CrossMark

Language Variation and Change, **28** (2016), 103–128.
© Cambridge University Press, 2016 0954-3945/16
doi:10.1017/S0954394515000174

Different registers, different grammars? Subject expression in English conversation and narrative

CATHERINE E. TRAVIS
Australian National University and
ARC Centre of Excellence for the Dynamics of Language

AMY M. LINDSTROM
University of New Mexico

ABSTRACT

As a so-called non-null subject language, it has been proposed that in English, unexpressed subjects occur only in registers that have specific grammatical properties. We test this hypothesis through a comparison of the conditioning of subject expression for third-person singular human specific subjects in English conversation and narrative. Despite a stark difference in the rates of nonexpression (4% in conversation vs. 22% in narratives), there is no evidence of different grammars across the registers—in both, outside of coreferential clauses conjoined with a coordinating conjunction, unexpressed subjects only occur in prosodic initial position in main clause declaratives. Within the variable context, in both registers, expression is sensitive to accessibility, priming, and temporal sequentiality. A register effect is, however, evident in the contextual distribution, with a larger proportion of the narrative tokens occurring in contexts propitious to unexpressed subjects, and it is this that accounts for the higher rate of nonexpression in this register.

It has been said that unexpressed subjects in English are "restricted to certain types of discourse and/or register" (Roberts & Holmberg, 2010:5), and that the constraints on subject expression in registers that do allow it in English (e.g., "diary drop") differ from those in canonical "null subject languages" (e.g., Haegeman, 2013:93–94). As yet, however, there has been no quantitative evidence adduced in support of these claims. Here we address this, through a comparison of the factors conditioning the variation in pronominal and unexpressed third-person singular (3sg) human specific subjects in two spoken registers of English: interactional conversation and monologic narratives.

Overall, the two registers studied present vastly different rates of expression: roughly 4% (153 of approx. 3500) of the 3sg human specific subjects are

We are extremely grateful to two anonymous *LVC* reviewers for their valuable feedback on this paper, and also to Naomi Shin, Diane Massam, and Timothy Stowell for comments on an earlier draft.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

unexpressed in the conversational registers, compared with 22% (165 of 748) in the narrative data. The very low rate of nonexpression in conversation is of little surprise for a language typically classified as having obligatory subject expression (e.g., Dryer, 2011; Roberts & Holmberg, 2010:4). The rate of nearly one quarter unexpressed subjects in the narratives is, however, remarkably high for such a characterization. If we were to categorize language types on the basis of rates alone, these figures might suggest that English narrative is a kind of "null subject" register, while conversation is not. They may also lead some to wonder whether these two registers are operating under different grammars, or as Weir (2012:116) put it, there is a "reranking" of constraints across the registers (cf. Walkden, 2013:174). Such an understanding would be in accordance with Biber and Conrad's (2001:176) proposal that there exist "systematic and important linguistic differences across registers." However, a difference in rates of use alone cannot be taken as evidence of differences in grammar, as rates are affected by extralinguistic factors, including, crucially for our purposes, register (e.g., Silva-Corvalán, 2001:163; Travis, 2007). Hypothesized register differences can instead be ascertained through comparisons of factors conditioning the variability (see Szmrecsanyi & Hinrichs, 2008, for an illustration).

Thus, here we look beyond rates to compare the conditioning of subject expression in the two registers. We consider first the contexts where unexpressed 3sg human specific subjects do not occur (that is, categorical constraints), and then examine the linguistic conditioning of expression in the remaining contexts where unexpressed and pronominal subjects exist in variation (probabilistic constraints). The results demonstrate that the two registers are subject to identical constraints, including those that are categorical and those that are probabilistic, thus supporting Guy's (2005:562) claim that "the grammar is unchanged in stylistic variation."

Furthermore, we compare these constraints with those that have been identified in so-called null subject languages, to examine whether unexpressed subjects in English have "special properties which distinguish them from the canonical null subjects" (Roberts & Holmberg, 2010:5). We find that subject expression in English has a narrower envelope of variation than null subject languages, with unexpressed subjects occurring only in coreferential clauses conjoined with a coordinating conjunction, and outside of that, in prosodic initial position in main clause declaratives. However, within the envelope of variation, English is not "special" at all, and the constraints in operation (accessibility, priming, temporal sequentiality, and coordination) are parallel to those widely described for null subject languages, suggesting a need for further exploration into the widely accepted distinction between "null" and "non-null" subject languages.

UNEXPRESSED SUBJECTS IN CONVERSATION AND NARRATIVE

The data for this study come from two spoken corpora representing distinct registers, interactional conversation from the Santa Barbara Corpus of Spoken American English (SBCSAE, Du Bois, Chafe, Myer, Thompson, Englebretson

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

& Martey, 2000–2005), and monologic narratives from the transcriptions of the Pear Stories in English (Chafe, 1980).

The SBCSAE is a publically available corpus of transcriptions of 60 recordings, made between 1988 and 1996. The recordings consist of spontaneous spoken interaction between people of different sociodemographic backgrounds from across the USA. Most of the transcriptions that make up the corpus involve face-to-face interaction among groups of two to four friends, family members, and acquaintances. The corpus consists of a total of 249,000 words; in this study, we make use of the 34 transcripts that comprise interactional discourse and have tokens of unexpressed 3sg human specific subjects. This subcorpus totals 162,000 words, with individual transcripts ranging from approximately 3000 to 8000 words.[1]

The Pear Story corpus consists of transcriptions from recordings made in 1975 of 20 participants (US undergraduate native speakers of English) recounting a 6-min silent film that they were shown and asked to describe. Participants narrate the story to an interviewer, who claims not to have seen the film. The corpus totals 16,000 words, with the individual narratives ranging from approximately 300 to 1750 words. All transcripts present tokens of unexpressed 3sg human specific subjects, and thus all were included in this study.

Both of these corpora have been transcribed following the principles of the method outlined in Du Bois, Schuetze-Coburn, Cumming, and Paolino (1993). Central to this method is the Intonation Unit (IU), characterized as "a stretch of speech uttered under a single, coherent intonation contour" (Du Bois et al., 1993:47). Each IU is presented on a single line in the transcription, with punctuation used to mark the prosodic contour (or transitional continuity) at the end of the unit (see the appendix for the transcription conventions for each dataset). The IU corresponds largely to the Intonational Phrase (Pierrehumbert & Hirschberg, 1990) or the tone group (Halliday, 1967), though it is defined on the basis of the boundaries of the unit rather than its internal structure (cf. Chafe, 1994:58–59). It has been argued to be a basic unit of spoken discourse, where the validity of the "sentence" has been questioned (cf. Croft, 1995:875; Izre'el, 2005). The IU is particularly important to the variable under study here, as an "IU-initial constraint" (Torres Cacoullos & Travis, 2014:28–29) has been put forward as an alternative account of the restriction of unexpressed subjects to "sentence-/clause-initial" position (reported in Harvie [1998:21]; Leroux & Jarmasz [2005:4]; Roberts & Holmberg [2010:90]; though see Napoli [1982] and Weir [2012] on a proposed "initial material" or "left-edge deletion" constraint).

The conversational data included here are characterized by a large amount of interaction (turn-taking, question-and-answer sequences, overlaps), as seen in (1). The narrative data, on the other hand, are entirely monologic, as participants recount the series of events that occurred in the film (with occasional metalinguistic commentary on the film), as seen in (2). All 3sg subjects have been marked in **bold,** and tokens of unexpressed subjects are indicated with Ø.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

(1) 1. Jeff: (H) ~Ro=b (H) … decided that **he just needed** a vacation or something?
2. Jill: Unhunh.
3. Jeff: So=%,
4. … (TSK) (H) **he= .. took off** for .. Big Bear.
5. Jill: … You're kidding.
6. Jeff: **Ø Had** no idea where he was going,
7. [**he just took**] --
8. Jill: [All by] himself?
9. Jeff: All by himself,
10. a=nd,
11. (H) **he found** a cabin,
12. **he called** me?

(SBCSAE 28:628–639)[2]

(2) 1. [6.35[1.1] U--m [.7] then [.4] u--h [2.1]] a .. girl on a bicycle,
2. [1.15] comes riding towards him,
3. .. in the opposite direction.
4. [.9] And as they pass,
5. [.45] **he sort of turns**,
6. and **Ø looks** at her,
7. [.6] and his hat flies off,
8. [2.1[.2] a--nd [1.2]] his [.3] bicycle hits a rock.
9. .. Because **he's looking** at the girl.
10. [.75] **Ø Falls over**,
11. 1.5[1.35] uh] there's no conversation in this movie.

(Pear 6:10.1–13.2)

Third-person singular human specific subjects, the focus of study here, represent the bulk of the subjects in the Pear narratives, and the bulk of the third-person unexpressed subjects in both datasets. While some studies have considered the third person alongside other persons (for English, see Cote, 1996; Harvie, 1998; Leroux & Jarmasz, 2005; Shin & Montes-Alcalá, 2014), focusing on just one person may permit a fuller understanding of the conditioning, as it is yet to be established whether all subject personal pronouns are affected by the same constraints. For example, issues related to accessibility are likely to apply differently to the third person as compared with discourse participants (first and second person) (cf. Chafe, 1994:79), and this may manifest itself in subject expression. In Finnish, for instance, first and second person, but not third person, are reported to be generally left unexpressed (Helasuvo, 2014:454). Third person itself has a wide range, including animate and inanimate, human and nonhuman, as well as specific and nonspecific referents, such that Scheibman (2001:163) suggested that "third person singular subject is not a coherent category in English conversation." Hence, we restrict the analysis to what can be assumed to be a coherent category: human specific subjects.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

We extracted all instances of finite verbs with an unexpressed 3sg human specific subject in both datasets. We do not count as unexpressed tokens that occurred in the context of repair (3), where there is a potential subject in a preceding IU (4), nor in cases of what could be considered "subject-aux deletion" (cf. Harvie, 1998:19) (for example, for *ask* in *he'll turn around and ask*). We do consider to be unexpressed, but exclude from the study, a small number of tokens where the transcript is unclear, or it is not possible to conclusively determine whether the unexpressed subject has a 3sg human specific referent.

(3) potential unexpressed subject in the context of repair:
And u--m[1.35]] this man came by,
[.3] **walked by**,
(Pear 14:17.1–17.2)

(4) potential subject produced earlier in the discourse:
Miles: Well **!Sue !Swing**,
I mean when she was teaching **was** definitely more sedate,
(SBCSAE 2:1079–1080)

Following these guidelines, unexpressed 3sg human specific subjects were exhaustively extracted from the two datasets, for a total of 153 in the conversational data (produced by 52 speakers) and 165 in the Pear Story narratives (by 20 speakers).

We can estimate an initial overall rate of expression by comparing this with the number of tokens of *he* and *she* in the data, assuming that the vast majority of these will be subject pronouns representing specific human referents (as the use of *he* and *she* for inanimate and generic uses would be rare). Based on counts in each corpus using the concordance program Monoconc (Barlow, 2004), the overall rate of nonexpression is 22% (165 of 748) in the Pear Stories, compared with just 4% (153 of approx. 3500) in SBCSAE (significantly lower, $p < .0001$).[3] Note that the rate for the conversational data is comparable to that observed for first-person singular (1sg) in this same corpus by Torres Cacoullos and Travis (2014:22) and is slightly lower than the rate for all persons reported by Shin and Montés Alcalá in data from the talk show *Larry King Live* (2014:88).

The difference in rates across the two registers is also to some degree comparable to other languages. For example, in Spanish Pear Stories produced by Peninsular Spanish speakers, Comajoan (2006:60) found a rate of unexpressed (vs. pronominal) 3sg specific subjects of 97% (275 of 285); in sociolinguistic interviews, a lower rate has been reported, ranging from 92% unexpressed in Madrid ($n = 286$, Cameron, 1992:233) to 60% in Barranquilla, Colombia ($n = 758$, Orozco, 2015:21); and in conversational Colombian Spanish, 68% ($n = 1413$, Torres Cacoullos & Travis, 2015b). Despite these very different rates of expression, however, the constraints identified across different varieties of Spanish correspond, and thus, as Torres Cacoullos and Travis (2015a:82) put it for 1sg, "the grammar is the same across Spanish varieties."

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

We focus the analysis here on the variability that exists between unexpressed and (personal) pronominal subjects, assuming these forms to be essentially equivalent ways of referring to given information, unlike full noun phrases (NPs), which are also used to introduce new information (in accordance with Givón's [1983:17] accessibility hierarchy; cf. also Chafe [1994:75–76]).

In order to identify the conditioning of this variation, we first need to determine in what contexts pronominal and unexpressed subjects exist as competing variants in English.

It has been widely recognized in English grammars that unexpressed subjects occur relatively freely in coordinated contexts (e.g., Biber, Johansson, Leech, Finegan & Conrad, 1999:156; Dixon, 2005:22; Quirk, Greenbaum, Leech & Svartvik, 1985:910). This has been referred to as "VP coordination" (or predicate —vs. clausal—conjunction), in which the second conjunct is not clausal, and thus does not have an unexpressed subject (e.g., Haspelmath, 2004:31; Huddleston & Pullum, 2002:238). VP coordination has not, however, been explicitly defined. It is generally applied to explain the occurrence of unexpressed subjects in the context of coreferential clauses (Quirk et al., 1985:948), usually conjoined with a coordinating conjunction such as *and, but,* or *or* (Quirk et al. 1985:910), though coordination without a conjunction ("asyndetic" coordination) has also been described (Biber et al., 1999:156; Haspelmath, 2004:4). Some scholars have also noted a role for prosody, specifically occurrence of the two conjuncts in a single prosodic unit (e.g., Mithun, 1988:333; Quirk et al., 1985:948). Line 3 in example (5), which exhibits all these characteristics, would seem to be a canonical case of VP coordination.

However, VP coordination does not explain the occurrence of unexpressed subjects in English, because, on the one hand, pronominal subjects also occur in canonical VP coordination contexts, seen in (6), and, on the other, unexpressed subjects also occur in noncoordinated contexts (in the narrative data, 45% of the time, a rate similar to that of the coordinated tokens, see Table 2). One example can be seen in (7), where there is no coordinating conjunction and one clause intervening between the target clause in line 5 and the previous clause with a coreferential subject in line 2. Thus, the variability in English subject expression that exists in both coordinated and noncoordinated contexts must be accounted for in order to obtain a full understanding of this phenomenon.

(5) and **he goes** and **Ø takes** the hat to the kid,

(Pear 10: 48.5)

(6) 1. Rebecca: [(H)] So this man **walked up to him** and **he said**,
2. <Q I'm standing close to you,
3. I'm looking at you,
4. does that make me a pervert Q>?

(SBCSAE 8:1224–1227)

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

(7) 1. Miles: then she went out there,
2. Ø got on her knee = s.
3. … in front of him.
4. … in this skirt that's this short,
5. … (H) **Ø takes** her hands,
6. .. and Ø goes along his toes,

(SBCSAE 2:1169–1174)

The envelope of variation in the coordinated and noncoordinated contexts is somewhat distinct, as was identified by Torres Cacoullos and Travis (2014:22–23) for 1sg subjects in the SBCSAE. In the context of coreferential-subject verbs which are syntactically conjoined with a coordinating conjunction (*and, but, or, and/but* + ADV), unexpressed subjects occur relatively freely. Outside of this context, they are restricted to prosodic-initial position, occurring only in main clause declaratives in absolute IU-initial position and therefore not in subordinate clauses nor interrogatives (cf. Haegeman, 2013:93–94; Harvie, 1998:21; Leroux & Jarmasz, 2005:4; Roberts & Holmberg, 2010:90; Weir, 2012). Unexpressed subjects with quotative verbs occur only in coordinated contexts, and there are no instances of unexpressed subjects with contracted auxiliaries.

With this now established, we can proceed with the extraction of pronominal subjects that fall within this envelope of variation.

Extraction of Pronominal Subjects

In order to obtain a comparable corpus of pronominal subjects, we utilized an artificial sampling method that has been applied to the study of subject expression in English, extracting the closest pronominal subjects preceding and following each unexpressed token (cf. Harvie, 1998:18; Leroux & Jarmasz, 2005:3; Torres Cacoullos & Travis, 2014:23). We extracted preceding and following tokens produced by the same speaker and aligning with the envelope of variation as described, that is, noncontracted forms occurring either in coreferential clauses with a coordinating conjunction, or in declarative, main clauses.[4] Thus, in (1) and (2), the pronouns in line 1 in (1) and in line 9 in (2) fall outside the variable context because they occur in non-IU-initial position (in subordinate clauses); the pronoun in line 9 in (2) is also contracted.

In the SBCSAE, this gave an initial total of 306 pronominal subjects (2 for each of the 153 unexpressed). This included non-IU-initial pronouns, which were subsequently excluded in all noncoordinated contexts ($n = 42$), leaving a total of 264 pronouns for inclusion in the analysis, and therefore an artificial rate of nonexpression of 37% (153 of 417). In the Pear Stories, this same method resulted in the extraction of all available pronominal tokens ($n = 307$), slightly short of the goal of two pronouns for each of the 165 unexpressed subjects. Of these, 54 pronouns occurring in non-IU-initial position were excluded, leaving 253 for analysis, and thus a rate of nonexpression of 39% (165 of 418).

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

Summary

To summarize this discussion of the variable context, we make three points. The first is that the nonvariable contexts identified here represent a sizable proportion of the pronominal data: in the Pear Stories, where all 3sg human specific pronouns were extracted, pronouns falling within the envelope of variation account for well under one half (253 of 583) of all pronouns in the data, a proportion similar to that which Torres Cacoullos and Travis (2014:22, 28) reported for 1sg pronouns. This means that the real rate of nonexpression in the Pear Stories is closer to 40% (165 of 418) (as opposed to 22%). Following these same calculations, we can estimate that the real rate of nonexpression in the SBCSAE data would be at around 10% (153 of 1515) (as opposed to 4%). Were the variable context to be overlooked, we would end up with a substantially different set of data for the analysis, leading possibly to quite distinct results.

Second, the envelope of variation in the two registers is identical, and thus here at least, we see no evidence for any difference in the grammars.

And third, these constraints are reported to not pertain in null subject languages (Haegeman, 2013:93–94). In Spanish, for example, unexpressed subjects occur in IU-medial position (cf. Torres Cacoullos & Travis, 2015a:93), and in interrogatives and subordinate clauses. Thus, in this sense, English would indeed appear to differ from the so-called canonical null subject languages.

Remaining to be answered, however, is whether, in the contexts where variability does exist, the two registers behave in the same way or not, and how this compares with patterns observed in languages with more widespread variability. It is to this that we now turn.

ENGLISH SUBJECT EXPRESSION IN THE VARIABLE CONTEXT

The wide body of literature on subject expression in null subject languages has identified a range of constraints that condition this variability, and the constraints we test for here are drawn from this literature.

Accessibility

Cross-linguistically, probably the most widely discussed constraint on subject expression relates to what has been termed "accessibility" (Ariel, 1994:2630; Givón, 1983:17), "activation" (Chafe, 1994:75), or "recoverability" (e.g., Haegeman, 2013:89; Weir, 2012:116), such that the greater the accessibility, activation, or recoverability of the subject referent, the greater the likelihood that the subject will be unexpressed. Quantitative studies have shown this to be one of the most robust constraints conditioning variable subject expression across a range of languages, including Arabic (Owens, Dodsworth, & Kohn, 2013:263), Australian Sign Language (McKee, Schembri, McKee & Johnston, 2011:388),

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

Bislama (a Vanuatan creole) and Tamambo (an indigenous language of Vanuatu) (Meyerhoff, 2009:308), and Spanish (e.g., Silva-Corvalán, 1982:104).

While there are different paths of accessibility—previous mention in the discourse, situational presence, shared knowledge (cf. Chafe, 1994: chap. 6; Prince, 1978)—previous mention in the discourse is that which is most readily operationalized. Here, we focus on previous mention as subject, as this appears to have a greater effect than a nonsubject mention: in both datasets, subjects are unexpressed approximately one half of the time when coreferential with the subject of the immediately preceding clause (150 of 296 in Pear and 123 of 220 in SBCSAE), but only 10% of the time when coreferential with a nonsubject mention in the immediately preceding clause (4 of 40 in Pear and 4 of 34 in SBCSAE). Also note that coreferential subject mentions in the immediately preceding clause are some seven times more frequent than nonsubject mentions in the immediately preceding clause (296 vs. 40 in Pear, and 220 vs. 34 in SBCSAE).

Table 1 presents the rate of unexpressed subjects in relation to the previous mention of that same referent as subject. "Coreferential" refers to contexts where the immediately preceding clause is produced by the same speaker and shares the same subject referent, as in lines 2 and 6 in (7). "Coreferential–interlocutor" (only relevant to the conversational data) refers to contexts where the immediately preceding clause is coreferential but was produced by the interlocutor, as illustrated in line 4 in (8). "Intervening non-coreferential subordinate" refers to contexts where the preceding main clause shares the same subject referent, but intervening between it and the target clause is a subordinate clause, or clauses, with a distinct subject referent as in line 5 in (7). And "non-coreferential" is where the immediately preceding main clause has a distinct subject referent, as lines 9 and 12 in (9), where the unexpressed mentions refer to discourse participants (Jenn and Lisbeth, respectively), as the speakers report to Bill on an incident that happened at the supermarket.

(8)
1. Jo: I don't think ~Kathy ever buys him socks.
2. Cam: … Oh.
3. Jo: … [You know],
4. Wess: [**Ø Buys** e]verything else,
5. but not socks.

(SBCSAE 59:206–210)

(9)
1. Lisbeth: And she specifically said,
2. .. not to put [2everything2] on the [3counter3].
3. Laura: [2Okay2].
4. [3Well then3],
5. I didn't,
6. [I didn't know that].
7. Bill: [Yeah=,
8. and so what] happened.
9. Lisbeth: .. (SNIFF) .. **Ø Starts putting** everything on the cou[2nter2].
10. Jenn: [2No=2],

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

TABLE 1. *Rate of unexpressed 3sg human specific subjects according to previous coreferential subject mention*

	Pear ($n = 418$)			SBCSAE ($n = 417$)		
	% Ø	*n*/Total	% Data	% Ø	*n*/Total	% Data
Coreferential	51	150/296	71	56	123/220	53
Coreferential–interlocutor	—	—		35	15/43	10
Intervening non-coreferential subordinate	38	12/32	8	33	5/15	4
Non-coreferential	3	3/90	21	7	10/139	33

11. … I w- --
12. … **Ø Acts** like I'm a child,
13. like I'm just throwing things on.

(SBCSAE 33:660–672)

Note that in both registers, accessibility functions as predicted: the highest rate of nonexpression is in coreferential contexts (51% in the Pear Stories and 56% in the SBCSAE), and the lowest in non-coreferential contexts (3% and 7%, respectively).[5] Intermediate rates of nonexpression are found in contexts where the preceding coreferential token was produced by an interlocutor (35% unexpressed), or where there is an intervening subordinate clause(s) (38% and 33%; cf. Labov and Waletzky (1997[1967]:14) who point out that subordinate clauses are irrelevant in narrative sequences).

Though the effect is parallel, there is a difference in the two registers in terms of the distribution of the data (indicated in the third column, labeled "% data"): while over two thirds (71%) of the narrative data occur in coreferential contexts, just over one half (53%) of the conversational data do, and similarly, under one quarter (21%) of the narrative tokens occur in non-coreferential contexts while a third (33%) of the conversational data do. This is an artifact of greater subject continuity in the monologic narrative data, where the speaker is recounting an event, uninterrupted, while the interactional nature of the conversational data disrupts the subject continuity. Thus, here we see a clear register effect that can partially account for the higher rate of nonexpression in the narrative dataset. That effect is not, however, one of different grammars, but rather of contextual distribution (mirroring what Travis [2007:130] found for subject continuity in a comparison of 1sg subject expression in Spanish conversation and narratives).

Given the strength of the accessibility effect in both of these datasets, we exclude from the subsequent analyses those tokens occurring in non-coreferential contexts, focusing on the categories in the top three rows in Table 1, that is, when the target clause has the same subject referent of the immediately preceding clause, be that produced by the interlocutor or the speaker, and when the non-coreferential intervening material consists solely of a subordinate clause (or clauses).

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

Variable subject expression: Beyond accessibility

Beyond accessibility, there is a wide range of factors that have been found to apply in null subject languages. Here, we consider the form of the previous coreferential subject, to test for a widely observed coreferential subject priming effect (e.g., Cameron & Flores-Ferrán, 2003; McKee et al., 2011:389; Torres Cacoullos & Travis, 2014:29–30; 2016; Travis, 2007:120–122; Travis & Torres Cacoullos, 2014:376–378); tense, which has been found to play a role in Spanish subject expression (e.g., Otheguy & Zentella, 2012:186; Shin, 2014:311; Silva-Corvalán, 2001:160–163; Travis & Torres Cacoullos, 2012:733–734); temporal relationship between clauses, which would appear to play a role in Spanish (Cameron, 1992:207), Brazilian Portuguese (Paredes Silva, 1993:43–45), and has also been suggested for English conversation (Oh, 2006:832); and coordination, which has been reported for English (as discussed in relation to the variable context) and Spanish (e.g., Otheguy & Zentella, 2012:190; Shin, 2014:311).

In order to identify the role these factors jointly play in conditioning the variation, we conduct logistic regression analyses (Sankoff, 1988), using Goldvarb LION (Sankoff, Tagliamonte, & Smith, 2012). Following the variationist comparative method (cf. Poplack & Tagliamonte, 2001), we conduct independent, but identical analyses, in each dataset. The results of these two analyses are presented in Table 2, with the Pear Story narratives on the left, and the SBCSAE conversation on the right.[6] The factor groups (independent or predictor variables) appear in the left-hand column. The first column of numbers depicts the Probability (or factor weight) with which the 3sg subject is unexpressed in the corresponding context. A value closer to 1 indicates a favoring of nonexpression, and closer to 0, a favoring of the pronoun. The second column indicates the percentage of unexpressed instances in that subcontext. The hierarchy of constraints that is captured in these relative weightings is considered to be representative of "the grammar underlying the variability" (Poplack & Levey, 2010:400). The third and fourth columns provide the distribution of the data across the subcontexts (the contextual distribution): the number of tokens (*n*) and the proportion (% data) that each factor accounts for.

These results show very similar effects in the narrative and conversational data. In both datasets, Previous Realization and Temporal Relationship (marked in bold) have a significant effect on the variation: unexpressed subjects are favored when the previous subject was also unexpressed and are disfavored when there is no temporal relationship between the target and the preceding coreferential subject. In neither dataset does Tense have a significant effect. The one apparent difference concerns Coordination: while in both datasets, *and*-coordinated contexts exhibit a higher rate of nonexpression than noncoordinated contexts do, it is only in the SBCSAE that this achieves significance, with a rate of nonexpression over twice as high in contexts of *and*-coordination (77% vs. 35%) compared with the very slight difference of 52% vs. 45% in the Pear narratives. We will now review these results in detail.[7]

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

TABLE 2. *Two independent Variable-rule analyses of the contribution of factors selected as significant to the choice of unexpressed 3sg human specific subjects, in Pear Story narrative and SBCSAE conversational data*

	Pear Story Narratives $n = 328$; Input: .51 (Overall rate: 49%)[a]				SBCSAE Conversation $n = 278$; Input: .62 (Overall rate: 51%)[b]			
	Prob	% Ø	*n*	% Data	Prob	% Ø	*n*	% Data
Realization of previous coref								
3sg subject								
Unexpressed	.75	76	63	25	.77	88	33	19
Pronoun (*he/she*)	.41	42	188	75	.43	50	141	81
Temporal relationship								
Sequential or simultaneous	.56	57	243	81	.72	85	98	47
No temporal relationship	.25	25	56	19	.31	35	111	53
Coordination								
And-coordinated	[.54]	52	201	61	.69	77	109	39
Non-coordinated	[.44]	45	127	39	.37	35	169	61
Tense								
Present	[.51]	50	253	78	[.51]	51	73	27
Past	[.47]	47	72	22	[.50]	52	201	73

Note: [a]All eligible tokens. [b]Sample of expressed subjects. Factor weights of nonsignificant groups, given in square brackets, are drawn from the first stepping down run, in which all factor groups are included.

Coreferential subject priming

Realization of the previous coreferential 3sg subject allows us to test for priming, whereby the same syntactic structure is repeated across clauses (cf. Bock & Griffin, 2000:177). Structural priming has been reported in experimental studies (cf. Pickering & Ferreira, 2008), as well as in corpus-based studies (e.g., Gries, 2005). In variationist work, priming has been found in virtually every study that has tested for it, including for the agentless passive in English (Weiner & Labov, 1983); variable plural (s) expression in both Spanish (Poplack, 1980) and Brazilian Portuguese (Scherre, 2001); and variable subject/verb and subject/predicate adjective agreement in Brazilian Portuguese (Scherre & Naro, 1991, 1992). Crucially, it has also been reported for subject expression in several languages, including Australian Sign Language (McKee et al., 2011:388), Bislama and Tamambo (Meyerhoff, 2009:308), and Spanish (cf. Cameron & Flores-Ferrán, 2003; Torres Cacoullos & Travis, 2016; Travis, 2007:120–122; Travis & Torres Cacoullos, 2012:730). For English 1sg subject realization, priming has been reported both for expression (Torres Cacoullos & Travis, 2014:29–30) and stress on the pronoun *I* (Travis & Torres Cacoullos, 2014:376–377).

In order to test for a priming effect, we coded for the form of the previous coreferential subject (be that the immediately preceding clause, or of the previous main clause, skipping over an intervening non-coreferential subordinate

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

clause or clauses). We excluded tokens where the previous coreferential subject was produced by the interlocutor, as cross-speaker priming may not function identically to same-speaker priming (see Travis, Torres Cacoullos, and Kidd [2015:Section 5] for discussion). We also excluded tokens for which the previous coreferential subject was realized as a full NP or as an indefinite pronoun, as in line 2 in (10), where the unexpressed subject is preceded by the full NP, *Ron* (the rate of unexpressed subjects is close to the average in this context, 46% [32 of 70] in Pear and 50% [25 of 50] in SBCSAE).

As seen in Table 2, unexpressed subjects are favored when the preceding subject was also unexpressed, with rates of 76% in the Pear data (probability .75) and 88% in the SBCSAE data (probability .77). The following examples (in which the targets are bolded and the primes underlined) illustrate this phenomenon, with strings of unexpressed subjects.

(10) 1. Lajuan: .. ~Ron f=lew to Dallas,
2. .. And Ø went to my brother's house,
3. and **Ø banged** on the door,
4. and **Ø said** I want to see him,
5. we need to talk,

(SBCSAE 44:867–871)

(11) 1. [.3] = and /you see he notices that the bushel's missing,
2. and Ø counts them,
3. [.3] **Ø can't figure out** what's happening,

(Pear 8:6.9–6.11)

The priming effect holds in both *and*-coordinated (lines 3 and 4 in (10)) and noncoordinated contexts (line 3 in (11); see the section on *and*-coordination for the operational definition applied here). This is depicted in Table 3, which shows that in both contexts, the rate of unexpressed subjects is higher when the previous coreferential subject was also unexpressed (in the first row) than when it was pronominal (second row). In Pear, this difference is significant in both *and*-coordinated ($p < .0001$) and noncoordinated contexts ($p = .0067$), and in SBCSAE is significant for the noncoordinated tokens ($p = .0004$), but not for the *and*-coordinated tokens ($p = .1677$). This corresponds with the weaker priming effect in *and*-coordinated contexts for English 1sg subjects reported in Torres Cacoullos and Travis (2014:30). We attribute this to the fact that the rate of unexpressed subjects in coordinated contexts in the conversational data is so high even following a pronoun that it simply cannot be affected as much by a previous unexpressed subject. In noncoordinated contexts, however, unexpressed subjects are sufficiently rare following a previous pronoun (with rates of 31% and 39%) that they rise substantially following an unexpressed subject, doing so by between 1.5 (from 39% to 65%) and 2.5 times as much (from 31% to 81%).

As we saw to be the case for Accessibility, here also we see largely parallel effects for priming across both datasets, but distinct contextual distributions.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

TABLE 3. *Rate of unexpressed 3sg human specific subjects according to priming and* and-*coordination*

	Pear				SBCSAE			
	And-coord		Noncoord		*And*-coord		Non-coord	
	% Ø	*n*/Total	% Ø	*n*/Total	% Ø	*n*/Total	% Ø	*n*/Total
Previous Ø	90	26/29	65	22/34	94	16/17	81	13/16
Previous pronoun	43	51/118	39	27/70	77	46/60	31	25/81
Total		77/147		49/104		62/77		38/97

Looking overall (Table 2), the difference is marginal: in the Pear Story narratives, 25% (63 of 251) of the tokens occur in the context of a previous unexpressed subject, while in the SBCSAE conversational data, 19% (33 of 174) do (p = .1961). Recall that we have excluded tokens where the previous realization was a full NP or an indefinite pronoun and where it was produced by the interlocutor; if we consider all tokens, the difference in the distribution is significant: 19% (63 of 328) in the Pear narratives and 12% (33 of 278) in the SBCSAE conversation occur with a previous unexpressed subject (p = .0199). Thus, once again, a higher proportion of the narrative data occurs in a context favorable to an unexpressed subject, namely following another unexpressed subject.

Comparing now across the coordinated and noncoordinated tokens (Table 3), we again see a significant difference in the contextual distribution: while the proportion of the data represented by a previous unexpressed subject is similar in coordinated instances, at around one fifth (29 of 147 in Pear, and 17 of 77 in SBCSAE), in noncoordinated contexts, a preceding unexpressed subject accounts for 33% (34 of 104) of the Pear data, but only 16% (16 of 97) of the SBCSAE data (p = .0091). That is, in the Pear data, there are more opportunities for unexpressed-to-unexpressed priming among the noncoordinated than the coordinated tokens (33% vs. 20%), which serves to raise the rate of nonexpression in noncoordinated contexts and lessen the difference with the *and*-coordinated contexts.

Temporal relationship

In her qualitative study of unexpressed subjects in English conversation, Oh (2006:832) proposed an association between temporality and unexpressed subjects. Such an association for Spanish was quantitatively verified by Cameron (1992:207) in his study of subject expression in San Juan, Puerto Rico, and Madrid, Spain, where he found a higher rate of unexpressed subjects in clauses that "stand in a temporal relationship to each other." It would seem that such an effect is relevant across coreferential mentions only, in accordance with Paredes Silva's (1993:43) finding for written Brazilian Portuguese of a favoring

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

of unexpressed subjects in "discourse connected" clauses expressing a "sequence of events centered on the same referent" (that is, with the same subject and the same tense). This is supported by Shin (2014:321), who reported no effect for temporal sequencing in switch reference contexts in narratives occurring as part of sociolinguistic interviews by Spanish speakers newly arrived in New York.

To test for such an effect in the data under study here, we coded all clauses for the absence versus presence of a temporal relationship. Two types of temporal relationship were identified. The first is temporally sequential events, where the event referred to in one clause follows temporally from that of the previous clause, as in lines 3 and 4 in (12) and in (10)) presented previously. The second is temporally simultaneous events, illustrated in (13) and in lines 2 and 3 in (12) (where, as the boy is putting the pears away, he picks up a pear—note the use of the progressive here indicating that an event [picking up a specific pear] overlaps with a situation [putting away the pears] [cf. Hopper, 1979:216]). We collapse these two temporal relationships in Table 2, because the effect for sequentiality and simultaneity is similar—in the SBCSAE data, the rates are identical at 85% (60 of 71 in sequential contexts and 23 of 27 in simultaneous contexts), in the Pear Stories, they are nearly so, 58% (117 of 202) for sequential and 54% (22 of 41) for simultaneous contexts. Note also that the temporally sequential clauses represent the majority of the temporally related data (approximately three quarters in both datasets).

(12) 1. and **he comes** back down,
2. [1.1[.7] a--nd] **he's putting** the rest of the pears away,
3. and **he picks** this pear up,
4. and **Ø polishes** it up,
5. and everything else.

(Pear 10:13.7–13.11)

(13) Beth: for one thing it bothered me,
because Anna takes ahold of something like that and **Ø doesn't let go**.

(SBCSAE 31:1447–1448)

The absence of a temporal relationship applies where one of the verbs occurs with habitual aspect or is a stative predicate (e.g., with modal *can* in *can't figure out* in (11)), or where the second clause is some kind of repetition, (14) or elaboration of a prior comment, (15) (where the progressive *he was walking* provides background information on *stops*).

(14) Mitchell: and- he % totally evaded me,
.. **he has** to this day **evaded** me,

(SBCSAE 53:677–678)

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

(15) and .. the boy on the bicycle s [.45] stops,
[2.15[.75] u--m [.8] tsk..]] **he was walking** it at that time,..
(Pear 13:20.2–20.3)

Excluded from coding were instances where the temporal relationship between clauses was not relevant, including where the previous coreferential subject was produced by the interlocutor ($n = 43$), where it occurred in quoted speech ($n = 7$), and where it occurred in a subordinate clause, as in (16) (in the temporal "before she left"; $n = 48$, consisting of 28 from Pear and 20 from SBCSAE). This is because, as noted by Labov and Waletzky (1997[1967]:14), "only independent clauses are relevant to temporal sequence."[8]

(16) Alice: .. One that she bought just before she left.
(H) .. **She couldn't get** it open.
(SBCSAE 43:44–45)

As we saw in Table 2, the presence of a temporal relationship has a significant effect in both registers. In the Pear Stories, there is a strong disfavoring of unexpressed subjects in nontemporal clauses (with a weight of .25), and in SBCSAE, we observe a similar disfavoring (.31), but also a favoring in temporal clauses (.72).

Once again, while the effect is similar, the distribution across the two datasets is distinct, with 81% of the narrative data occurring in temporally related contexts compared with less than half of the conversational data (seen in column 4 in Table 2). Remembering that the majority of the temporally related clauses are sequential, this result is not surprising. The Pear Story corpus consists of narratives that recount a series of events that the speaker had just seen, in temporal sequence, on a video. The nature of the task thus ensures frequent occurrence of temporally sequential clauses. The conversational data is less event focused, evidenced in a higher proportion of stative verbs than in the narrative data (25% of all verbs in SBCSAE vs. 11% in Pear). As stative verbs tend to occur in nontemporal contexts, this raises the proportion of the nontemporal data. Furthermore, the interactional nature of the conversational data results in a higher number of (nontemporal) repetitions and elaborations than in the monologic narratives of the Pear data.

The patterning for temporal relationship holds across *and*-coordinated and noncoordinated contexts: in both datasets, the rate of nonexpression is higher when there is a temporal relationship between clauses than when there is no such relationship. However, once again the distribution of the temporally and nontemporally related tokens according to coordination is somewhat different, as illustrated in Figure 1. In the Pear data, there is a higher proportion of temporally related clauses in both *and*-coordinated and noncoordinated contexts (considering just that portion of the data to which it is possible to apply the notion of a temporal relationship, 84% [157 of 186] of the *and*-coordinated tokens are temporal and 76% [86 of 113] of the noncoordinated tokens are). In the SBCSAE, in *and*-coordinated contexts, temporally related clauses remain in

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

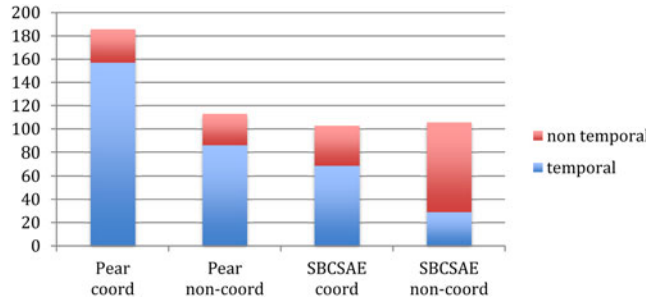


FIGURE 1. Distribution of Temporal and Non-Temporal clauses according to *and*-Coordination in Pear and SBCSAE (*n* of tokens in temporal/nontemporal contexts).

the majority (accounting for two thirds of the data [69 of 103]), but in noncoordinated contexts, temporally related clauses account for just one quarter of the data (29 of 106). Thus, in the conversational data, the noncoordinated tokens occur primarily in an environment that disfavors unexpressed subjects, while in the narrative data, the noncoordinated tokens (like the coordinated tokens) occur primarily in a context that favors unexpressed subjects. This distribution, then, lowers the rate of nonexpression in noncoordinated contexts in the SBCSAE data (heightening the difference between *and*-coordinated and noncoordinated) and raises it in the Pear data (lessening the difference).

It is worth briefly noting here that Tense is not entirely independent from Temporal relationship. While we would expect the Past tense to be more associated with temporal sequentiality than the Present tense (cf. Hopper 1979:216), this is not the case in the Pear Story narratives, where both Past and Present tense tokens occur in temporally related contexts approximately 80% of the time. This is because, given the nature of this dataset, all Present tense tokens are in fact the Historical Present (illustrated in the Pear examples given thus far). In the conversational data, Historical Present tokens account for over one third (28 of 73) of the Present tense tokens (examples can be seen in (7) and (9)), and while over one half (57%, 16 of 28) of the Historical Present occur in temporally sequential contexts, just over one tenth (5 of 45) of the remaining Present tense tokens do. Correspondingly, we find a higher rate of unexpressed subjects with the Historical Present (64%) than the other Present tense tokens (42%). Thus, while we do not see an effect for Tense itself, we do see an effect for discourse function, such that those tenses with a higher occurrence in temporally related contexts may have a higher rate of unexpressed subjects (similar to what Silva-Corvalán [2001:160–163] proposed for Spanish).

And-*coordination*

And-coordination is used here to refer to strictly coreferential clauses (produced by the same speaker, with no intervening subordinate clauses) conjoined by *and*; all

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

TABLE 4. *Rate of unexpressed 3sg human specific subjects in strictly coreferential contexts according to presence and type of conjunction*

	Pear		SBCSAE	
	% Ø	*n*/Total	% Ø	*n*/Total
And	52	105/201	77	84/109
Other conjunction (*but, or, and/but* + Adv)	30	9/30	48	10/21
No conjunction	55	36/65	35	29/84

other clauses are considered noncoordinated. We limit the definition of coordination to strictly coreferential clauses because, as was seen in Table 1, these have the highest rate of nonexpression. Table 4 gives the breakdown of the tokens in strictly coreferential contexts according to presence and type of conjunction: conjoined with the preceding clause by *and*; with another conjunction (*but, or, and/but* + ADV [e.g., *and just, and apparently*]); or without a conjunction. As can be seen, *and* is overwhelmingly the most frequent conjunction in both datasets, occurring over five times as often as the other coordinating conjunctions combined. *And* also occurs with a significantly higher rate of unexpressed subjects than other coordinating conjunctions: 52% vs. 30% in the Pear narratives ($p = 0.0305$), and 77% vs. 48% in SBCSAE ($p = 0.0142$)). With no conjunction, we find a still lower rate of expression in SBCSAE, of less than half that of clauses conjoined with *and* (35%). In the Pear data, however, the rate of nonexpression without a conjunction is similar to that with *and* at 55%.

While some might interpret this as evidence of asyndetic coordination in the narrative, but not the conversational, data, it can instead be accounted for by the contextual distribution in each register according to temporality. In the narratives, we find a high proportion of temporally related tokens both with *and* and without—78% (157 of 201) of the tokens conjoined with *and*, and 60% (39 of 65) of the tokens with no conjunction, are temporally related. This results in a high rate of nonexpression in both environments in the Pear Stories data. In the conversational data, however, while 63% (69 of 109) of the tokens conjoined with *and* are temporally related, just 18% (15 of 84) of those with no conjunction are, thus lowering the rate of unexpressed subjects in contexts where there is no conjunction. The higher rate of temporal sequentiality both with and without a conjunction in the Pear Story data is a reflection of the very high proportion of the data overall that occurs in temporally related contexts (81%, as we saw in Table 2). In order to disentangle coordination from temporal sequentiality, then, we restrict coordination to the presence of *and*.

As seen in Table 2, in the conversational data, unexpressed subjects occur over twice as often in *and*-coordinated contexts than in noncoordinated contexts (77% vs. 35%), while in the narrative data, there is only a marginal, and nonsignificant, difference between the two, of 52% versus 45%.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

In order to test whether this might be evidence of a real grammatical difference between the two registers, we conducted separate analyses of tokens in *and*-coordinated and noncoordinated contexts in each of the two datasets utilizing the same factor groups as those included in the analyses reported in Table 2, and one further factor group only applicable to the *and*-coordinated tokens, namely, the occurrence of the two conjuncts in the same IU. The same constraints emerged in all analyses: in both datasets and in both registers, unexpressed subjects are favored when the preceding subject is unexpressed, disfavored when there is no temporal relationship, and unaffected by tense. The priming effect does not reach significance for the coordinated tokens in the SBCSAE, which is consistent with the weaker effect for priming in coordinated contexts in this dataset seen in Table 3. And in all analyses, unexpressed subjects are favored when the two conjuncts occur in the same IU, as in (17) and (18), more than when they occur across IUs, seen in (19) (as Torres Cacoullos and Travis [2014:27–28] found for 1sg subjects).

(17) Tom 2: (H) we got a new guy k-,
named Mason **called up** and **Ø said**,
hey,
would you join,
(SBCSAE 32:963–966)

(18) [.5] And the kid turns around,
and **he goes** and **Ø takes** the hat to the kid,
[2.2[.85] a--nd um [.65]] the kid gives him three pears.
(Pear 10:48.4–48.6)

(19) Marie: (TSK) **this man went** to Wal-Mart.
And **Ø just shot**.
(SBCSAE 36:906–907)

While these same-IU occurrences have been interpreted as cases of so-called VP coordination, pronominal subjects also occur in such contexts, as noted previously (and illustrated in example (6)). Torres Cacoullos and Travis (2015a:93–95) argued that VP coordination does not exist as a discrete category, but rather as a continuum from conjuncts that are more to less tightly linked, with linking being based on occurrence in the same IU, the same tense and constituting a single event (cf. Barth-Weingarten & Couper Kuhlen, 2011:280; Croft, 1995:846; Mithun, 1988:335). Tightly linked conjuncts tend to favor unexpressed subjects more than less tightly linked conjuncts do.

This is supported by the data examined here. Table 5 shows the rate of unexpressed subjects in *and*-coordinated contexts, according to occurrence in the same IU versus across IUs. In both datasets, the subject of the second conjunct is unexpressed more often when the two conjuncts occur in the same IU than

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

TABLE 5. *Rate of unexpressed 3sg human specific subjects in* and*-coordinated contexts according to occurrence in the same IU vs. across IUs*

	Pear Stories			SBCSAE		
	% Ø	*n*/Total	% Data	% Ø	*n*/Total	% Data
Same IU	84	21/25	12	94	30/32	29
Across IUs	48	84/176	88	70	54/77	71
Total		201			109	

when they occur across IUs (with rates of 84% vs. 48% in the Pear Story data, and 94% vs. 70% in the SBCSAE data).

However, there is once again a difference in terms of the contextual distribution. In the Pear Story data, just 12% (25 of 201) of the coordinated tokens occur in the same IU, while in the SBCSAE conversational data, close to one third (32 of 109) do ($p = .0004$). This, then, raises the rate of nonexpression in *and*-coordinated contexts in the conversational data and heightens the difference in rates of nonexpression between the *and*-coordinated and noncoordinated tokens.

Among these single-IU occurrences, for 1sg, Torres Cacoullos and Travis (2014:31) identified two set constructions, both of which occur in the 3sg data studied here, namely with a motion verb as the first conjunct ([MOTION VERB$_i$ *and* Ø VERB$_i$]) (cf., Brinton, 1990:124–125; Hopper, 2002:152; Quirk et al., 1985:978), and with a verb of speech as the second conjunct [VERB$_i$ *and* Ø VERB-OF-SPEECH$_i$]. These constructions together account for close to two thirds of the same-IU coordinated tokens in the SBCSAE (including 11 tokens of the motion verb construction, illustrated in (5), 8 with the speech verb construction, as in (17), and two where the two occur together, as in (6)). In the Pear data, the single-IU speech verb construction does not occur (not surprisingly, as this is a silent film), and the motion verb construction (seen in (18)) accounts for just 40% of the same-IU tokens. Thus, we find not only a higher proportion of single-IU *and*-coordinated tokens in the conversation than in the narrative data, but within those single-IU *and*-coordinated tokens, we find a higher proportion of set constructions in the conversational data than in the narrative data (66% vs. 40%, nearly significant, $p = .0658$).

The lack of a significant effect for *and*-coordination in the narrative data, in contrast to the conversational data, is therefore not reflective of differing constraints in the two registers. Rather, it is attributable to the higher proportion of the narrative noncoordinated tokens occurring in contexts favorable to unexpressed subjects (following a preceding unexpressed subject and in temporally sequential clauses) and a higher proportion of the conversational *and*-coordinated tokens occurring in contexts favorable to unexpressed subjects (in temporally related contexts and in single-IU coordinated constructions, in particular with certain set constructions).

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

CONCLUSIONS

Here, we have provided the first empirical test of the hypothesis that different English registers have distinct grammars for subject expression, through examination of the variability between pronominal and unexpressed 3sg human specific subjects in monologic narrative and interactional conversation. One very notable difference is in the proportion of unexpressed subjects in each dataset: just 4% in conversation, but close to one quarter (22%) in the narratives (or, taking into account the variable context, 10% in conversation and 40% in the narratives), strikingly high for a so-called non-null subject language. Despite this stark disparity in rates, examination of the factors conditioning this variation across the two registers provides no evidence for grammatical differences between them, but demonstrates that, instead, the disparity is attributable to diverging contextual distribution of the tokens in each dataset.

The envelope of variation in both registers is identical: unexpressed subjects exist in variation with pronominal subjects only in coreferential clauses conjoined with a coordinating conjunction, or outside of that, in noncontracted declarative main clauses in IU-initial position. Such constraints have not been reported for canonical null subject languages, and thus English would appear to differ from such languages in this restricted set of variable contexts.

The probabilistic constraints in operation within the variable context are also identical in both registers: 3sg human specific subject expression in English narratives and conversation is sensitive to accessibility (unexpressed subjects are favored in coreferential contexts), priming (favored when the preceding coreferential subject was also unexpressed), and temporal relationship (disfavored when there is no temporal relationship with the preceding clause). These probabilistic constraints parallel those reported for null subject languages. That is, the variable grammar of subject expression in English corresponds to that of other languages, suggesting that the proposed typology of languages according to rates of subject expression deserves further exploration, in particular in terms of potentially shared constraints on speaker choice of expression versus nonexpression.

The explanation for the difference in overall rates of expression is found in the contextual distribution, which is itself attributable to the nature of the registers studied. In the narrative data, a higher proportion of the tokens occur in environments that favor nonexpression, namely coreferential contexts (Table 1), following a previous unexpressed subject (Table 2) and in temporally related clauses (Figure 1). This is to be expected, given that the Pear Story data comprise monologic narratives that recount a series of events centered around a small number of participants. In the conversational data, there is a higher rate of switching of subjects (non-coreferentiality) and of clauses expressing a nontemporal relationship, consistent with the interactional nature of this data.

The effect of coordination is the one apparent difference across the two registers. Though in both registers unexpressed subjects occur proportionally more in *and*-coordinated than noncoordinated contexts, this effect only reaches significance

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

in the conversational data. Here again the explanation can be found in the contextual distribution. In the narrative data, a higher proportion of the noncoordinated tokens occur in environments that favor nonexpression (following a preceding unexpressed subject [Table 3] and in temporally related clauses [Figure 1]), whereas in the conversational data, it is the *and*-coordinated tokens that tend to occur most in environments favorable to unexpressed subjects (in temporally related clauses [Figure 1], and in same-IU contexts [Table 5], in particular with motion and speech verb constructions.

We conclude that, while observing overall rates and drawing hypotheses about differences in function are important foundations for the study of grammatical patterning, grammatical similarities and differences can best be elucidated through comparisons of the probabilistic constraints operating in variable contexts. Such comparisons should be conducted in full cognizance of possible effects of the contextual distribution of the data, as differing relative proportions of occurrences in favoring and disfavoring contexts can mask shared patterns and create the illusion of different grammars.

NOTES

1. Transcripts included in the study are SBCSAE nos. 1–6, 8, 11, 13–15, 19, 23, 28, 31–37, 43–45, 47, 49, 50, 52, 53, 55–57, 59, 60; all 3sg human specific unexpressed subjects in these recordings are included ($n = 153$). We exclude 27 unexpressed 3sg human specific subjects from four more monologic transcripts (sermons recorded in nos. 20, 21, and 30 and a lecture recorded in no. 25). The remaining 22 transcripts had no tokens of unexpressed 3sg human specific subjects.

2. SBCSAE examples are reproduced verbatim from the Santa Barbara Corpus of Spoken American English (SBCSAE; http://www.linguistics.ucsb.edu/research/santa-barbara-corpus), and the Pear Story examples from the narratives supplied by Du Bois (personal communication). The transcripts have been modified only with the addition of Ø to ease identification of the unexpressed subjects, **bolding**, and underlining to highlight the aspects of interest in any particular example, and line numbering where relevant. Information given in parentheses indicates the corpus, the transcript number, and the line numbers of the example.

3. All *p*-values given in the paper were determined using Fisher's exact test.

There are a total of 572 tokens of *he* and 11 tokens of *she* in the Pear Stories, and approximately 2100 tokens of *he* and 1300 tokens of *she* in the 34 SBCSAE files used for this study. Note that the rate of nonexpression remains at approximately 4% in the SBCSAE if we consider (i) all dialogic transcripts in the corpus, at 153 of 4000 (excluding the monologic transcripts, nos. 20, 21, 25, 27, 30, 38, 40, 54, which together account for 10% of the total number of words in the corpus); and (ii) the entirety of the corpus (including all 60 transcripts) at 180 of 4500 (including the 27 unexpressed subjects that occur in the monologic transcripts, see Note 1).

4. When two or more unexpressed tokens occurred sequentially, we extracted the two closest preceding and following tokens from the same speaker. When there were no available tokens from the same speaker in the transcript, in SBCSAE, we extracted from another speaker in that same transcript who produced at least one unexpressed token and, in the Pear Stories, we extracted from another transcript.

5. For 1sg in the SBCSAE data, Torres Cacoullos and Travis (2014:30) found little effect for coreferentiality, with unexpressed subjects occurring at a marginally higher rate in coreferential than non-coreferential contexts (48%, 48 of 101, vs. 36%, 8 of 22 respectively; $p = .4792$). This may be because 1sg (as a discourse participant) can be assumed to be accessible (or "active"), independently of explicit mentions of it (cf. Chafe, 1994:79).

6. Many thanks to Rena Torres Cacoullos for advice on the analysis and interpretation of the results presented in this paper.

7. Two other factors tested are Turn position and Polarity. A higher rate of pronominal subjects in turn-initial (vs. noninitial) position in the conversational data was found not to be related to floor taking (cf. Aijón Oliva & Serrano, 2010:16; Davidson, 1996:561; Martínez Lara, 2013:88), but to a lower rate of occurrence of *and*-coordinated tokens in turn-initial position. And though negative polarity occurs with a lower rate of unexpressed subjects than positive polarity in both datasets, in neither is this difference

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

significant (in Pear 51%, 154 of 304, for affirmative vs. 33%, 8 of 24 for negative [$p = .1371$], and in SBCSAE, 52%, 128 of 244, for affirmative vs. 44%, 15 of 34, for negative [$p = .4643$]).

8. Where a non-coreferential subordinate clause intervenes, we look back to the previous coreferential main clause to determine the temporal relationship.

REFERENCES

Aijón Oliva, Miguel Angel, & Serrano, María José. (2010). El hablante en su discurso: Expresión y omisión del sujeto de *creo*. *Oralia* 13:7–38.

Ariel, Mira. (1994). Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics* 30(1):3–42.

Barlow, Michael. (2004). *MonoConc Pro (MP2.2)*. Houston: Athelstan.

Barth-Weingarten, Dagmar, & Couper Kuhlen, Elizabeth. (2011). Action, prosody and emergent constructions: The case of *and*. In: P. Auer & S. Pfänder (eds.), *Constructions: Emerging and emergent*. Berlin: de Gruyter. 236–292.

Biber, Douglas, & Conrad, Susan. (2001). Register variation: A corpus approach. In D. Schiffrin, D. Tannen, & H. E. Hamilton (eds.), *The handbook of discourse analysis*. Oxford: Blackwell Publishers. 175–195.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Finegan, Edward, & Conrad, Susan. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Bock, J. Kathryn, & Griffin, Zenzi M. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General* 129(2):177–192.

Brinton, Laurel. (1990). The development of discourse markers in English. In J. Fisiak (ed.), *Historical linguistics and philology*. Berlin: Mouton de Gruyter. 45–71.

Cameron, Richard. (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Ph.D. dissertation, University of Pennsylvania.

Cameron, Richard, & Flores-Ferrán, Nydia. (2003). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1):41–65.

Chafe, Wallace (ed.). (1980). *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*. Norwood: Ablex.

_____ (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Comajoan, Llorenc. (2006). Continuity and episodic structure in Spanish subject reference. In J. C. Clements & J. Yoon (eds.), *Functional approaches to Spanish syntax: Lexical semantics, discourse and transitivity*. New York: Palgrave MacMillan. 53–79.

Cote, Sharon A. (1996). *Grammatical and discourse properties of null arguments in English*. Ph.D. dissertation, University of Pennsylvania.

Croft, William. (1995). Intonation units and grammatical structure. *Linguistics* 33:839–882.

Davidson, Brad. (1996). "Pragmatic weight" and Spanish subject pronouns: The pragmatic and discourse uses of *tú* and *yo* in spoken Madrid Spanish. *Journal of Pragmatics* 26(4):543–565.

Dixon, R. M. W. (2005). *A semantic approach to English grammar*. 2nd ed. Oxford: Oxford University Press.

Dryer, Matthew S. (2011). Expression of pronominal subjects. In M. S. Dryer & M. Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. Chap. 101. Available at: http://wals.info/chapter/101 Accessed October 13, 2015.

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. Philadelphia: Linguistic Data Consortium.

Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, & Paolino, Danae. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (eds.), *Talking data: Transcription and coding in discourse*. Hillsdale: Lawrence Erlbaum Associates. 45–89.

Givón, T. (1983). Topic continuity in discourse: An introduction. In: T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-linguistic study*. Amsterdam: John Benjamins. 1–41.

Gries, Stefan Th. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4):365–399.

Guy, Gregory. (2005). Grammar and usage: A variationist response. *Language* 81(3):561–563.

Haegeman, Liliane. (2013). The syntax of registers: Diary subject omission and the privilege of the root. *Lingua* 130:88–110.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

Harvie, Dawn. (1998). Null subject in English: Wonder if it exists? *Cahiers Linguistiques d' Ottawa* 16:15–25.
Haspelmath, Martin. (2004). Coordinating constructions: An overview. *Typological Studies in Language* 58:3–40.
Helasuvo, Marja-Liisa. (2014). Searching for motivations for grammatical patternings. *Pragmatics* 24 (3):453–476.
Hopper, Paul J. (1979). Aspect and foregrounding in discourse. In T. Givón (ed.), *Discourse and syntax*. New York: Academic Press. 213–241.
_____ (2002). Hendiadys and auxiliation in English. In J. Bybee & M. Noonan (eds.), *Complex sentences in grammar and discourse*. Amsterdam: John Benjamins. 145–174.
Huddleston, Rodney, & Pullum, Geoffrey K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Izre'el, Shlomo (2005). Intonation units and the structure of spontaneous spoken language: A view from Hebrew. In C. Auran, R. Bertrand, C. Chanet, A. Colas, A. Di Cristo, C. Portes, A. Reynier, & M. Vion (eds.), *Proceedings of the IDP05 International Symposium on Discourse-Prosody Interfaces*. CD ROM. Available at: http://www.tau.ac.il/~izreel/publications/IntonationUnits_IDP05.pdf%3E. Accessed November 8, 2014.
Labov, William, & Waletzky, Joshua. (1997 [1967]). Narrative analysis: Oral versions of personal experience. *Journal of Narrative and Life History* 7(1–4):3–38.
Leroux, Martine, & Jarmasz, Lidia-Gabriela. (2005). A study about nothing: Null subjects as a diagnostic of the convergence between English and French. *University of Pennsylvania Working Papers in Linguistics* 12(2):1–14.
Martínez Lara, José Alejando. (2013). Estudio del sujeto variable de 3ª persona en el habla actual de Caracas: Un análisis cuantitativo. *Núcleo* 25(30):69–96.
McKee, Rachel, Schembri, Adam, McKee, David, & Johnston, Trevor. (2011). Variable "subject" presence in Australian Sign Language and New Zealand Sign Language. *Language Variation and Change* 23:375–398.
Meyerhoff, Miriam. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21(3):297–317.
Mithun, Marianne. (1988). The grammaticization of coordination. In J. Haiman & S. A. Thompson (eds.), *Clause combining in grammar and discourse*. Amsterdam: John Benjamins. 331–359.
Napoli, Donna Jo. (1982). Initial material deletion in English. *Glossa* 16:85–111.
Oh, Sun-Young. (2006). English zero anaphora as an interactional resource II. *Discourse Studies* 8 (6):817.
Orozco, Rafael. (2015). Pronominal variation in Colombian Costeño Spanish. In A. M. Carvalho, R. Orozco, & N. L. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown: Georgetown University Press. 17–37.
Otheguy, Ricardo, & Zentella, Ana Cecilia. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
Owens, Jonathan, Dodsworth, Robin, & Kohn, Mary. (2013). Subject expression and discourse embeddedness in Emirati Arabic. *Language Variation and Change* 25(2):255–285.
Paredes Silva, Vera Lucia. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5(1):35–49.
Pickering, Martin J., & Ferreira, Victor S. (2008). Structural priming: A criticial review. *Psychology Bulletin* 134(3):427–459.
Pierrehumbert, Janet, & Hirschberg, Julia. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (eds.), *Intentions in communication*. Cambridge: MIT Press. 271–311.
Poplack, Shana. (1980). The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic Press. 55–67.
Poplack, Shana, & Levey, Stephen. (2010). Contact-induced grammatical change: A cautionary tale. In P. Auer & J. E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation*. Vol. 1. *Theories and methods*. Berlin: Mouton de Gruyter. 391–419.
Poplack, Shana, & Tagliamonte, Sali. (2001). *African American English in the diaspora*. Malden: Blackwell Publishers.
Prince, Ellen F. (1978). On the function of existential presupposition in discourse. *Papers from the Regional Meeting of the Chicago Linguistic Society* 14:362–376.
Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, & Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. London: Longman.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

Roberts, Ian, & Holmberg, Anders. (2010). Introduction: Parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (eds.), *Parametric variation: Null subjects in minimalist theory*. Cambridge: Cambridge University Press. 1–57.
Sankoff, David. (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society*. Vol. 2. Berlin: Walter de Gruyter. 984–997.
Sankoff, David, Tagliamonte, Sali, & Smith, Eric. 2012. Goldvarb LION: A variable rule application for Macintosh. University of Toronto. Available at: http://individual.utoronto.ca/tagliamonte/goldvarb.html. Accessed January 15, 2014.
Scheibman, Joanne. (2001). Local patterns of subjectivity in person and verb type in American English conversation. In J. Bybee & P. J. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. 61–89.
Scherre, Maria Marta Pereira. (2001). Phrase level parallelism effect on noun phrase number agreement. *Language Variation and Change* 13(1):91–107.
Scherre, Maria Marta Pereira, & Naro, Anthony J. (1991). Marking in discourse: "Birds of a feather." *Language Variation and Change* 3(1):23–32.
_____ (1992). The serial effect on internal and external variables. *Language Variation and Change* 4(1):1–13.
Shin, Naomi Lapidus. (2014). Grammatical complexificatoin in Spanish in New York: 3sg pronoun expression and verbal ambiguity. *Language Variation and Change* 26(3):303–330.
Shin, Naomi Lapidus, & Montes-Alcalá, Cecilia. (2014). El uso contextual del pronombre sujeto como factor predictivo de la influencia del inglés en el español en Nueva York [English influence on Spanish in New York: Evidence from subject pronouns in context]. *Sociolinguistic Studies* 8 (1):85–100.
Silva-Corvalán, Carmen. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elías Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*. New York: Cambridge University Press. 93–120.
_____ (2001). *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.
Szmrecsanyi, Benedikt, & Hinrichs, Lars. (2008). Probabilistic determinants of genitive variation in spoken and written English. In T. Nevalainen, I. Taavitsainen, P. Pahta, & M. Korhonen (eds.), *The dynamics of linguistic variation: Corpus evidence on English past and present*. Amsterdam: John Benjamins. 291–309.
Torres Cacoullos, Rena, & Travis, Catherine E. (2014). Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63:19–34.
_____ (2015a). Foundations for the study of subject pronoun expression in Spanish in contact with English: Assessing interlinguistic (dis)similarity via intralinguistic variability. In A. M. Carvalho, R. Orozco, & N. L. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown: Georgetown University Press. 83–102.
_____ (2015b). Variationist typology: Structure of variable subject expression in English and Spanish. Paper presented at the New Ways of Analyzing Variation 44 Conference, Toronto, October 22–25.
_____ (2016). Two languages, one effect: Structural priming in code-switching. *Bilingualism: Language and Cognition*. doi:10.1017/S1366728914000406.
Travis, Catherine E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2):101–135.
Travis, Catherine E., & Torres Cacoullos, Rena. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4):711–748.
_____ (2014). Stress on *I*: Debunking unitary contrast accounts. *Studies in Language* 38(2):360–392.
Travis, Catherine E., Torres Cacoullos, Rena, & Kidd, Evan. (2016). Cross-language priming: A view from bilingual speech. In G. J. Kootstra & P. Muysken (eds.), Special issue, *Bilingualism: Language and Cognition*. doi:10.1017/S1366728915000127.
Walkden, George. (2013). Null subjects in Old English. *Language Variation and Change* 25(2):155–178.
Weiner, E. Judith, & Labov, William. (1983). Constraints on the agentless passive. *Journal of Linguistics* 19(1):29–58.
Weir, Andrew. (2012). Left-edge deletion in English and subject omission in diaries. *English Language and Linguistics* 16(1):105–129.

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press

APPENDIX: TRANSCRIPTION CONVENTIONS

SBCSAE (DU BOIS ET AL., 1993)

Carriage return	IU break
.	final intonation contour
,	continuing intonation contour
?	appeal intonation contour
…	medium pause (> .7 sec)
..	short pause (about .5 sec)
--	truncated intonation contour
-	truncated word
=	lengthening
[ ]	speech overlap
[N N]	used to distinguish adjacent overlaps
(H)	in-breath
(TSK)	click
%	glottal stop
!	booster: emphatic speech
<Q Q>	quoted speech, produced with marked voice quality
~	name is a pseudonym

PEAR STORIES (CHAFE, 1980)

Note that the conventions are slightly different for the Pear Stories and SBCSAE, because the method had not been fully developed at the time the former were transcribed, and the sound files are not available to us to standardize the transcriptions.

.	sentence-final falling intonation
?	sentence-final level or rising intonation
,	clause-final but not sentence-final intonation
/?/	portion unintelligible
--	lengthened segment
•	break in timing too short to be measured as pause
[X]	pause lasting *x* seconds
[X[Y]a--nd u--m [Z]]	a sequence lasting a total of *x* seconds, consisting of a pause of *y* seconds, "a--nd u--m," and a pause of *z* seconds

https://doi.org/10.1017/S0954394515000174 Published online by Cambridge University Press