

Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation[☆]



Rena Torres Cacoullos^{a,*}, Catherine E. Travis^b

^a Department of Spanish, Italian, and Portuguese, The Pennsylvania State University,
University Park, PA 16802, USA

^b School of Language Studies, The Australian National University, Canberra, ACT 0200, Australia

Received 9 April 2013; received in revised form 31 July 2013; accepted 2 August 2013

Abstract

Unexpressed subjects, though rare, do occur systematically in English. In this study, we seek to answer the question of what motivates speaker choice between expressed and unexpressed first singular subjects (i.e. *I* vs. an unexpressed, or null, pronoun) in a corpus of conversational American English. We find that the apparently widespread cross-linguistic constraint of subject continuity is bound to coreferential coordinating constructions with *and*, including lexically particular constructions ($[I \text{ VERB}_{1\text{sgi}} \text{ and } \emptyset \text{ QUOTATIVE VERB}_{1\text{sgi}}]$, $[I \text{ GO}_{1\text{sgi}} \text{ and } \emptyset \text{ VERB}_{1\text{sgi}}]$), and to an overarching priming constraint, whereby coreferential unexpressed mentions tend to cluster together. A pivotal restriction is prosodic, such that, outside of coordinating constructions, unexpressed 1sg subjects occur only in Intonation-Unit initial position. We therefore find that variable *I* expression is sensitive to factors operative in subject expression in other languages and in language variation more generally, though paramount are prosodic considerations and particular constructions that may be specific to English.

© 2013 Elsevier B.V. All rights reserved.

Keywords: Structural priming; Prosody; Constructions; Variation; Subject expression; English

1. Introduction

The topic of subject expression and ellipsis has been explored extensively, and several factors have been proposed to operate widely across languages, including cognitive factors, such as discourse cohesion and referent activation or accessibility (e.g., Ariel, 1988; Chafe, 1994; Givón, 1983b; Levinson, 1987), alongside more interactional or pragmatic factors, related to the kind of action the utterance performs (e.g., Fox, 1987; Oh, 2005, 2006; Ono and Thompson, 2003). Quantitative analyses have confirmed accessibility effects in languages where subject expression is noticeably variable, or so called “null-subject languages” (such as Spanish, Portuguese, Italian, Russian, Cantonese, Japanese, cf. Clancy, 1980; Nagy et al., 2011; Paredes Silva, 1993; Silva-Corvalán, 2001), and have also revealed the mechanical factor of priming (Cameron and Flores-Ferrán, 2003:50–54; Travis, 2007:120–121), as well as lexically particular constructions with distinct subject expression patterns (Posio, 2013:286; Travis and Torres Cacoullos, 2012:741). We know much less about the

[☆] The order of authors is alphabetical; both contributed equally to this work. We gratefully acknowledge the support of the National Science Foundation (BCS 1019112/1019122 [2010–2013] “Collaborative Research: Evaluating convergence via code-switching: Cross-linguistic priming, rates and the structure of subject expression”); <http://nmcode-switching.la.psu.edu/>.

* Corresponding author. Tel.: +1 814 865 1153.

E-mail addresses: rena@psu.edu (R. Torres Cacoullos), Catherine.Travis@anu.edu.au (C.E. Travis).

constraints on subject expression in “non-null subject languages” such as English (though see, e.g., Clancy, 1980; Cote, 1996; Harvie, 1998; Leroux and Jarmasz, 2005; Oh, 2005, 2006). We address this gap here, to empirically test the hypothesis that “such null subjects [. . .] have special properties that distinguish them from the canonical null subjects” (Roberts and Holmberg, 2010:5).

We focus on first singular subjects (1sg), because we cannot assume that all persons will pattern in the same way; first and second person, for example, have different information status from 3rd person, and 1sg has a unique role in interaction as a reference to the speaker. We refer to the instances of 1sg exemplified in lines 2 and 5 in example (1) theory-neutrally as unexpressed subjects, which we indicate with a bolded \emptyset for ease of recognition.

- (1)
1. Angela: . . . (TSK) (H) and I put some onion powder . . . in the mayonnaise,
 2. . . . \emptyset Opened em up,
 3. (H) and I didn't stuff the eggs.
 4. (H) I just put that (H) mayonnaise on top.
 5. .. (H) and \emptyset put it on some .. boiled eggs.
- (11 This retirement bit: 759–763)¹

In this study we are concerned with quantitative patterns, in order to account for the range of data in a corpus of spontaneous language use. Crucially, we consider not just unexpressed *I*, as in lines 2 and 5 in (1), but its competing, or alternative form, expressed *I*, as in lines 1, 3 and 4. In confronting this variation, we adhere to the methodological *principle of accountability*. This requires that an analysis account for not only cases in which the linguistic form of interest *was* used, but also all cases where it *could have* been used, but was not (Labov, 2005:7). This is key, because in order to demonstrate that a form serves a given function, it must be shown that when that function is evidenced, speakers use the specified form rather than the other variant(s) at a relatively higher rate.

Much research on unexpressed subjects, though quantitative, neglects precisely the “non-occurrences” (Labov, 2005:7). For instance, it has been proposed that one of the functions of unexpressed subjects in English is to mark a resumption after a parenthetical insert (Oh, 2006:822–830) (cf. also Tao, 1996:507, for a related notion in Chinese). Although Oh does not offer a replicable operationalization of “parenthetical insert”, in (2) the unexpressed subject in line 7 could be seen as signaling a resumption following the parenthetical insert in lines 2–6; note also the repetition of the verb, which has been interpreted as a further signal to the listener of this resumption.

- (2)
1. Arianna: And I got a whole case[=],
 2. Kelly: [I'm so]= excit[2ed2].
 3. Arianna: [2Like2] the companies s- --
 4. Like Pepsi was one of the sponsors.
 5. Kelly: Unhunh.
 6. Arianna: So they sent all this stuff?
 7. .. \emptyset Got a whole case.
 8. . . . Of apple-@cinnamon uh . . . rice cakes,
- (50 Just wanna hang: 255–265)

However, we find very similar examples with an expressed subject, as illustrated in (3), where the *I* in line 4 also occurs in the context of a resumption after a parenthetical insert, with repetition of the verb. There are no apparent different pragmatic motivations for the unexpressed and expressed subjects in (2) and (3). We cannot rely on the argument that (2) signals a resumption but (3) does not on the basis of the unexpressed subject itself (to conclude that unexpressed subjects mark a resumption), because that would be circular. In order to claim such a resumptive function for unexpressed subjects, rather than proffering interpretations of selected examples one by one, it would have to be shown that, in the aggregate, speakers are more likely to opt for an unexpressed subject in such resumption-after-a-parenthetical-insert contexts than in other contexts.

¹ This information gives the transcript name and the line numbers from which the example is drawn. Examples are reproduced verbatim from the transcripts (transcription conventions are presented in Appendix), with the exception of overlapped speech in the first or last line of the example, where only the speaker's contribution is given.

- (3)
1. Alan: Well,
 2. ... I keep a diary- --
 3. course I .. we.. we haven't traveled much lately,
 4. I keep a diary,
 5. ... when I,
 6. ... go out of the country.

(60 Shaggy dog story: 700–706)

Here we supply independent evidence of proposed functions, speaker motivations or constraints through examining co-occurrence and distribution patterns of variable subject expression to discover the linguistic contexts in which speakers are more likely to choose an unexpressed subject over an expressed one. In order to do this, we first identify the locus of variability, or the set of contexts in which speakers have a choice between these two forms. In the next section, we describe the protocol we followed for extracting unexpressed subjects, and then our protocol for extracting a comparable sample of expressed subjects. In section 3 we put forward operationalizations of hypotheses about subject expression, based on contextual features, and submit all tokens of unexpressed and expressed subjects to multivariate analysis in section 4. Quantitative arguments for particular constructions are offered in section 5.

2. Data and method: the variable context for English subject expression

The data for this study are drawn from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al., 2000–2005) a publically available corpus of audio recordings of naturally occurring interaction, primarily face-to-face conversation. The corpus comprises 60 transcripts, totaling approximately 249,000 words.

2.1. Extracting unexpressed 1sg subjects

All finite verbs that occurred with no subject (such as those marked with \emptyset in (1)) were extracted.² We did not extract tokens of non-finite verbs, such as *give* in (4), where there is presumably also a missing auxiliary (here *would*) because we cannot assume that patterns of expression of a sequence of subject plus auxiliary (or other groups of items) would be the same as those for the subject alone.

- (4) Angela: ... I thought I'd pull em off,
just **give em** to her if she liked em that well.

(11 This retirement bit: 290–291)

We do not count cases of repair ($N = 25$) (cf. Fox and Jaspersen, 1995), which occur either with a modification of what was said, or with truncation (indicated in the transcription by --), as in (5); cases that were unclear as to who the subject was, where 1sg was a possibility ($N = 11$), as in (6), where the unexpressed subject could be first person singular or plural; cases where the sound quality rendered the content unclear ($N = 5$); and one token of a book title ("Don't Know Much About History" 52, Cindy: 981).

- (5) Phil: I did not --
.. wasn't really in support of one.

(10 Letter of Concerns: 384–385)

- (6) Marcia: cause all we did was,
... I was in here and \emptyset put a bandage on (Hx).

(18 Vet morning: 255–256)

² Unexpressed *I* appears in 39 of the transcripts (#1–3, 6–8, 11, 15, 17, 18, 21–23, 27, 29, 31, 32, 34, 35, 38, 39, 41–53, 55–57, 59, 60), produced by 60 speakers, most of whom produce just one ($N = 34$) or two ($N = 14$) tokens. Tokens were extracted by Jenny Dumont and Amy Lindstrom. Excluded were tokens produced by two speakers who code-switch to Spanish and show variable subject-verb agreement (Julia and Dolores) ($N = 2$, for 1sg).

This procedure yielded 151 tokens of unexpressed *I*. The rate of expressed *I* in the transcripts with at least one unexpressed *I* is thus roughly 98% (out of some 6600 instances of *I*), and somewhat higher considering all 60 SBCSAE transcripts (out of approximately 9000 tokens of *I*).

The overall rate, taken on its own, however, is not enlightening. The high rate of 98% in English compares with reported rates of expression for 1sg subjects in spoken language of 21% in Polish (Chociej, 2011:52), 28% in Italian (Nagy p.c., cf. Nagy et al., 2011), 33% in Javanese (Ewing this volume), 47% for Cantonese (Nagy p.c., cf. Nagy et al., 2011), 49% in Auslan (Australian Sign Language, McKee et al., 2011:388) and 89% in Finnish (Helasvuo this volume)—a wide range indeed for presumed null subject languages of some sort. Rates differ even within a single language, for example ranging from approximately one-fourth to one-half in spoken Spanish, across dialects and genres (cf. Torres Cacoullós and Travis, forthcoming). The present study seeks to make a contribution toward answering the question of whether, despite glaring differences in overall rates, there are cross-linguistic similarities in the constraints on variable subject expression, or whether these in fact differ, as Roberts and Holmberg (2010:5) suggest.

2.2. Defining the variable context

To answer this question, it is necessary to first identify the *variable context*, the broadest domain in which speakers have a choice between variants, here expression or non-expression, in order to determine what factors may influence that choice. This has not been empirically delineated for English subject expression beyond observations in grammars of English that the subject can be left unexpressed in coordinated contexts (e.g., Biber et al., 1999:156), in initial position (Quirk et al., 1985:896 (§12.46)) or “in casual style” (Stirling and Huddleston, 2002:1540). Scrutiny of the data reveals that English *I* is categorically expressed in interrogatives, subordinate or relative clauses; with discourse markers and (non-conjoined) quotatives; and with contracted auxiliaries. These three sets of contexts together represent nearly one half of the tokens initially extracted in a random sample of expressed *I* (155/320),³ and thus their inclusion in the analysis of variation would have provided very different, and in fact misleading, results. Given the impact of this circumscription of the variable context, we describe each of these excluded contexts below.

First, we found no cases of unexpressed *I* in interrogatives ($N = 0/11$) (as also reported by Harvie, 1998:19). With respect to the clausal status of the verb, there were no cases in a relative clause ($N = 0/20$) nor in a subordinate (complement, adverbial, or *if*) clause ($N = 2/25$), other than ones involving coordination, as in *if I go out and Ø ask for it* (17 Jim: 7). This accords with formalist characterizations of the null subject as a “root phenomenon” (Haegeman and Ihsane, 1999:126) and judgments of unexpressed subjects as ungrammatical in subordinate clauses (Thrasher, 1977:29) (though their occurrence in this environment has been reported for some varieties; see Ihalainen (1991:204) for Somerset English and Haegeman (1999:128–132) on diary writing). Thus, a first circumscription of the variable context is to declarative main clauses.

Second, formulaic constructions (including discourse markers and quotatives) were excluded from the variable context. *I* is not left unexpressed in discourse-marker uses of *I mean*, *I guess*, *I think*, *I (don't) know*, *I remember* and *I'm sure* ($N = 1/24$).⁴ Discourse-marker status is operationalized based on position in the clause and in the Intonation Unit (see section 3.3). We coded as discourse markers occurrences of these collocations that did not occur as canonical main clauses, being produced either as parentheticals (when they appear between the subject and verb or following the verb), as in (7), or produced prosodically independently from other clausal material (that is, when they occur strictly alone in their own Intonation Unit or preceded only by a conjunction, cf. Thompson, 2002:143–145; Travis, 2005:48–52), as in (8). While we do not observe variation in *I* expression in these formulaic constructions, instances of these seven collocations that occur with clausal material in the same Intonation Unit do show variation, as illustrated in examples (9) and (10), and thus are included in the variable context.

(7) Alan: ... You know I've had that painting for ... thirty-five years **I guess**.

(60 Shaggy dog story: 253–254)

³ This initial sample consisted of the 151 tokens of unexpressed *I* and 320 tokens of unstressed *I* extracted by Leah Houle, drawn evenly from 16 transcripts that have at least one unexpressed 1sg token, for which tokens of *I* had previously been tagged as stressed or unstressed (Travis and Torres Cacoullós, in preparation).

⁴ Out of 340 occurrences of *I don't know* in the entire SBCSAE, there is only one first singular *don't know* (apart from song title “Don't know much”), and although it occurs on its own IU, it is an answer to a question and really means not knowing, so is not formulaic.

Dana: Where's your class.

Nancy: ... Ø Don't know.

(50 Just wanna hang: 521–522)

- (8) Alan: ... But she nudged me when she got back,
but of course,
... **I guess**,
I haven't done it recently,
but over the years I like to tell her this story.

(60 Shaggy dog story: 601–605)

- (9) Marcia: **Ø Guess** they gotta make money somehow,

(18 Vet morning: 565)

- (10) Alan: **I guess** it probably was,

(60 Shaggy dog story: 790)

As with these discourse markers, *I* with quotative verbs *say*, *be like*, *go* and *think* can be considered formulaic. *I* is only unexpressed in such contexts where the quotative was coordinated with another verb ($N = 9/25$), that is, where it was conjoined with *and*, as in *I phoned her and Ø said* (11 Angela: 954). (See section 5 for discussion of the conjoined quotative as a lexically specific construction.)

A third consideration is auxiliary contraction, as in (11), where contracted *had* occurs with *I* but an unexpressed 1sg occurs with non-contracted *have to*. According to Akmajian et al.'s (2001:290) rule of "Tag-Controlled Deletion", if the auxiliary is "contractible", it must be contracted onto the subject, whereas if it is not, the subject can be deleted. In our sample, non-contracted forms of "contractible" auxiliaries were few ($N = 6$). Contracted forms of auxiliaries *be*, *will*, *have*, *had*, *would* did not once appear in the absence of expressed *I* ($N = 0/50$) (i.e., there were no cases of *'m*, *'ll*, *'ve*, *'d* (pace Weir, 2012:107)), and so *I'm*, *I'll*, *I've*, *I'd* were excluded from the variable context.

- (11) Stephanie: <<STOMPING<MRC **I'd** better get twelve hundred on these.
Ø Have to get twelve hundred MRC>STOMPING>>.

(35 Hold my breath: 39–40)

Adhering to the *principle of accountability*, we proceeded to extract a sample of expressed *I* excluding any tokens that occur in the three environments that fall outside the envelope of variation detailed above, to include only those tokens of expressed *I* that can be considered "non-occurrences" (Labov, 2005:7) of an unexpressed 1sg subject.

2.3. Extracting a sample of expressed 1sg subjects

Given the preponderance of expressed 1sg subjects in English (section 2.1), only a sample of expressed *I* was extracted (modeled on the method of extraction followed by Harvie, 1998; Leroux and Jarmasz, 2005). In order to account for variable subject expression in all speakers who had at least one null subject occurrence, for each unexpressed *I* we extracted the closest preceding expressed coreferential *I* (i.e., produced by the same speaker) occurring with a finite verb. In cases of sequences of two or more unexpressed, we went back to the next closest previous eligible token; in a few cases where there was no preceding instance (for example, when the null subject is at the start of a transcript), we took the second closest following. We skipped over any tokens occurring in the three non-variable contexts noted in section 2.2, as well as *Is* occurring in quoted speech, since these would be uncodable for the predictors of distance and previous realization (discussed below), and *Is* occurring in a separate IU from the verb (as in line 1 in (22) below), since this could never be the case for an unexpressed mention.

The following two examples illustrate the application of this protocol. In (12), for the unexpressed subject in line 2, the *I* in line 1 was extracted, as the closest eligible preceding coreferential *I*; in (13), for the unexpressed subject in line 3, we skipped the *I* in line 2 which occurs outside the variable context (a subordinate clause) and extracted the one in line 1.

- (12)
1. Curt: .. And then I put them in there.
2. ... **Ø S=quashed** em down,

(42 Stay out of it: 226–227)

- (13)
1. Lajuan: ... (H) I don't like it,
2. like and when I was ho=me,
3. .. **Ø just went** home to Indiana.
4. (H) I went to hug my sister,

(44 He knows: 161–164)

Following these protocols we extracted a total of 151 tokens of expressed *I*.⁵ The rate of unexpressed subjects in this database is thus an artificial 50% (151/302) (since an expressed subject was extracted for every unexpressed one). While the overall rate is artificial, the data as extracted allows us to compare the frequency of the variants in sets of linguistic sub-contexts, as we outline in the following section.

3. Operationalizing hypotheses: coding for contextual features

All 302 tokens were coded for several contextual features, including subject continuity, form of the previous coreferential 1sg subject, position, and *and*-coordination. Each represents a replicable operationalization of a hypothesized constraint on subject expression.

3.1. Subject continuity

According to Givón's (1983b) cross-linguistic topic continuity/accessibility hierarchy, more continuous or accessible participants receive less linguistic coding, and less continuous or accessible participants receive more coding (cf. Ariel, 1988:79; Chafe, 1994:74; Levinson, 1987:384). Here, we operationalize continuity/accessibility in terms of recency of mention, as measured in intervening human subjects between the target token and the previous coreferential mention as subject.

Counted as the previous coreferential "mention" were 1sg subjects produced with a finite verb, including those that occurred in truncation with an auxiliary without a main verb (e.g. *I should --, I didn't --, I'd --*), but not those that occurred without a verb (e.g. *I just --*). Because of the dubious nature of the referentiality of *I* in the formulas identified in section 2.1 (discourse markers and quotatives), these were not considered to be coreferential mentions.

In order to obtain a measure of "recency" of previous mention, we then proceeded to count intervening human subjects. Although distance is typically measured in clauses (e.g. papers in Givón, 1983a), accessibility may not be affected the same way by all clauses, in particular those with non-referential, or even non-human subjects. In Travis and Torres Cacoulllos (2012) we found for Spanish that the presence of Intervening Human Subjects is a better measure than Switch Reference, i.e. when there is a different subject referent—of any kind—in the preceding clause (e.g. Cameron, 1994:28–32; Silva-Corvalán, 1982:104). The two measures often coincide, as illustrated below in (20) for coreferential subjects (no intervening human subjects, and no intervening clauses) and in (21) for non-coreferential subjects (one intervening human subject and one intervening clause). However, this is not always the case. In (14) for the token in line 4, there are two intervening clauses, but no intervening human subjects (*this* being non-referential, and *they* being generic).⁶

(14)

1. Tom3: Ø went through law school at the University of New Mexico,
2. (H)= this was just shortly after they had started law school [down he]re.
3. Tom2: [Mhm].
4. Tom3: ... (H) And Ø have basically been here ever since.

(32 Handshakes all around: 1713–1717)

We apply Intervening Human Subjects here, and refer to contexts where there are no intervening human subjects as *coreferential* (i.e. the subject of the target clause is coreferential with that of the previous human subject) and those where there is one or more intervening human subjects as *non-coreferential* (i.e. there has been a change in subject between the target clause and the previous clause with a human subject).

3.2. Priming

Structural priming (also known as perseveration or persistence) is a speaker's repetition of a preceding form. One of the earliest reports of this phenomenon was that of Weiner and Labov (1983) in their study of constraints on the English agentless passive (*The liquor closet was broken into* vs. *They broke into the liquor closet*) in spontaneous discourse. They

⁵ Although it has been proposed that unexpressed mentions are deletions of unstressed sentence material (Biber et al., 1999:157; Napoli, 1982:99; Quirk et al., 1985:896 (§12.46); Zwicky and Pullum, 1983:159), the sample is not limited to unstressed tokens of *I*, because *I* has not been tagged for stress in all transcripts. However, we can assume that the great majority of the expressed tokens are unstressed (in the vicinity of 87%, 1651/1861 (Travis and Torres Cacoulllos, in preparation)).

⁶ Though not illustrated here, in coding for recency, cases where the predicate was produced in overlap with the target clause were not counted as intervening clauses.

found that the passive variant was most strongly favored by the presence of a passive in the preceding five clauses (1983:52). A strong priming effect has been observed for subject expression in Spanish (Cameron and Flores-Ferrán, 2003:50–54; Travis, 2007:120–121; Travis and Torres Cacoullos, 2012:729–733), leading Cameron to propose that “pronouns lead to pronouns, and null subjects lead to null subjects” (1994:40). Examples of priming are seen below, where the speaker produces a pair of unexpressed subjects in (15), and of expressed *I*s in (16). While these examples illustrate priming from immediately preceding to following clause, here we test this effect through consideration of the form of the previous coreferential subject occurring within the preceding five clauses.

- (15) Wess: and \emptyset **put** it in there,
 Cam: [@ @@@]
 Wess: [and] \emptyset **beat** it a little bit,

(59 You baked: 1228–1230)

- (16) Wess: and then **I put** it on,
 and **I boil** it.

(59 You baked: 1211–1212)

Whereas a subject continuity effect would be motivated by considerations of accessibility, a priming constraint on English subject expression would be a mechanical one, that is, “the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of an utterance,” as observed in psycholinguistic experiments (Bock and Griffin, 2000:177).

3.3. Positioning of the subject

A number of reference grammars and generative treatments have proposed that unexpressed subjects are elided initial material. But the unit by which initial position is identified remains poorly defined. For instance, Quirk et al. (1985:896 (§12.46)) describe elided words as those that “normally occur before the onset of a tone unit [. . .] and hence have weak stress and low pitch”; Napoli (1982:99) proposes a phonological rule deleting unstressed material “not preceded by any other phonetic material”; and Roberts and Holmberg (2010:5) claim that unexpressed referential subjects in English “have to be strictly sentence initial”. To locate this general “left-edge” deletion (Weir, 2012) for *I* expression, we take two measures, namely the Intonation Unit (IU) (Chafe, 1994:Ch. 5), described by Croft (1995:875) as the “most plausible basic unit of the grammar of spoken language”, and the “prosodic sentence”, following Chafe (1994:139).

The IU is characterized as “a stretch of speech uttered under a single, coherent intonation contour” (Du Bois et al., 1993:47). It is associated with such features as an initial pitch reset (or rise in fundamental frequency); a change in duration, with faster speech early in the IU, and then lengthening of the final syllable; and pausing between IUs (cf. Chafe, 1994:58–60; Du Bois et al., 1993:47). These features are best considered cues that tend to occur at IU boundaries; none individually define the IU (note, for example, that pauses also occur IU internally), and none are required for the identification of an IU boundary. Each IU is presented on a different line in the transcription, and punctuation is used to mark the terminal pitch; for example, a comma represents continuing and a period final intonation contour.

Instances of 1sg subjects were classified as occurring or not in absolute-initial position in the IU. IU-initial tokens were those that occurred without any preceding material on the line (e.g. *I went* in line 5 in (17)). Non-IU initial tokens followed clausal material (as for coordinated \emptyset *saw* in (17)) or a discourse marker or conjunction (e.g. *so I eat* in (18)) on the same line.

- (17)
1. Alina: He doesn't really wan[na do] TV.
 2. Lenore: [(H)]
 3. Alina: (H) (SNIFF)
 4. Lenore: Hm.
 5. Alina: .. **I went** and \emptyset **saw** their house the other night.

(06 Cuz: 1470–1474)

- (18)
- Ken: (H) So **I eat** the local food,

(15 Deadly Diseases: 587)

On the other hand, the sentence, which has featured in publications on English “null” subjects based on grammaticality judgments (e.g. Haegeman, 2002; Napoli, 1982), is not straightforwardly a unit of speech (Croft, 1995:841; Miller,

1995:132). For the sake of replicability, we delimit the sentence as a section of speech that is prosodically bound by final or rising intonation (marked by a period or a question mark), following Chafe's (1994:139) interpretation. This is based on an understanding that "The period intonation . . . serves to express a speaker's judgment that he or she has completed the verbalization of some coherent unit of content" (Chafe, 1994:143). This correspondence between prosody and syntax has been substantiated by Ford and Thompson (1996:155) who, in approximately 20 minutes of American English conversation, found that 99% (428/433) of all points of "prosodic completion" corresponded with "syntactic completion".

Thus defined, the "prosodic sentence" may consist of a single IU (as in line 3 in (19)), or two or more IUs with continuing intonation (marked by a comma) for all but the final IU (as for lines 1/2 and 4/5 in (19)). Where the speaker is initiating a new turn after contributions by other speakers, the beginning of the sentence was the beginning of the speaker's turn.⁷ Also considered to be the beginning of the sentence are those cases where the period or question mark does not appear in the immediately preceding IU but in the intervening IU there is a conjunction or a discourse marker with a comma (e.g. *uh, oh, but uh, yeah, so, well, okay*; for example, *it* in line 2 in (19) would be coded as sentence initial).

(19)

1. Curt: . . . Well,
2. it's all packed up in . . . garbage bags now.
3. I didn't even touch it.
4. . . . Ø Used a hoe,
5. Ø stuffed it X.

(42 Stay out of it: 209–213)

3.4. Coordinating constructions with *and*

Unexpressed subjects in coordinated clauses have been noted in English grammars (e.g., Biber et al., 1999:156; Dixon, 2005:74), and have been considered distinct from instances of English unexpressed subjects in non-coordinated contexts. In particular, verbs conjoined with *and* have been viewed as involving a single clause with two predications rather than two clauses with an unexpressed subject in the second (cf. Huddleston, 2002:238 (§3.1); Quirk et al., 1985:942 (§13.44)).

Here, we define coordinated clauses narrowly as clauses conjoined by *and* with coreferential subjects and occurring either adjacent to each other, as in the two bolded tokens in (20), or with a single subordinate clause intervening, as in (21). We include tokens of *and* with a filler (e.g. *and uh, and um*), but not *and* with more substantial material (e.g. *and then, and like, and at that point*).⁸ There were no tokens with more than one subordinate clause (and no main clause) intervening between coreferential mentions.

(20)

Harold: I went and Ø **got** a wet rag and Ø **wiped** it off the car=.

(02 Lambada: 308)

(21)

Alice: . . . (SNIFF) So I **sat** over here,
.. before we went over to Diane's,
and Ø **explained** the recipe to em,

(43 Try a couple of spoonfuls: 484–486)

These operationalizations of hypothesized constraints based on contextual features define factors for multivariate analysis. All factors coded for are considered simultaneously and those making a significant contribution to variant choice are selected, allowing us to ascertain the structure of variability.

⁷ There were no cases of co-constructions (cf. Lerner, 1991), where the speaker was clearly finishing off someone else's sentence.

⁸ As to other so-called coordinating conjunctions, there were no cases of *or*; of the eight tokens of *but* none were straightforwardly a coordinating conjunction (because they occur in non-coreferential contexts, with other substantial material (e.g. *but then*) or in prosodic sentence-initial position (that is, at the beginning of an IU following final intonation contour).

Table 1
Variable-rule analysis of the contribution of factors selected as significant to the choice of unexpressed *I* in conversational American English.

	Prob.	% Ø	N	% data
<i>N</i> = 302; Input: .55 (overall rate: 50%)				
Coordinating construction				
Coordinated verbs with <i>and</i>	.96	88%	73/83	28%
Non-coordinated verbs	.24	36%	78/219	72%
Prosodic position				
IU-initial	.69*	46%	88/191	63%
Non-IU-initial	.21	57%	63/111	37%
Priming: realization of previous coref 1sg subject**				
Unexpressed	.86	84%	27/32	11%
Pronoun (<i>I</i>)	.43	50%	103/206	68%
Subject continuity				
Coreferential contexts	[.55]	59%	121/206	70%
Non-coreferential contexts (Intervening Human Subj)	[.39]	29%	26/90	30%

Non-significant factor groups: subject continuity (indicated by square brackets above), Position in speaker turn, Polarity, Semantic class of verb, Tense. In Spanish, expressed first singular subject pronoun *yo* is favored by cognition verbs (e.g., Bentivoglio, 1987:60; Silva-Corvalán, 1994:162; Travis, 2007:116–117) and, with this class of verbs, in the first IU of a speaker turn (Travis and Torres Cacoullos, 2012:736–737); also by backgrounding Tense-Aspect-Mood forms (such as past imperfective) (Silva-Corvalán, 1997). Negative polarity favors stress on *I* (Travis and Torres Cacoullos, in preparation).

* The mismatch between the relative ordering of probabilities and percentages here is discussed in section 4.2.

** Realization of previous coreferential 1sg subject applies to coreferential subjects that occur within the five preceding clauses (*N* = 238).

4. The structure of variable first singular subject expression

The results of the multivariate analysis are given in Table 1, which depicts the factors contributing to the choice of an unexpressed subject, based on a logistic regression (Variable-rule) analysis (Sankoff, 1988), using Goldvarb Lion (Sankoff et al., 2012). Of the hypothesized constraints, the three that were retained through the analysis as significant are Coordinating construction, Position in prosodic unit and Previous realization. Each predictor (factor group) has two levels (factors), listed on the left (e.g., Coordinating construction: verbs coordinated with *and*, and non-coordinated verbs). Shown in the first column of numbers are the Probabilities, with values between 0 and 1: within each group, the factor with a Probability closer to 1 can be said to *favor*, and that with a Probability closer to 0, to *disfavor* an unexpressed 1sg subject. The table also shows in subsequent columns, for each factor, the rate (%) of unexpressed 1sg, the number of tokens, and the percentage of the data the factor makes up. We discuss each significant effect in turn, and then one non-significant effect, that of subject continuity.

4.1. 'And' coordinating constructions: prosodic and discourse connectedness

As shown in Table 1, unexpressed subjects are far more likely in coordinated clause constructions (Prob. .96, rate 88%) than with non-coordinated verbs (Prob. .24, rate 36%). Coordinated verbs are those that are conjoined to a previous one with *and* in both non-coreferential contexts where there is only one subordinate clause intervening (as in (21) above; *N* = 5), and in coreferential contexts (as in the bolded tokens in (22) below; *N* = 78), which are far more frequent. The schema [*I* VERB_{1sg} *and* Ø VERB_{1sg}] serves to represent such coreferential *and*-coordinating constructions.

These coordinating constructions may be more, or less, tightly linked, as measured prosodically by occurrence in the same, or in different, IUs, in accordance with a syntax-prosody mapping (see Chafe (1994:Ch. 9) and Croft (1995)). Close to 60% (46/78) of the time, the schema [*I* VERB_{1sg} *and* Ø VERB_{1sg}] occurs across IUs, as seen in (22): in both lines 2–3 and lines 3–4 the two conjuncts occur in different IUs; in line 3 the subject is unexpressed and in line 4 it is expressed. The rate of unexpression in coordinating constructions produced across IUs is 83% (38/46).

(22)

1. Mary: .. then I,
2. ... shut the hood,
3. **and Ø got** back in,
4. **and I started** up the engine and,
5. (H) both !Gary and !Rita were sitting on the edges of their seat.

6. ... (SWALLOW) And I turned around **and I looked**,
 7. .. **and I said**,
 ... <Q did I scare you kids Q>?

(07 A Tree’s Life: 546–553)

Less frequently (32/78), both verbs in [*I* V_{ERB1sg} *and* \emptyset V_{ERB1sg}] appear in the same IU (as seen in (20) and (17) above). In our sample, these tokens display syntactic integration in that they all (100%, *N* = 32) have an unexpressed subject in the second conjunct. Based on other syntactic criteria (such as the possibility of inserting adverbs between the conjuncts or negation as a complex unit) in their study of *and* coordination, Barth-Weingarten and Couper Kuhlen (2011:279) also find a correlation between prosodic and morphosyntactic integration in *and* coordinated clauses. Following this, we interpret these same-IU *and*-conjuncts as a “single event” (cf. Givón’s (1993:7) “Temporal contiguity and event integration” or Hopper’s (2001:153) “hendiadic” verbal expressions). Note that the same-IU criterion cannot be substituted by sameness of tense: when coordinated verbs are in the same IU they appear in the same tense (32/32), but this is overwhelmingly also the case when they are in different IUs (45/46).

Exceptions to the tendency for the second conjunct of *and*-coordinated verbs in the same IU to have an unexpressed subject may be associated with discrete situations. For example, in line 6 of example (22) above, the situations are sequential: “I looked” is temporally subsequent to “I turned around”. In (23) below, while the situations presented in the first IU may be temporally simultaneous, the speaker elaborates on *I see* in the following IU, suggesting that “seeing” may be an event in its own right.

- (23) Alina: .. (H) So **I walk in and I see**,
 .. two of the .. paddlers I definitely don’t wanna see.

(06 Cuz: 920–921)

In summary, *and* coordination very strongly favors unexpressed *I*, overwhelmingly so when the conjuncts are presented in the one IU as a single, hendiadic construction.

4.2. Prosodic position constraint for non-coordinated verbs: IU-initial position

The Probabilities for Prosodic position in Table 1 also indicate that unexpressed *I* is favored when the verb appears at the beginning of the prosodic unit (.69 in IU-initial position vs. .21 in non-IU-initial position). However, note that this is the reverse of the relative ordering of the percentages (lower for IU-initial than for non-IU initial position). Through this mismatch between the Probability (based on multivariate analysis) and percentages (based on univariate analysis), the Variable-rule analysis alerts us that prosodic position is interacting with another variable.

Table 2, which presents a cross-tabulation of Position in IU and Coordinating construction, explains this interaction. We see that the favoring effect of IU-initial position does not apply to coordinated verbs (on the left of the table): the rate of unexpressed subjects is similarly high (86% and 88%) in IU-initial position (when the *and* is in the preceding IU) as in non-IU-initial position (either when the verb is preceded by *and* alone (*N* = 37) or by the first conjunct on the same IU (*N* = 32) (as discussed in the preceding section)).

For non-coordinated verbs, however, depicted on the right-hand side of the table, *I* is unexpressed 43% of the time in IU-initial position, but a mere 5% of the time in non-IU initial position. Note that most of these expressed 1sg subjects in non IU-initial position occur following *so* (*N* = 11), as in (23), (*and*) *then* (*N* = 7), or *but*, *okay*, *no*, *well*, *cause*. The only two tokens of non-initial unexpressed *I* with non-coordinated verbs are preceded by *and* in the IU,

Table 2
 Rate of unexpressed 1sg subject by coordination and position in prosodic unit (Intonation Unit) (*N* = 302).

	Coordinated verbs (<i>and</i> V)		Non-coordinated verbs	
	%	<i>N</i> \emptyset /total	%	<i>N</i> \emptyset /total
Position in IU				
IU-initial	86%	12/14*	43%	76/177
Not IU-initial	88%	61/69	5%	2/42**
TOTAL		83		219

* *and* occurs in the preceding IU.

** The two unexpressed tokens here are preceded by *and* on the IU and are cases of coordination in a broad sense (with intervening main clauses).

and could in fact be considered cases of coordination in a broader definition than the one adopted here (coreferential verbs conjoined with *and* with no more than a subordinate clause separating them, section 3.1); one example is given in (14). Both tokens occur in the same conversation, by the same speaker. We can say then that, outside the context of coordinated verbs following *and*, unexpressed 1sg subjects occur virtually only in absolute initial position of the prosodic unit.

Might this restriction to IU-initial position be masking a syntactic constraint? A constraint of “sentence”- or “clause”-initial position has been reported in previous quantitative analyses, such that unexpressed subjects are disfavored in the presence of “hesitations, subordinate clauses, or other elements that precede the subject” (Harvie, 1998:20–21; Leroux and Jarmasz, 2005:7). However, this effect is weaker than the IU-position restriction: when Harvie (1998:23) analyzed non-conjoined sentences separately, the rate of unexpressed subjects was approximately twice as high in “clause”-initial than non-initial position (36:21), while the results in Table 2 demonstrate that the rate of unexpressed subjects is about eight times greater in IU-initial than non-initial position (43:5). This suggests the primacy of the prosodic unit in subject expression.

Cross tabulations between sentence-initial and IU-initial position confirm this. Applying the notion of the “prosodic sentence” outlined above (section 3.3), we find firstly that IU-initial position corresponds just one half of the time to sentence-initial position in our data: 49% (94/191 of the IU-initial first-singular subjects are also sentence initial (as in line 4 in (19) above).⁹ The other half of the time, the IU-initial (expressed or unexpressed) *I* is not sentence-initial, but follows a continuing intonation contour, as in line 5 in (19). More importantly for present purposes, there is no difference between the rate of expression of IU-initial-*I* in “sentence-initial” and “non-initial” position thus defined (46% for both, at 43/94 and 45/97 respectively). This bolsters the claim that the positioning constraint for subject expression is indeed best defined in terms of prosodic units.

In summary, for non-coordinated verbs, a prosodic constraint is such that unexpressed 1sg subjects occur (in variation with expressed 1sg) in absolute IU-initial position only. In short, virtually the only time we get an unexpressed 1sg subject IU-internally is with coordinated verbs following the conjunction *and*.

4.3. Coreferential priming

Finally, the multivariate analysis reveals an effect for the realization of the previous coreferential first singular subject. When we consider the form of the previous mention within up to five preceding clauses, we observe that greatly conducive to unexpressed *I* is previous realization as unexpressed (seen in the high Probability of .86).¹⁰

Given the low rate of unexpressed subjects in the data overall (see section 2.1), the environment of a previous unexpressed mention is much less frequent than that of a previous expressed mention. Most speakers in fact produced only one unexpressed 1sg subject per transcript. But when more than one unexpressed subject is produced, they tend to cluster together. This is illustrated in (24), with IU-initial non-coordinated verbs. In the earlier pair of examples (15) and (16) (section 3.2), with coordinated (*V and V*, different-IU) verbs, we have a near “minimal pair”, in that the same speaker selects *I and I* in the first example, \emptyset and \emptyset in the second, in accordance with priming.¹¹

- (24) Miles: (H) But it was like I went [to] Bahia,
 Jamie: [What's] --
 Miles: ... last Sunday,
 (H) ... \emptyset got there at eight,
 (Hx) .. \emptyset left a te=n,
 ... \emptyset dropped this person off at home,
 in Foster City,

(02 Lambada: 848–854)

⁹ A sentence-initial subject, as sentence is defined here, is necessarily IU-initial. The count of 191 IU-initial instances includes 14 conjoined verbs (2 of which are sentence initial).

¹⁰ Our token numbers are insufficient to examine the dissipation of coreferential subject priming effects with distance (such as was observed for Spanish subject expression in Travis and Torres Cacoullós, 2012).

¹¹ A reviewer objects that examples such as (24) are cases of rhetorical chains or lists of similar events presented by the speaker as connected through unexpressed subjects. But evidence independent of the form of the subject itself would be required, since events are not obviously any less chain-like in cases where both expressed and unexpressed subjects appear, for example, lines 1–4 in (22).

Table 3

Rate of unexpressed 1sg subject by previous realization and *and*-coordination ($N = 238$).^{*}

	Coordinated verbs (<i>and</i> V)		Non-coordinated verbs	
	%	$N \emptyset$ /total	%	$N \emptyset$ /total
Previous realization				
Unexpressed (\emptyset)	100%	12/12	75%	15/20
Expressed (<i>I</i>)	86%	60/70	32%	43/136
TOTAL		82		156

^{*} Excluded are tokens where the previous coreferential subject occurred at a distance of greater than five clauses ($N = 57$), those occurring in the context of quotation ($N = 2$), or unclear content ($N = 5$).

Table 4

Rate of unexpressed 1sg subject by previous realization and subject continuity, for IU-initial (non-coordinated) verbs ($N = 123$).^{*}

	Coreferential		Non-coreferential ^{**}		Total	
	%	$N \emptyset$ /total	%	$N \emptyset$ /total	%	$N \emptyset$ /total
Previous realization						
Unexpressed (\emptyset)	88%	14/16	NA	0/1	82%	14/17
Expressed (<i>I</i>)	40%	34/85	38%	8/21	39%	43/106
TOTAL	48%	48/101	36%	8/22		

^{*} Excluded are tokens where the previous coreferential subject occurred at a distance of greater than five clauses ($N = 48$), those occurring in the context of quotation ($N = 2$), or with unclear content ($N = 4$).

^{**} By the measure of Intervening Human Subjects (section 3.1).

Priming is thus operative for both coordinated and non-coordinated verbs, which are presented separately in Table 3. The priming effect is more striking for non-coordinated verbs on the right of the table, with a rate of unexpressed *I* when the previous realization was unexpressed that is twice as high as that when the previous realization was expressed (75% vs. 32%) (Chi-square test $p < .01$).¹²

4.4. The lack of a subject continuity (coreferentiality) effect

What of the cross-linguistically applicable subject continuity constraint amply reported in the literature? Though in the predicted direction, this was not retained as a significant contributor to speaker choice of an unexpressed *I* when all factors were considered together in the multivariate analysis shown in Table 1. For coordinated verbs, coreferentiality is nearly moot: as noted above (section 4.1), most coordinated instances (94%, 78/83) occur in coreferential contexts ($VERB_{1sgI}$ and $VERB_{1sgI}$), the cases of an intervening subordinate clause being very few.

For IU-initial non-coordinated verbs, subject continuity is not really an independent effect. Table 4 gives the rate of unexpressed *I* by previous realization and coreferentiality. From the Totals row, the tendency appears to be the predicted one, with a lower rate of unexpressed *I* in non-coreferential contexts (36% vs. 48%). However, with the cross-tabulation we see that priming operates in coreferential contexts (on the left), where 1sg is more likely to be unexpressed if the preceding subject was also unexpressed (88% vs. 40%) (Chi-square test $p < .01$). Further, we see from the first row that (though the data are sparse), unexpressed-to-unexpressed priming tends to be limited to coreferential contexts: most (16 of 17) cases of a previous unexpressed occur in coreferential contexts. On the other hand, when the previous realization is expressed *I*, seen in the second row, coreferentiality makes no difference, with a rate of expression of 38–40% in both coreferential and non-coreferential contexts. Thus, the appearance of an effect for subject continuity is bound to unexpressed-to-unexpressed priming; since priming tends to occur in coreferential contexts, the rate of unexpressed 1sg subjects is raised in such contexts.

¹² Separate Variable-rule analysis of IU-initial tokens only of non-coordinated verbs confirms the strength of the priming effect, which is the only significant predictor ($N = 177$, Input .46, Previous realization unexpressed .84 (14/17) – Previous realization expressed .43 (42/106)). Subject continuity was not selected, though it did show the predicted direction of effect; also not significant but without a discernable direction of effect were Polarity, Turn position, Semantic class of verb, Tense.

5. Particular constructions

In ascertaining the structure of variability, we have seen that *I* expression is conditioned by coordination, prosody and priming, while coreferentiality does not have an effect independent from coordination and priming. These conditioning environments can be depicted schematically, as given below. Unexpressed *I* is favored by verbs coordinated with *and* (25); in IU-initial position (indeed, the only cases of unexpressed *I* in non-initial prosodic position are following *and*) (26); and when the previous coreferential mention was also unexpressed (a constraint which holds for both coordinated and non-coordinated verbs) (27).

(25)

Coordinating constructions: $[I \text{ VERB}_{1\text{SGI}} \text{ and } \emptyset \text{ VERB}_{1\text{SGI}}]$
(conjuncts on the same or different IUs)

(26)

IU-initial position: $[\emptyset \text{ VERB}_{1\text{SG}} \dots]_{\text{IU}}$

(27)

Priming: $[\emptyset \text{ VERB}_{1\text{SGI}} (\text{and}) \emptyset \text{ VERB}_{1\text{SGI}}]$

We can think of these general schemas as *constructions*—form-function pairings, where function includes stored information about contexts of use (for an overview, see, for example, Croft and Cruse (2004:225–290)). Within these general schemas, more particular constructions emerge in the data.

For example, earlier (section 4.1), we observed the local pattern of two verbs with a coreferential 1sg subject in a coordinating structure in the same IU, in which the subject of the second verb is overwhelmingly unexpressed. Irrespective of whether prosodic connectedness is viewed as merely reflecting discourse connectedness, or whether this is also related to conceptual connectedness, the prosodic—same IU—constraint on *I* expression for coordinated verbs is strong, and can be viewed as defining a particular construction, contributing to the more general schema in (25). The form of this particular construction would be as depicted in (28), paired with the meaning of discourse (event) connectedness.

(28) coreferential same-IU, coordinating construction:

$[I \text{ VERB}_{1\text{SGI}} \text{ and } \emptyset \text{ VERB}_{1\text{SGI}}]_{\text{same IU}}$

Associated with this construction are two more lexically specific constructions which together, make up 88% (29/32) of the $[I \text{ VERB}_{1\text{SGI}} \text{ and } \emptyset \text{ VERB}_{1\text{SGI}}]_{\text{same IU}}$ occurrences. One involves a quotative or a verb of speech, and the other a verb of motion (also noted in older English varieties for *go* and *take* (Brinton, 1990:124–125) and characterized by Hopper (2002:152) as “core” hendiadic verbal expressions).

Cases of the quotative construction, depicted in (29), add up to close to one half (14/32) of the occurrences of the coreferential same-IU construction depicted in (28). Other examples are *I came home and \emptyset told you.* (42, Kendra: 111), *(H) I wrote home to my family and \emptyset said,* (32, Tom_2: 511), and, with quoted thought, *I look at that and \emptyset think,* (02, Miles: 526).

(29) $[I \text{ VERB}_{1\text{SGI}} \text{ and } \emptyset \text{ QUOTATIVE VERB}_{1\text{SGI}}]_{\text{same IU}}$

Angela: (H) I phoned her and \emptyset said,

(11 This retirement bit: 954)

Cases with a motion verb occur most frequently with the verb *go* (with or without a particle) ($N = 9$), as is depicted in (30). Other examples are *Ten after two I went over and \emptyset got her,* (43, Alice: 328), *And even if I go out and \emptyset a = sk for it.* (17, Jim: 7), *That's why I went out and \emptyset bought the coffee.* (49, John: 1176). There are a further six tokens with other verbs of motion (such as *come*, *run*, *turn around*), for example, *so I ran and \emptyset put em on,* (52, Darlene: 282). This may indicate a schematic or productive construction with a verb-of-motion in the first verb slot, the category of which is built around the more frequent verb (cf. Bybee, 2010:Ch. 4), *go*.

(30) $[I \text{ GO}_{1\text{SGI}} \text{ and } \emptyset \text{ VERB}_{1\text{SGI}}]_{\text{same IU}}$

Harold: I went and \emptyset got a wet rag and \emptyset wiped it off the car=.

(02 Lambada: 308)

These local patterns accord with a construction-based view of grammar, in which it is profitable to identify lexically particular constructions, which both contribute to and deviate from more general patterns.¹³

6. Conclusion

Based on low rates of non-expression, we could say that English is a “non-null subject language”. How, then, does the patterning of 1sg expression in English compare with that of “null-subject languages”? The answer to this is found in the structure of variability, which can only be ascertained through definition of the *variable context* and *operationalization* of hypothesized generalizations or constraints. In probing the variable context, we find that the locus of speaker choices is only a subset of the contexts in which expressed *I* appears, namely declarative main clauses, further excluding discourse formulas and contracted auxiliary forms. In our operationalization of hypothesized constraints, the factors included in the multivariate analysis are replicably defined based on elements of the context of occurrence.

First, we identify two environments—coordinated and non-coordinated contexts, that are distinct not solely because of the doubly high rate of unexpressed *I* in the former. While the difference has been viewed in terms of underlying syntactic structure (cf. Huddleston, 2002:238 (§3.1); Quirk et al., 1985:942 (§13.44)), the data show these contexts to be subject to a different prosodic constraint. With coordinated verbs, the rate of unexpressed *I* is even higher when the second conjunct occurs in the same IU as the first. With non-coordinated verbs, the variability is restricted to IU-initial position (in non-IU-initial position, 1sg is near-categorically expressed in our sample). The role of these prosodic constraints highlights the interplay between prosody and grammar, something which can readily be observed and measured following the IU-based transcription used here.

What of priming and subject continuity? A mechanical priming constraint applies across the board (for coordinated and, especially, non-coordinated verbs), such that unexpressed subjects are favored when they are immediately preceded by another unexpressed coreferential subject. In other words, although unexpressed *I* is rare, when it does occur, it tends to do so in clusters. The commonly reported cross-linguistic subject continuity constraint on subject expression appears to apply (also to both coordinated and non-coordinated verbs), as evidenced by a higher rate of unexpressed subjects in coreferential contexts (in the absence of intervening human subjects). But this is not an independent effect. For non-coordinated, prosodically-initial subjects, this is due to unexpressed-to-unexpressed priming, which is overwhelmingly applicable in coreferential contexts, while for coordinated verbs, it is bound to the fact that most of the time they occur in coreferential contexts.

Finally, the data suggest several different schemas, of varying degrees of specificity, ranging from the general schema for coordinated verbs [*I* VERB_{1sg} and \emptyset VERB_{1sg}], to the more particular manifestation where this occurs in the one IU representing the one event. Within this we find the lexically specific constructions, with a quotative as the second verb, [*I* VERB_{1sg} and \emptyset QUOTATIVE VERB_{1sg}], and with a motion verb as the first verb, [*I* GO_{1sg} and \emptyset VERB_{1sg}]. The crosslinguistic generalization lies in the existence of such constructions, in that particular constructions have also been revealed to play a role in subject expression in other languages (see Travis and Torres Cacoulllos, 2012:738–742 on Spanish), but at the same time, the constructions themselves are language specific.

Thus, on the one hand, despite the extremely low rate of unexpressed subjects in English, the structure of the variation follows some general cross-linguistic tendencies, such as priming. On the other hand, some constraints, such as that of prosodic-initial position, would appear to be particular to English, while the subject continuity constraint reported for other languages is bound to *and*-coordinating constructions and to priming. The systematic quantitative analysis of variation in speech thus enables shared and language-specific patterns to be discerned. More such studies are called for to allow for the characterization of viable cross-linguistic generalizations on subject realization.

Appendix. Transcription conventions (Du Bois et al., 1993)

.	final intonation contour	(H)	in-breath
,	continuing intonation contour	(Hx)	out-breath
?	appeal intonation contour	(TSK)	click
...	medium pause (>0.7 secs)	X	one syllable of unclear speech
..	short pause (about 0.5 secs)	@	one syllable of laughter
--	truncated intonation contour	@word	word produced while laughing
=	lengthening	%	glottal stop

¹³ Other potential candidates for particular constructions favoring unexpressed *I* in IU-initial position are with preverbal adverbials, [\emptyset *still* VERB-1sg] (e.g., *he took us .. around the city, ... \emptyset still got his car=d somewhere.* (60 Alan: 161)) and [\emptyset *just* VERB-1sg], the former with 3/3, the latter with 4/5 tokens occurring with an unexpressed subject.

!	booster: emphatic speech	<MRC MRC>	marcato voice
[]	speech overlap (numbers used to distinguish consecutive overlaps)	<< >>	action performed while speaking (e.g. stomping)
(CAPS)	vocal noise (e.g. clearing the throat, sniff, swallow)	<Q Q>	speech produced with quotative speech quality

References

- Akmajian, Adrian, Demers, Richard A., Farmer, Ann K., Harnish, Robert M., 2001. *Linguistics: An Introduction to Language and Communication*, 5th ed. MIT Press, Cambridge, MA.
- Ariel, Mira, 1988. Referring and accessibility. *J. Linguist.* 24 (1), 65–87.
- Barth-Weingarten, Dagmar, Couper Kuhlen, Elizabeth, 2011. Action, prosody and emergent constructions: the case of *and*. In: Peter, Auer, Pfänder, Stefan (Eds.), *Constructions: Emerging and Emergent*. de Gruyter, Berlin, pp. 236–292.
- Bentivoglio, Paola, 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Universidad Central de Venezuela, Caracas.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Finegan, Edward, Conrad, Susan, 1999. *The Longman Grammar of Spoken and Written English*. Longman, London.
- Bock, J. Kathryn, Griffin, Zenzi M., 2000. The persistence of structural priming: transient activation or implicit learning. *J. Exp. Psychol. Gen.* 129 (2), 177–192.
- Brinton, Laurel, 1990. The development of discourse markers in English. In: Jacek, Fisiak (Ed.), *Historical Linguistics and Philology*. Mouton de Gruyter, Berlin/New York, pp. 45–71.
- Bybee, Joan, 2010. *Language, Usage and Cognition*. Cambridge University Press, Cambridge.
- Cameron, Richard, 1994. Switch reference, verb class and priming in a variable syntax. In: *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on Variation in Linguistic Theory*, vol. 30(2) pp. 27–45.
- Cameron, Richard, Flores-Ferrán, Nydia, 2003. Perseveration of subject expression across regional dialects of Spanish. *Spanish Context* 1 (1), 41–65.
- Chafe, Wallace, 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago.
- Chociej, Joanna, 2011. *Polish Null Subjects: English Influence on Heritage Polish in Toronto*. Department of Linguistics, University of Toronto, Ms.
- Clancy, Patricia M., 1980. Referential choice in English and Japanese discourse. In: Wallace, Chafe (Ed.), *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Ablex, Norwood, NJ, pp. 127–202.
- Cote, Sharon A., 1996. *Grammatical and Discourse Properties of Null Arguments in English*. University of Pennsylvania, Philadelphia, PA, (Thesis Type thesis).
- Croft, William, 1995. Intonation units and grammatical structure. *Linguistics* 33, 839–882.
- Croft, William, Cruse, D. Alan, 2004. *Cognitive Linguistics*. Cambridge University Press, Cambridge.
- Dixon, R.M.W., 2005. *A Semantic Approach to English Grammar*, 2nd ed. Oxford University Press, Oxford.
- Du Bois, John W., Chafe, Wallace L., Myer, Charles, Thompson, Sandra A., Englebretson, Robert, Martey, Nii, 2000–2005. *Santa Barbara Corpus of Spoken American English, Parts 1–4*. Linguistic Data Consortium, Philadelphia.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, Paolino, Danae, 1993. Outline of discourse transcription. In: Jane, Edwards, Lampert, Martin (Eds.), *Talking Data: Transcription and Coding in Discourse*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 45–89.
- Ford, Cecilia E., Thompson, Sandra A., 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Elinor, Ochs, Schegloff, Emanuel A., Thompson, Sandra A. (Eds.), *Interaction and Grammar*. Cambridge University Press, Cambridge, pp. 134–184.
- Fox, Barbara, 1987. *Discourse structure and anaphora: written and conversational English*. Cambridge University Press, Cambridge.
- Fox, Barbara, Jasperson, Robert, 1995. A syntactic exploration of repair in English conversation. In: Philip, W. Davis (Ed.), *Alternative Linguistics: Descriptive and Theoretical Modes*. John Benjamins, Amsterdam, pp. 77–134.
- Givón, T. (Ed.), 1983a. *Topic Continuity in Discourse: A Quantitative Cross-Linguistic Study*. John Benjamins, Amsterdam.
- Givón, T., 1983b. Topic continuity in discourse: an introduction. In: Givón, T. (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Linguistic Study*. John Benjamins, Amsterdam, pp. 1–41.
- Givón, T., 1993. *English Grammar: A Function-Based Introduction*. John Benjamins, Amsterdam/Philadelphia.
- Haegeman, Liliane, 2002. Non-overt subject pronouns in written English. In: Sybil, Scholz, Klages, Monika, Hantson, Evelyn, Römer, Ute (Eds.), *Language, Context and Cognition: Papers in Honour of Wolf Dietrich Bald's 60th Birthday*. Langenscheidt-Longman, Munich, pp. 135–149.
- Haegeman, Liliane, Ihsane, Tabae, 1999. Subject ellipsis in embedded clauses in English. *Engl. Lang. Linguist.* 3 (1), 117–145.
- Harvie, Dawn., 1998. Null subject in English: wonder if it exists? *Cahiers Linguistiques d' Ottawa* 16, 15–25.
- Hopper, Paul J., 2001. Grammatical constructions and their discourse origins: prototype or family resemblance? In: Martin, Pütz, Niemeier, Susanne, Dirven, René (Eds.), *Applied Cognitive Linguistics I: Theory and Language Acquisition (Cognitive Linguistics Research)*. Mouton de Gruyter, Berlin/New York, pp. 109–129.
- Hopper, Paul J., 2002. Hendiadys and auxiliiation in English. In: Joan, Bybee, Noonan, Michael (Eds.), *Complex Sentences in Grammar and Discourse*. John Benjamins, Amsterdam, pp. 145–174.
- Huddleston, Rodney, 2002. The clause: complements. In: Rodney, Huddleston, Pullum, Geoffrey K. (Eds.), *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, pp. 213–322.
- Ihalainen, Ossi, 1991. The grammatical subject in educated and dialectal English: comparing the London-Lund Corpus and the Helsinki Corpus of Modern English Dialects. In: Stig, Johansson, Stenström, Anna Brita (Eds.), *English Computer Corpora: Selected Papers and Research Guide*. (Topics in English Linguistics) Mouton de Gruyter, Berlin, pp. 201–214.

- Labov, William, 2005. Quantitative reasoning in linguistics. In: Ulrich, Ammon, Dittmar, Norbert, Mattheier, Klaus J., Trudgill, Peter (Eds.), *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society*, 2nd ed., vol. 1. Mouton de Gruyter, Berlin, pp. 6–22.
- Lerner, Gene H., 1991. On the syntax of sentences-in-progress. *Lang. Soc.* 20, 441–458.
- Leroux, Martine, Jarmasz, Lidia-Gabriela, 2005. A study about nothing: null subjects as a diagnostic of the convergence between English and French. *Univ. Penn. Work. Pap. Linguist.* 12 (2), 1–14.
- Levinson, Stephen C., 1987. Pragmatics and the grammar of anaphora: a partial pragmatic reduction of binding and control phenomena. *J. Linguist.* 23 (2), 379–434.
- McKee, Rachel, Schembri, Adam, McKee, David, Johnston, Trevor, 2011. Variable “subject” presence in Australian Sign Language and New Zealand Sign Language. *Lang. Var. Change* 23, 375–398.
- Miller, Jim, 1995. Does spoken language have sentences? In: Palmer, F.R. (Ed.), *Grammar and Meaning: Essays in Honour of Sir John Lyons*. Cambridge University Press, Cambridge, pp. 116–135.
- Nagy, Naomi G., Aghdasi, Nina, Denis, Derek, Motut, Alexandra, 2011. Null subjects in heritage languages: contact effects in a cross-linguistic context. *Univ. Penn. Work. Pap. Linguist.* 17 (2), Article 16.
- Napoli, Donna Jo., 1982. Initial material deletion in English. *Glossa* 16, 85–111.
- Oh, Sun-Young, 2005. English zero anaphora as an interactional resource. *Res. Lang. Soc. Interact.* 38 (3), 267–302.
- Oh, Sun-Young, 2006. English zero anaphora as an interactional resource II. *Discourse Stud.* 8 (6), 817.
- Ono, Tsuyoshi, Thompson, Sandra A., 2003. Japanese (*w)atashi/ore/boku* ‘I’: they’re not just pronouns. *Cog. Linguist.* 14 (4), 321–347.
- Paredes Silva, Vera Lucia, 1993. Subject omission and functional compensation: evidence from written Brazilian Portuguese. *Lang. Var. Change* 5 (1), 35–49.
- Posio, Pekka, 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: semantic roles and formulaic sequences. *Folia Linguist.* 47 (1), 253–291.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, Svartvik, Jan, 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Roberts, Ian, Holmberg, Anders, 2010. Introduction: parameters in minimalist theory. In: Theresa, Biberauer, Holmberg, Anders, Roberts, Ian, Sheehan, Michelle (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge/New York, Cambridge University Press, pp. 1–57.
- Sankoff, David, 1988. Variable rules. In: Ulrich, Ammon, Dittmar, Norbert, Mattheier, Klaus J. (Eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*, vol. 2. Walter de Gruyter, Berlin/New York, pp. 984–997.
- Sankoff, David, Tagliamonte, Sali, Smith, Eric, 2012. Goldvarb LION: A Variable Rule Application for Macintosh. University of Toronto. <http://individual.utoronto.ca/tagliamonte/goldvarb.htm>.
- Silva-Corvalán, Carmen, 1982. Subject expression and placement in Mexican-American Spanish. In: Jon, Amastae, Olivares, Lucia Elías (Eds.), *Spanish in the United States: Sociolinguistic Aspects*. Cambridge University Press, New York, pp. 93–120.
- Silva-Corvalán, Carmen, 1994. *Language Contact and Change: Spanish in Los Angeles*. Clarendon Press, Oxford.
- Silva-Corvalán, Carmen, 1997. Avances en el estudio de la variación sintáctica: La expresión del sujeto. *Cuaderno del Sur: Letras, Homenaje a Beatriz Fontanella de Weinberg* 27, 35–49.
- Silva-Corvalán, Carmen, 2001. *Sociolingüística y pragmática del español* (Georgetown Studies in Spanish Linguistics). Georgetown University Press, Washington, DC.
- Stirling, Lesley, Huddleston, Rodney, 2002. Deixis and anaphora. In: Rodney Huddleston, Pullum, Geoffrey K. (Eds.), *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, pp. 1449–1564.
- Tao, Liang, 1996. Topic discontinuity and zero anaphora in Chinese discourse: cognitive strategies in discourse processing. In: Barbara, Fox (Ed.), *Studies in Anaphora*. John Benjamins, Amsterdam/Philadelphia, pp. 487–513.
- Thompson, Sandra A., 2002. ‘Object Complements’ and conversation: towards a realistic account. *Stud. Lang.* 26 (1), 125–163.
- Thrasher, Randolph, 1977. *One Way to Say More by Saying Less: A Study of So-Called Subjectless Sentences*, vol. 11. Kwansai Gakuin University Monograph.
- Torres Cacoulllos, Rena, Travis, Catherine. Subject pronoun realization in Spanish and English: Assessing inter-linguistic functional equivalence via intra-linguistic inherent variability. In: Ana, M. Carvalho, Orozco, Rafael, Lapidus Shin, Naomi (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective*. Georgetown University Press, Washington, DC (forthcoming).
- Travis, Catherine E., 2005. *Discourse Markers in Colombian Spanish: A Study in Polysemy* (Cognitive Linguistics Research). Mouton de Gruyter, Berlin/New York.
- Travis, Catherine E., 2007. Genre effects on subject expression in Spanish: priming in narrative and conversation. *Lang. Var. Change* 19 (2), 101–135.
- Travis, Catherine E., Torres Cacoulllos, Rena, 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cog. Linguist.* 23 (4), 711–748.
- Travis, Catherine E., Torres Cacoulllos, Rena. On the patterning of stressed / in conversation (in preparation).
- Weiner, E. Judith, Labov, William, 1983. Constraints on the agentless passive. *J. Linguist.* 19 (1), 29–58.
- Weir, Andrew, 2012. Left-edge deletion in English and subject omission in diaries. *Engl. Lang. Linguist.* 16 (1), 105–129.
- Zwicky, Arnold M., Pullum, Geoffrey K., 1983. Deleting named morphemes. *Lingua* 59, 155–175.

Rena Torres Cacoulllos is a Professor of Spanish and Linguistics at the Pennsylvania State University. Adopting a usage-based variationist approach, she studies language variation, processes of grammaticalization and the linguistic outcomes of language contact.

Catherine Travis is a Professor of Modern European Languages at the Australian National University. Her research focuses on the way in which grammar emerges through discourse, which she investigates through the study of variation in spontaneous speech, in monolingual and bilingual contexts.