1 Introduction

In this chapter we present a perspective that places language use in discourse at the forefront of syntax. The central postulate of USAGE-BASED theory is that the basis for grammatical knowledge is speakers’ linguistic experience – the frequency and contexts of use of forms (Bybee 2010), in contradistinction to the view that language use and knowledge about language (or “performance” and “competence” cf. Chomsky 1965) are independent. These two perspectives on syntax have been broadly characterized as FUNCTIONAL (according to which grammatical structures conventionalize out of discourse patterns) and FORMAL (according to which explanations of linguistic structure are based on grammatical form, independent of function) (cf. Croft (1995) vs. Newmeyer (1998)).

The discourse-based perspective on syntax sees the observation of spontaneous language use, rather than reflection about language use, as the appropriate way to obtain data for linguistic analysis. Conversation receives particular primacy in this area, as the form of language that we spend most of our time using, that which is acquired first and without any formal instruction, and that which provides the most systematic data for the analysis of grammar as it is produced in real time (Labov 1984; Chafe 1994; Ono and Thompson 1995; Briz 1998; Du Bois 2003; Schegloff 2007). Community-based research has further highlighted the need to discover patterns in sociolinguistically-sampled speech corpora rather than collections of individual data (e.g., Labov 1984).

Functionalist analysis has given rise to a questioning of some basic formalist assumptions, such as the discreteness of categories, the notion that syntax, the lexicon, and phonology are separate components, and that grammar consists of generalizations made about abstract features. From a functionalist perspective, the following claims are made: linguistic categories (e.g., word classes, transitivity, subordination, etc.) are continua, with a prototype or exemplar structure; grammatical constructions combine syntactic, semantic, and phonological information;
and grammatical generalizations derive “from the repetition of many local events” in discourse (Bybee 200: 714) (see Hopper and Thompson 1980, 1984; Du Bios 1987; Hopper 1998; Croft and Cruse 2004; Goldberg 2006; Langacker 2008).

Here we present an overview of work in Hispanic Linguistics that approaches syntactic patterning from the perspective of discourse function, with a particular focus on community-based studies of spoken corpora. We begin Section 2 with information flow considerations in the realization and distribution of NPs. Next, in Section 3, we discuss transitivity as a scalar phenomenon. Nonreferential uses of lexical Noun Phrases in discourse are reviewed in Section 4. In Section 5, we consider constructions as the basic unit of grammar, including both particular and general expressions. Finally, in Section 6, we present variable first-person singular subject expression as a case study of discourse effects in grammar.

2 Information flow

One of the richest areas of study within Hispanic Linguistics from the functionalist approach has been that of Noun Phrase realization and syntactic role. Such work has shown that the form of the subject (unexpressed/null, pronoun or full lexical NP), the syntactic role an NP takes (subject, direct object), and the positioning of NPs in relation to the verb are driven by several issues. These include the referents’ relationship with the preceding discourse (whether they are opening a new sequence or continuing an old one), their topicality and, especially, information flow properties, as explained below (e.g. Bentivoglio 1983; Bolinger 1991; Myhill 1992; Ocampo 1992, 1993; Silva-Corvalán 1994; Morris 1998; López-Meiramás 2006; García Salido 2008).

A wide body of research on Spanish and other languages has demonstrated that the information flow properties of the referent affect the way it is linguistically expressed. Specifically, a referent that is being introduced to the discourse for the first time is more likely to be expressed with “lexically heavy” material (full lexical NPs) than a referent that has already been talked about, which tends to be realized with less material (pronouns, unexpressed mentions) (e.g., Prince 1981; Givón 1983; Chafe 1994; Bentivoglio and Sedano 2007. This is illustrated in the following example, where Don Julio Restrepo is introduced into the discourse with a proper name, then referred to by a demonstrative pronoun, then a personal pronoun, followed by an unexpressed mention (represented by Ø) (see Dumont 2006: 290).

(Notice also a similar pattern for the mentions of mi tía Juanita.)

(1) C: Recuerdas los nombres de otros o otras curanderas, o curanderos.
V: Don Julio Restrepo curaba de brujería.
Ese era de Santa Cruz.
Ese curaba de brujería una vez.
a mi tía Juanita le hicieron mal,
y él vino y Ø la curó y Ø sanó.
de brujería.

[NMCOSS: 4]
C: ‘Do you remember the names of other (female) healers, or (male) healers.’

V: ‘Mr Julio Restrepo cured (people) from witchcraft. He was from Santa Cruz. He cured of witchcraft once. they put a spell on my Aunt Juanita, and he came and (he) cured her and (she) got better, of witchcraft.’

It is not just syntactic form, but also syntactic role that interacts with discourse function. Specifically, NPs introducing referents into the discourse for the first time tend not to occur as subjects of transitive clauses, and instead speakers prefer the roles of subjects of intransitive clauses or objects of transitive clauses. This pattern has been dubbed Preferred Argument Structure, and its attestation in a wide range of typologically diverse languages questions the notion of “subject” and “object” as categories even for traditionally nominative/accusative languages (see Du Bois 2003 for a review).

In Spanish, the tendency has been extensively tested, and found to hold (Bentivoglio 1993, Bentivoglio and Ashby 2003). In an analysis of the distribution of full NPs, pronouns, verbal clitics, and unexpressed mentions extracted from sociolinguistic interviews of speakers in Caracas, Venezuela, Ashby and Bentivoglio (1993: 65) found that while full NPs occurred much less frequently than pronominal and unexpressed NPs overall (representing just 28% of the total number of NPs in the data, 591/2,121); their distribution was skewed along the lines predicted by Preferred Argument Structure. Full NPs occurred proportionately less as subjects in transitive clauses (6% vs. 94% pronominals and unexpressed mentions), followed by subjects in intransitive clauses (22%), and only for objects in transitive clauses did they reach a majority (60%). A similar skewing was found for information status: approximately one-third (200/591) of all full NPs expressed “brand new information” (that is, following Prince (1981), those referents that had not occurred in the preceding discourse nor were they identifiable through a frame or schema), and this was distributed differently across the syntactic roles: 42% of full NP direct objects expressed brand new information compared with 26% of full NP subjects of intransitive clauses and just 6% of full NP subjects of transitive clauses (1993: 71). This pattern can be seen in example (2) below (cf. Dumont 2006: 288). The referents of the underlined full NPs are all introduced into the discourse for the first time, el marido ‘my husband’ and un muchito ‘a little boy’ as subjects of intransitive clauses, and su grandma ‘his grandmother’ as the direct object of the verb tener ‘have.’ Note that while all three referents are introduced into the discourse for the first time, un muchito is unidentifiable, or brand new, while el marido can be considered identifiable through a familial relationship with the speaker, and su grandma identifiable through a familial relationship with un muchito, who was introduced in the previous clause (cf. Thompson 1997). This is reflected in their morphosyntax, with the speaker expressing the identifiable referents in the form of definite NPs in their canonical positions (preverbal for subjects and postverbal for objects),
and the unidentifiable referent as an indefinite NP in postverbal noncanonical subject position.

(2) Ahí iba yo sola.
porque el marido en veces estaba trabajando.
y en ve-
luego,
me acompañaba un muchito.

que tenía su abuela allá en Park View. [NMC02: 8]

'I would go there alone,
because my husband was sometimes working.

and some- 
later,
a little boy would accompany me.
who had his grandmother in Park View.'

Thus, here we see patterning in grammatical structures (NP realization, distribution across syntactic roles, and word order) with a parallel patterning in discourse function (information status of the referents), an illustration of how grammar is shaped by language use, or the way in which "grammars code best what speakers do most" (Du Bois 2003: 49).

3 Transitivity

The distribution of new information in the clause challenges not only the categories of "subject" and "object", but also the traditional distinction between "transitive" and "intransitive" verb classes. Ashby and Bentivoglio (1993), for example, observe distinct patterning with different kinds of verbs within each class. In particular, they find that the copulas ser and estar pattern more similarly to transitive verbs in terms of their discourse pragmatics in that they disfavor new information more than do other intransitive verbs (though less than transitive verbs as a class) (1993: 71). Furthermore, they find that Spanish speakers use the single argument of haber and the direct object of tener (as in (2) above) in existential-presentational constructions to introduce new information significantly more than the other roles favored for the introduction of new information, subjects of other intransitive clauses, and objects of transitive clauses (1997: 22). We find in operation, then, a local strategy that functions at the level of constructions with specific verbs, regardless of the class they pertain to, thus giving cause to question the discreteness of transitivity (see also Section 5).

Hopper and Thompson’s (1980) work presents cross-linguistic evidence to challenge this widely accepted dichotomy, proposing that transitivity is not based on the mere presence of a direct object, but is manifested to different degrees depending on the “effectiveness” of the action, as reflected in components (or parameters) such as aspectual features of the verb, agentivity and volitionality
of the subject, and affectedness and individuation of the object. Hopper and Thompson (1980) hypothesize that these components of scalar transitivity covary such that features indicative of higher transitivity tend to co-occur in the clause. Testing this hypothesis in Spanish, Vázquez Rozas (2006: 101–108) argues that verbs like *atraer* ‘attract,’ *fascinar* ‘fascinate,’ and *interesar* ‘interest’ (which may appear either with a direct object or as *gustar* ‘please’-type verbs with an indirect object experiencer) tend to be used in the higher transitivity direct object construction when the object is physically affected, the subject agentive, and the situation has an inherent endpoint (i.e., is telic), but in the lower transitivity indirect object construction with aspectually stative situations, an inanimate subject, and an object that is affected psychologically. In the following pair of examples, the direct object experiencer construction in (3) involves an activity of the stimulus participant (the growls of the lion), while the indirect object experiencer construction in (4) involves a stative property of the stimulus (the leather binding of the book).

(3) Los rugidos del león *atrajeron* al cazador, quien [. . .] lo metió en una gran jaula.

‘The growls of the lion attracted the hunter, who [. . .] put it in a big cage.’

(Vázquez Rozas 2006: 104)

(4) El libro estaba encuadernado en piel [. . .], pero a ella no le *atraía*.

‘The book was bound in leather [. . .], but it didn’t attract her.’

(Vázquez Rozas 2006: 104)

A further challenge that presents itself to the traditional understanding of transitivity concerns how to deal with different kinds of “objects”. “Objects” with low-content verbs, often in conventionalized combinations (cf. Ashby and Bentivoglio 1993: 67–68; Chafe 1994: 111–113), may actually form unitary predicates, or “V-O compounds”, for example, *have fun, make no sense, get sleep* (Thompson and Hopper 2001: 37). Examples from Spanish are given in (5), (6) (Torres Cacoullos and Aaron 2003: 321–322), and (7) (Burgos 2007: 29). Note that in (7) the predicate is jointly produced by the two interlocutors, evidence that this is a conventionalized schema which they share (cf. Lerner 2002).

(5) hago escultura. [NMCOSS: 102]

‘I do sculpture.’

(6) Y le puse *complaint* a ese chota. [NMCOSS: 219]

‘and I filed a complaint against that cop’

(7) P: no me gusta mucho trabajar,
porque le sacan --
A: Ay,
la paciencia. [CCS: ESP]
P: ‘I don’t like much to work, because they try --’
A: ‘Oh, your patience.’

Sentential objects may also be of issue for transitivity. Rather than viewing the
material following creo ‘I think’ in (8) below as a subordinate clause (specifically an
object complement), Thompson (2002) argues that apparent biclausal combinations
often behave as single clauses which include a speaker-stance discourse formula.
That is, certain apparent complement-taking predicates such as yo creo and yo (no) sé
(see Section 6 below) may serve as frames of speaker stance toward the content of
their putative subordinate clause, which actually carries out “the work that the
utterance is doing” (Thompson 2002: 155). Evidence comes from their distributions
as well as pragmatics; for example, (yo) creo ‘I think’ shows the same kind of
variability in grammatical and prosodic position as epistemic/evidential adverbial
expressions such as a lo mejor ‘probably’ by being variable in position, occurring
preceding in (8), following in (9), or even within, the material it marks. It is also
often prosodically independent, i.e., occurring in an Intonation Unit of its own, as in
(9) (cf. Travis 2006). (Following the Du Bois et al. (1993) method, Intonation Units
are represented on distinct lines in the transcription; see Ono and Thompson 1995;
Sánchez-Ayala 2001; Shenk 2006, on the use of the Intonation Unit as a meaningful
unit for syntactic analysis.)

(8) Yo,
     Creo que estoy --
     casi que seguro que puede ser eso. [CCS: breakfast]
     I,
     think that I’m --
     almost sure that it might be that.’

(9) Él es de --
     Friburgo,
     creo. [CCS: Tumaco]
     He is from --
     Freiburg,
     I think.’

In sum, this body of research has shown that transitivity is usefully viewed as a
continuum which is subject to local effects of frequently used discourse
patterns. These considerations reinforce our understanding of the discourse
functions of transitivity. Namely, speakers use higher transitivity for fore-
grounded material, especially in a narrative (Hopper and Thompson 1980:
280–294), and lower transitivity as a reflection of subjective “talk about how
things are from our perspective”, especially in conversation (Thompson and
4 Referentiality in discourse

Another illustration of grammatical patterns driven by discourse function is found in the area of referentiality. Discussions of referentiality often conflate specificity and discourse referentiality, which are correlated but not coterminous. On the one hand, specificity has to do with the way an NP refers to entities. Specific (particular) NPs refer to unique people or things that are not considered to be interchangeable (10), while nonspecific (general) NPs indicate any member of a class of entities (11) (cf. Ashby and Bentivoglio 1993: 69–70). A third type is generic NPs, which refer to an entire class of entities (12) (Torres Cacoullos and Aaron 2003: 308). In (New Mexican) Spanish this functional distinction is grammatically encoded: nonspecific NPs are more likely than specific NPs to appear bare (without a determiner) but, unlike in English, generic NPs line up with specific NPs in disfavoring bare forms.

(10) Y luego una vez que se me % --
  .. quemó el generador de la troca mía,
  'and then once when --
  the generator of my truck burnt out,'

(11) Lo que me ayudó a mí,
  que yo arrió quince años troca.
  'what helped me,
  Is that I drove trucks for fifteen years.'

(12) Pero el nombre que la gente usaba aquí pa’ los --
  pa’ los buzzards,
  ... I can’t remember.
  ‘But the name that people used here for --
  for buzzards,
  I can’t remember.’

Discourse referentiality, on the other hand, has to do with tracking participants. Discourse referential (Tracking) NPs are used “to speak about an object as an object, with continuous identity” in the discourse (Du Bois 1980: 208). In Hopper and Thompson’s terms, Tracking NPs are used for “manipulable” discourse participants, people, or things about which something is said (1984: 711). In contrast, non-Tracking NPs serve to classify (as predicate nominals) or to orient (as adverbials, giving the location or time of an entity or situation) (Thompson 1997: 69). For example, in (13) the unexpressed subject is a discourse-referential NP but the predicate nominal train master is nonreferential, serving to characterize the subject. The proper noun Corpus Christi in the en locative prepositional phrase is also used nonreferentially, serving to orient the predication rather than to introduce the city in order to talk about it.
es train master en Corpus Christi.  
‘He’s a train master in Corpus Christi.’

A third kind of non-Tracking NP concerns lexical NPs used in the “V-O compounds” discussed in Section 3. The underlined NPs in examples in (5), (6), (7), and also (11) are not used referentially as arguments (objects) but rather as “Predicating NPs [which] function as part of naming a type of event, activity, or situation” (Thompson 1997: 71–72): ‘sculpturing,’ ‘accusing,’ ‘trying (one’s) patience,’ and ‘truck driving.’

Applying this discourse-based understanding of referentiality to English-origin nouns that are not established loanwords in New Mexican Spanish, Torres Cacoullos and Aaron (2003) are able to account for a slightly higher rate of bare (determinerless) forms in English-origin nouns than in native Spanish ones (36%, 98/270 vs. 30%, 413/1,386). These scholars show that such nouns are English in origin only, following Spanish grammatical patterns for predicate nominals in a classifying function, especially ones designating occupations, as in (13), and for Predicating NPs in combinations with low content verbs such as hacer ‘do’ and poner ‘put,’ seen in (5) and (6) above. The higher rate of bare forms is thus due not to lack of grammatical integration but to the use by bilinguals of English-origin nouns as non-Tracking NPs.

In sum, studies have profited from distinguishing discourse referentiality and specificity in characterizing NPs. Specificity in particular has been found to affect the form of objects (e.g., García and Putte 1989; Schwenter 2006; Reig Alamillo 2009; Balasch (to appear). This is evidence, once again, of the role of discourse considerations in linguistic structures.

5 Constructions and prefabs

Constructions have been viewed as the units of grammar on the basis of observations about the idiomaticity of language (Fillmore et al. 1988), the interaction of syntax and lexicon (Langacker 2008), cross-linguistic variation (Croft 2001), language acquisition and learning (Tomasello 2003; Goldberg 2006), and grammaticalization (Bybee and Torres Cacoullos 2009). Constructions are conventionalized form-function pairings, ranging from fixed expressions (including lexical items, idioms and prefabs: see below) to productive morphosyntactic structures that are at least partially schematic (for an overview, see Croft and Cruse 2004: 225–290; Goldberg 2006: 3–17). For example, the idiom subir a la cabeza ‘get intoxicated’ (lit. ‘go to your head’) would be considered a construction, as would the more schematic subir ‘go up’ STAIRS (ex. escaleras, peldaños) and subirse ‘climb’ TREE (ex. higuera, mezquite). Torres Cacoullos and Schwenter (2008) present evidence for constructional status in terms of the differential patterning with (middle) se: in their data subir was overwhelmingly not marked with se for stairs (3/66), but was se-marked for trees (9/9) (cf Maldonado 1999; Aaron 2004; Clements 2006 for other functionalist accounts of se).
Social and pragmatic use is included as part of the meaning of a construction. Consider, for example, the variation between *estar* verb-*ndo* and *andar* verb-*ndo* as expressions of progressive and other imperfective aspects in Mexican varieties, as illustrated in (14) and (15).

(14) estás hablando de una forma de vida
‘you are talking about a lifestyle’

(Lope Blanch 1971: 261)

(15) ando buscando unas tijeras
‘I am looking for some scissors’

(Lope Blanch 1971: 415)

Torres Cacoullos (2001) shows that this variation does not reflect aspectual differences, as is often assumed, but collocational routines: *estar hablando* ‘speaking’ is the conventional way to express ‘be speaking’, and *andar buscando* the conventional way to ‘be looking for something.’ Auxiliary *estar* is generally more likely to co-occur with verbs of speech (*platicando* ‘chatting’), as well as verbs denoting perceptible bodily activities (*llorando* ‘crying’) and mental activities (*pensando* ‘thinking’), whereas auxiliary *andar* is more likely with verbs denoting motion such as *dando la vuelta* ‘going around’ and physical activities, particularly outdoor ones, such as *trabajando* ‘working’ (*en las pizcas* ‘in the harvest,’ *en el campo* ‘in the country’). In tracing the evolution (grammaticalization) of the two constructions, Torres Cacoullos (2001) suggests that this distribution is the residue of source-construction meanings, ‘be located (stationary)’ for *estar* and ‘go around’ for *andar*.

Moreover, the *andar* verb-*ndo* construction has social meaning in that its relative frequency is higher in popular than educated speech. This social meaning is related to the lexical associations observed, since rural activities in large spaces are more compatible with ‘going around’. Thus, the educated/popular social stratification of *estar/andar* verb-*ndo* may derive from particular instances of use, developing from an indoor–outdoor and urban–rural difference.

Collocational patterns, or lexical associations, are also shown in the combinations of adjectives with verbs of becoming (change of state) such as *quedarse*, *ponerse* and *volverse*, all meaning something similar to English ‘become.’ Studies attempting to establish generalizations about abstract features (such as the duration of the state or the degree of involvement of the subject) have failed to predict which adjectives go with which verbs. Bybee and Eddington (2006) argue that speakers’ choices of verb–adjective combinations are shaped by more local exemplar-based categorization. For example, they propose several categories of adjectives that occur with *quedarse* ‘become (lit. remain + reflexive).’ Importantly, these categories are based on clusters of adjectives which are semantically related to particular lexical types that occur frequently with this same verb in oral and written corpora. For instance, with high frequency *quedarse solo* ‘end up alone’ are grouped *soltera* ‘single,’ *aislado* ‘isolated,’ and other semantically related adjectives, whereas *parado* ‘stopped’ and *tieso* ‘stiff,’ for example, are grouped with high frequency *quedarse inmóvil* ‘immobile.’ In acceptability judgments, Bybee and Eddington found that high frequency verb–adjective combinations were ranked as more acceptable than those
with low corpus frequency. Furthermore, of the low-frequency items, those that were semantically related to a high-frequency phrase (e.g., quedarse parado ‘be stopped,’ related to quedarse inmóvil) were rated as more acceptable than those that had no such semantic relation (e.g., quedarse orgullosismo ‘be proud’). Thus, the most frequent quedarse ADJECTIVE collocations form the basis for different groups of adjectives through analogical comparison. These results suggest that highly frequent central members of a category attract like members, such that category formation is based on local exemplars.

Usage frequencies affect the formation of constructions, as frequent items may become automatized as single units (Bybee 2003) (cf. Company Company 2006). For example, the collocation yo no sé ‘I don’t know’ (cf. Travis 2006) need not be produced by speakers through the application of grammatical rules to the individual items that make it up, but rather may be produced (and stored) as a conventionalized single unit. Evidence is found in its entrenchment or relative fixedness in form (e.g., the limited variability in subject expression, with a strong tendency for yo to be expressed), and its autonomy from other uses of the verb saber as seen in its semantics (e.g., yo no sé does not necessarily refer to lack of knowledge, but is also used as a general mitigating device). Constructions that occur with high (token and relative) frequency with particular lexical types (such as andar buscando ‘be looking for,’ quedarse solo ‘end up alone,’ and yo no sé ‘I don’t know’) are known as prefabs or “word sequences that are conventionalized, but predictable in other ways” (Bybee 2006: 713). Thus, in usage-based theory, both idioms with an unpredictable meaning (such as subirse a la cabeza ‘get intoxicated’) and prefabs may be stored in memory as units. Should prefabs be relegated to a lexical component that is viewed as separate from a syntactic component as in the formalist perspective, or should prefabs be considered part of grammar? From a usage-based functionalist perspective, the lexically particular and the grammatically general interact, existing as points along a continuum. Diachronic evidence for the interaction of the particular and the general can be seen in the evolution of the estar VERB -ndo Progressive construction. Bybee and Torres Cacoullos (2009) find that the prefab estar hablando ‘be talking’ takes a lead in the semantic-structural changes observed in the grammaticalization of this construction, and may serve as the center of a verbs-of-speech subclass of estar + VERB -ndo, attracting more lexical types and thereby contributing to the productivity of the general Progressive construction. Wilson (2009) shows similarly that quedarse ‘become’ prefabs have persisted over the centuries, with gradient categories of adjectives clustering around them.

The role of prefabs in language variation and change belies a lexicon–syntax dichotomy, as does gradualness in constituency change. Complex prepositions provide evidence that changes in constituency, or cohesion among units, are gradual. For example, concessive a pesar de ‘in spite of’ developed from a meaning of sorrow > opposition > contradiction, increasing in relative frequency from 2% (4/199) of all occurrences of pesar ‘sorrow, regret’ in Old Spanish to 96% (167/174) in twentieth-century data (Torres Cacoullos 2006). As the erstwhile independent item pesar was absorbed into the new unit a pesar de, it underwent
decategorization (Hopper 1991: 22), shedding its nominal trappings, as measured by the loss of plural marking and a drop in the use of determiners and adjectival modification. Rather than the abrupt reanalysis illustrated in (16), where X is at first an NP and over time also an infinitive or a finite clause, gradualness in loss of analyzability is shown by changing distribution patterns.

(16) from \([a + \text{pesar}] + [\text{de} + X]\) to \([a + \text{pesar} + \text{de}] + [X]\) \(\text{in + spite} + [\text{of} X]\) ‘in spite of X’

For example, in coordinated adnominal NPs, repetition of de ‘of’ for each adnominal NP, as in (17), suggests the relative independence of de from a + pesar; the alternative configuration, a single a pesar de with scope over two NPs, as in (18), has increased in relative frequency over time.

(17) algo de atrevido y varonil en todo el ademán, a pesar del recogimiento y de la mansedumbre clericales
   (nineteenth century, Juan Valera, Peppita Jiménez: 316)
   ‘something bold and virile in his whole look, in spite of the withdrawal and of the tameness of the cleric’

(18) olía a lavanda y espliego, pero por debajo del perfume olía como yo, la fisiología nos igualaba a pesar de los potingues y las abluciones diarias
   (twentieth century, Juan Manuel de Prada, La tempestad: 135)
   ‘she smelled of lavendar, but underneath the perfume she smelled like me, our physiologies making us equals in spite of the concoctions and the daily ablutions’

Thus, in a usage-based view, constituent structure crystallizes from the frequent co-occurrence of items (Bybee 2010). In this section we have seen the gradience of constituent structure (e.g., a pesar de), the contributions of prefabs to grammaticalization and their centrality in synchronic categories (estar hablando, quedarse solo), and the social meaning of grammatical expressions (andar + VERB -ndo). All are compatible with usage-based constructions – on a lexicon–syntax continuum – as units of grammar.

6 Exemplifying discourse syntax: variable first-person singular subject expression

The discovery of grammatical patterning in discourse has been the preoccupation of cumulative research following the Variationist method over the last four decades. Variationist research has examined the form–function asymmetry widespread in speech, that is, the variation among different forms serving generally similar discourse functions. This variation has been shown to be structured, as
observed in the linguistic conditioning of variant forms, that is, probabilistic statements about linguistic contexts that differ significantly in the relative frequencies of the forms (Labov 1969; Sankoff 1988). In the Variationist method, hypotheses about constraints on usage are operationalized based on contextual elements with which the variants co-occur and are then tested as factors in multivariate analysis. This has been profitably applied to the study of variable subject expression in Spanish (for reviews see Silva-Corvalán 2001: 154–169), revealing a range of lexical and discourse conditioning factors of the nature discussed above. Note that while distinct genres such as conversation and narrative have been found to differ in the token frequencies of subject expression, the linguistic conditioning has been shown to be similar (Travis 2007) (cf. Poplack and Tagliamonte 2001, for comparable findings in relation to other linguistic variables in African American speech).

Our own research on New Mexican Spanish (Travis 2007; Torres Cacoullos and Travis 2010a, 2010b) has identified very similar effects as studies for first-person singular expression in other varieties of Spanish. Table 30.1 shows the results of a multivariate (Variable-rule) analysis (Sankoff et al. 2005) of the contribution of environmental factors selected as significant to the choice of expressed yo in a corpus of sociolinguistic interviews of New Mexican Spanish. In the first column, the numbers represent the probability (or factor weight) that each factor contributes to the occurrence of the variant: the closer to 0, the less likely that yo will be

Table 30.1 Factors contributing to the choice of expressed yo in New Mexican Spanish speakers.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Probability</th>
<th>% yo</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous realization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressed</td>
<td>0.68</td>
<td>45%</td>
<td>968</td>
</tr>
<tr>
<td>Not expressed</td>
<td>0.39</td>
<td>20%</td>
<td>610</td>
</tr>
<tr>
<td>Semantic class of verb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td>0.70</td>
<td>51%</td>
<td>290</td>
</tr>
<tr>
<td>Other</td>
<td>0.46</td>
<td>28%</td>
<td>1543</td>
</tr>
<tr>
<td>Subject continuity (switch reference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different subject</td>
<td>0.58</td>
<td>39%</td>
<td>986</td>
</tr>
<tr>
<td>Same subject</td>
<td>0.41</td>
<td>23%</td>
<td>847</td>
</tr>
<tr>
<td>Ambiguity of verb morphology (person)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>0.61</td>
<td>37%</td>
<td>503</td>
</tr>
<tr>
<td>Not ambiguous</td>
<td>0.46</td>
<td>29%</td>
<td>1330</td>
</tr>
</tbody>
</table>

Factors not selected as significant

Clause type

Position in turn

N = 22.
expressed and the closer to 1, the more likely that it will be. The second column shows the percentages of expressed yo and the third column the total number of tokens in the environment defined by each factor. The factor groups which jointly account for the largest amount of variation in a statistically significant way are previous realization, semantic class of verb, subject continuity, and ambiguity of verb morphology (shown in bold in Table 30.1).

We begin our discussion with a very strong lexical effect observed, namely that implicated in the semantic class of verb. These results indicate that yo is most favored by psychological (0.70) as opposed to other semantic classes of verbs (0.46), a result also reported in numerous previous studies (e.g., Enríquez 1984: 240; Bentivoglio 1987: 60; Silva-Corvalán 1994: 162; Travis 2007: 115). However, close analysis of this class reveals that one half of the tokens occurring with psychological verbs (49%, 143/290) occur in just two particular discourse formulas or constructions, namely yo creo ‘I think’ and yo (no) sé ‘I (don’t) know,’ both of which show much higher rates of expression than the overall average of 32% (yo creo as high as 87%, 46/53, and yo (no) sé at 52%, 47/90). The behavior of these two highly frequent constructions seems to be pulling up the rate of expression for the class as a whole, suggesting that the effect may be one of particular constructions rather than of a general class of psychological verbs. This is not unique to cognitive verbs; García-Miguel (2005) makes a similar point regarding the high frequency of verbs of sight (in particular ver ‘see’) within the class of perception verbs and the different constructions in which they occur in Spanish written and spoken data.

Consider, next, what appears to be a morphosyntactic effect, that is, the favoring of yo expression by verb forms which are morphologically ambiguous in person marking (0.61) (for example in first- and third-person singular Imperfect forms such as yo tenía cuidado de ellos ‘I would take care of them’ in (19) below). It has been proposed that subjects are expressed with morphologically ambiguous verbs in order to resolve the ambiguity (e.g., Hochberg 1986; García Salido 2008), but it has also been found that true ambiguity is rare in natural discourse, as it is typically resolved by context (e.g., Bentivoglio 1987: 45; Silva-Corvalán 1994: 154). In relation to this controversy, it is interesting to note that this morphosyntactic feature is affected by lexical patterning: in our data, psychological verbs are rarely ambiguous (9%, 25/290, compared with 31%, 478/1,543 of other verbs), and we find no ambiguity effect for psychological verbs, with the rate of yo expression being nearly identical in ambiguous and unambiguous contexts (48%, 12/25 and 51%, 136/265, respectively). As an alternative to the notion of resolving ambiguity, Silva-Corvalán (2001: 161–163) proposes that it is the discourse function of the different tense, moods and aspects that motivates their use with expressed or unexpressed subjects, a hypothesis supported by the present data: Preterit (perfective) forms, which have a foregrounding function, show a significantly lower yo rate (24%, 114/478) than Imperfect, Conditional, and Subjunctive forms, which have a backgrounding function (37%, 187/499).

We now move on to clear discourse effects, beginning with that of subject continuity. The subject continuity, or switch reference, effect follows the predicted pattern noted above (Section 2): a subject is less likely to be expressed when it is
coreferential with the immediately preceding subject, and is more likely to be expressed when it is not. This can be seen in (19), where yo is expressed when there is a switch in reference with respect to the subject of the preceding clause (mi mamá, ellos), but is left unexpressed in the last clause where it is coreferential with the preceding subject.

(19) y mi mamá se alistaba pa' el trabajo. 
y yo tenía cuidado de ellos, 
ellos se iban a la escuela y yo me quedaba alzando la casa y, 
y, 
(Q) me quedaba en la casa con los niños chiquitos. [NMC OSS: 117] 'and my mom would get ready for work. 
and I would take care of them, 
they would go to school and I would stay maintaining the house and, 
and, 
(I) would stay at home with the little kids.'

Finally, the results for previous realization exhibit another discourse effect. What we observe here is that yo is more likely to be expressed when the preceding coreferential subject is also expressed (factor weight 0.68) than when it is unexpressed (0.39), a constraint which has been observed for subject expression in several dialects of Spanish (Cameron 1994; Flores-Ferrán 2002; Travis 2007). This can be seen in example (20) below, where we have strings of expressed subjects in (a–c) and strings of unexpressed subjects in (d–f).

(Speech produced in English is in small caps in both the original and the translation).

(20) a. O: yo les planchaba porque en ese tiempo no había AUTOMATIC WASHERS, okay. 
A: Hm. 
@@ < @ No @ >.
O: so=,
b. Yo tenía una RINGER TYPE MACHINE, 
c. so yo me les vinía por la ropa ONCE A WEEK, OR TWICE A WEEK y=,
d. (Ø) la llevaba pa' mi casa, 
e. y (Ø) se las lavaba, 
f. (H) (Ø) se las planchaba y (Ø) se las traiba. [NMC OSS 117] 
a. O: I would iron for them because at that time there were no AUTOMATIC WASHERS, okay. 
A: ‘Hm. 
@@ < @ No @ >.’ 
O: ‘so, 
b. I had a RINGER TYPE MACHINE, 
c. So I would come for the clothes ONCE A WEEK, OR TWICE A WEEK and,
This finding is consonant with a robust constraint uncovered in sociolinguistic and psycholinguistic research known as structural priming (or parallel structure or perseveration) (cf. Labov 1994: 547–568 for a review). Structural priming refers specifically to “the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of an utterance” (Bock and Griffin 2000: 177). This effect illustrates that each clause is not constructed independently, but is shaped by what precedes in the discourse. As Travis (2007) concludes, the finding that priming has an effect in spontaneous language use has profound implications for our view of grammar, as it indicates that the grammar of discourse is developed on-line and in real time as discourse is constructed.

In sum, like the other grammatical features discussed above, variable subject expression in Spanish is not determined by abstract syntactic rules. Rather, it is shaped by a confluence of lexical and discourse factors. We stress that the effect of such factors can only be observed through empirical analysis of spontaneous discourse data.

7 Conclusion

In this chapter we have presented a sample of studies of discourse, in which grammatical structure is manifested in, and seen as deriving from, patterns of use. We have reviewed empirical studies on information flow, transitivity, discourse referentiality and constructions, all of which support a discourse basis for syntax. We look forward to the continued growth of this line of research in Hispanic linguistics, both in terms of deeper analyses of problems in Hispanic linguistics under the lens of discourse function and further contributions of empirical studies of Spanish-speaking speech communities to cross-linguistic insights about usage-based grammar.

APPENDIX

Transcription conventions (Du Bois et al. 1993)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LETTER:</td>
<td>speaker label</td>
</tr>
<tr>
<td>Carriage return:</td>
<td>new Intonation Unit</td>
</tr>
<tr>
<td>Capital initial restart</td>
<td>-- truncated intonation contour</td>
</tr>
<tr>
<td>@</td>
<td>one syllable of laughter</td>
</tr>
<tr>
<td>@@ @@</td>
<td>words spoken while laughing</td>
</tr>
<tr>
<td></td>
<td>final intonation contour</td>
</tr>
<tr>
<td>,</td>
<td>continuing intonation contour</td>
</tr>
<tr>
<td>%</td>
<td>glottal stop</td>
</tr>
<tr>
<td>-</td>
<td>truncated word</td>
</tr>
<tr>
<td>=</td>
<td>lengthened syllable</td>
</tr>
</tbody>
</table>
NOTES

1 This information gives the corpus from which the examples are drawn, and the number or name of the specific transcript. Those marked as NMCOSS come from the New Mexico Colorado Spanish Survey (cf. Bills and Vigil 2008); those marked CCS come from the Corpus of Colombian Spanish (cf. Travis 2005); others come from the respectively cited literature.

2 Muchito is a variation of the form muchachito.

3 Thanks to Bill Croft for pinpointing the semantic difference between (3) and (4).

REFERENCES


