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

Abstract

Unexpressed subjects, though rare, 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. Analyzing variation in main, declarative clauses—the locus of variability—we find that the apparently widespread cross-linguistic discourse constraint of subject continuity (switch reference) is bound to coordinating constructions with and, including particular constructions ([V and V-of-speech], [GO/V-of-motion and V]), and to an overarching priming constraint, whereby unexpressed mentions tend to cluster together. A pivotal language-particular restriction is prosodic, such that, outside of coordinating constructions, unexpressed subjects occur only in Intonation-Unit initial position. We therefore find that variable I expression is sensitive to discourse factors operative in subject expression in other languages and in language variation more generally, though paramount are prosodic considerations and local patterns that may be specific to 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 activation or accessibility (Ariel 1988:79; Chafe 1994:74; Givón 1983; Levinson 1987:384), alongside more interactional or pragmatic factors, related to the kind of action the utterance performs (Fox 1987; Oh 2005, 2006; Ono and Thompson 2003, inter alia). Quantitative analyses have confirmed accessibility effects in languages where subject expression is noticeably variable (such as Spanish, Portuguese, Italian, Russian, Cantonese, Javanese, Indonesian, Japanese, cf. Clancy 1980; Nagy et al. 2011; Paredes Silva 1993; Silva-Corvalán 2001 and papers in this volume), and have also revealed the mechanical factor of priming, or the perseveration of the same syntactic structure across utterances (Cameron and Flores-Ferrán 2003:50-54; Travis 2007:120-121; Travis and Torres Cacoullos 2012:729-732), as well as lexically particular constructions, with distinct subject expression patterns (Travis and Torres Cacoullos 2012-741). We know much less about languages where unexpressed subjects are comparatively rare, such as Swedish and Finnish, which are addressed in this volume, and English (though see, e.g., Clancy 1980; Cote 1996; Harvie 1998; Leroux and Jarmasz 2005; Oh 2005, 2006), which we consider in detail here.
2 Data for the study of I expression

The data for this study are drawn from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois 2000; Du Bois et al. 2003; Du Bois and Englebretson 2004, 2005), a publicly available corpus of recordings of naturally occurring face-to-face spoken interaction. The corpus comprises 60 transcripts, totaling approximately 249,000 words.

The data is transcribed in accordance with the method outlined in Du Bois et al. (1993), fundamental to which is the Intonation Unit (IU), characterized as “a stretch of speech uttered under a single, coherent intonation contour” (1993:47). Each IU is presented on a different line as illustrated in (1), where a comma represents continuing and a period final intonation contour; see the Appendix for transcription conventions.

Although the sentence has featured in publications on English unexpressed (or “null”) subjects based on grammaticality judgments (e.g. Haegeman 2002; Napoli 1982), not only is it questionable as a unit of speech (cf. Miller 1995:132), but our study will reveal that the prosodic unit is crucial to an understanding of the patterns of I expression.

We refer to the instances exemplified in lines 2 and 5 in (1) as unexpressed first singular subjects, which we indicate with a bolded Ø for ease of recognition.

(1)
1. Angela: ... (TSK) (H) and I put some onion powder ... in the mayonnaise,
2. ... Ø 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 Ø put it on some .. boiled eggs.

(11 This retirement bit: 759-763)

2.1 Extracting unexpressed 1sg subjects

All finite verbs that occurred with no subject (such as those marked with Ø in (1)) were extracted. We did not extract tokens of non-finite verbs, such as give in (2), 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.

(2) 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 Jasperson 1995), which occur either with a modification of what was said, or with truncation, indicated by --, as in (3); cases that were unclear as to who the subject was, where 1sg was a possibility (N=11) as in (4), where the subject could be first person singular or plural; cases where

---

1 Unexpressed I appears in 39 of the transcripts, produced by 60 speakers, most of whom produce one (N = 34) or two (N = 14) tokens. A total of 724 tokens of unexpressed subjects (all grammatical persons) from 123 speakers were extracted from the entirety of the SBCSAE 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).
the sound quality was unclear (N=5); and one token of a book title (“Don’t Know Much About History” 52, Cindy: 981).

(3) Phil: I did not --
.. Ø wasn’t really in support of one. 

(10 Letter of Concerns: 384-385)

(4) Marcia: cause all we did was,
... I was in here and Ø put a bandage on (Hx).

(18Vet morning: 255-256)

This procedure yielded 151 tokens of unexpressed I. The rate of expressed I in the transcripts with at least one unexpressed I is approximately 97.5% (out of some 6,600 instances of I), and just above 98% in all 60 SBCSAE transcripts (out of approximately 9,000 tokens of I). This compares with rates of expression for 1sg subjects of 21% in Polish (Chosiej 2011), 33% in Javanese (Ewing this volume), and between 34% and 48% in Spanish (Travis 2007:113). But, despite glaring overall rate differences, are there cross-linguistic similarities in patterns of speaker choices of an unexpressed vs. expressed subject, or the constraints on variable subject expression?

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. As far as we know, this has not been empirically determined 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 or (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), 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

---

2 This initial sample consisted of the 151 tokens of unexpressed I and 320 tokens of unstressed I extracted by Leah Houle, drawn evenly from16 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 Cacoullos Submitted).
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 was never 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).3 Discourse-marker status is operationalized based on position in the clause and in the Intonation Unit (as identified in the transcription). We coded as discourse markers occurrences of these collocations as parentheticals (when they appear between the subject and verb or following the verb), and tokens that were produced prosodically independently from other clausal material (those tokens that occurred strictly alone in their own Intonation Unit or were preceded only by a conjunction, cf. Thompson 2002:143-145; Travis 2005:48-52). Not only is I invariably expressed but, with the exception of I don’t know, it is generally unstressed (96%, 106/110) in these collocations when they are used as discourse markers (Travis and Torres Cacoullos Submitted). The lack of variation in the realization of I in these seven constructions supports their formulaic status. In contrast, instances of these collocations that occur with clausal material in the same IU show variation in I expression (and stress on I, Travis and Torres Cacoullos Submitted), as illustrated in examples (5) and (6) and thus are included in the variable context.

(5) Marcia: Ø Guess they gotta make money somehow, (18 Vet morning: 565)

(6) Alan: I guess it probably was, (60 Shaggy dog story: 790)

As with 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). (Likewise, Travis and Torres Cacoullos (Submitted) found this to be an environment that patterned like other formulaic contexts in disfavoring stress on I).

A third consideration is auxiliary contraction, as in (7), 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.

3 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)
The following two examples illustrate the application of this protocol. In (8), for the unexpressed subject in line 2, the I in line 1 was extracted, as the closest eligible preceding coreferential I; in (9), 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.

(8) 1. Curt: .. And then I put them in there.
2. .. Ø S—quashed em down,

(9) 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,
Following these protocols we extracted a total of 151 tokens of *I*. The rate of unexpressed subjects in this data set 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 set as extracted allows us to compare the frequency of the variants in sets of linguistic sub-contexts, which operationalize the hypothesized constraints on variable subject expression, as we outline in the following section.

3 Operationalizing hypotheses: coding for contextual features

3.1 Coordinated constructions with and

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

Here, we define coordinated clauses narrowly as coreferential clauses conjoined by *and* and occurring either adjacent to each other, as in (10), or with a single subordinate clause intervening, as in (11). We include tokens of *and* with a filler (e.g. *and uh, and um*) and one token of *and* repeated (*and, and*), but not *and* with more substantial material (e.g. *and then, and like, and at that point*). Other so-called coordinating conjunctions were far less frequent and may behave differently. In all cases where there was an intervening subordinate clause, the conjoined clause continues on from what was being said in the coreferential main clause. Marking such a continuation is one function for unexpressed subjects that has been proposed in the literature (e.g., Oh 2006:822ff, on English; Tao 1996:507, on Chinese). There were no tokens with more than one subordinate clause (and no main clause) intervening between coreferential mentions.

(10)

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

(02 Lambada: 308)

4 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 not all transcripts have *I* tagged for stress. However, we can assume that the great majority of the expressed tokens are unestressed, since this is the most frequent realization of expressed *I* (87%, 1,651/1,861 (Travis and Torres Cacoullos Submitted)).

5 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, or with other substantial material (e.g. *but then*), or at the beginning of an IU following final intonation contour, e.g., line 13 in (13)).
(11) 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)

3.2 Subject continuity: coreferential vs. switch reference contexts

According to Givón’s (1983) cross-linguistic topic continuity/accessibility hierarchy, the most continuous or accessible participants receive less linguistic coding, and the most discontinuous or least 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 the number of intervening clauses between the previous coreferential mention as subject and the target token.

Counted as the previous coreferential mention as subject were Is 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 discourse formulas identified in Section 2.1 (discourse markers and quotatives), these were not considered to be coreferential mentions. For example in (12) for the token Ø stop in line 5, the previous mention is I get out in line 4; in turn, for the token I get out, we skip quotative I’m like in line 1 and go to I went out as the previous mention.

(12)
1. Lajuan: (H) [so I went out in the] morning I’m like,
2. Cam: [Oh <% Go=d %>].
3. Lajuan: <VOX (H)= VOX>,
4. .. so I get out of my car,
5. Ø stop in the driveway @@@,

(44 He knows: 34-38)

In order to obtain a measure of “recency” of previous mention, we then proceeded to count the number of intervening clauses, counting all finite verbs, except for: the discourse formulas noted above; other fixed expressions with questionable clausal status (such as you know, let’s see, you bet; it’s like, it’s really, that’s right, it’s because with non-referential subjects); and first-person cognitive verbs (I know, I guess, I see) in their own IU uttered by the interlocutor as a backchannel. Note that, in some cases, this results in a markedly different clause count than if all clauses were counted equally. For example, in (12), distance from the previous mention for both tokens, Ø stop and I get out, is 0 intervening clauses (not counting the formula I’m like). In (13), for the token I really enjoy that in line 13, the distance from the previous mention, I get in line 1, is three intervening clauses, not counting it’s because (line 2), And that’s, (line 6), or you know (line 8). Clauses produced in overlap with the target clause were not counted as intervening clauses. For example, in (14), the token I wrote in line 4 was coded as occurring at a distance of 0 intervening clauses from the previous mention in line 1, I stayed, because the interlocutor’s clause overlaps with I wrote, as indicated by the [ ].
We refer to contexts of 0 intervening clauses (thus defined) as coreferential and those of 1 or more intervening clauses as switch reference (following what has been applied to Spanish, cf. Cameron 1994:28-32; Silva-Corvalán 1982:104).

3.3 Positioning of the subject in the prosodic unit

A number of reference grammars and generative treatments have proposed a prosodic constraint on the occurrence of unexpressed subjects. For instance, Quirk et al. (1985:896 (§12.46)) describe “elided” material as occurring in utterance-initial position with weak accent and stress, and Napoli (1982:99) proposes a phonological rule deleting unstressed material “not preceded by any other phonetic material”. To test whether this general “left-edge” deletion (Weir 2012) plays a role in I expression, we used the Intonation Unit (IU), as marked in the transcription.

Instances were classified as occurring or not in absolute-initial position in the IU. Non-IU initial tokens were those that occurred following clausal material (as for coordinated Ø saw in (15)), or following a discourse marker or conjunction (e.g. so I eat in (16), and Ø explained in (11)). IU-initial tokens were those that occurred without any preceding material (e.g. I went in line 7 in (15), and Ø stop in line 5 in (12)).
3.4 Priming

A strong priming effect, whereby speakers tend to repeat the same form they have used previously, has been observed for subject realization, both for patterns of stress on I in English (Travis and Torres Cacoullos Submitted), and for 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). Here we test this effect for coreferential subjects that occur at a distance of up to five intervening clauses,6 through consideration of the form of the previous coreferential subject. Examples of priming are seen in (22) – (24) below.

3.5 Other factors tested

Finally, other contextual features that are known to play a role in the patterning of subject expression were examined. We considered semantic class, since cognitive verbs have been found to strongly favor expressed first singular subject in varieties of Spanish (e.g., Bentivoglio 1987:60; Silva-Corvalán 1994:162; Travis 2007:116-117); tense, since backgrounding TAMs (such as past imperfective) have been found to favor expression (Silva-Corvalán 1997); position in the turn, a speaker-turn effect observed with first-person singular cognitive verbs in Spanish (Travis and Torres Cacoullos 2012:736-737); and polarity, in order to compare with stress on I, which is favored in denials (Travis and Torres Cacoullos Submitted). None of these turn out to be statistically significant (as seen below in Table 1).

4 The structure of variable first singular subject expression

We ascertain the structure of variation in first singular subject expression by submitting the data to multivariate analysis, which considers together the predictors (factor groups) outlined in the preceding section. Table 1 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). The three predictors that were retained through the analysis as significant are coordinating construction, position in prosodic unit and previous realization. Each predictor 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 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.

---

6 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 Cacoullos 2012).
Table 1: Variable-rule analysis of the contribution of factors selected as significant to the choice of unexpressed I in conversational American English

<table>
<thead>
<tr>
<th></th>
<th>Prob</th>
<th>% Ø</th>
<th>N</th>
<th>% data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coordinating construction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coordinated verbs with <em>and</em></td>
<td>.96</td>
<td>88%</td>
<td>73/83</td>
<td>28%</td>
</tr>
<tr>
<td>Non-coordinated verbs</td>
<td>.24</td>
<td>36%</td>
<td>78/219</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Prosodic position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU-initial</td>
<td>.69</td>
<td>46%</td>
<td>88/191</td>
<td>63%</td>
</tr>
<tr>
<td>Non-IU-initial</td>
<td>.21</td>
<td>57%</td>
<td>63/111</td>
<td>37%</td>
</tr>
<tr>
<td><strong>Realization of previous coref 1sg subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpressed</td>
<td>.86</td>
<td>84%</td>
<td>27/32</td>
<td>11%</td>
</tr>
<tr>
<td>Pronoun (I)</td>
<td>.43</td>
<td>50%</td>
<td>104/209</td>
<td>69%</td>
</tr>
</tbody>
</table>

Non-significant factor groups: Coreferentiality, Turn position, Polarity, Semantic class of verb, Tense.

*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 at a distance of five or fewer intervening clauses (N = 241)

4.1 and coordinated 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%). As detailed above, we consider to be coordinated those verbs that are conjoined to a previous one with *and* in both coreferential contexts (as in the bolded tokens in (17)), and in non-coreferential contexts where there is only one subordinate clause intervening, seen in (11) above. All but five of the coordinated verbs have coreferential subjects (i.e. have no intervening subordinate clause), that is they occur in the schema \([V-1SG_i \text{ and } V-1SG_i]\).

These coordinated constructions may be more, or less, tightly linked, as measured prosodically by occurrence in the same, or in different, IUs (on the relationship between syntactic, semantic and prosodic integration, see Chafe 1994: Ch. 9; Ono and Thompson 1995: 233-246; Sánchez-Ayala 2001). Close to 60% (46/78) of the time, the schema \([V-1SG_i \text{ and } V-1SG_i]\) occurs across IUs, as seen in (17): in both lines 2-3 and lines 3-4 the first conjunct occurs in a different IU; in line 3 the subject is unexpressed and in line 4 it is expressed. (Similar variability can be seen in lines 3 and 5 in example (1)). The rate of unexpression in this context is 80% (41/51), substantially higher than the overall average.

(17)
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*,

(07 A Tree’s Life: 546-552)
For the remaining 32 \([V-1\text{SG}, \text{and} V-1\text{SG}]\) tokens, both verbs appear in the same IU (as seen in (10) above). In our sample, all 32 tokens have an unexpressed subject in the second conjunct. These conjuncts could be interpreted as a “single event” (cf. Givón’s (1993:7) “Temporal contiguity and event integration”). Outside our sample (see Section 2.3), we do find tokens with expressed subjects in this context, but in this case, the two V slots appear to represent distinct events. For example, in line 6 of example (17) above, the situations are sequential: “I looked” is temporally subsequent to “I turned around”. In (18) below, while the situations presented in the first IU may be temporally simultaneous, the speaker elaborates on \(\text{I see}\) in the following IU, indicating that ‘seeing’ is an event in its own right.

(18) Alina: .. (H) So \textbf{I walk in and I see},
.. two of the .. paddlers I definitely don't wanna see.

(06 Cuz: 920-921)

In summary, \textit{and} coordination very strongly favors unexpressed \(I\), overwhelmingly so when the second conjunct is in the same IU as the first.

4.2 \textit{Prosodic constraint for non-coordinated verbs: 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 explains this interaction, through a cross-tabulation of Prosodic position and Coordinating construction. 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 \textit{and} is in the preceding IU) as in non-IU-initial position (either when the verb is preceded by \textit{and} alone (N=37) or by the first conjunct (N=32) on the same IU (as discussed in the preceding section)).

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

<table>
<thead>
<tr>
<th>Position in IU</th>
<th>Coordinated verbs (\textit{and} V)</th>
<th>Non-coordinated verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% N unexpressed (total)</td>
<td>% N unexpressed (total)</td>
</tr>
<tr>
<td>IU-initial</td>
<td>86% 12/14*</td>
<td>42% 76/177</td>
</tr>
<tr>
<td>Not IU-initial</td>
<td>88% 61/69</td>
<td>5% 2/42**</td>
</tr>
<tr>
<td>TOTAL</td>
<td>83</td>
<td>219</td>
</tr>
</tbody>
</table>

*\textit{and} occurs in the preceding IU
**Both unexpressed tokens are preceded by \textit{and} on the IU and are cases of coordination in a broad sense (with intervening main clauses).

For non-coordinated verbs, however, depicted on the right-hand side of the table, \(I\) is unexpressed 42% of the time in IU-initial position, but a mere 5% of the time in non-
IU initial position. The two tokens of unexpressed I with non-coordinated verbs in non-IU-initial position are preceded by and in the IU, and could in fact be considered cases of coordination in a broader definition than the one adopted here (Vs conjoined with and with no more than a subordinate clause separating them, Section 3.1), in that the intervening material could be interpreted as a parenthetical insert (cf. Oh 2006:823). Both tokens occur in the same conversation, by the same speaker; one example is given in (19).

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

(32 Handshakes all around: 1713-1717)

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. Note that 1sg subjects in non-IU-initial position represent less than one-fifth (42/219) of all 1sg occurrences in non-coordinated contexts. Most of these non IU-initial tokens (virtually always expressed) occur following so (N = 11), as in (18), (and) then (N = 7), as in line 4 in (14), but (then, it was like) (22), as in line 13 in (13), or okay, no, well, cause.

Might this overwhelming IU-initial prosodic constraint be masking a syntactic one? That the constraint is prosodic rather than syntactic may be inferred from reports that clause-initial position favors unexpressed subjects, but as a variable constraint (Harvie 1998:23; Leroux and Jarmasz 2005:7), not a (near-)categorical restriction, as we find here for IU-initial position. To examine the relationship between initial position with respect to units of prosody and units of content, we coded all non-coordinated instances (N=219) for whether the beginning of the IU was also the beginning of a sentence. For the sake of replicability in delimiting sentences, the sentence was operationalized, following Chafe (1994:139), as a section of speech that is prosodically bound by final or rising intonation (marked by a period or a question mark). Thus defined, the “prosodic sentence” may consist of a single IU (as in line 3 in (20)), or a series of IUs with continuing intonation (marked by a comma) for all but the final IU (as for lines 1 / 2 and 4 / 5 in (20)). Where there were substantive contributions by other speakers, the beginning of the sentence was the beginning of the speaker’s turn. Also considered to be cases where the beginning of the IU coincides with the beginning of the sentence are those 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 (20) would be coded as sentence initial). 7

We find that IU-initial position for first-singular subjects coincides with sentence-initial in only approximately half of the IU-initial cases (52%, 92/177). 8 The other half of the time, the IU-initial (expressed or unexpressed) I is not sentence-initial, but follows continuing intonation contour, as in line 5 in (20), and line 5 in (21). Furthermore, the

7 There were no cases of co-constructions (cf. Lerner 1991), where the speaker was clearly finishing off someone else’s sentence.
8 A sentence-initial subject, as sentence is defined here, is necessarily IU-initial.
rate of unexpressed 1st IU-initially is not significantly different following period intonation from its rate following continuing intonation contour (46%, 42/92 and 40%, 34/85 respectively, \( p = .5434 \)). This bolsters the view that the positioning constraint is indeed prosodic rather than syntactic.

\[(20)\]
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. \[\text{(42 Stay out of it: 211-213)}\]

\[(21)\]
1. Paige: Mhm.
2. ... ØOh,
3. Kristin: \([\text{U=m}],\)
4. Paige: \([\text{that was just cause}] \) I was like,
5. ... Ø couldn't eat dinner (Hx). \[\text{(41 X units of insulin: 429-433)}\]

In summary, for non-coordinated verbs, a prosodic constraint is such that unexpressed 1st subjects occur (in variation with expressed 1st) in absolute IU-initial position only. Virtually the only time we get an unexpressed 1st subject IU-externally is with coordinated verbs following the conjunction \textit{and}, whether the first conjunct is in the same or a preceding IU.

\subsection*{4.3 Priming}

Finally, from the multivariate analysis we observe an effect for the realization of the previous coreferential mention: highly conducive to unexpressed 1st is previous realization as unexpressed (seen in the high Probability of .86). This is a structural priming or perseveration effect, pervasive in linguistic variation, whereby speakers tend to repeat the same structure that has occurred previously (cf. Bock 1986; Labov 1994: 547-568; Poplack 1980; Scherre and Naro 1991; Weiner and Labov 1983, inter alia). The finding here mirrors the effect noted above (Section 3.4) for subject realization in Spanish (in terms of expression) and English (in terms of stress).

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 produce only one unexpressed 1st subject per transcript. But when more than one unexpressed subject is produced, they tend to cluster together. This is illustrated in (22), with non-coordinated IU-initial verbs, and in examples (23) and (24), with coordinated (V and V, different IU) verbs. These are a near “minimal pair”, in that the same speaker selects \textit{I and I} in the first example, Ø and Ø in the second, in accordance with priming.
(22) Miles: (H) But it was like I went [to] Bahia, 
     JAMIE: [What’s] --

... last Sunday,
(H) ... Ø got there at eight,
(Hx) .. Ø left a te=n,
... Ø dropped this person off at home, in Foster City,

(02 Lambada: 848-854)

(23) Wess: and then I put it on,
     and I boil it.

(59 You baked: 1211-1212)

(24) Wess: and Ø put it in there,
     Cam: [@ @@@] Wess: [and] Ø beat it a little bit,

(59 You baked: 1228-1230)

Thus, priming is operative for both coordinated and non-coordinated verbs, which are presented separately in Table 3. As can be seen, for coordinated verbs, I is categorically unexpressed (12/12) when the previous coreferential 1sg subject was also unexpressed, compared with 86% when it was expressed. For non-coordinated verbs, the priming effect is more striking, with a rate of unexpressed I of 75% when the previous realization was unexpressed but only half that (32%) when the previous realization was expressed (though the numbers are sparser in this context).

Table 3: Rate of unexpressed 1sg subject by previous realization (N=241)

<table>
<thead>
<tr>
<th>Previous realization</th>
<th>Coordinated verbs (and V)</th>
<th>Non-coordinated verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>N unexpressed (total)</td>
</tr>
<tr>
<td>Unexpressed (Ø)</td>
<td>100%</td>
<td>12/12</td>
</tr>
<tr>
<td>Expressed (I)</td>
<td>86%</td>
<td>60/70</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>82</td>
</tr>
</tbody>
</table>

4.4 The lack of a coreferentiality effect

What of the cross-linguistically applicable switch reference constraint amply reported in the literature? This was not retained as a significant contributor to speaker choice of an unexpressed I when all predictors were considered together in the multivariate analysis shown in Table 1. This is because for coordinated (V and V) verbs switch reference is nearly moot: as noted above (Section 4.1), most coordinated instances (94%, 78/83)

9 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 .85 (14/17) – Previous realization expressed .43 (43/110)). Coreferentiality 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.

9
occur in coreferential contexts ($V$-1$\text{SG}_i$ and $V$-1$\text{SG}_i$), even when in a different IU from the first conjunct (92%, 47/51).

Switch reference is not really an independent effect for non-coordinated IU-initial verbs either. Table 4 gives the rate of unexpressed $I$ by previous realization and coreferentiality. From the Totals column, the tendency appears to be the predicted one, with a lower rate of unexpressed $I$ under switch reference (37% vs. 51%). However, with the cross-tabulation we see that priming operates in coreferential contexts, where 1$\text{SG}$ is more likely to be unexpressed if the preceding subject was also unexpressed (93%, 14/15) than if it were expressed (40%, 24/60) ($p = 0.0003$). Further, we see from the left-hand column that (though the data are sparse), unexpressed-to-unexpressed priming tends to apply in coreferential contexts: most (15 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 right-hand column, coreferentiality makes no difference, with a rate of expression of 38-40% in both switch-reference and coreferential contexts. Thus, the appearance of an effect for switch reference is bound to unexpressed-to-unexpressed priming; since this tends to occur in coreferential contexts, the rate of unexpressed 1$\text{SG}$ subjects is raised in this context.

Table 4: Rate of IU-initial unexpressed 1$\text{SG}$ subject by previous realization and switch reference, for IU-initial (non-coordinated) verbs (N=127)

<table>
<thead>
<tr>
<th>Realization of previous coreferential 1$\text{SG}$</th>
<th>Unexpressed (N=17)</th>
<th>Expressed ($I$) (N=110)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreferential</td>
<td>93%</td>
<td>14/15</td>
<td>40%</td>
</tr>
<tr>
<td>Switch reference**</td>
<td>0/2</td>
<td>38%</td>
<td>19/50</td>
</tr>
<tr>
<td>TOTAL</td>
<td>82%</td>
<td>14/17</td>
<td>39%</td>
</tr>
</tbody>
</table>

**Regardless of distance from the previous coreferential first singular subject as measured in the number of intervening clauses (Section 3.2).

5 \hspace{1cm} \textbf{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 \textit{and} (25); in IU-initial position (indeed, the only cases of unexpressed $I$ in non-initial prosodic position are following \textit{and}) (26); and when the previous coreferential mention was also unexpressed (a constraint which holds for both coordinated and non-coordinated verbs) (27).
Coordinated constructions: \[ [I\ V-1SG, and \ Ø\ V-1SG_i] \] (conjuncts on the same or different IUs)

IU-initial position: \[ [Ø\ V-1SG \ldots.]_{IU} \]

Priming: \[ [Ø\ V-1SG_i\ (and)\ Ø\ V-1SG_i] \]

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 3.1), we observed the local pattern of two verbs with a coreferential 1sg subject in a coordinated 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.

coreferential same-IU, same tense coordinating construction:
\[ [I\ V-1SG, and\ Ø\ V-1SG_i]_{same\ IU, same\ tense} \]

Associated with this construction are two more lexically specific constructions which together, make up 88% (28/32) of the \([I\ V-1SG_i\ and\ Ø\ V-1SG_i]_{same\ IU, same\ tense}\) constructions. One involves a quotative or a verb of speech, and the other a verb of motion.

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, same tense construction depicted in (28). Other examples are I came home and Ø told you. (42, Kendra: 111), (H) I wrote home to my family and Ø said, (32, Tom_2: 511), I look at that and Ø think, (+ quoted thought; 02, Miles: 526).

Angela: (H) I phoned her and Ø said,

Cases with a motion verb occur most frequently with the verb go (N = 9), as is depicted in (30). Other examples are Ten after two I went over and Ø got her, (43, Alice: 328), And even if I go out and Ø a=sk for it. (17, Jim: 7), That’s why I went out and Ø 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 Ø 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: Chapter 4), go.
Finally, there also exist potential candidates for particular constructions which appear to favor unexpressed I in IU-initial position, based on certain preverbal adverbials (though the numbers are sparse). One such candidate is [Ø just V-1sg], with four of five tokens occurring with an unexpressed subject (3/3 IU-initially, 1/2 with a conjoined verb) (31). Another is [Ø still Verb-1sg], with all three IU-initial tokens occurring with an unexpressed subject (32). In contrast, the one other adverbial with more than one token, never (even), appears with an unexpressed subject in only three of seven cases (2/5 IU-initial, 1/2 with a conjoined verb).

(31)  
[Ø just V-1sg]_IU  
Dana: I don't feel like eating a sandwich.  
Nancy: I am [so full].  
Dana: [Ø just ..] .. bought all this food.  

(50 Just wanna hang: 720-722)

(32)  
[Ø still V-1sg]_IU  
Alan: he took us .. around the city,  
... Ø still got his car= elsewhere.  

(60 Shaggy dog story: 160-161)

6 Conclusion

Unexpressed subjects are indeed rare in English. We identified here an overall rate of expression of approximately 97.5% when comparing all tokens of unexpressed I with all tokens of expressed I in the same transcripts. However, going beyond rates to examine the structure of variability, we established 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-marker uses of highly frequent collocations, such as I mean, and contracted auxiliary forms. As far as we are aware, this is the first study to explore and empirically determine the variable context for subject expression in English yet, as we have shown, this is fundamental if we are to understand the variability in English, and from there to compare it with cross-linguistic tendencies in subject expression.

So, how does the patterning of 1sg expression in English compare with that in other languages? We find that English is sensitive to more generally applicable constraints related to priming and, at first blush, to the constraint on subject expression related to coreferentiality, widely reported for other languages with more robust variation. When we look more closely, we find that these general constraints apply in quite specific ways.

Firstly, we identify two distinct environments—coordinated and non-coordinated contexts. Beyond the doubly high rate of unexpressed I in coordinated contexts, these contexts are subject to a different prosodic constraint. With coordinated verbs, the rate of unexpressed I is even higher when the conjoined clauses occur in the same IU than when they occur in different IUs. 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 that could only have been identified through the prosodic detail captured in the SBCSAE transcriptions.

What of priming and coreferentiality? A mechanical priming constraint applies across the board, 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 cross-linguistic switch reference constraint on subject expression appears to apply, with a higher rate of unexpressed subjects in coreferential contexts. But the multivariate analysis shows that 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 V-1SGi and Ø V-1SGi], to the more particular manifestation where this occurs in the one IU and representing the one event. Within this we find the lexically specific constructions, with quotative as the second verb, [I V-1SGi and Ø V-of-speech-1SGi], and with motion verb as the first verb, [I GO (+ particle)-1SGi and Ø V-1SGi]. 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. Particular constructions have also been revealed to play a role in subject expression in other languages (see Travis and Torres Cacoullos 2012:738-742 on Spanish), though 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 coreferentiality constraint reported for other languages is bound to and-coordinating constructions and to priming for IU-initial (non-coordinated) verbs. Further empirical studies establishing the structure of variability are therefore needed to enable cross-linguistic generalizations about subject expression.
Appendix

Transcription Conventions (Du Bois et al. 1993)

- final intonation contour (H) in-breath
- continuing intonation contour (Hx) out-breath
- appeal intonation contour ! booster: emphatic speech
- medium pause (> 0.7 secs) X one syllable of unclear speech
- short pause (about 0.5 secs) @ one syllable of laughter
- truncated intonation contour % glottal stop
- lengthening [ ] speech overlap
- (TSK) click <MRC MRC> marcatto voice
- (CAPS) vocal noise (e.g. clearing the throat, sniff, swallow) <VOX VOX> speech produced with marked voice quality
- <<CAPS action performed while speaking (e.g. stomping)

References


Thrasher, Randolph. 1977. One way to say more by saying less: A study of so-called subjectless sentences, vol. 11. Kwansei Gakuin University Monograph.


