, increase the likelihood of reference via a non-anaphoric, full lexical NP (the most discontinuous NP being a referential indefinite NP, e.g. *a man*). Bentivoglio found that maximum continuity of reference to topics in Spanish was expressed by null subjects (her “verb agreement”) and unstressed clitics, while maximum discontinuity was indicated by the use of “existentials and presentatives” with indefinite NPs (e.g. *había un hombre, estaban unas tablas*) (1983:279), or by definite NPs modified by relative clauses.

Morales (1997) analyzes subject pronoun use in oral interviews from San Juan, Buenos Aires and Madrid within a variationist approach, but argues that discourse, cognitive, and communicative factors also influence subject expression. She observes that overt subject pronouns are more likely to be used with verbs of “mental activity” (e.g. *creer, pensar*), especially when the speaker (*yo*) is the subject (see also Enríquez 1984; Davidson 1996; Quesada and Blackwell 2009; and Posio 2011).⁷

Travis (2007) examines the effects of structural priming on first-person, singular subject expression in New Mexican Spanish narratives and Colombian Spanish conversation and finds that “genre has a profound effect on language patterns, and must be taken into account in order to better understand the grammar of language in use” (2007:103). Although her study deals exclusively with first-person, singular subject expression, the findings show that patterns of subject expression are conditioned by different genre needs. Our study examines the effects of the cognitive status of the referent on subject expression in the genre of film-retell narratives, which may be different for other types of discourse.

More recently, Filiaci (2010) compares null and overt subject reference in Spanish and Italian in order to test the validity of the Position of Antecedent Strategy (PAS), a parsing strategy proposed by Carminati (2002) for intrasentential anaphora resolution in Italian that is largely motivated by pragmatic principles, particularly Ariel’s (1990) Accessibility Theory principles (Filiaci 2010:172). Ariel’s theory (e.g. 1988, 1990) is intended to account for the use of referring expressions in discourse. It posits that each NP type is associated with a specific degree of accessibility of its antecedent in the memory of the addressee and that some potential antecedents are more accessible than others, which affects the speaker’s choice of referring expressions.⁸ Most relevant to the present study is the PAS

⁶ Both aboutness, i.e. the object the discourse is about at a given point, and antecedent saliency are directly related to the notions of “topic” and “givenness” (see, e.g. Chafe 1976; Givón 1983; Reinhart 1981, 1983).

⁷ Davidson (1996) and Posio (2011) focus on first and second-person, singular subject pronoun expression and omission in Peninsular Spanish, while Quesada and Blackwell (2009) deal only with first-person *yo* in native Mexican and L2 learner Spanish. It remains to be seen whether or not the same factors found to influence first and second-person subject use play a similar role in conditioning third-person subject use, where there are many more types of nominal expressions one can choose.

⁸ These factors include: 1) recency (Ariel’s “distance factor”); 2) the relative saliency of a potential antecedent when compared with other potential antecedents; 3) the degree of semantic connectivity between the antecedent

hypothesis⁹ “that the prominence of the antecedent, at least in the type of discourses examined, is determined by its syntactic position: the higher the structural position, the more prominent the expression” (Filiaci 2010:173).¹⁰ Filiaci assumes that preverbal subjects occupy the most prominent position in the sentence and are therefore relatively more accessible than their object competitors; thus, Italian and Spanish null subjects should be better at retrieving subject antecedents than object antecedents. The PAS claim of the relatively greater saliency of grammatical subjects supports the proposal (discussed below) that expressions in subject position are usually more likely to bring a referent into “in focus” cognitive status and maintain this status, than are references to entities in syntactically lower (non-subject) positions.

3. The Givenness Hierarchy

Gundel et al. pose the question, “what do speakers/writers know that enables them to choose an appropriate form to refer to a particular object and what do hearers/readers know that enables them to identify correctly the intended referent of a particular form?” (1993:274). Building on their previous work, and that of Chafe (1976, 1987), Prince (1981), Grosz (1981), Givón (1983), and Ariel (1990), among others, Gundel et al. (1993) distinguish six cognitive statuses: in focus, activated, familiar, uniquely identifiable, referential and type identifiable. Each is defined in terms of the necessary conditions for the use of different types of referring expressions in terms of the “degree of activation” or “referential givenness” of the intended referent at the point in discourse where the expression is used. Gundel et al. propose that these statuses are related implicationally along the following continuum, dubbed “the Givenness Hierarchy” (Gundel et al. 1993:275, 284):

The Givenness Hierarchy

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
<i>it</i>	{that/this/this N}	{that N}	{the N}	{indefinite this N}	{a N}
{Ø/él}	{ÉL/éste/ése/aqué/este N}	{ese/aquel N}	{el N}	{Ø N/ un N}	

According to this framework, in using a particular form, a speaker “signals that she assumes that the associated cognitive status is met and, since each status entails all lower statuses, she also signals that all lower statuses (statuses to the right) have been met” (1993:275-276). For instance, the referent of unstressed *it* must be not only in the interlocutors’ short-term memory (and thus “activated”), but also “in focus”, i.e. at the current center of their attention, in order for this pronoun to be felicitously used.

“Activated” referents are currently represented in the addressee’s short-term memory on account of having been introduced in the immediate linguistic or extralinguistic context. In the following example (Gundel et al. 1993:278, italics ours),

(5) I couldn’t sleep last night. *That* kept me awake.

the demonstrative pronoun *that* is possible only if the referent is present during the speech event or has been introduced in the immediate linguistic context. Gundel et al. maintain that the status “activated” is necessary for the appropriate use of all pronominal forms and that it is a sufficient status for the use of Spanish stressed pronouns (e.g. *ÉL*) and demonstrative pronouns (*éste*, *ése*, *aqué*), as well as a demonstrative NP (*este N*) (1993:284).

and the anaphor; 4) the number of competitors for the role of antecedent; and 5) the degree of “unity” in the discourse context, e.g., whether the antecedent is considered within the same world, frame, or point of view, or is found in the same discourse or text segment (e.g. paragraph) as the anaphor.

⁹ Similar hypotheses based on the syntactic salience of the antecedent have been formulated in earlier work on anaphora (see e.g., Huang 1994, 2000b).

¹⁰ Carminati tested the validity of the PAS in a series of reading experiments and sentence-completion tasks that manipulated variables such as the structural position of the antecedent, their number, and the use of gender, number and person cues for anaphora resolution (Filiaci 2010:172).

“Familiar” referents can be identified by the addressee “because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has)” (1993:278). According to Gundel et al., a speaker can use the demonstrative NP *that dog* only if the hearer already knows that the speaker’s neighbor has a dog (1993:278):

(6) I couldn’t sleep last night. *That dog* (next door) kept me awake.

A referent is “uniquely identifiable” if the addressee can identify it on the basis of the nominal expression alone. Gundel et al. propose that this status is necessary for all definite reference and thus for the use of the definite article *the* + *N*. However, they note that “identifiability does not have to be based on previous familiarity if enough descriptive content is encoded in the nominal itself” (1993:277). Thus, in

(7) I couldn’t sleep last night. *The dog* (next door) kept me awake.

the addressee does not need to have previous familiarity with *the dog next door* (e.g. on account of prior mention of this entity) to conjure up and identify that same dog (Gundel et al. 1993:277).

Finally, for entities that are “referential” and “type identifiable”, the addressee need only be able to access a representation of the type of object described by the referring expression. However, for referents with the status “referential”, the speaker intends to refer to a particular object, whereas for “type identifiable” entities, the hearer must only access a representation of the type of object being described (e.g. *a dog*, *a car*). These statuses distinguish between the following uses in English, though this distinction is unnecessary in Spanish. Specifically, *this dog* in (8) is “referential” (referring to a particular instance of a dog), whereas *a dog* in (9) requires that the addressee only access a representation of the “type” of animal described by the indefinite NP (Gundel et al. 1993:277, 276).

(8) I couldn’t sleep last night. *This dog* (next door) kept me awake.

(9) I couldn’t sleep last night. *A dog* (next door) kept me awake.

In Spanish either status, “type identifiable” or “referential”, which entails type identifiable, is necessary for the use of any nominal expression and is a sufficient condition for the use of the indefinite article + *N*. For Spanish, the Givenness Hierarchy predicts that an entity with either of these statuses will be referred to with an indefinite NP, with or without an indefinite article. For instance, in *Luis tiene coche*, the referent of *coche* is type identifiable for the addressee; but in *Luis conduce un coche nuevo hoy*, the speaker refers to a particular car (i.e. it has the status “referential”), yet the hearer must only access a representation of the “type of object” referred to by the indefinite NP.

4. Hypotheses based on the Givenness Hierarchy

We formulated the following hypotheses based on the six cognitive statuses in the Givenness Hierarchy. They reflect revisions to the original framework in terms of the predicted correspondences between cognitive status and type of referring form in Spanish. Thus, while Gundel et al. (1993) propose the same hierarchy to account for the distribution of referential expressions in Chinese, English, Japanese, Russian and Spanish, we argue that a revised set of statuses is needed in order to better account for third-person subject references in Spanish discourse.

Our first hypothesis is based on Gundel et al.’s status “in focus”. Following their proposal, we consider a referent to be in focus when it is the current center of attention and the most salient entity at a given point in the discourse; but, whereas they predict that this status is necessary for null subjects and unstressed pronouns including unstressed pronominal subjects, we predict that “in focus” status is necessary only for the use of null subjects but not for overt subject pronouns:

Hypothesis 1: “In Focus” (INF): Speakers will use null subjects to refer to utterance topics that are “in focus”, i.e., they are the current center of attention, the most salient entity from the previous utterance and the topic of the current utterance.

With regard to topic, Gundel et al. explain:

[t]he entities in focus at a particular point in the discourse will be that partially-ordered subset of activated entities which are likely to be continued as topics of subsequent utterances. Thus, entities in focus generally include at least the topic of the preceding utterance, as well as any still-relevant higher-order topics. (1993:279)

Gundel et al. define topic as “what the speaker intends the sentence to be about” (1993:279); however such a definition is subjective. Taboada and Wiesemann (2010) point out that unlike subjecthood, which is determined grammatically (by verb agreement in Spanish), “the definition of topic relies on context” (2010:1817). As noted earlier, the saliency of referents is typically determined by the grammatical function of their corresponding referring expressions. Also, Gundel et al. suggest that “subjects and direct objects of matrix sentences are highly likely to bring a referent into focus” (Gundel et al. 1993:279). However, this generalization leaves identification of the topic of an utterance up to one’s subjective impressions. Taboada and Wiesemann (2010) propose the following scale based on Di Eugenio’s (1998) ranking for Italian, where the experiencer in psychological predicates is the highest-ranked entity in terms of salience and thus the most likely to be assigned “topichood”, though it is not a subject (Taboada and Wiesemann 2010:1818):

Experiencer > Subj > Animate IObj > DObj > Impersonal/Arbitrary pronouns

Although the subject as the topic is a grammaticalized way to encode salience, the fact that the subject is not always sufficiently salient to become a topic is illustrated by Taboada and Wiesemann, who note that “some utterances encode a subject that contains very little information, such as impersonal subjects” (2010:1824); furthermore, “[i]mpersonal subjects and expletives such as *it* and *there* in English typically do not become [topics] in a sentence. Equivalent expressions exist in Spanish, typically realized in pro-drop form” (2010:1824). For instance, in (10), the pronoun *they* is the subject but does not become a topic, since the speaker (A) clearly wants to keep talking about herself (...*me*) (Taboada and Wiesemann 2010:1824):¹¹

- (10) B: a. wait a second
 b. this is being recorded, Michelle
 A: c. I don’t care
 d. they don’t know me
 B: e. oh okay
 A: [laughter]

The following example from the native speaker film narratives analyzed for this study illustrates how we operationalized the status “in focus”. Here, the subject, *Charles*, upon mention, brings this actor into focus, and the subsequent coreferential null subjects were coded as being “in focus” or INF:

- (11) ...pues, Charles llegó, Ø vio a la muchacha que estaba sola; Ø se sienta a platicar con ella pero... bueno, pues casi, casi Ø tropezó con ella; (G6,P1)

The second most restrictive status in the Givenness Hierarchy is “activated”. Our hypothesis based on this status is tri-partite, and, in contrast with the proposed implicational relationship between the

¹¹ The “in focus” constraint is also accounted for in Centering Theory by “the Pronoun Rule”, which “captures the fact that a topic that is continued from a previous utterance does not need to be signaled by more explicit means than a pronoun (or a zero pronoun, in languages that allow them)”, and, “the most salient entity must be realized by the least marked referring expression” (Taboada and Wiesemann 2010:1818). In other words, topic (and subject) continuity of reference predict the use of a null subject in Spanish “as the optimal form of expression for the most salient entity in the discourse” (Taboada and Wiesemann 2010:1818).

statuses in the Givenness Hierarchy, the categories of “activated” defined in Hypotheses 2A, 2B and 2C are mutually exclusive.

Hypothesis 2A is proposed to account for cases where the intended referent is not currently the most salient one (i.e., it is not in focus), yet reference via a null subject is recoverable on account of disambiguating factors including grammatical agreement, the hearer’s knowledge based on the previous discourse, and the lexical semantics of the corresponding verb. This hypothesis (like Hypothesis 1) is based on native speakers’ tendency toward minimization of linguistic form:

Hypothesis 2A: “Activated + Recoverable with a null subject” (ACTR): Speakers will use a null subject to refer to a referent that is in short-term memory due to being the topic of the immediately preceding event sequence, as long as the intended antecedent is recoverable such that it can be inferred from the context alone in the absence of an overt referring expression.

For example, in (12),

(12) ...y \emptyset_j llega y todo y \emptyset_j está probando a ver en donde \emptyset es más profundo. Mientras tanto **el esposo_i** se da cuenta mientras \emptyset_i está pensando que \emptyset_j era la muchacha, \emptyset_j se deja y todo, pero sintiendo **el esposo_i a Charles_j**, \emptyset_i se da cuenta de que \emptyset_j tiene bigote, ... (G6,P1)

the first two null subjects refer to the discourse topic in focus, *Charles* (introduced earlier in the narrative); next, mention of *el esposo* brings the husband into focus, and whereas the subsequent null refers to him, the following one refers to *Charles*, an activated referent in the local discourse due to being the topic of the preceding event sequence. The intended referent of this null subject (\emptyset_j era...) is recoverable due to its activation, the semantics of the verb phrase (i.e., the man could not be thinking that he himself was the girl) and knowledge of the previous discourse. This null was therefore coded as ACTR. The next null, \emptyset_j se deja, refers again to *Charles*, who was brought into focus by the previous null (\emptyset_j era la muchacha). The subsequent full NP, *el esposo*, brings the husband back into focus, and the next (fourth) null (\emptyset_i se da cuenta) refers to him (encoding in focus status), while the final null, which we coded as ACTR, refers again to *Charles*, a referent that is activated and recoverable from the discourse context, but no longer in focus at this point.

Hypothesis 2B is proposed to account for reference to activated entities where use of a null subject is not recoverable from the discourse context. Again, this prediction is based on native speakers’ tendency toward minimization of linguistic form:

Hypothesis 2B: “Activated” (ACT): Speakers will use an overt pronominal expression (minimally, a definite subject pronoun) to refer to referents in short-term memory due to prior mention in the immediately preceding event sequence when the intended referent (antecedent) is not the topic in focus and NOT recoverable from the context alone when a null subject is used.

This hypothesis accounts for the tendency by native speakers to use the most (semantically and phonologically) minimal expression available to express coreference in contexts where the referent is activated but where use of a null subject would not pick out the intended antecedent or would be rendered ambiguous. For instance, in example (13),

(13) Entonces \emptyset_i jala al señor de los ojos vendados_j, y cuando **él_{i/j}** se da cuenta de que \emptyset_i no es la muchacha, que \emptyset_i es un señor, \emptyset_j se quita la venda que \emptyset_j trae en los ojos y Chaplin_i aprovecha... (G6, P4)

the first null (\emptyset_i jala) refers to Charlie Chaplin, the “in focus” referent, but in the next clause, the pronoun *él* refers to the activated object in the preceding clause, *el señor de los ojos vendados*, where a null would encode “in focus” status and reference to Chaplin. We therefore coded this token of the pronoun *él* as ACT.

Hypothesis 2C is intended to account for the use of referring expressions that are semantically general¹² yet marked (more prolix, less usual, and less minimal than a definite pronoun) in contexts where 1) the intended referent is activated in the immediately preceding event sequence, and 2) a definite pronoun would not successfully pick out the intended antecedent:

Hypothesis 2C: “Activated but Non-Recoverable” (ACTNR): Speakers will use a semantically general NP expression to refer to referents that are activated in short-term memory due to prior mention in the immediately preceding event sequence when the intended referent (antecedent) is NOT recoverable from the context alone when a definite pronoun is used.

This hypothesis accounts for the tendency by native speakers to use semantically general, non-pronominal expressions (e.g. *los dos*, *las dos mujeres*, *el otro*, *uno de los hombres*, *uno de ellos*) to refer to activated referents, where a pronoun would not felicitously refer to the intended one. In example (14), the speaker uses *uno de ellos* to encode reference to one of the men who are activated at this point in the narrative due to the introduction of “another two men” (*otros dos señores*) in the immediately preceding clause. We coded *uno de ellos* in the data as ACTNR:

(14) Aparecen otros dos señores y se sientan con Chaplin en la banca, pero **Chaplin** empieza a tomar de su bebida mientras ellos están distraídos. **Uno de ellos** se da cuenta y empieza otra vez otro pleito. (G6,P4)

In this context, either a null subject or the pronoun *él* would refer to *Chaplin*, the “in focus” referent at the point where the verb phrase *se da cuenta* is used.

Hypothesis 3 is proposed to account for subject references to entities that are familiar to the addressee on account of their having been previously introduced in the discourse context. These referents can be introduced by the speaker producing the narrative, the addressee, or may have been referred to in the instructions to the participants in the study or in the film they were asked to retell:

Hypothesis 3: “Familiar” (FAM): Speakers will use a definite expression, including a definite NP (*el N*) or a proper name (*Charlie*), when the addressee can identify the intended referent because s/he has a representation of it in short-term memory due to prior mention in the discourse.

In (15), the speaker uses the NP *la muchacha* in the middle of her narrative, which in turn signals to the hearer that s/he can identify this referent on account of it having been introduced earlier in the narrative—here, via the indefinite NP *una muchacha*. We coded this definite NP as FAM:

(15) ...pasa una muchacha_i que le coquetea al esposo_j, este señor_j se va atrás de la muchacha_i ... (several scenes later)... pero **la muchacha_i** le pide que Ø_j se tape los ojos. (G6,P4)

Hypothesis 4 is formulated to account for references to entities that the addressee can identify based on the NP expression alone or because the identity of the referent is “inferrable” with no prior mention of it in the discourse. According to Prince (1981), “[a] discourse entity is inferrable if the speaker assumes the hearer can infer it, via logical—or, more commonly, plausible—reasoning, from discourse entities already Evoked or from other inferrables” (1981:236). The referent of an inferrable is linked to some other entity in the immediate discourse context through a contextual assumption, i.e. a “bridging inference” (Clark and Haviland 1977).¹³ The ability to construct an appropriate referent

¹² Levinson (1987:100, 102) refers to both the pronoun *he* and the NP *the man* in the sentence *John turned the switch and {Ø/he/the man} started the motor* as being “semantically general”, explaining that both expressions have roughly the same sense (e.g. third-person, singular, masculine), but that they contrast on the level of markedness (*the man* being more “prolix” and thus the marked form). Levinson argues that a speaker would not use the marked yet semantically general NP, if coreference were intended with the subject, *John*.

¹³ Michaelis and Francis (2007), following Prince (1981), refer to NPs triggering bridging inferences as “frame inferrables”. These include “those NPs whose referents are identifiable by virtue of belonging to a semantic frame

for the inferrable depends on its plausibility and on the degree to which the addressee's attention is focused on the entity in the discourse context to which the inferrable is linked (Gundel 1996). Our hypothesis accounts for instances where, for example, a speaker introduces an entity via an indefinite NP such as *una familia* and then refers to entities via definite NPs such as *el padre/el hombre* or *la madre/la mujer* because these entities can be inferred as being specific members of the already introduced set, “a family”:

Hypothesis 4: “Uniquely Identifiable” (UNI): Speakers will refer to entities via definite NPs (e.g. *el N*) when the entities are already existing in the addressee’s memory and thus s/he can identify the speaker’s intended referent (1) on the basis of the referring expression alone (e.g. *Charlie Chaplin, el sol, la luna, el perro del vecino*), or (2) because a particular entity can be inferred by association with another entity that has been activated in recent discourse.

In (16), the referent encoded by the definite NP *los actores* can be identified by the hearer as “the actors who appear in the film being narrated” on account of the fact that the film (*la película*) has been introduced in the recent discourse. We therefore coded the NP *los actores* as UNI:

(16) La película comienza como todas, con algunas letras las cuales expican, me imagino, la secuencia que tenían ahí **los actores**... (G6, P8)

Hypothesis 5 is proposed to account for third-person references to entities that are introduced into the discourse for the first time. For example, if a speaker says *Había una familia*, the addressee will be able to understand the “type” of entity referred to by the NP *una familia*.

Hypothesis 5: “Referential”/“Type Identifiable” (TYP): Speakers will use an indefinite NP to refer to an entity when the addressee can be assumed to understand the type of thing that the NP describes.

For instance, mere mention of *una muchacha* in (17) is sufficient to enable the hearer to understand the type of referent being described (“a girl”). *Una muchacha* in the following sequence was therefore coded as TYP:

(17) ...pasa **una muchacha** que le coquetea al esposo, este señor se va atrás de la muchacha ... (G6,P4)

While this hypothesis predicts the use of indefinite NPs to refer to entities that are type identifiable, in some cases entities with this status may be referred to via a structurally definite NP, e.g. *la policía*, as in presentational utterances like *Llegó la policía*, to refer generically to any or all members of a group.

5. The present study

In order to test the proposed hypotheses, we examined the oral narratives of native Spanish speakers and three groups of L2 learners and compared their use of third-person subject expressions in terms of the cognitive status of the entity to which each expression referred.

5.1. Participants

For the present study, we analyzed 40 film retell narratives produced by learners of Spanish from

that is currently active” (2007:40). To illustrate this notion, they cite a passage where the NP *the prosecuting attorney* is used with no prior mention to this referent, but after a reference to *the jury system*. As these authors explain, the entity denoted by *the prosecuting attorney* is “highly recoverable by virtue of its relationship to the previously evoked frame, the litigation frame” (2007:40).

three proficiency levels and native speakers from Mexico.¹⁴ All learners were enrolled in first and second-year Spanish and upper-level Spanish linguistics courses at a major U.S. university. Level 1 was comprised of beginning students in the third trimester. Eight of the ten had had two to three years of high school Spanish, but placement exams indicated a beginning level of Spanish. Level 2 consisted of intermediate students enrolled in the second trimester of the second year. All but one had had two to four years of Spanish in high school, and one had taken French for six years. The students in Level 3 were taken from upper-division Spanish linguistics courses; all had had two to four years of Spanish in high school and two had also studied French for four or six years. There were no participants who had studied abroad and no heritage speakers of Spanish. A subsequent error analysis was performed on the learners' narratives and participants were divided into three groups according to their use of correct verbal morphology in obligatory contexts. Ten participants each were placed into three different groups. The error analysis revealed that for these participants, level of study was a good indicator of linguistic competence: seven first-year learners were assigned to Group 1 (beginner), two were placed into Group 2 (intermediate) and one participant into Group 3 (advanced); seven second-year learners were assigned to Group 2 and three were placed into Group 1; nine of the upper-division learners were placed in Group 3 and one into Group 2. The speakers in Group 4 were Mexican students enrolled in an undergraduate degree program in Modern Languages at a Mexican university.

5.2. Instruments

For the narrative task, participants viewed a five-minute segment of the Charlie Chaplin silent film, *A Woman*, and were asked to narrate the film to a research assistant who they had been told had not seen the film. Previously, participants had provided information on their foreign language background, study abroad experience, and major, and had signed a consent form agreeing to participate in the study. The learners were given instructions for the task in English and some basic vocabulary in Spanish related to the film; Spanish speakers were provided with the same instructions in Spanish but were not given vocabulary. The narratives were transcribed using conventional orthography and included pauses, false starts, and verbalizations such as 'um' and 'er'.¹⁵ If several forms were uttered, as is customary in second language acquisition, the last uttered form was the one considered for analysis. We chose a silent film narrative for this study so that we could control for both content and discourse genre. Furthermore, this task elicited numerous references to third persons, as there are several protagonists in the film who carry out sequential actions and interact with each other in rapidly changing scenes.

5.3. Analysis

The narratives were tagged for all subject expressions, including lexical, pronominal and null subjects. We then generated collocations of the tags with the concordance program *WordSmith*. The subjects analyzed included null pronouns, overt pronouns, full noun phrases with definite, indefinite and demonstrative determiners, definite, indefinite and demonstrative pronouns, quantifiers and proper names. We eliminated subjects that were obligatorily null (e.g. *Ø es muy extraño, los señores que Ø llegan con Chaplin*), indefinite pronouns (e.g. *no pasó nada*), and subjects of dative verbs (e.g. *a Charlie le gusta la mujer*). Although we observed that dative constructions often activated a following subject, most learners used these *gustar*-type verbs as if they were transitive (e.g. *Mis amigos gustan mi padre porque mi padre está chistoso*) and therefore, they were eliminated from the analysis. The collocations generated by *WordSmith* were used as the basis for analysis but we also consulted the full text of the transcriptions in order to determine the cognitive status of each referential subject. Each type of subject was analyzed for its cognitive status based on our revised version of Gundel et al.'s Givenness Hierarchy, including "in focus", "activated" (three types including "not recoverable with a

¹⁴ Although speakers of Mexican Spanish are known to employ high rates of null subject pronouns in comparison with speakers of other varieties of Spanish, the purpose of this paper is to examine all subject use according to the cognitive statuses discussed and not the comparison of null and overt subject pronouns.

¹⁵ In many of our examples, we have eliminated these elements when they are not relevant to the analysis.

null”, “recoverable with a null” and “not recoverable with an overt pronoun”), “familiar”, “uniquely identifiable”, and “type identifiable”. Each author analyzed all 40 narratives individually and then collaboratively. When discrepancies occurred, the researchers discussed the viability of the two analyses, consulted the full transcription of the text, and came to an agreement. On only two occasions the researchers did not agree on the cognitive status of the referential subject – the context and the narrator’s choice of subject form did not make clear who the subject was – and those two contexts were eliminated. The first named author had not viewed the film and analyzed the subject types used based only on the forms, the context of the narrative and the hypotheses of the cognitive statuses. The second named author had seen the film and could use that background knowledge as well.

Following the analysis of the narratives, the distribution and frequencies were calculated for each group’s third-person subject forms used according to cognitive status. Subsequently, chi-square and, in some instances, Fisher’s Exact tests were run for categories of subject form that were relevant for the cognitive status; that is, if there were 0-3 tokens across all groups, this category was eliminated from the chi-square. Because there often were too few tokens for subject types, all learner groups were combined for the chi-square tests. If there were still too few cells, Fisher’s Exact test was subsequently carried out. Finally, logistic regressions were performed for any subject form within any cognitive status where it was possible that there were significant differences between subject form choice according to learner level (comparing beginning, intermediate and advanced speakers). In these cases we performed six tests (comparing beginners to intermediate, advanced and native speakers, comparing intermediate to advanced and native speakers, and comparing advanced to native speakers). We compared these pairwise p-values to a significance level lower than 0.05. In Section 6 we will make clear when the logistic regression revealed significant differences between groups for each subject form according to cognitive status.

6. Results and discussion

The analysis was based on a total of 1,902 third-person subject expressions among the four groups. Table 1 shows the distribution of these expressions by cognitive status. The data show that the participants in this study tend to select subject forms that were predicted by our hypotheses, particularly for the cognitive statuses that are found on either extreme of the Givenness Hierarchy.

Table 1. Frequency and distribution of subject forms by cognitive status (all groups)

Frequency Percent	INF	ACTR	ACT	ACTNR	FAM	UNI	TYP
Null	682 76.03	108 61.71	9 3.90	1 0.80	3 1.11	0 0.00	0 0.00
Overt	141 15.72	34 19.43	81 35.06	2 1.60	1 0.37	1 1.14	1 0.87
GenNP	6 0.67	8 4.57	5 2.16	42 33.60	19 7.01	0 0.00	2 1.74
Dem	2 0.22	2 1.14	5 2.16	0 0.00	0 0.00	0 0.00	0 0.00
Def NP	31 3.46	7 4.00	88 38.10	31 24.80	155 57.20	57 64.77	13 11.30
Prp N	34 3.79	16 9.14	41 17.75	47 37.60	87 32.10	30 34.09	0 0.00
Ind NP	1 0.11	0 0.00	2 0.87	2 1.60	6 2.21	0 0.00	99 86.09
Total N	897 100.00	175 100.00	231 100.00	125 100.00	271 100.00	88 100.00	115 100.00

For the INF status, participants overall selected a null form in 76% of the cases and an indefinite noun phrase in over 86% of the instances for the TYP status. For the ACTR status, participants

selected a null form in almost 62% of the cases, and for the UNI and FAM statuses they preferred a definite noun phrase approximately 65% and 57% of the time, respectively. In the following sections we discuss the selection of subject forms by cognitive status and, where relevant, the differences among the groups.

6.1. The “in-focus” (INF) cognitive status

We had hypothesized that speakers would use a null subject to refer to referents that are “in focus”, i.e., the current center of attention, the most salient entity from the previous utterance and the topic of the current utterance.

As the results in Table 2 demonstrate, all groups prefer a null pronoun to refer to a referent that is in focus, although none of the learner groups do so to the same extent as the native speakers (NSs), who choose a null almost 91% of the time, compared to between 61 and 65% for the learner groups. Chi-square tests were carried out on the categories of null, overt, definite NP and proper name and results reveal that there is a highly significant difference between all learner groups combined and the NS group in the use of subject forms when the referent is in focus.

Table 2. Subject forms for cognitive status: INF (expect NULL)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Null	70 (63)	107 (61)	120 (65)	385 (90.5)
Overt	27 (24)	42 (24)	49 (26)	23 (5)
GenNP	1 (1)	0	2 (1)	3 (1)
Dem	0	0	0	2 (.5)
Def NP	8 (7)	9 (5)	6 (3)	8 (2)
Prp N	6 (5)	16 (9)	8 (4)	4 (1)
Ind NP	0	1 (1)	0	0
Total	112	175	185	425

¹⁶ $df=3, X^2=100.2005, p<.0001$

The logistic regression performed for the Null, Overt and PrpN categories for the TYP status confirmed that there were no significant differences among learner groups but a highly significant difference between each learner group and the NS group; p-values ranged from 0.0117 to <.0001.

Indeed, as far as INF referents are concerned, although learners mainly use the expected null, as seen in (18), they also tend to overuse the unexpected overt and PrpN forms, as in (19) and (20).

(18) Charlie Chaplin_i vea la chica y Ø_i siéntense¹⁷ al lado de la chica (G1, P2)

(19) y Charlie Chaplin_i entró y él_i se tropezó en un aspersor (G1, P10)

(20) Charlie Chaplin gana con su bastón y **Charlie Chaplin** busca la padre (G1, P9)

The NSs also use overt pronouns or definite referring expressions for in-focus referents, but to a lesser degree; however, a few examples will illustrate that although the referents are in focus in the narrative, other pragmatic factors can constrain or trigger the use non-null forms. We find full noun phrases, proper names or overt stressed pronouns to signal the beginning of a new scene or action, reiteration of a current topic, contrastive focus, or another type of emphasis. In (21), although *el señor* is also the subject of the two previous clauses, the narrator repeats the full NP to signal the beginning of his contact with the girl. In (22) and (23), the proper name and overt pronoun are used in situations

¹⁶ In each case, chi-square tests are given beneath the tables and refer to tests carried out on subject form by combining all learner groups and contrasting them with native speakers. If 25% of the cells had expected counts less than 5, Fisher’s Exact test was also performed.

¹⁷ Although this is the command form, it is clear from the context that the beginning learner intended it to be a third-person singular form with Charlie Chaplin as its null subject.

of contrastive focus, and in (24), the narrator emphasizes the same subject of the second clause with an overt stressed pronoun to highlight the referent's (Charlie Chaplin's) craftiness at slinging the girl to the ground so that he can then help her. Finally, in (25) the narrator interrupts himself to try to remember details of the film, and afterwards, in order to pick up where he left off, uses an overt pronoun to signal his reiteration of the topic. In each case, the referent is already in focus, but the narrative structure appears to require a type of emphasis to carry the story-line forward.

- (21) este señor_i se va atrás de la muchacha y Ø_i empieza como que a cortejarla. Entonces **el señor**_i le invita a algo de tomar (G4, P4)
 (22) entonces Ø_i lo_j empuja. Mientras **Charlie Chaplin**_i se está riendo porque el otro_j está en el agua... aparece_k un policía (G4, P2)
 (23) el papá_i está volteando para otro lado y en ese momento cuando **él**_i está sentado en la banca, va pasando una muchacha_j (G4, P2)
 (24) la_j Ø_i tira al suelo y ya casualmente **él**_i está ahí como para ayudarla_j (G4, P8)
 (25) Después llegaron dos señores_i... ¿Cómo me pareció que eran?... pero, bueno ahí **ellos**_i estaban tomando agua, cerveza o lo que sea... (G4, P1)

We do not observe these functions of fuller (i.e. non-null) forms to refer to entities that are “in focus” in the learner narratives.

6.2. The “activated and recoverable” (ACTR) cognitive status

Table 3 presents the results of the analysis of the use of subject forms for the cognitive status, “activated and recoverable”. We had hypothesized that, although a referent is not the topic and not “in focus”, it can be completely recoverable from the context, and therefore a null pronoun is the expected form. Our results reveal that this is the tendency for the NSs, who employ a null in 82% of such cases, and, to a lesser extent, for the more advanced learners, who use nulls in 59% of these instances. Group 2 prefers an overt pronoun in these cases (41%), and Group 1, a proper name (52%). Eliminating the demonstrative (Dem) form, the chi-square and Fisher's confirm a significant difference in use of subject forms between learner groups combined and NSs. Although for the INF status there was no significant difference among learner groups, for the ACTR status we note that only the advanced group prefers a null in the majority of the cases and an overt pronoun in second place. The intermediate learners select more often overt pronouns and the beginners, a proper name.

Table 3. Subject forms for cognitive status: ACTR (expect Null)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NSs
Null	2 (10)	3 (19)	24 (59)	79 (82)
Overt	5 (24)	7 (41)	14 (34)	8 (9)
GenNP	0	4 (23)	0	4 (4)
Dem	0	0	0	2 (2)
DefNP	3 (14)	1 (6)	1 (2)	2 (2)
PrpN	11 (52)	2 (12)	2 (5)	1 (1)
Total	21	17	41	96

$df=4$, $X^2=45.2529$, $p<.0001$; Fisher's Exact Test $p=.00015$

Logistic regressions were performed on differences between groups for null, overt and PrpN forms only. The results revealed that for the null form there was a highly significant difference between learner groups (except between the beginners vs. intermediate learners with a p-value of 0.4675) and between each group and NSs (p-values ranged from 0.0080 to <.0001); NSs use nulls, learners do not, and beginners use them the least of all. The difference between the groups is less significant as proficiency level increases. The logistic regression for the use of overt forms revealed significant differences only for intermediate vs. NSs ($p=0.0009$) and advanced vs. NSs ($p=0.0004$);

finally, the regression for PrpN forms showed a highly significant difference between beginners and advanced learners ($p=0.0003$) and between beginners and NSs ($p<.0001$). There were no significant differences between the intermediate and advanced groups and between the advanced and NSs. In sum, the beginners in this study consistently prefer the PrpN, intermediates, the overt pronoun, and advanced and NSs select the expected null, although advanced learners not to the same extent as NSs.

In (26) below, the advanced learner uses the expected null form in the third clause to refer back to Charlie who was the focus of attention in the first clause, but it changes to both Charlie and the man in the second clause. Context and reversion to a singular verb (*Ø usa*) after a plural subject pronoun and verb (*ellos van*) make it clear that it is Charlie who uses his cane to play with the man. In (27), the learner from Group 2 uses an overt pronoun to refer to the woman, who has been activated in the previous clause, and it is clear that she is with Charlie, not the man, who gets mad; a null would have also referred to the woman. Finally, the beginning learners from Group 1 use a DefNP in (28) and a PrpN in (29) to refer to referents that, although they are not in focus, are activated and recoverable with a null due to the context.

(26) y Charlie_i usa el un el bastón para jugar más con el hombre_j...ellos_j van a caminar...Ø_i usa el bastón para como engañar, burlar el hombre (G3, P2)

(27) y cuando el otro hombre_i regresa a la mujer_j, él_i se enoja porque ahora **ella_j** es con Charlie.... (G2, P 7)

(28) el padre_i coquetea con la coqueta_j y **el padre y la coqueta_j** caminan por un parque. (G1, P9)

(29) Cuando Chaplin_i está listo a empujar el hombre_j en el lago, el hombre_j sienta su bogota_(i) [bigote] y sabe_j que **Chaplin_i**, no es la coqueta_k (G1, P6)

Whereas the lower proficiency learners tend to use a more elaborate form when the referent is activated yet recoverable (PrpN or overt), more advanced learners and NSs will tend towards the null as (30) and (31) illustrate. In (30) the husband is the focus of attention (realizing and thinking); the second null refers to Charlie, who is on the scene and is activated. The referent of the subject of *se queda inconsciente* in (31) is also activated in the previous clauses, and because of the hit he receives from Charlie, it is obvious that he is the one who ends up unconscious (not Charlie, who does the hitting). NSs speakers apparently do not need to specify reference to activated referents via overt expressions when the context, including verb semantics, makes it clear.

(30) el esposo_i se da cuenta mientras Ø_i está pensando que Ø_j era la muchacha (G4, P1)

(31) Charlie Chaplin_i, en un descuido que tienen los dos_{j,k}, toma su_j botella y Ø_i lo_j golpea y Ø_j se queda inconsciente (G4, P2)

The difference between INF and ACTR and the choice of subject forms may be due to English transfer and interference. Learners must learn that they do not need an overt subject when the referent is activated and the context makes it clear about whom we are speaking.

6.3. The “activated” (ACT) cognitive status

For the cognitive status “activated” where we predicted that speakers would select an overt pronoun to pick out a referent which is not recoverable from context with a null, our data reveal some interesting results. NSs only selected overt pronouns in about 41% of the instances, selecting a DefNP in second place (26%) and a PrpN in third (13%). Surprisingly, they chose nulls in 12% of the instances which, examples will demonstrate, lead to a breakdown in communication. For this cognitive status, all learner groups preferred a DefNP in first place and overts in second, although both the chi-square and the logistic regression (not including the IndNP category) show no significant differences between learners and NSs or between each of the pairs of groups. Although we observe that the lower proficiency learner groups choose from a more limited number of subject forms compared to the advanced and NS groups, again the logistic regression reveals these differences are not significant.

Table 4. Referring expressions for cognitive status: ACT (expect overt pronoun)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Null	0	0	1 (2)	8 (12)
Overt	14 (37)	22 (30)	17 (33)	28 (41)
GenNP	0	2 (3)	0	3 (4)
Dem	0	0	2 (4)	3 (4)
DefNP	17 (45)	30 (42)	23 (44)	18 (26)
PrpNP	7 (18)	16 (22)	9 (17)	9 (13)
IndNP	0	2 (3)	0	0
Total	38	72	52	69

$df=5$, $X^2=24.9720$, $p<.0001$; Fisher's Exact Test $p=.00012$

The examples below illustrate the subject forms selected:

(32) cuando $\acute{e}l_i$ no vio, Charlie Chaplin $_j$ bebió la soda de $\acute{e}l_i$, y $\acute{e}l_i$ vio (G1, P10)

(33) la coqueta $_i$ y el primer hombre $_j$ deciden jugar a las escondidas y **la coqueta** $_i$ venda a los ojos con una pañuela (G2, P9)

(34) y dos hombres $_i$ con bebidas se sienten con Charlie $_j$ y **Charlie** $_j$ bebe la bebida...la bebida...de uno hombre (G3, P4)

In (32), the beginning learner uses the expected overt pronoun to pick out the activated, but not in focus, referent. In (33), the learner from Group 2 repeats the full noun phrase to describe the woman who does the blindfolding when the overt pronoun, *ella*, would have picked her out just as well; likewise in (34), the pronoun *él* instead of the second *Charlie* would have been sufficient to pick out *Charlie* as the antecedent. As noted in Section 2, Luján (1986) shows how an overt subject pronoun picks out an object antecedent, where a null would select the subject (see example 1).

(35) el hombre $_i$ no trae un refresco pero ya, se ve que ya \emptyset_i lo trae otra vez, pero bueno, entonces, \emptyset_j le tapa los ojos y ella $_j$ se esconde (G4, P2)

(36) Como no \emptyset_i le toma importancia, \emptyset_j regresa y \emptyset_j le vuelve a tirar la falda y **la señora** $_i$ sigue indiferente (G4, P8)

(37) Aparecen otros dos señores $_i$ y \emptyset_i se sientan con Chaplin $_j$ en la banca, pero **Chaplin** $_j$ empieza a tomar de su $_i$ bebida mientras ellos $_i$ están distraídos (G4, P4)

In (35), the null subject refers to the girl, but the first author, not having seen the film, could not tell who was covering up the man's eyes; the second author, having seen the film, knew that it was the girl. Here, communication breaks down and without sufficient contextual information, we do not know who the intended referent of the null subject is. The same occurs in (36) where the null of the second clause would refer to the same subject, the girl, as in the first clause, but only by knowing the story, do we know that it is Charlie who returns and again tangles the girl's skirt. Also, in (36), in the fourth clause of the example, the narrator selects a full noun phrase, *la señora*, when the overt pronoun, *ella*, would refer to the girl. Finally, in (37) Chaplin is already activated in the sequence and the narrator could select an overt pronoun, *él*, to refer to him; however, she selects the proper name (*Chaplin*) to refer to him. Although NSs do use nulls 12% of the time to refer to activated referents that are not, at least for the analysts, recoverable with a null, the learners do not (except in one case). There were no instances in the data where the intended antecedent was referred to via a null in a contrastive construction. As noted earlier, overt subjects are usually considered to be obligatory in contrastive constructions, though nulls may be rendered acceptable when other linguistic devices such as adverbials are used to mark contrastiveness (see Section 2).

6.4. The “activated-not recoverable” (ACTNR) cognitive status

For the cognitive status, “activated-not recoverable”, we had predicted that speakers would use semantically general NP expressions (*los dos, las dos mujeres, el otro, uno de los hombres, uno de ellos*) to refer to referents that are activated in short-term memory when the intended referent is not recoverable with an overt pronoun. This hypothesis is upheld for NSs less than half of the time (with 44% use), and oddly enough, it is more strongly supported by the results from the intermediate group (with 67% use), although with only 15 occurrences for this group, these conclusions are only tentative. Surprisingly, this hypothesis does not hold for the advanced group who selects a GenNP in only 10% of the instances; this group prefers to select the proper name (*Charlie Chaplin*) to distinguish between the protagonists of the film when several of them, usually male, are activated in the narration. Among the three learner groups, there is less variety in the number of minimal expressions used, including *uno/a (de ellos/as), uno, los dos/tres, ambos, y todos*, while the NSs use these and others, including *el/la otro/a, el otro hombre/señor, la otra mujer, y los dos señores/las dos mujeres*, which may show more complexity in the types of forms used by NSs to refer to the ACTNR referents, or simply may be a product of the overall higher number of subjects used in their narratives.

Table 5. Referring expressions for cognitive status: ACTNR (expect GenNP)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Null	0	0	1(2.5)	0
Overt	1 (5)	0	0	1 (2)
GenNP	6 (28)	10 (67)	4 (10)	22 (44)
DefNP	4 (19)	2 (13)	10 (26)	15 (30)
PrpN	9 (43)	3 (20)	23 (59)	12 (24)
IndNP	1 (5)	0	1 (2.5)	0
Total	21	15	39	50

$df=2, X^2=7.6051, p=0.0223$

The chi-square test calculated for the GenNP, DefNP and PrpN forms reveals a significant difference between the learners and NSs; NSs choose the GenNP form whereas the non-native speakers (NNSs) more often select PrpNs, followed by GenNPs. However, the logistic regression between pairs for the use of GenNPs shows significant differences between beginners vs. intermediates ($p=0.0275$), between intermediate vs. advanced ($p=0.0002$), and between advanced vs. native speakers ($p=0.0013$). There are no significant differences between the other pairs. For the PrpN forms there were also differences between intermediate vs. advanced ($p=0.0155$) and between advanced and NSs ($p=0.0011$). These results confirm that learners and NSs are performing differently for this cognitive status; but then there is also a difference between the intermediates and the other learner groups. We believe again that the low number of tokens makes it difficult to explain these findings. The logistic regression shows that there are no differences between groups in the selection of DefNPs.

Below are examples from both learners and NSs for one of the scenes. The girl blindfolds the man (telling him that they are going to play hide-and-peek), then steals his wallet and leaves the scene. Then Charlie comes back on the scene, and the blindfolded man thinks that Charlie is the girl, but Charlie thinks he can play a joke on him, too. The learner in (38) refers to the man with a full NP, *el hombre*, to distinguish him from Charlie, rather than a more minimal NP, *el otro*. In (39), the NS refers to him with a more minimal NP, *el otro*.

(38) el hombre_i piensa que Charlie_j es la muchacha_k que regresó para jugar con él_i, pero Charlie_j, en una manera, tiene el poder porque ahorita **el hombre**_i está detrás de él_j (G3, P3)

(39) [la chica]_i no quiere estar con él_j y le Ø_i dice que Ø_{ij} jueguen a las escondidas y le Ø_i tapa los ojos, le Ø_i roba la cartera y ya y Ø_i se va, y en eso llega Chaplin_k pero **el otro**_j piensa que Ø_i sigue jugando (G4, P9)

In another scene, Charlie is with two other men and he drinks from one of their beverages. When they find out, they hit Charlie on the head and he falls down and is apparently unconscious. Again, the learner uses a proper name in (40) and the NS a less specific, semantically less informative NP in (41).

(40) y entonces **Charlie**_i es confundido porque \emptyset _i se cayó (G3, P3)

(41) y **el otro** queda inconsciente

6.5. The “familiar” (FAM) cognitive status

Our prediction that speakers will use a definite expression including definite NPs (*el N*) and proper names (*Charlie*) when the speaker can identify the intended referent due to prior mention in the discourse is born out for all groups. All groups use definite NPs and PrpN combined at least 90% of the time. Results of this analysis are summarized in Table 6. Again the chi-square ($p=0.0003$) for the GenNP, DefNP, PrpN and IndNP forms reveals a highly significant difference in how NSs and learners select subject forms for this cognitive status; although both groups select DefNPs more often than other forms, NSs by far select more DefNPs than NNSs and NNSs select a higher proportion of PrpNs. The logistic regression for the selection of PrpN confirms this finding and reveals a difference between each learner group vs. NSs (p -values ranging between 0.0043 and 0.0009).

Table 6. Referring expressions for cognitive status: FAM (expect DefNP or PrpN)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Null	0	0	0	3 (4)
Overt	0	0	0	1 (1)
GenNP	3 (5)	4 (6)	2 (3)	10 (12)
DefNP	30 (55)	31 (45)	39 (60)	55 (67)
PrpN	22 (40)	28 (40)	24 (37)	13 (16)
IndNP	0	6 (9)	0	0
Total	55	69	65	82

$df=3$, $X^2=19.0302$, $p=0.0003$

Questions arise as to why learners would use indefinite expressions for familiar referents (in the cases of Groups 2 and 3) and why NSs would use null and overt pronouns for referents with this cognitive status. In the case of the learners, the use of indefinites is clearly unexpected, as the example in (42) shows. In the first clause, the boy and the two women have been introduced; after other background information is presented, the narrator refers to these referents with indefinite NPs, whereas the expected form would have been either a DefNP or a PrpN.

(42) Hay un muchacho_i y dos mujeres_{s_{jk}} en el parque, se sentieron_{ijk} en el parque...**dos** **mujeres**_{jk} están aburrida y **un muchacho**_i lea un papel (G2, P10)

In examples (43) and (44), NSs use null and overt pronouns to refer to familiar referents. These examples show how lack of the use of a definite expression leads to a breakdown in communication. Only the researcher who had seen the film could identify the referents in the following passages. In (43) there are three men involved, Charlie Chaplin and two others who fight with Charlie when he drinks one of their beverages. The null pronouns in the last two clauses of the passage appear to pick out *el hombre* as the referent, but in the film, it is Charlie who walks and passes over the fallen, unconscious man. When the speaker produces the utterance in (44), s/he has just narrated two scenes involving a man and a woman and subsequently two other men with beverages. The focus is then on Charlie Chaplin who was unconscious but then sits up. It is not clear if the use of *ellos* in the next clause refers to the man and woman or the other two men. Only the researcher who knew the film could understand the intended referent of the pronoun used, *ellos*. In both (43) and (44), the use of a definite noun phrase in these cases would have disambiguated the intended referents.

- (43) El otro_i se le pone al brinco pero cuando ve_i que Chaplin_j se le va a enfrentar, sale_i corriendo. Mientras el hombre_k está tirado en el piso, Ø_k queda inconsciente, y Ø_j camina, Ø_j pasa sobre él_k (G4, P2)
- (44) estaba Charlie Chaplin_i inconsciente y Ø_i se para, y ellos_j se sientan junto a él (G4, P10)

6.6. The “uniquely identifiable” (UNI) cognitive status

Our fourth hypothesis predicts that speakers will refer to entities with definite expressions when the referent exists in memory on the basis of the referring expression alone or because it can be inferred by association with one that has already been activated in recent discourse. Our data support this prediction for all groups who use a definite expression in almost 100% of the instances. Learners either acquire this constraint for Spanish or their knowledge of English transfers to their L2 system. Nonetheless, the chi-square ($p=0.0167$) and Fisher’s ($p=0.0183$), not including the overt form, indicate that there is a significant difference between learners’ and NSs choices; NSs are more likely to pick a DefNP whereas learners will select PrpN more frequently, although the logistic regression did not reveal significant differences among pairs of groups.

Table 7. Referring expressions for cognitive status: UNI (expect DefNP or PrpN)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Overt	1 (4)	0	0	0
DefNP	15 (60)	9 (53)	9 (53)	24 (83)
PrpN	9 (36)	8 (47)	8 (47)	5 (17)
Total	25	17	17	29

$df=1$, $X^2=5.7237$, $p=0.0167$; Fisher’s Exact test $p=0.0183$

- (45) Hay un familia_i en el parque y los señoritas_(part of i) estaba durmiendo cuando el señor_(part of i) conozó una señorita_j; (G1, P1)
- (46) entonces, entra **Charlie**_i quien_i cámina como un payaso (G3, P3)
- (47) Interviewer: ¿Qué pasó en el video?
NNS: Ok. Pues, **el título** es “Una Mujer”... (G2, P5)

In (45), the referent, a family, is introduced and later, the narrator refers to the individuals in the family with a DefNP; we can uniquely identify them as part of the family. Also in (46), there is no need to refer to Charlie Chaplin with anything else than his name; the film is about him and he is (almost) universally known. Finally, in (47), the narrator and the interviewer know that the topic of the narration is a film and that films have titles, so we can identify the referent uniquely as belonging to the film being narrated.

6.7. The “type identifiable” (TYP) cognitive status

We had hypothesized that a speaker will use an indefinite NP (with or without a determiner) to refer to an entity when the addressee can be assumed to understand the type of thing that the expression describes. As Table 8 demonstrates a “type identifiable” cognitive status is a good indicator that an indefinite expression will be used across all groups although it is not selected 100% of the time, even among NSs. Results of chi-square and Fisher’s Exact tests, calculated only with DefNP and IndNP categories for learners combined and NSs, and the logistic regression ($p=0.4832$) for all groups revealed no significant differences among levels.

Table 8. Referring expressions for cognitive status: TYP (expect IndNP)

Frequency (Percent)	Group 1 Beginners	Group 2 Intermediate	Group 3 Advanced	Group 4 NS
Overt	0	0	0	1 (2)
GenNP	0	2 (9)	0	0
DefNP	3 (11)	2 (9)	5 (23)	3 (7)
IndNP	25 (89)	19 (82)	17 (77)	38 (91)
Total	28	23	22	42

$df=1$, $X^2=1.1602$, $P=0.2814$; Fisher's Exact test $p=0.3674$

The learners' use of a DefNP when the entity has not been introduced in the discourse or is uniquely identifiable runs counter to our hypothesis. In (48), the learner uses a DefNP to introduce the flirty woman in the first clause but an IndNP would be expected here; she has not been introduced into the discourse previously. In addition, although no police (i.e. no group of policemen) has been introduced into the narrative, the learner refers to the entity with a DefNP in the second clause of (49). The policeman that arrives on the scene in the film is always introduced by the NS participants with an indefinite determiner, as shown in (50). When the NSs selected a definite expression to introduce a type identifiable referent, it was of the sort illustrated in (51) below, where the NP refers to a generic entity, "the ideal woman", rather than a particular, uniquely identifiable individual.

(48) y luego **la coqueta**_i camina enfrente de él_j y ella_i coqueta con el hombre_j. (G2, P8)

(49) está nadando_i en el agua y **la policía**_j viene (G3, P3)

(50) entonces llega **un policía** por atras de Charlie Chaplin (G4, P10)

(51) y bueno, aparece **la mujer ideal**_i para eso, aparece una chica_i con una actitud (G4, P6)

In sum, we found that the cognitive status of the subject referent in many cases determines which subject form will be used and this constraint was strongest at the extremes of our 'revised' Givenness Hierarchy. For the INF status, there was a clear tendency for all groups to use null subject pronouns although that tendency was greatest for NSs. At the other end of the Hierarchy, for the TYP status, we also observed a tendency among all groups to select an indefinite NP, although again for the NSs the tendency was strongest. On the right end of the Hierarchy, for the UNI and FAM statuses, we observe differences between NSs and learners; although all groups tend towards the preference for definite NPs for both statuses as predicted by our hypotheses, the differences here are reflected in the NSs using a wider range of subject forms, a finding that confirms the Geeslin and Gudmestad (2008) study for third-person subjects. It was in the middle of the Hierarchy that we saw the greatest differences between learners and NSs and among the three learner groups. For the ACTR status we observed a tendency for advanced learners and NSs to prefer null followed by overt pronouns; the intermediate group preferred overt pronouns and beginners selected proper names in most of the cases. The data reveal that this is perhaps the most difficult cognitive status for learners; they consistently do not select the subject form that NSs would select in most cases. For the ACT status, all learner groups performed in a similar fashion and preferred definite NPs whereas NSs selected more overt pronouns. More differences between the learner groups and the NSs were seen in the selection of subject forms for the ACTNR status; NSs chose general NPs and the learners (except for the intermediates) selected proper names.

For all cognitive statuses, with the exception of TYP (and to a lesser extent INF), there are significant differences in the choice of subject form between learners and NSs. In every case, NSs are more likely than learners to use a "less specific" form, that is, a null or overt pronoun as opposed to a definite NP or proper name.

7. Conclusions

It is clear from the data presented here that subject expression for both NSs and learners at all levels is constrained by the cognitive status of discourse entities. The NSs demonstrate a strong

tendency to use the most minimal form possible, e.g. a null subject to express coreference with the most salient NP in the local discourse (i.e. the referent “in focus”). This tendency was also observed by Blackwell (2003) in her study of anaphoric reference in both elicited film narratives and conversational discourse in Spanish, in which speakers’ minimization in their use of referring expressions was viewed as an instantiation of Gricean principles of informativeness and economy (e.g. say no more than you must, i.e. choose the most minimal form possible to express anaphoric reference). In the present study, the use of minimal (lexically and phonologically reduced) referring expressions was seen to be promoted by other, disambiguating factors besides the cognitive status of the referent, such as agreement between an anaphoric expression and its antecedent and assumptions of mutual knowledge of the previous discourse. Such additional information makes it possible for speakers to use minimally informative null subjects to refer to entities that are not “in focus” but are activated in the immediate discourse context. However, more informative, less minimal expressions (pronouns and reduced NPs such as *el otro*) are chosen for reference to entities that are activated but in contexts where a null subject would be referentially ambiguous.

As learners’ proficiency develops, so too does their sensitivity to the cognitive status of referents introduced into the discourse. Whereas learners choose more elaborate forms, native speakers choose more minimal forms; as proficiency develops, learners gradually replace full NPs with pronominal forms and overts with nulls. L2 subject referential marking is a process of replacing fuller NPs with more minimal ones within the cognitive status hierarchy.

While we find strong evidence that the cognitive status of subject referents significantly conditions the choice of third-person subject forms in our data, we do not intend to suggest that we can always predict which form will be used, especially in conversation, or that with respect to subject referents, it is the only conditioning factor that comes into play. Indeed other studies have shown that discourse genre, switch/same reference, tense, aspect and mood, person and number, etc. all play a key role. Future research will need to look at the interaction of these factors and cognitive status, and verb semantics in order to have a clearer picture of the complexity of subject expression choices made by both native speakers and learners in several types of discourse.

References

- Al-Kasey, Tamara & Pérez-Leroux, Ana Teresa. (1998). Second language acquisition of Spanish null subjects. In Suzanne Flynn, Gita Martohardjono, & Wayne A. O’Neil (Eds.), *The generative study of second language acquisition* (pp. 161-183). Hillsdale, NJ: Lawrence Erlbaum.
- Ariel, Mira. (1988). Referring and accessibility. *Journal of Linguistics*, 24, 65-87.
- Ariel, Mira. (1990). *Accessing noun-phrase antecedents*. New York: Routledge.
- Bentivoglio, Paola. (1983). Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish. In Talmy Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 255-311). Amsterdam: John Benjamins.
- Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Blackwell, Sarah E. (2003). *Implicatures in discourse: The case of Spanish NP anaphora*. Amsterdam: John Benjamins.
- Cameron, Richard. (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Doctoral dissertation, University of Pennsylvania.
- Cameron, Richard. (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics*, 6/7, 1-27.
- Cameron, Richard & Flores-Ferrán, Nydia. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context*, 1, 41-65.
- Carminati, Maria Nella. (2002). *The processing of Italian subject pronouns*. Doctoral dissertation, University of Massachusetts, Amherst.
- Chafe, Wallace L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (Ed.), *Subject and topic* (pp. 25-55). New York: Academic Press.
- Chafe, Wallace L. (1987). Cognitive constraints on information flow. In Russell S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21-51). Amsterdam: John Benjamins.

- Clark, Herbert H. & Haviland, Susan E. (1977). Comprehension and the given-new contract. In Roy O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum.
- Davidson, Brad. (1996). Pragmatic 'weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish. *Journal of Pragmatics*, 26, 543-565.
- Enríquez, Emilia V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: CSIC.
- Filiaci, Francesca. (2010). Null and overt subject biases in Spanish and Italian: A cross-linguistic comparison. In C. Borgonovo, M. Español-Echevarría & P. Prévost (Eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium* (pp. 171-182). Somerville, MA: Cascadilla Proceedings Project.
- Geeslin, Kimberly & Gudmestad, Aarnes. (2008). Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In Melissa Bowles, Rebecca Foote, Silvia Perpiñán & Rakesh Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 69-85). Somerville, MA: Cascadilla Proceedings Project.
- Givón, Talmy. (1983). Topic continuity in discourse: An introduction. In Talmy Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1-41). Amsterdam: John Benjamins.
- Grosz, Barbara. J. (1981). Focusing and description in natural language dialogues. In Aravind K. Joshi, Bonnie L. Webber & Ivan A. Sag (Eds.), *Elements of discourse understanding* (pp. 84-105). Cambridge: Cambridge University Press.
- Gundel, Jeanette K. (1996). Relevance theory meets the givenness hierarchy: An account of inferrables. In Thorstein Fretheim & Jeanette K. Gundel (Eds.), *Reference and referent accessibility* (pp. 141-153). Amsterdam: John Benjamins.
- Gundel, Jeanette K., Bassene, Mamadou, Gordon, Bryan, Humnick, Linda, & Khalfaoui, Amel. (2010). Testing predictions of the givenness hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics*, 42, 1770-1785.
- Gundel, Jeanette K., Hedberg, Nancy, & Zacharski, Ron. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274-307.
- Gundel, Jeanette K., Hegarty, Michael, & Borthen, Kaja. (2003). Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information*, 12, 281-299.
- Hedberg, Nancy & Zacharski, Ron. (2007). Introduction. In Nancy Hedberg & Ron Zacharski (Eds.), *The grammar-pragmatics interface: Essays in honor of Jeanette K. Gundel* (pp. 1-15). Amsterdam: John Benjamins.
- Huang, Yan. (1994). *The syntax and pragmatics of anaphora: A study with special reference to Chinese*. Cambridge: Cambridge University Press.
- Huang, Yan. (2000a). Discourse anaphora: Four theoretical models. *Journal of Pragmatics*, 32, 151-176.
- Huang, Yan. (2000b). *Anaphora: A cross-linguistic study*. Oxford: Oxford University Press.
- Isabelli, Casilde A. (2003). The acquisition of null subject parameter properties in SLA: Some effects of positive evidence in a natural learning context. *Hispania* 87, 150-162.
- Levinson, Stephen C. (1987). Minimization and conversational inference. In Jef Verschueren & Marcella Bertuccelli-Papi (Eds.), *The pragmatic perspective* (pp. 61-129). Amsterdam: John Benjamins.
- Luján, Marta. (1985). Binding properties of overt pronouns in null pronominal languages. In William H. Eilfort, Paul D. Kroeber, & Karen L. Peterson (Eds.), *Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistic Society* (pp. 424-438). Chicago: CLS.
- Luján, Marta. (1986). Stress and binding of pronouns. In Anne M. Farley, Peter T. Farley & Karl-Erik McCullough (Eds.), *Papers from the Parasession on Pragmatics and Grammatical Theory at the Twenty-Second Regional Meeting of the Chicago Linguistic Society* (pp. 248-262). Chicago: CLS.
- Luján, Marta. (1999). Expresión y omisión del pronombre personal. In Ignacio Bosque & Violeta Demonte (Eds.), *Gramática descriptiva de la lengua española, Vol. I* (pp. 1275-1315). Madrid: Espasa Calpe.
- Liceras, Juana M. (1989). On some properties of the pro-drop parameter: Looking for missing subjects in non-native Spanish. In Susan Gass & Jacquelyn Schacter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 109-133). Dordrecht: Foris.
- Liceras, Juana M. & Díaz, Lourdes. (1998). On the nature of the relationship between morphology and syntax: Inflectional typology, *f*-Features and null/overt pronouns in Spanish interlanguage. In Maria-Luise Beck

- (Ed.), *Morphology and its interfaces in second language knowledge* (pp. 307-338). Amsterdam: John Benjamins.
- Liceras, Juana M. & Díaz, Lourdes. (1999). Topic-drop versus pro-drop: Null subjects and pronominal subjects in the Spanish L2 of Chinese, English, French, German, Japanese and Korean speakers. *Second Language Research*, 15, 1-40.
- Lozano, Cristóbal. (2002). Knowledge of expletive and pronominal subjects by learners of Spanish. *ITL Review of Applied Linguistics*, 135/6, 37-60.
- Martín Rojo, Luisa & Meeuwis, Michael. (1993). Referentes del sujeto pronominales y tácitos en la conversación en español: Un enfoque pragmático. In Henk Haverkate, Kees Hengeveld & Gijs Mulder (Eds.), *Aproximaciones pragmalingüísticas al español 12* (pp. 87-118). Amsterdam: Editorial Rodopi.
- Mayol, Laia. (2010). Contrastive pronouns in null-subject Romance languages. *Lingua*, 120, 2497-2514.
- Michaelis, Laura & Francis, Hartwell. (2007). Lexical subjects and the conflation strategy. In Nancy Hedberg & Ron Zacharski (Eds.), *The grammar-pragmatics interface* (pp. 19-48). Amsterdam: John Benjamins.
- Montalbetti, Mario M. (1984). *After binding*. Doctoral dissertation. MIT.
- Morales, Amparo. (1997). La hipótesis funcional y la aparición de sujeto no nominal: El español de Puerto Rico. *Hispania*, 80, 153-65.
- Pérez-Leroux, Ana Teresa & Glass, William. R. (1997). OPC effects on the L2 acquisition of Spanish. In Ana Teresa Pérez-Leroux & William R. Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish, Vol. 1: Developing grammars* (pp. 149-165). Somerville, MA: Cascadilla Press.
- Pérez-Leroux, Ana Teresa & Glass, William. R. (1999). Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research*, 15, 220-249.
- Posio, Pekka. (2011). Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics*, 43, 777-798.
- Prince, Ellen F. (1981). Toward a taxonomy of given-new information. In Peter Cole (Ed.), *Radical Pragmatics* (pp. 223-255). New York: Academic Press.
- Quesada, Margaret Lubbers & Blackwell, Sarah E. (2009). The L2 acquisition of null and overt Spanish subject pronouns: A pragmatic approach. In Joseph Collentine, Maryellen Garcia, Barbara Lafford & Francisco Marcos Martín (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp.117-130). Somerville, MA: Cascadilla Proceedings Project.
- Ranson, Diana L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change*, 3, 133-152.
- Reinhart, Tanya. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27, 59-94.
- Reinhart, Tanya. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
- Rothman, Jason. (2009). Pragmatic deficits with syntactic consequences?: L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics*, 41, 951-973.
- Silva-Corvalán, Carmen. (1982). Subject variation in spoken Mexican-American Spanish. In Jon Amastae & Lucía Elías-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93-120). Cambridge: Cambridge University Press.
- Taboada, Maite & Wiesemann, Loreley. (2010). Subjects and topics in conversation. *Journal of Pragmatics*, 42, 1816-1828.
- Travis, Catherine E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19, 101-135.

Selected Proceedings of the 14th Hispanic Linguistics Symposium

edited by Kimberly Geeslin
and Manuel Díaz-Campos

Cascadilla Proceedings Project Somerville, MA 2012

Copyright information

Selected Proceedings of the 14th Hispanic Linguistics Symposium
© 2012 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-450-8 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Blackwell, Sarah E. and Margaret Lubbers Quesada. 2012. Third-Person Subjects in Native Speakers' and L2 Learners' Narratives: Testing (and Revising) the Givenness Hierarchy for Spanish. In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, ed. Kimberly Geeslin and Manuel Díaz-Campos, 142-164. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2662.