Stress on *I*

Debunking unitary contrast accounts

Catherine E. Travis and Rena Torres Cacoullos*
Australian National University, Pennsylvania State University

Much previous work on stress describes its function as being that of marking contrast. While some evidence has been adduced in experimental studies, work on spontaneous speech data has been plagued by a lack of operational definitions. To address this, we examine approximately 1,500 tokens of the English first singular subject pronoun in a corpus of conversational American English. Independently motivated operationalizations of contrast fail to support an overarching contrastive function of stress on *I*. Rather, examining co-occurrence patterns through multivariate analysis, we find that, besides chunked units (including discourse formulae as delimited by frequency and positioning), patterns of stress are subject to context-dependent discourse factors: accessibility (measured in distance from the previous mention), in tandem with coreferential priming (a tendency to repeat a preceding coreferential stressed *I*), as well as turn taking (an initial-position effect), and contrast in a semantic sense (manifested in higher rates of stress under negative polarity).

Keywords: stress, subject realization, variation, contrast, accessibility, priming, constructions

1. The discourse function of stress

Prosodic stress in English is variable; that is, speakers have a choice as to whether or not they stress a word in a given discourse context. In this paper we ask what factors influence a speaker's choice to stress the English subject pronoun *I*, as in (1), where the *I* in *I think* (line 4) is unstressed and the *I* in *I know* (line 6), is stressed, as indicated by the asterisk.

(1) 1. Ron: You can't live off of it. ((referring to being a musician))
2. Frank: .. No no.
3. It's it- --

Studies in Language 38:2 (2014), **360–392**. DOI 10.1075/sl.38.2.04tra
ISSN 0378–4177 / E-ISSN 1569–9978

4. I think,
5. … for .. per[formances],
6.→ Brett: [I* know somebody who makes a] living as a d-
musician.
(19 Doesn't Work in This Household: 365–370)[1]

Stress is used to "render salient the material with which [it is] associated" (Pierrehumbert & Hirschberg 1990: 288), and is widely interpreted as a marker of contrast. However, this interpretation is typically drawn from the analyst's intuitions (e.g., Bolinger 1961; Chafe 1976; Givón 2001; Halliday 1994), or from experimental data alone (e.g., Krahmer and Swerts 2001). Furthermore, the notion of "contrast" is ill-defined (Myhill & Xing 1996: 308), a problem which extends to studies examining stress in more naturalistic spoken data (e.g., Dehé 2009; Dehé & Wichmann 2010a, 2010b).

In order to address these issues, we first configure two operational definitions of contrast (one related to converse predicates (cf. Myhill & Xing 1996), and the other related to the marking of contrastive attitudes (e.g., Dehé & Wichmann 2010a), which we apply to the subject pronoun *I* in a corpus of conversational American English. We find that neither explains the patterning of stressed *I*, and we therefore turn from such unitary contrast accounts to exploration of the set of factors jointly conditioning the variation, from which we are able to offer a more complete account of actual usage.

2. Data: Stressed *I* in spontaneous speech

The data for this study are drawn from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al. 2000–2005). The SBCSAE is a corpus of recordings of naturally occurring spoken interaction which have been carefully transcribed to include prosodic information. Central to this transcription method is the Intonation Unit (IU), each of which is represented on a separate line. The IU is a prosodic unit defined on the basis of contour boundaries (Chafe 1994: 58–59), as "a stretch of speech uttered under a single, coherent intonation contour" (Du Bois et al. 1993: 47). The transcription also includes the annotation of pauses and interactive features such as overlap, as outlined in the Appendix. It does not, however, include the placement of stress, and thus we began the study by determining the status of stress on tokens of the first-person singular subject pronoun *I*.

For this study, we use the published transcripts of 20 conversations (approximately 85,000 words, drawn from SBCSAE I, II, IV).[2] These represent 60 speakers, 35 females and 25 males from across the United States, ranging from 11 to 76

years of age (though primarily between the ages of 20 and 60), the majority Anglo-Americans with a high school or college education.

All tokens of *I* as a first person singular subject in these 20 conversations were extracted (N = 2,893). Four linguistics graduate students, native speakers of American English, identified stress on *I* auditorily by listening to the IU in which the token of *I* appeared, using the program *Transcriber*, where the transcription of the target IU was linked with the corresponding audio.[3] The excerpts consisted primarily of only the IU in which the token of *I* appeared so that the perception of stress would rely as much as possible on auditory impressions of that individual IU, rather than on expectations based on lexical or grammatical features of the fuller discourse context.

The taggers were instructed to classify the *I* as stressed or unstressed, attending to features identified in prior work as contributing to the perception of stress, namely vowel quality and length, syllabic status and amplitude (Pierrehumbert & Hirschberg 1990: 272) (cf. also Beckman 1986; Fry 1958). Any tokens for which the tagger felt she couldn't make a determination as to the status of stress were coded as "undeterminable" (as was sometimes the case, for example, in cases of overlap, soft speech, marked voice quality, or poor sound quality).

Each tagger listened to subsections of the data, such that each token was independently coded by three different taggers. Tokens for which all three taggers agreed on classifying as stressed or unstressed totaled 1,538 (53% of all extracted instances of *I*). There were a further 336 tokens for which there was no disagreement between the taggers, in the sense that either two taggers agreed and the third tagger classified the status of the token as undeterminable (N = 323); or the three taggers agreed that stress was undeterminable (N = 13). The taggers disagreed about the status of stress for approximately one third of the tokens (1,019/2,893); it is unknown what might account for the high disagreement rate.

Acoustic analysis based on objective measures of difference verified that the auditory classification of the agreed upon tokens was reliable. All eligible "stressed" tokens (N = 142) and a corresponding sample[4] of the "unstressed" tokens were submitted to acoustic analysis in *Praat* (Boersma & Weenink 2011). The mean duration of stressed *I* was found to be longer than that of unstressed *I* (.15 vs. .09 sec respectively; paired $t(141) = 14.03$, $p < .0001$), the mean intensity louder (71.17 Db vs. 66.73 Db respectively, $t(141) = 5.69$, $p < .0001$), and the mean peak pitch higher (225.17 Hz vs. 187.97 Hz respectively, $t(141) = 3.91$, $p < .0001$).

The present analyses are based on the agreed upon tokens only, thus allowing us to ensure that our comparison sets are as distinct from each other as possible. Of these, 174 were classified as stressed and 1,364 as unstressed.

3. Circumscribing the envelope of variation: The behaviour of discourse formulas

In order to identify the factors motivating speakers to stress *I*, we adopt the *principle of accountability* (Labov 2005: 7) and take into consideration both the form under examination and the form with which it co-varies to calculate their relative frequencies. It is only by doing so that we are able to test whether an ascribed function is, or is not, associated with stress on *I*, as if there is such an association, then *we should observe a higher rate of stressed vs. unstressed I when the ascribed function is evinced than when it is not*. Thus, we need to identify the contexts constituting the envelope of variation, that is, when speakers have a choice.

First, we exclude clauses with interrogative syntax; not only is stress on *I* near-categorically absent in such a context (only 1 stressed *I* in 34 interrogatives), but also the factors affecting stress on *I* may differ in interrogatives and declaratives.

Second, after careful demarcation, we set aside a number of formulaic units. It has been proposed that there is an association between stress and the discourse-marker status of expressions such as *I think* and *I believe*. For example, based on 41 "sentence-initial" tokens of each of *I think* and *I believe* drawn from the "public domain" spoken portion of the British Component of the International Corpus of English (which includes formal dialogues, scripted and unscripted monologues, and broadcast news), Dehé & Wichmann propose that stress on *I* in these collocations serves to indicate that they are functioning as main clauses, expressing "true speaker attitude, opinion or belief"; lack of stress on *I*, on the other hand, is associated with discourse marker use (2010b: 63).

But how do we know whether a given instance of an expression is functioning as a discourse marker or not? That is, how can more propositional, or "clausal", instances be replicably distinguished from less propositional, or more formulaic "discourse-marker" instances, independently of their prosody (in order to allow any prosodic correlations to be identified)? Dehé & Wichmann identify clausal (vs. discourse marker) instances of *I think* by way of traditional tests based on the analyst's intuition: clausal *I think* cannot be omitted; can be replaced by *it is my view that*; and can take adverbial modification (e.g. *I really think*) (Dehé & Wichmann 2010b: 55–56). Due to the unreliability of such tests and the indeterminacy of the status of a collocation in any one specific instance as a discourse marker (cf. Thompson 2002), we seek out an operational definition of discourse markers, based on frequency and positioning.

We begin with a set of *I* + verb collocations which have been characterized independently as formulaic based on patterns of complementizer *that* — *I think, I mean*, *I guess, I remember* and *I'm sure* (Tagliamonte & Smith 2005: 299; Thompson & Mulac 1991: 244; Torres Cacoullos & Walker 2009: 21),[5] or based on phonetic

Table 1. Discourse marker status of frequent collocations (in parenthetical or prosodically independent uses)

Collocation	N collocation	N Verb + 1sg subject	% of verb represented by collocation	Discourse marker uses		% of collocation represented by discourse marker uses
				N parenthetical	N prosodically independent	
I mean	188	188	100% (188/188)	3	88	48% (91/188)
I guess	36	36	100% (36/36)	5	8	36% (13/36)
I'm sure	16	16	100% (16/16)	2	1	19% (3/16)
I remember	12	18	66% (12/18)	0	2	17% (2/12)
I think	87	157	55% (87/157)	6	12	21% (18/87)*
I don't know	59	114	52% (59/114)	1	18	32% (19/59)
I know	35	114	31% (35/114)	0	24	69% (24/35)
TOTAL	433	529**	82% (433/529)	17	153	39% (170/433)

*There are a further 11 tokens of *I think* as a quotative, also considered formulaic (see below).
**Counting *know* once.

reduction, in the case of *I don't know* (Scheibman 2000: 120). We also include *I know*, due to its high frequency, and because it provides a valuable comparison with *I don't know*.[6] Each of these fixed collocations (i.e. with no variation in polarity, tense, and modal use) represents a high proportion of the occurrences of the corresponding verb with a 1sg subject, ranging from 100% in the case of *I mean, I guess,* and *I'm sure*, to close to one third in the case of *I know*, as shown in the first three columns in Table 1.

We then distinguish discourse marker from clausal uses of these frequent collocations based on positioning, both in the clause and in the Intonation Unit. In terms of position in the clause, all cases of these collocations that occurred as parentheticals (that is, when they appear between the subject and verb or following the verb, as in (2)), were coded as discourse markers. Despite the great attention such parenthetical forms have received in the literature (e.g., Dehé & Wichmann 2010a; Thompson 2002; Traugott 1995), this use is in fact quite rare in the aggregate, accounting for just 4% (17/433) of all tokens of these collocations (with the highest proportion for *I guess*, at 15% (5/36)), as shown in the fourth column of Table 1.

(2) Corinna: … he's gonna divorce her I think.

(45 The classic hooker: 519)

In terms of position in an Intonation Unit, we coded as discourse markers tokens that were produced prosodically independently from other clausal material. This is based on the understanding that material that occurs in the one IU tends

to have a tighter syntactic relationship than material that occurs across different IUs. Accordingly, direct objects (including clausal complements) tend to occur in the same IU as their verb. In English Pear Story narratives, for example, Croft (1995: 849) found that 82% (179/219) of clausal complements are realized in the same IU as the verb. Thus, if a verb is functioning as a canonical main clause, then it would be more likely to occur in the same IU as its complement; when it occurs in a separate IU, we draw an association between prosodic and syntactic independence, and treat it as a discourse marker (for discussion of the prosodic independence of discourse markers, see Redeker (1991: 1166), Tao (2001: 129), Thompson (2002: 143–145), and Travis (2005: 48–52)).[7] We considered to be prosodically independent those tokens that occurred strictly alone in their own Intonation Unit, as for *I think* in (1) and *I don't know* in (3),[8] or preceded by a conjunction (e.g. *and I think*). Unlike parenthetical use, prosodically independent use is robust, accounting for approximately one third (153/433) of all tokens of the collocations, as shown in the fifth column of Table 1.

(3) Pamela: Well,
I don't know,
I guess it must,
(05 A book about death: 237–239)

We classified as clausal, on the other hand, all other tokens of these collocations that do not occur on their own in an IU, most commonly introducing a clause in the same IU, such as *I guess* in (3) and (5) below, and occasionally occurring with a nominal direct object, as in *I know* in (1). Also classified as clausal were all other subject+verb combinations. We underline that our operationalization is based on the understanding that instances of these constructions on their own in an IU (or as parentheticals with respect to the clause) tend to function as discourse markers, while instances occurring in an IU with other material are unlikely to do so.[9] Our strict criteria yield a conservative measure of discourse marker status, but this is necessary in order to operationalize the notion of formulaicity in a replicable way (see Thompson 2002 for a less conservative application of the term, and correspondingly higher proportions of formulaic uses).

As seen in the last column in Table 1, the proportion of use of these collocations as discourse markers thus defined ranges from approximately one fifth to two thirds, with the most frequent being *I mean*.

Also classified as formulaic were quotatives with *say, be like, go, ask,* and *think* introducing quoted speech or internal thought, as in examples (4) and (5) below (N = 81).

(4) Alina: I said,

<VOX oh,
you want a matching set VOX>?

(06 Cuz: 1282–1284)

(5) Miles: And I'm thinking,
we=ll,
… I guess that's her husband,

(02 Lambada: 1313–1315)

In sum, there are a total of 251 instances of *I* + verb which we classify as having formulaic status thus defined, or 17% of the data. As can be seen in Table 2, these formulaic instances show a strong disfavoring of stress, with a rate of just 4%, compared with 13% for other tokens. This distinction can be seen clearly in the case of *think*, for which the rate of stressed *I* is halved in its formulaic use, either as a discourse marker, as in (1) and (2), or as a quotative, as in (5), as opposed to its other, more clausal, occurrences (7%, 2/29, vs. 16%, 21/128, respectively). On the other hand, *I mean* and *I guess* are virtually categorically unstressed, regardless of whether they are formulaic or not — 1/188 *I mean* and 0/36 *I guess* tokens are stressed.

Given the minimal variability for discourse formulae and the (virtual) lack of variability for all occurrences of *I mean* and *I guess*, we exclude all such tokens from the analysis of variation. This leaves us with a total of 1,133 tokens, with a rate of stress of 14% (163/1,133), as depicted in Figure 1. What, then, motivates a speaker to stress *I*?

Table 2. Stressed *I* according to clausal status of verb (N = 1,504)

	% stress (N)	% data
Discourse formula	4% (10/251)	17%
Clausal	13% (163/1,253)	83%

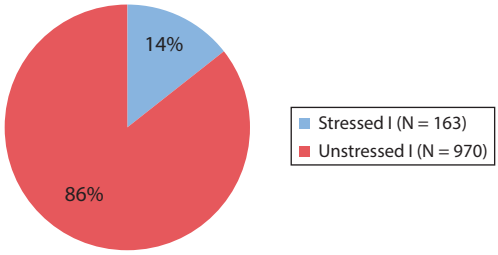

Figure 1. Rate of stress on *I* in the variable context (N = 1,133)

4. The quest for a unitary account of stress as marking contrast

It is widely understood that stress functions to highlight "salient" information (cf. Pierrehumbert & Hirschberg 1990: 288). This salience is typically associated with discourse features, such that information which is either newly introduced into the discourse and/or expresses contrast receives stress. Thus, just as new information is more likely to be introduced with full NPs rather than pronouns in accordance with the cross-linguistic tendency whereby less accessible, or less continuous, referents receive more linguistic coding than more accessible referents (Ariel 1990; Chafe 1994; Du Bois 1987; Givón 1983b; Levinson 1987; Prince 1981; inter alia), so do stressed rather than unstressed forms signal new information (Givón 1983b: 17; Hirschberg & Pierrehumbert 1986: 141). Complementing this accessibility explanation is a widespread understanding that stress marks information that is contrastive (e.g., Bolinger 1961: 84), in particular for elements such as pronouns which tend to represent given information (e.g., Chafe 1976: 35; Givón 2001: 251; Halliday 1994: 298).

The association between prosodic and discourse salience has been probed experimentally. For example in a dialog game for Dutch, in which speakers describe colored figures (red triangle, red square, blue cylinder), Krahmer & Swerts (2001: 398) found that a word was more likely to be stressed if it referred to a property or object not mentioned previously in the game ("new information"), or if a different value had been mentioned in the previous turn (red triangle > blue triangle; red triangle > red square) ("contrast"). The authors of this study note that "the data thus obtained allow an unambiguous operationalization of the relevant contexts" (2001: 394). Such unambiguous instances of "contrast" and "new" information are, however, much more difficult to obtain in spontaneous discourse.

Here we test operationalizations of contrast applicable to natural speech. We begin with an operationalization of contrast in a semantic sense, that of "converse predicates", based on Myhill & Xing (1996), and then consider contrast in an interactional sense, related to the expression of speaker opinion vis-à-vis the interlocutor (Dehé & Wichmann 2010a, 2010b), by examining turn position in relation to cognition verbs and the preceding clause subject.

4.1 A semantic test of contrast: Converse predicates

The notion of contrast based on converse predicates derives from an operationalization developed by Myhill & Xing to test for a contrastive effect of object position in narratives in Chinese and Biblical Hebrew. Utilizing what they term "double contrast" (a term also used by Chafe 1976: 35; Mayol 2010: 2499–2501), Myhill & Xing propose that "in contrastive pairs there are two or more elements which

are different in two clauses (either verbs with opposite meanings or nonverbal elements in a set relationship)" (1996: 314). Following Travis & Torres Cacoullos (2012: 715–718), we identify such contrastive 1sg subjects as those which occur in a pair of clauses at a distance of no greater than three clauses that differ, first, in the subjects, and second, in the predicates, which must be related but in some sense converse, e.g., negated (*believe / not believe*), doing vs. not doing (*take along / leave behind*), opposite direction (*take / give*), and so on.

The following two examples illustrate our application of this operationalization. In (6), in the target clause marked with the arrow, *I* (referring to Harold) can be understood to contrast with 'he' in line 1 (referring to Pete), and the negated verb *wasn't* with the preceding affirmative use of this same verb in the predicate 'was my fashion consultant'. Thus, here we have converse predicates established via negation. In (7), the converse predicates refer to distinct preferred travel destinations (Mexico and Guatemala); the subjects of these two clauses differ as they are non-coreferential 1sg tokens (the first referring to Joanne and the second to Ken). Note that in both these examples, the target *I* (indicated with the arrow) is stressed.[10]

(6) Miles: … He was my fashion consultant today.
Pete: Right.
… @@@
Miles: Hm.
Pete: @ (H)
→ Harold: … I* wasn't?

(02 Lambada: 185–190)

(7) Joanne: I'd ra[ther go to] Mexico though.
Ken: [I] --
Lenore: … @[2=
Joanne: [2@@@@2]
Lenore: Not quite2] the same.
Joanne: [3@3]
→ Ken: [3(H)3] I*'d kind of like to go to Guatemala.

(15 Deadly Diseases: 246–252)

We examined this operationalization on a subset of the data, obtained by extracting 80 consecutive tokens of *I* from the first seven transcripts of the SBCSAE, which, following exclusions, left a total of 320 tokens for analysis.[11] Table 3 gives the distribution of stressed *I* according to converse predicates in these data. Note that the rate of stress in contrastive contexts thus operationalized is over twice that when there is no such contrast (40% vs. 19%), suggesting a correlation between stress and contrast in the sense of converse predicates.

Table 3. Stressed *I* according to converse predicates in the three preceding or following clauses (N = 320)

	% stress (N)	% data
Contrastive (converse predicates)	40% (4/10)	3%
Non-contrastive	19% (59/310)	97%

One problem with this operationalization is that, notwithstanding the occurrence of some apparently clear-cut examples such as (6) and (7), it includes tokens for which the context does not support a contrastive meaning, and also excludes some tokens that do appear to be contrastive, as also noted by Travis & Torres Cacoullos (2012: 717–718) and Myhill & Xing (1996: 305). What renders converse predicates inadequate as an account of stress on *I*, however, is the exceedingly low number of tokens that meet this operational definition (N = 10), accounting for just 3% of all tokens of *I* and 6% (4/63) of all tokens of stressed *I*. This low rate is similar to that found by Travis & Torres Cacoullos (2012: 718, 724) for first-person singular subjects in Spanish conversational data, and by Paredes Silva (1993: 41–43) applying a slightly broader definition to include "contrast" and "emphasis" with first singular subjects in Brazilian Portuguese letters.[12]

In sum, contrast as defined in terms of converse predicates applies to only a minimal proportion of the data; this notion, however, points to negative polarity as a feature linked to contrast in a semantic sense, which we return to in Section 5.4 ahead.

4.2 An interactional test of contrast: Turn-initial position with cognition verbs and non-coreferential 1sg preceding subjects

An alternative interpretation of contrast is an interactional one, and this has been ascribed specifically to stressed *I* in the discourse formulas *I think* and *I believe* by Dehé & Wichmann, who propose that "If … the accent is on the pronoun …, the focus is on the speaker's opinion and may involve a contrast to his/her interlocutor's attitude" (2010a: 18). If this is the case, then we may expect stressed *I* to be favored in turn-initial position with verbs of cognition such as *I think*, which express speaker attitude, or, alternatively, when the context sets up an interactional contrast, such as when the preceding clause subject is a non-coreferential *I*, that is, when the interlocutor has ended their turn with a self-reference.

We code as turn-initial any tokens that occurred in a turn-initial IU (as in line 4 and 9 in (8)), or, in some cases, in the second IU in a turn, where the first IU is made up of non-substantial material (such as a minimal response with continuing intonation, represented by a comma, as in line 4 in (9)).

(8) 1. BERNARD: It was horrible.
2. .. Yeah[2=2].
3. FRAN: [2Yeah2]=.
4.→ SEAN: And so I went to .. the design center,
5.→ and I went to all the places,
6. and,
7. … that was the one.
8. FRAN: .. Yeah.
9.→ ALICE: … Why do we have to look at the one I*'m % @on.
(51 New Yorker's Anonymous: 1128–1136)

(9) 1. JAN: You argued with me long enough,
2. I changed my mind.
3. MELISSA: .. Okay,
4.→ .. I=* retract .. all my arguing.
(19 Doesn't work in this household: 110–113)

In order to distinguish interactional considerations from possible prosodic ones (due to the position of the *I* in the prosodic unit itself), we also code for position in the IU. IU-initial position is exemplified in the two tokens of *I* in (9). IU-medial position includes tokens following a conjunction (most frequently *and,* N = 142, as in line 5 in (8)), an adverbial (e.g. *all of a sudden*), or clausal material, as in line 9 in (8).

Table 4 gives the rate of stress according to turn and IU position. There is a strong tendency for *I* to be stressed proportionally more in absolute-initial than non-initial position, both in terms of the turn (22% vs. 12%, $p < 0.0001$ by Fisher's exact test), shown in the bottom row, and the IU (18% vs. 10%, $p < 0.0005$), shown in the rightmost column. What is important is that, as indicated in this cross-tabulation, in IU-initial position (the first row) the turn-position effect is maintained (with a rate of stress of 24% turn-initially vs. 15% non-turn initially; $p = 0.0205$); furthermore, in turn-initial position (the first column), the difference by IU position does not achieve significance (24% IU-initially and 18% non-IU initially; $p = 0.2752$). In other words, there is a genuine effect on stressed *I* attributable to

Table 4. Stressed *I* according to position in the turn and Intonation Unit (N = 1,129)

Position in the IU	**Position in the turn**		**TOTAL**
	Turn initial **% stress (N)**	**Non-turn initial** **% stress (N)**	**% stress (N)**
IU initial	24% (43/177)	15% (65/430)	18% (108/607)
Non-IU initial	18% (16/90)	9% (38/432)	10% (54/522)
TOTAL	22% (59/208)	12% (103/862)	

turn position. In fact, it is in initial turn- and IU-position (as in both tokens in (7), and line 4 in example (9)) where we observe the highest rate of stress (24%). Thus, in the following we focus on the effects on stress of absolute (turn- and IU-) initial position.

Is the higher rate of stress in initial-position to be interpreted straightforwardly in terms of turn-taking or more indirectly in terms of contrasting opinions, as proposed by Dehé & Wichmann (2010a: 18)? If it were related to the expression of opinions, we would expect to find an even higher rate of stress with cognition verbs such as *think, know, remember, be sure*. Overall, we do observe a higher rate of stress with cognition verbs than with other verbs (18%, 54/304 vs. 13%, 105/811, $p = 0.0437$) (excluding, the reader is reminded, the use of these verbs in formulaic expressions). But while the position effect for non-cognition verbs is strong (with a rate of stress of 26% (25/98) in initial position and just 7% (25/360) in non-initial position), for cognition verbs, though in the same direction, there is no significant difference (with rates of 22% (17/77) and 18% (12/67) respectively). The initial-position effect, then, is not tied to cognition verbs, thus failing to support a contrastive-opinion function for stressed *I*.

Alternatively, if the initial-position effect were due to a contrast with the interlocutor's attitude (e.g., Dehé & Wichmann 2010a: 18), then we would expect a higher rate of stress when the preceding clause subject refers to the interlocutor, as in (9) above. However, the rate of stress on *I* in initial-position contexts where the preceding clause subject was a non-coreferential 1sg is just 15% (5/33), not significantly different from when the preceding clause had a non-coreferential 3sg subject, as in (6) (19%, 97/517) ($p = 0.8173$). Thus, stress on *I* according to grammatical person of the preceding clause subject also fails to support an interpretation of the initial-position effect as a manifestation of the expression of speaker stance vis-à-vis the interlocutor. As we will see in 5.1 below, the higher rate of stressed *I* overall in non-coreferential contexts is best construed as a graded effect of distance from the previous mention, a manifestation of considerations of accessibility (Givón 1983b).

In sum, the rate of stress on *I* tends to be higher in initial than non-initial position overall, but as this is not particularly the case with cognition verbs rather than non-cognition verbs nor with an immediately preceding non-coreferential 1sg rather than 3sg subject, it cannot be interpreted as evidence of an interactional contrastive function for stress on *I*.

4.3 The lack of a unitary contrast account for stress on *I*

The two operationalizations tested here lend little support for an overarching contrastive function of stress on *I*, whether in a semantic sense, in terms of converse

predicates (which cover a miniscule portion of the 1sg subject data), or in an interactional sense, in terms of initial position with cognition verbs or with a preceding non-coreferential *I* (neither of which correlates with a higher rate of stressed *I*). However, we have noted several linguistic features that may correlate with a higher rate of stress, namely negative polarity, absolute-initial position, cognition verbs, and distance from the previous mention.

Unlike in an experimental set up (e.g., Krahmer & Swerts 2001), in natural language use, these contextual features occur together. For example, in (10), both tokens of *I* occur in negative polarity contexts, and neither occurs with a cognition verb; but the first token occurs in absolute-initial position, while the second does not; and the first occurs in a non-coreferential context, while the second occurs in a coreferential context. A further factor worthy of note that has not arisen as yet is repetition, as with the *I* in *I've never seen* which occurs with another stressed *I* immediately preceding it, a widely found conditioning factor in variation.

(10) Pamela: we think of her,
(H) you know,
smoking (H) .. cigarette smoke into the faces of .. William Holden
and,
(H) and the like.
Darryl: I* don't,
I*'ve never seen those movies.
(05 A book about death: 260–266)

Recognizing that all these factors work together in natural discourse to jointly condition the variation, we abandon the quest for a unitary, overarching account of stress on *I*, and instead focus our attention on accounting for the multiplicity of factors that play a role. For this, we apply the variationist approach (Labov 1969; Sankoff 1988a), as we outline in the following section.

5. A variationist approach to stress on *I*

Linguistic variability is structured, conditioned by elements of the linguistic and extralinguistic context, which contribute to speaker choice among variant forms that have overlapping functions, here, stressed and unstressed first-person singular subjects, which in broad terms function to index the speaker. The variationist method yields the *linguistic conditioning* of variant selection in the terms of probabilistic statements about the co-occurrence of variant expressions and elements of the linguistic context in which they appear. We use a kind of multivariate analysis known as Variable-rule analysis (Sankoff 1988b), in order to account for

Table 5. Variable-rule analysis of the contribution of factors selected as significant to the choice of stressed *I* in conversational American English (Santa Barbara Corpus of Spoken American English, SBCSAE)

N = 1,133; Input: .11 (Overall rate: 14%)				
	Prob	**% *I* stress**	**N**	**% data**
Distance from previous mention as subject*				
Greater distance (2+ intervening clauses)	.60	20%	487	46%
Lesser distance (0–1 intervening clauses)	.41	9%	574	54%
Realization of previous mention as subject**				
Stressed *I*	.69	21%	67	13%
Unstressed *I*	.47	10%	464	87%
Turn and IU position				
Initial (Turn and IU)	.61	24%	177	16%
Other	.54	16%	520	46%
Medial (Turn and IU)	.41	9%	432	38%
Polarity				
Negative	.61	21%	216	19%
Affirmative	.47	13%	917	81%
Semantic class of verb				
Cognition	[.49]	18%	307	27%
Other	[.50]	13%	810	73%

Non-significant factor group (Probabilities given in square brackets): Semantic class of verb.[14]
*Excluding cases in which the target token or the previous coreferential mention occurs as part of quoted speech (N = 63), or there is unclear speech intervening between the coreferential mentions (N = 9).
**Excluding tokens in which the target token occurs at a distance of over five intervening clauses.

the variation in the data by identifying a statistically significant subset of the factor groups operationalizing hypotheses about the function(s) of stressed *I*. *The particular function(s) ascribed to variant forms, or the lack thereof, are thus quantitatively ascertainable in their linguistic conditioning* (Poplack & Malvar 2007: 137–143; Poplack & Tagliamonte 2001: 88–94; Silva-Corvalán 2001: 133–138).

Table 5 depicts a multivariate model of contextual effects on speakers' choice of stressed subject pronoun *I*. The linguistic sub-contexts, or factor groups (predictors or independent variables), which we hypothesize to condition stress on *I*, appear in the left-hand column. Shown in the first column of numbers are the Probabilities: factors with a Probability closer to 1 can be said to *favor*, and those with a Probability closer to 0 to *disfavor*, speakers' selection of stressed *I*.[13] Subsequent columns show, for each factor (linguistic sub-context), the rate of

stressed *I*, the number of tokens, and how much of the data in that factor group these tokens constitute.

These results indicate that stressed *I* is favored when the previous coreferential *I* occurred at a distance of two or more clauses and when the previous mention was realized as stressed, when the target token is in absolute-initial position, and in negative polarity contexts. Semantic class of the verb, however, does not make a significant contribution when the effect of these factor groups is considered together (unlike what we saw in the univariate analysis above, Section 4.2). We will now go on to discuss each of these constraints.

5.1 Distance from the previous coreferential 1sg subject: A measure of accessibility

Givón's (1983a) influential cross-linguistic volume on topic continuity established that less continuous referents (or referents that are lower on the accessibility scale) tend to be associated with more linguistic coding, noting that "the more disruptive, surprising, discontinuous, or hard to process a topic is, the more *coding material* must be assigned to it" (1983b: 18). Conversely, Levinson (1987: 384) observes that "the more 'minimal' the form, the stronger the preference for a coreferential reading" (cf. also Ariel 1988: 79). "More" linguistic coding (vs. a more "minimal" form) is understood to refer to stressed (vs. unstressed) pronouns, pronouns (vs. unexpressed mentions), or full Noun Phrases (vs. pronouns) (Givón 1983b: 18). In this sense, stressed pronouns in a language like English are thought to correspond to expressed pronouns in a language like Spanish (cf. Payne 1997: 43). Accordingly, the prediction is that stressed pronouns would be more likely to occur with less accessible referents.

How might this apply to the first person, which, as a discourse participant, is considered to maintain a position high on the accessibility scale? According to Chafe, the pronoun *I* may be stressed when speakers "bring the idea of themselves back into the active consciousness of the listeners" (1994: 87).

In reports on variable subject pronoun expression in so-called null subject or pro-drop languages, what has been widely found, including for first person, is a switch-reference effect, whereby an expressed pronoun is more likely when the immediately preceding clause subject is not coreferential (see Silva-Corvalán 2001: 154–169, for a review of studies of Spanish).

Here, we take into account the distance between coreferential mentions to operationalize accessibility. Distance was counted in clauses[15] from previous mention as subject, as referents occurring as subjects tend to be more "topical" than those occurring in other syntactic roles (Givón 1983b: 22). Furthermore, for first singular, subjects are the great bulk of mentions in the data.[16] Thus, coreferential

tokens are those that occur with no clauses intervening, illustrated in (11); switch reference tokens are those with one or more clauses intervening, and here we consider different degrees of distance. For example, in (12) and (13), while both qualify as switch reference contexts, they differ in terms of distance; in (12) there is only one, and in (13) there are three intervening clauses (indicated with dotted underlining). We find that, for subject pronoun stress, what counts is distance, not switch reference.

(11) coreferential / 0 intervening clauses
Joanne: (H) that's what I'd do,
→ I'd pick it up the night before.
(15 Deadly Diseases: 1742–1743)

(12) switch reference / 1 intervening clause
Alina: Well I didn't get along with !Dennis at all,
he was a jerk.
→ .. I did not like him.
… Period.
(6 Cuz: 937–940)

(13) switch reference / 2 (or more) intervening clauses
Jamie: (TSK) (H) No=,
.. I was there before,
it was fine.
.. it was [really fun].
Harold: [You were the=re] before?
Jamie: Yeah=.
→ .. I went out there before.
(02 Lambada: 1280–1286)

Figure 2 below shows a breakdown of the rate of stress by the number of intervening clauses. Note that there is no coreferentiality effect here, with the rate of stress being identical (9%) at zero (coreferential) and one intervening clause (switch reference); from two clauses upwards (also switch reference), there is a gradual increase in stress. For this reason, we applied this split in the multivariate analysis depicted in Table 5, and configured the Distance factor group such that instances of the variable in which the previous coreferential *I* occurred at a distance of zero or one intervening clauses ("lesser distance") (approximately one half of the data) are opposed to instances in which the previous coreferential *I* was two or more clauses away ("greater distance"). With this split, we observe an accessibility effect in the predicted direction, with stress on *I* more favored at the greater degree of distance than at lesser distance.[17]

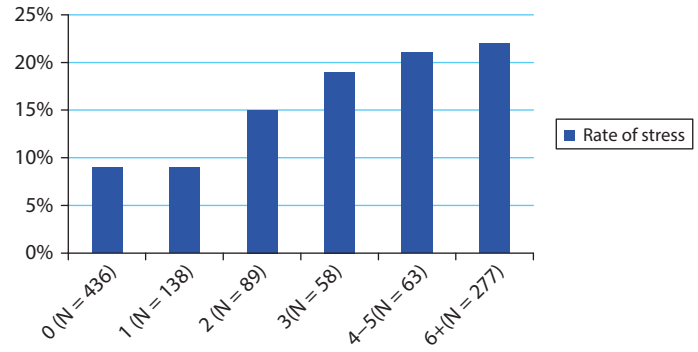

Figure 2. Rate of stress on *I* according to distance in clauses from previous mention (N = 1,061)

Accessibility in terms of distance has received much less attention in the literature than has switch reference; nevertheless, this distance effect accords with the same general tendency for less accessible material to receive more linguistic coding. Thus, it would appear that subject continuity / accessibility can be more or less local. More locally, what may count is the subject of the immediately preceding clause, that is, whether the preceding clause subject was coreferential or not (as is the case, for example for Spanish subject pronoun expression (Torres Cacoullos & Travis forthcoming)). Less locally, what may matter is distance, that is, how much discourse has passed since the last mention of the coreferential subject. This is the case here, where we have established that relevant for stress on *I* is an operationalization of accessibility at a distance of two or more clauses from the previous mention.

5.2 Realization of previous coreferential *I*: Local priming

Favoring stressed *I* in the multivariate analysis is realization of the previous coreferential 1sg subject as a stressed *I* — a *coreferential priming* effect. Priming is where the use of a certain structure in one utterance favors choice of that same structure in a subsequent utterance, including intonational patterns (e.g., Swerts et al. 2002: 638). Priming, or perseveration, has been characterized as "mechanical" rather than "functional" (Labov 1994: 547–568), being defined by Bock & Griffin (2000: 177) as "the unintentional and pragmatically unmotivated tendency to repeat" a preceding structure. Coreferential priming at a low degree of distance — no intervening clauses, a context that otherwise disfavors stressed *I*, as discussed above — is illustrated in line 6 in (14) below, and in (10) above.

(14) 1. Mary: … (TSK) (H) She was probably lonely when she was doing it,
2. you know that?
3. … She probably was.

4. … [X] --
5. Alice: [I*] sat up with her,
6.→ and I* was talking to her,
7. she was doing all the decorating, (07 A Tree's Life: 328–334)

Priming, as far as we know, is found in just about every study that tests for it (cf. Bock 1986; Labov 1994: 547–568; Poplack 1980; Scherre & Naro 1991; Weiner & Labov 1983, inter alia), and coreferential priming has been found to condition patterns of subject expression both in English (Torres Cacoullos & Travis 2013) and in Spanish (Cameron & Flores-Ferrán 2003; Torres Cacoullos & Travis 2011; Travis & Torres Cacoullos 2012). The question then becomes the coverage of this effect: how close to the target *I* does the previous mention have to be for priming to be observed, and how much of the data does priming apply to?

While the life span of structural priming remains under debate, there is some evidence from laboratory-based research that it can be maintained for up to 10 intervening sentences for some constructions, such as the prepositional vs. the double-object dative (Bock & Griffin 2000). In spontaneous discourse, for Spanish first singular subject expression, Travis (2007: 121) found that a significant coreferential priming effect was maintained at a distance of zero or one clause in interactive conversational data, but at distances of up to five clauses in more monologic narratives, which exhibited greater subject continuity.

As shown in Table 6, at a distance of zero clauses (the first row), the rate of stressed *I* is 22% when the previous mention was realized with stress but 5% when the previous realization was unstressed. Thus, priming of stressed *I* by a previously stressed *I* is strong in coreferential contexts, and at a distance of one clause we find the same tendency (20% vs. 9%). At distances of two or more clauses from the previous coreferential *I*, the priming effect has disappeared, with virtually identical rates of stress in both environments. Coreferential priming, therefore, has limited coverage, both in terms of its duration (it applies up to at most one intervening

Table 6. Rate of stress on *I* according to Distance and Realization of previous mention as subject (N = 531)*

Distance in clauses from previous coreferential *I*	**Previous realization**				**TOTAL**
	Stressed		**Unstressed**		
	% *I**	**N**	**% *I****	**N**	
0 (Coreferential)	22%	(9/40)	5%	(13/258)	7% (22/298)
1 (Local switch reference)	20%	(2/10)	9%	(7/76)	10% (9/86)
2 + intervening cl.	18%	(3/17)	19%	(25/130)	19% (28/147)
TOTAL	21%	(14/67)	10%	(45/464)	11% (59/531)

*See note notes on Table 5 for exclusions.

clause), and its pertinence (previous realization as stressed is an infrequent environment, given the low rate of stressed *I* overall).[18]

The strong priming effect in coreferential contexts yields a local pattern of clusters of "coreferential stressed *I* + stressed *I*" across consecutive clauses, as illustrated in (10) and (14). It is important to note that this effect confounds that of distance: while, in accordance with the received accessibility account, overall, the non-reduced form (stressed *I*) occurs proportionally more at greater than at lesser distances (20% vs. 9%, Table 5), we observe that when the preceding subject was a stressed *I*, the rate of stress is just as high at lesser distances (22%–20% with 0 or 1 intervening clauses, Table 6) as it is at greater distances (19% at two clauses and above).

5.3 Turn position: A turn-taking role for stressed *I*

We noted above the favoring of stressed *I* in initial position, and this is borne out in the multivariate analysis, with absolute initial position (initial in the speaker turn and in the prosodic unit), illustrated in line 5 of (14), shown to be the most favorable position for stress. Our examination of the initial-position effect both with respect to the semantic class of the verb and the grammatical person of the preceding subject (Section 4.2 above) led us to reject an interpretation of this effect as supporting a role for stressed *I* expressing a contrast between the speaker and the interlocutor. Instead, we interpreted it as having to do with turn-taking per se. However, given considerations of accessibility, as operationalized by distance, we are obliged to consider here how turn position may be related to the robust distance effect. This we examine in Table 7.

Not surprisingly, a greater proportion of initial instances have the previous coreferential *I* at greater distances, i.e., two or more intervening clauses, in comparison with non-initial instances (approximately double, at 82%, 141/173

Table 7. Rate of stress on *I* according to Distance and Position (N = 1,059)

Distance in clauses from previous coreferential *I*	Position				TOTAL
	Initial (absolute) (Turn and prosodic unit)		All other		
	% *I**	N	% *I**	N	
Lesser distance (0–1 intervening clause)	22%	(7/32)	8%	(44/541)	9% (51/573)
Greater distance (2+ intervening clauses)	24%	(34/141)	19%	(65/345)	20% (99/486)
TOTAL	24%	(41/173)	12%	(109/886)	14% (150/1,059)

for initial vs. 39%, 345/886, for all non-absolute-initial *I*). The position effect is, however, independent of the distance effect: the tendency toward higher stress in initial position obtains at lesser distances of zero or 1 intervening clauses (22% vs. 8%, $p = 0.0172$), and goes in the same direction at greater distances (24% vs. 19%). Thus, we interpret the favoring of stress on *I* in initial position as a genuine turn-taking effect, related to issues associated with taking the floor.

5.4 Polarity: Stressed *I* in denials

Finally, we observe in the multivariate analysis (Table 5) a favoring effect of negative polarity contexts. We noted above that negative polarity is one way in which converse predicates can be established, which may suggest a contrastive function in a semantic sense for stressed *I*. However, might the negation effect also be associated with an interactional contrastive function, in terms of a contrast vis-à-vis the interlocutor (Dehé & Wichmann 2010a: 18)? If this were the case, we would expect to find a higher rate of stress in negative polarity contexts that are turn-initial.

We consider the relationship between these two environments in Table 8. What we see is that the position effect does not obtain at all in negative polarity contexts, with the same rate of stress on negated *I* (21%) regardless of position, not what an interactional function of negated stressed *I* would predict. (The position effect is found in positive polarity contexts, where the rate of stress is twice as high in initial position as in non-initial position; 25% vs. 11%). This suggests, then, that the favoring of stress in negative polarity contexts is not associated with a contrast with the interlocutor's opinion.

The token numbers in Table 8 indicate, furthermore, that negative polarity contexts of *I* are not much more likely than positive ones to be in initial position (20% (43/215) of negated instances are initial, compared with 15% (134/914) of positive instances). Although it was claimed in the early literature (cf. Givón 1979: 104; Horn 1985: 143) that negation corresponds to something that was previously uttered or is presupposed, in conversational English, Tottie finds most (67%,

Table 8. Rate of stress on *I* according to Polarity and Position (N = 1,129)

Polarity	Position				TOTAL
	Initial (absolute) (Turn and prosodic unit)		All other		
	% *I**	N	% *I**	N	
Positive	25%	(34/134)	11%	(83/780)	13% (117/914)
Negative	21%	(9/43)	21%	(36/172)	21% (45/215)
TOTAL	24%	(43/177)	12%	(119/952)	14% (162/1,129)

N = 427) negative clauses to be denials of "something which has not been explicitly asserted"(1991: 21, 35) and Thompson observes that, unlike interrogative clauses, negative clauses do not participate in adjacency pairs, concluding that, though negatives are mostly denials, "what they deny is typically not explicitly present in the conversation" (1998: 325). In sum, while we can say that, in general, speakers are more likely to stress *I* in denials than in affirmations, the effect of negation cannot be interpreted as supporting a contrastive role for stressed *I* in an interactional sense.

We propose instead that the negative polarity effect is one of contrast in a semantic sense. This is in accordance with Sun & Givón (1985: 346) and Myhill & Xing (1996: 342), both of whom consider negation to be a contrastive element. Some support for this interpretation is found in the patterning of stress on *I* with other, related, morphological expressions including the adversative conjunction *but* and focus particles *also, even, just, only,* and *still* (cf. Myhill & Xing 1996: 342; Sun & Givón 1985: 346, who consider these to be markers of "contrast" or "emphasis"). Out of a total of 27 such tokens in the sub-corpus described in Section 4.1 (N = 320), the rate of stress on *I* was high, at 30% (8/27) (compare Table 3). As with converse predicates, the token numbers are exceedingly low, but we nevertheless interpret the patterning of such items alongside that of negation as indicative of contrast in a semantic sense.

A further point to make in regards to negation is that the most frequent verb to be negated in the data is *know*, comprising approximately one quarter of the negated tokens (52/216, excluding 19 discourse marker tokens). Of these, over three quarters (40/52) represent the expression *I don't know* (that is, tokens of *know* in the present tense, negated with *no*, and without a co-occurring modal), which we now look at in detail.

5.5 Particular constructions: From fixed to schematic

Not significant in the multivariate analysis was semantic class. When we take into account the effect of other co-occurring factors, verbs of cognition do not favor stressed *I* in the aggregate, as we might expect if stress on *I* is a marker of speaker attitude or opinion. In fact, as can be seen in the left half of Table 9, cognition verbs do not behave uniformly in terms of rate of stressed *I*: *know* has by far the highest rate of stress at 27%, *think* is close to the average at 16%, and *remember, be sure, mean,* and *guess* have the lowest rates (these six verbs constitute more than three fourths (348/427) of all tokens of cognition verbs, as indicated in Table 9). There is thus no evidence from patterns of stressed *I* for a distinctive first person singular construction based on a class of cognition verbs, as is the case for Spanish 1sg subject pronoun expression (Travis & Torres Cacoullos 2012: 734–742).

Table 9. Rate of stress on *I* with cognition verbs (N = 427)

Verb	% stressed *I* (non-formulaic only)*	N Verb + 1sg subject (non-formulaic only)*	Collocation	% stressed *I*	N
know	27%	19/71	*I don't know*	20%	12/59
			I know	14%	5/35
think	16%	21/128	*I think*	8%	7/87
remember	6%	1/16	*I remember*	8%	1/12
mean	0	0/97	*I mean*	0%	1/188
guess	0	0/23	*I guess*	0%	0/36
be sure	0	0/13	*I'm sure*	0%	0/16
Other (18 types)	18%	14/79			

* Excluding discourse marker (parenthetical and prosodically independent) uses of *I don't know*, *I know, I think*, *I remember*, *I mean, I guess, I'm sure* depicted in Table 1 and quotative uses of *I* + *think*.

There is, however, evidence for lexically-particular configurations, or constructions, with tendencies for higher or lower rates of stressed *I*. Constructions are form-function pairings, where function includes stored information about linguistic and extralinguistic contexts of use (for an overview, see Croft & Cruse 2004: 225–290; Goldberg 2013). The right half of Table 9 shows rates of stress on *I* in fixed collocations, which, as shown earlier in Table 1, represent from one third to 100% of the corresponding verb with a 1sg subject in the present data. Of these, *I mean* and *I guess* stand out with respect to stress: as seen on the right half of Table 9, these two fixed collocations (nearly) categorically appear with unstressed *I*. *I mean* and *I guess* are thus candidates for particular constructions, and may be treated as prefabs, or prefabricated units (Bolinger 1976: 1), rather than as combinations of subject and verb.

Also showing notably low rates of stress, as we noted earlier, are what we identified as formulaic expressions, that is, discourse marker (prosodically independent or clausally parenthetical) and quotative uses of the collocations listed in Table 9. Formulaic uses have an aggregate stressed *I* rate of 4% (compared with 13% in clausal instances of *I* + verb) (see Table 2 above, Section 3). These too may be considered prefabs.

In usage-based theory, Bybee (2010: 34) proposes that the cognitive basis for linguistic units is "the chunking of sequential experiences that occurs with repetition", and such "chunking" of frequent sequences of words can result in a loss of compositionality and analyzability (Bybee 2010: Ch. 3). This has been noted by Traugott (1995: 39) who proposes that, in *I think*, "the subject is losing referential (objective) properties, and becoming simply the starting-point of a perspective"

(cf. also Dehé & Wichmann 2010a: 5; Thompson 2002: 146). That is, in these formulas, the *I* is not referential but has been absorbed into a chunk; in such a chunk, it is less likely to be stressed.

In this, *I don't know* presents idiosyncratic behavior. While formulaic chunks with a non-referential *I* generally disfavor stress, this one particular construction favors stress. Furthermore, *I don't know* favors stress regardless of its formulaic status (with a rate of 21%, 4/19 as a discourse formula, i.e., in prosodically independent or clausally parenthetical uses, and 20%, 8/40 in the more clausal use), in contrast to all other discourse markers identified in Table 2 above (including *I know*, which follows the expected pattern (with a rate of 8% (2/24) when used as a formula, but 27% (3/11) in its non-formulaic use)). Indeed, the already low aggregate rate of stressed *I* in the discourse-marker and quotative formulas drops even further when we set aside *I don't know*, from 4% (10/251) to just 2.5% (6/232).

It has been argued that a low rate of stress on discourse markers is because the accent on the pronoun is an "expression of true speaker attitude, opinion or belief", while the lack of the accent on the pronoun is associated with a loss of "semantic content" (Dehé & Wichmann 2010b: 63). We emphasize that a disfavoring of stress cannot be attributed to a lack of expression of speaker opinion or speaker commitment, or some other motivation an analyst might attribute to a speaker, but is an element of the formulaicity itself. Evidence is precisely the behavior of *I don't know*: in favoring stress on *I* it is unlike *I think* (which presumably is semantically similar) and is more like non-formulaic *I know* (which presumably is semantically different).

We can think of *I*$^{(*)}$ *don't know* (CLAUSE) as a lexically particular construction with some schematicity, where *I*$^{(*)}$ indicates that there is a favoring of stressed *I,* and the (CLAUSE) in parentheses indicates that the construction has an open slot that may be filled by a clause, as in (15), or left unfilled, as in example (3) above. Of the 59 tokens of *I don't know*, 85% (all but six) occur in this construction, namely either on their own in an Intonation Unit (N = 28) or with a clause initiated in the same Intonation Unit (N = 25).[19]

(15) Miles: I* don't know why they did it that way. (02 Lambada: 1036)

The status of *I don't know* as a particular construction is appreciable when we compare *I don't think*. While *I don't know* has a higher rate of stressed *I* than affirmative *I know* (20%, 12/59 vs. 14%, 5/35) and has a higher token frequency (almost double, at 59 to 35), *I don't think* is far less frequent than *I think* (at a ratio of 14 to 87). Given its high frequency relative to other cases of negation and relative to its affirmative counterpart, the *I*$^{(*)}$ *don't know* (CLAUSE) construction cannot be dismissed as simply a manifestation of the general favoring effect of negation. This is consistent with observations in the literature that lexically particular

constructions are not impervious to more general patterns evident in the language (Torres Cacoullos & Walker 2009: 29–33; Travis & Torres Cacoullos 2012: 741; Van Bogaert 2010: 417).

Another schematic construction with idiosyncratic behavior with respect to stressed *I* is conditional (non-habitual) *I* would* (NEG) VERB. Conditional *would*, as in example (16) below, shows a high rate of stressed *I* (43%, 17/40, including negated forms), in distinction from habitual *would*, which is categorically unstressed in these data (N = 24). This is undoubtedly related to the relatively low rate of contraction of conditional (non-habitual) *would*: not surprisingly stressed *I* is less likely in contracted than non-contracted forms of *be, have, will, would* (10%, 22/214 vs. 41%, 17/41), and conditional *would* is less likely than other auxiliaries-modals to contract (less than one half (15/34) of the tokens where contraction was possible were contracted). However, rather than viewing the higher rate of stressed *I* in conditional *I would* as being directly attributable to less frequent contraction, in a construction-based view we can interpret these co-occurring patterns — disfavoring contraction and favoring stress on *I* — as part of a single conditional (non-habitual) *would* construction, namely *I*$^{(*)}$ *would* (NEG) VERB. Like *I*$^{(*)}$ *don't know* (CLAUSE), this particular construction displays a higher than average rate of stressed *I* beyond the predictions of the general constraints identified in the multivariate analysis.

(16) Alina: .. I* would let him run editorial.
I* would let him do anything he wanted to do in editorial.
(H) They clipped his wings.
They wouldn't let him do anything. (06 Cuz: 1445–1448)

6. Summary and conclusion

In this paper, we have examined stress on *I* by identifying a number of contextual features associated with speakers' choice of stressed *I*, based on co-occurrence patterns in conversational data. These contextual features serve as a means to operationalize and test hypotheses about the function of stress (for example, polarity as a measure of a semantic notion of contrast, distance as a measure of accessibility). The patterns observed demonstrate support for some conjectures on the discourse functions of stress, but not for others.

We fail to find evidence of an overarching contrastive role for stress on *I*. We studiously applied two independently motivated operationalizations of contrast but neither offers a meaningful account of stress on *I*. On the one hand, converse predicates are highly infrequent for 1sg subjects. On the other, when we consider initial position in relation to cognition verbs and non-coreferential 1sg preceding

subjects — the contexts most compatible with the speaker contrasting his/her opinion with that of the interlocutor — we do not find higher rates of stressed *I* than other turn-initial contexts. In sum, coding tokens one by one directly for contrast as a unitary global concept, even when done in a replicable manner through the operationalization of converse predicates (cf. Myhill & Xing 1996), proved a fruitless exercise and the results of the indirect tests for "contrasting opinions" (Dehé & Wichmann 2010a, 2010b) did not go in the predicted direction.

Abandoning the quest for a unitary contrast account and looking beyond univariate tests, we utilize multivariate analysis to identify the joint effect of a set of co-occurring features, that is, the linguistic conditioning of variably stressed *I* in the spontaneous conversational data studied here. Our analysis demonstrates that, while *contrast* as a general abstract notion is unverifiable, the favoring of stressed *I* at greater distances, in turn-initial position and under negative polarity, indicate specifiable functions of stress on *I*.

Accessibility, an information flow property, is relevant to stress on *I*, based on distance between mentions. What matters is greater (two or more clauses) versus lesser (zero or one intervening clause) distance. It is notable that this distance effect for *stress* found here differs from the switch reference (between adjacent clauses) effect found for pronoun *expression* in languages with variable subject expression. Thus, the assumed cross-linguistic equivalence between expression and stress does not quite hold. Accessibility apparently operates differently in the realization of English *I* as stressed or unstressed from how it operates in the realization of subject pronouns as expressed or unexpressed cross-linguistically.

Confounding the accessibility effect is the fact that distance interacts with coreferential priming: while, following the received generalization, the reduced form (unstressed *I*) is preferred in coreferential contexts when the preceding subject was unstressed, the non-reduced form (stressed *I*) is favored in coreferential contexts when the preceding subject was stressed. Thus a local pattern due to priming is clusters of "coreferential stressed *I*+stressed *I*" across consecutive clauses. This result corresponds with a wide body of work that has found priming to play a role not just in subject expression, but in grammatical variation in general, though it had not previously been identified for stress.

Also contributing to speakers' choice to stress *I* is a position effect, whereby stressed *I* is favored in absolute initial position — of the speaker turn and of the prosodic unit. This effect is not subsumable under that of accessibility, as it holds at lesser distances between mentions of coreferential *Is*, and thus it can be interpreted as a turn-taking function. Finally, a negative polarity effect, whereby we see a higher rate of stress on *I* in denials, provides a measure of a contrastive function in a semantic sense.

In accordance with a construction-based view of grammar, lexically particular constructions are discernable, which both contribute to and deviate from more general patterns. Two collocations, *I mean* and *I guess*, occur nearly categorically with unstressed *I*. In addition, we find that stressed *I* is overwhelmingly absent in discourse formulae — discourse markers as delimited in terms of frequency and positioning *(I think, I don't know* etc. appearing on their own in the Intonation Unit or in parenthetical position in the clause) together with quotatives (*be like, say*). For example, in *I think*, *I* is stressed one half as often in its formulaic use as opposed to its more clausal use. One exception to this, however, is *I*$^{(*)}$ *don't know* (CLAUSE), which shows a higher than average rate of stress, and this regardless of its formulaic status. Another particular construction which also favors stress is conditional (non-habitual) *I* would* (NEG) VERB.

We conclude that questions about the functions of linguistic forms can be empirically addressed by accounting for the variation in the data of spontaneous speech. Patterns of stress can be seen here to be subject to both context-dependent discourse factors on the one hand and chunked units on the other. Rather than ascribing speaker motivations on an example-by-example basis, we hope to have illustrated that *qualitative* interpretation is inseparable from *quantitative* linguistic analysis, through the operationalization and testing of hypotheses.

Notes

* Though the order of authors is not alphabetical, both contributed equally to this work, which was which was made possible by funding from the National Science Foundation (1019112/1019122, http://nmcode-switching.la.psu.edu/). We are grateful to three *SL* reviewers, and to the Language and Culture Research Centre at James Cook University (in particular, Sasha Aikhenvald) for support to Catherine Travis.

1. All examples are from the Santa Barbara Corpus of Spoken American English (SBCSAE) and are reproduced verbatim from the transcripts, with the exception of the asterisk to mark stressed *I* and underlining to highlight relevant segments of examples given. See the Appendix for a list of the transcription conventions.

2. The conversations from which tokens were extracted are SBCSAE 01, 02, 03, 04, 05, 06, 07, 15, 16, 17, 18, 19, 45, 46, 47, 48, 49, 50, 51, 52. All are face-to-face conversations, except for transcript 52, which is a phone conversation (N = 64; 1.5% of data).

3. We thank Kellie Baker, Susan Brumbaugh, Janalyn Byrne, and Shelece Easterday for this tagging.

4. The corresponding sample was made up of tokens that were matched by speaker in all cases; by contraction type (contracted vs. uncontracted) in 88% of the cases; by voice quality of the following phonological segment (voiced vs. unvoiced) in 93% of cases; and by following phone

in 62% of cases. Thanks to Colleen Balukas, Nicole Benevento, Tim Poepsel, and Chip Gerfen for help with the acoustic analyses.

5. Other collocations that have been classified as formulaic based on complementizer *that* absence (Torres Cacoullos & Walker 2009: 21) occur infrequently in these data, namely *I + find* (N = 2), *I + hope* (N = 7), *I + wish* (N = 4); no tokens of these collocations occur in what we have defined as discourse-marker contexts (that is, either parenthetically or on their own in an Intonation Unit). There are ten tokens of *I + believe*, two of which meet this categorization of discourse markers.

6. *I don't think* is not included; it accounts for just 9% (14/157) of all tokens of *think*.

7. But on the use of *I think* as a discourse marker occurring in the same IU as the material which it can be considered to mark, see Kärkkäinen (2007: Ch. 5) and Ono & Thompson (1995: 240–242).

8. We do not consider to be discourse markers those tokens of *I don't know* that are prosodically independent but which answer a question (N = 10), as in the following example. We note that none of these ten tokens occur with stress on *I*.

Dave: … Will you play later?
Dan: … I don't know. (49 Noise Pollution: 1280–1281)

9. Though agreeing that discourse markers are correlated with a low rate of stress (see Table 2 below and Dehé & Wichmann (2010b: 61)), our proposal is contrary to Dehé & Wichmann's (2010b: 61) claim that there is an association between prosodic independence and full clausal status, and prosodic integration and discourse marker status. This difference may be due to the nature of the prosodic units on which our respective analyses are based, however: the Intonational Phrase, a unit which is defined in terms of having at least one pitch peak (Dehé & Wichmann 2010b: 40; Pierrehumbert & Hirschberg 1990: 220), versus the Intonation Unit, which has no such restriction (but is, instead, defined on the basis of the boundaries of the contour (Chafe 1994: 58–59)). Thus, it is not possible for an entirely unstressed *I think*, for example, to occur on its own in an Intonational Phrase, but it can occur on its own in an Intonation Unit (cf. Du Bois et al. 1993: 57). We also note that Dehé & Wichmann's (2010a, 2010b) lack of an operational definition of discourse markers renders their claim ultimately untestable.

10. In examples (6) and (7), the target clause is the second clause in the "contrastive" pair, as was the case in 8 of our 10 contrastive tokens, which accords with prior reports (cf. Myhill & Xing 1996: 325; Silva-Corvalán 2003: 853).

11. Extraction was done beginning with IU #10 of the transcript; we excluded 14 interrogatives, 13 tokens of *I guess,* 67 tokens of *I mean*, and 59 formulaic tokens of *I remember, I'm sure, I think, I know,* and *I don't know* (see Section 3).

12. It is notably lower than Myhill & Xing found for Hebrew and Mandarin preverbal objects in narrative data, approximately one quarter of which met the operationalization (our recalculations: 31/116 for Hebrew and 12/50 for Mandarin, Myhill & Xing 1996: 324–325, Tables 321 and 322 and 329, Table 327).

13. Goldvarb Lion (Sankoff et al. 2012) was used for a logistic regression to perform binomial multivariate analysis (here, choice of stressed *I* vs. unstressed *I*). The procedure determines the factor groups that together account for the largest amount of variation, in terms of stepwise

increase of log likelihood, discarding the remaining factor groups whose addition does not significantly increase the fit to the model (Sankoff 1988b).

14. Probabilities for the non-significant Semantic class factor group are from Goldvarb's first "step down" run, in which all groups are included in the regression.

15. Counted as clauses were all finite verbs, regardless of formulaic status (Section 3).

16. Based on counts in the SBCSAE using the concordance program *Monoconc* (Barlow 2004), 82% (9,036/11,042) of all first person singular mentions are the subject pronoun *I* (counts for other forms are *me* = 1,021, *my* = 954, and *mine* = 31).

17. We note that this is a genuine distance effect, and not due to the presence of intervening human subjects, as might be expected from Travis & Torres Cacoullos' finding that for Spanish subject expression, Human Switched Reference is the pertinent measure (2012: 728). In the data studied here, 62% (85/138) of the tokens with one intervening clause had no intervening human subject, but the rate of stress at this distance remains nearly identical whether there are none (9%, 8/85) or there is one (8%, 4/53) intervening human subject.

18. Dissipation with increasing distance from the previous coreferential *I* confirms that the priming effect here is in speakers' production of stressed *I*, and not in the raters' perception of stress, since in coding for the perception of stress the raters listened to the IUs in which the tokens occurred consecutively, regardless of the distance between them (see Section 2 on the coding procedure).

19. *I don't know* occurrences in their own IU include 18 formulaic tokens and ten answers to questions. Of the remaining six not appearing in $I^{(*)}$ *don't know* (CLAUSE), one is a parenthetical, two occur with a nominal complement, and three with a wh-complement (e.g. *I don't know why*).

References

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24(1). 65–87. DOI: 10.1017/S0022226700011567

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. New York: Routledge, Chapman & Hall.

Barlow, Michael. 2004. MonoConc Pro.

Beckman, Mary E. 1986. *Stress and non-stress accent* (Netherlands Phonetic Archives No. 7). Dordrecht: Foris Publications. DOI: 10.1515/9783110874020

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–387. DOI: 10.1016/0010-0285(86)90004-6

Bock, J. Kathryn & Zenzi M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General* 129(2). 177–192. DOI: 10.1037/0096-3445.129.2.177

Boersma, Frederic, J. & D. Weenink. 2011. *Praat: Doing phonetics by computer*[Computer Software]. Amsterdam: Department of Language and Literature, University of Amsterdam. Retrieved from http://www.praat.org/.

Bolinger, Dwight. 1961. Contrastive accent and contrastive stress. *Language* 37(1). 83–96. DOI: 10.2307/411252

Bolinger, Dwight. 1976. Meaning and memory. *Forum Linguisticum* 1(1). 1–14.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511750526

Cameron, Richard & Nydia Flores-Ferrán. 2003. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1). 41–65. DOI: 10.1075/sic.1.1.05cam

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.

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

Croft, William. 1995. Intonation units and grammatical structure. *Linguistics* 33: 839–882. DOI: 10.1515/ling.1995.33.5.839

Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511803864

Dehé, Nicole. 2009. Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics* 45. 569–615. DOI: 10.1017/S002222670999003X

Dehé, Nicole & Anne Wichmann. 2010a. The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary. *Functions of Language* 17(1). 1–28. DOI: 10.1075/fol.17.1.01deh

Dehé, Nicole & Anne Wichmann. 2010b. Sentence-initial *I think (that)* and *I believe (that)*: Prosodic evidence for uses as main clause, comment clause and discourse marker. *Studies in Language* 34(1). 39–74. DOI: 10.1075/sl.34.1.02deh

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855. DOI: 10.2307/415719

Du Bois, John W., Wallace L. Chafe, Charles Myer, Sandra A. Thompson, Robert Englebretson & Nii Martey. *2000-2005. Santa Barbara corpus of spoken American*, Parts 1-4. Philadelphia: Linguistic Data Consortium.

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

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1(2). 126–152.

Givón, T. 1979. *On understanding grammar*. New York: Academic Press.

Givón, T. (ed.). 1983a. *Topic continuity in discourse: A quantitative cross-linguistic study*. Amsterdam: John Benjamins. DOI: 10.1075/tsl.3

Givón, T. 1983b. Topic continuity in discourse: An introduction. In T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-linguistic study*, 1–41. Amsterdam: John Benjamins. DOI: 10.1075/tsl.3.01giv

Givón, T. 2001. *Syntax: An introduction*, vol. 1, 2 vols., 2 edn. Amsterdam: John Benjamins.

Goldberg, Adele E. 2013. Constructionist approaches. In T. Hoffman & G. Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 15–31. Oxford: Oxford University Press.

Halliday, M. A. K. 1994. *An introduction to functional grammar*. London: Edward Arnold.

Hirschberg, Julia & Janet Pierrehumbert. 1986. The intonational structure of discourse, *Proceedings of the Twenty-fourth Annual Meeting of the Association for Computational Discourse*, 136–144. New York, NY.

Horn, Laurence R. 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61(1). 121–174. DOI: 10.2307/413423

Kärkkäinen, Elise. 2007. The role of *I guess* in conversational stancetaking. In Robert Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 183–220. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/pbns.164.08kar

Krahmer, Emiel & Marc Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* 34. 391–405. DOI: 10.1016/S0167-6393(00)00058-3

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4). 715–762. DOI: 10.2307/412333

Labov, William. 1994. *Principles of linguistic change: Internal factors*, vol. 1, 3 vols. Oxford: Basil Blackwell.

Labov, William. 2005. Quantitative reasoning in linguistics. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society*, vol. 1, 2nd edn, 6–22. Berlin: Mouton de Gruyter.

Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics* 23(2). 379–434. DOI: 10.1017/S0022226700011324

Mayol, Laia. 2010. Contrastive pronouns in null subject Romance languages. *Lingua* 120. 2497–2514. DOI: 10.1016/j.lingua.2010.04.009

Myhill, John & Zhiqun Xing. 1996. Towards an operational definition of contrast. *Studies in Language* 20(2). 303–360. DOI: 10.1075/sl.20.2.04myh

Ono, Tsuyoshi & Sandra A. Thompson. 1995. What can conversation tell us about syntax? In Philip W. Davis (ed.), *Alternative linguistics: Descriptive and theoretical modes*, 213–271. Amsterdam: John Benjamins.

Paredes Silva, Vera Lucia. 1993. Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5(1). 35–49. DOI: 10.1017/S0954394500001381

Payne, Thomas E. 1997. *Describing morphosyntax: A guide to field linguists*. Cambridge: Cambridge University Press.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press.

Poplack, Shana. 1980. The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In William Labov (ed.), *Locating language in time and space*, 55–67. New York: Academic Press.

Poplack, Shana & Elisabete Malvar. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19. 121–169. DOI: 10.1515/PROBUS.2007.005

Poplack, Shana & Sali Tagliamonte. 2001. *African American English in the diaspora*. Malden, MA: Blackwell Publishers.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic press.

Redeker, Gisela. 1991. Linguistic markers of discourse structure [review of *Discourse Markers* by Deborah Schiffrin]. *Linguistics* 29. 1139–1172.

Sankoff, David. 1988a. Sociolinguistics and syntactic variation. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge survey* (Vol. 4, Language: The socio-cultural context), 140–161. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620577.009

Sankoff, David. 1988b. Variable rules. In Ulrich Ammon, Norbert Dittmar & Klaus J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society*, vol. 2, 984–997. Berlin & New York: Walter de Gruyter.

Sankoff, David, Sali Tagliamonte & Eric Smith. (2012). *Goldvarb LION: A variable rule application for Macintosh*. University of Toronto. URL http://individual.utoronto.ca/tagliamonte/goldvarb.htm.

Scheibman, Joanne. 2000. *I dunno*: A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32(1). 105–124. DOI: 10.1016/S0378-2166(99)00032-6

Scherre, Maria Marta Pereira & Anthony J. Naro. 1991. Marking in discourse: 'Birds of a feather'. *Language Variation and Change* 3(1). 23–32. DOI: 10.1017/S0954394500000430

Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español* (Georgetown Studies in Spanish Linguistics). Washington, DC: Georgetown University Press.

Silva-Corvalán, Carmen. 2003. Otra mirada a la expresión del sujeto como variable sintáctica. In Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, María Luz Gutiérrez Araua, María Vaquero & César Hernández (eds.), *Lengua, Variación y contexto: Estudios dedicados a Humberto López Morales*, vol. 2, 849–860. Madrid: Arco Libros.

Sun, Chao Fen & T. Givón. 1985. On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language* 61(2). 329–351. DOI: 10.2307/414148

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654. DOI: 10.1006/jpho.2002.0178

Tagliamonte, Sali & Jennifer Smith. 2005. No momentary fancy! The zero 'complementizer' in English dialects. *English Language and Linguistics* 9(2). 289–309. DOI: 10.1017/S1360674305001644

Tao, Hongyin. 2001. Discovering the usual with corpora: The case of *remember*. In Rita C. Simpson & John M. Swales (eds.), *Corpus linguistics in North America*, 116–144. Ann Arbor: University of Michigan Press.

Thompson, Sandra A. 1998. A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. In Anna Siewierska & Jae Jung Song (eds.), *Case, typology and grammar: In honor of Barry J. Blake*, 309–341. Amsterdam: John Benjamins.

Thompson, Sandra A. 2002. 'Object Complements' and conversation: Towards a realistic account. *Studies in Language* 26(1). 125–163. DOI: 10.1075/sl.26.1.05tho

Thompson, Sandra A. & Anthony Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15. 237–251. DOI: 10.1016/0378-2166(91)90012-M

Torres Cacoullos, Rena & Catherine E. Travis. 2011. Using structural variability to evaluate convergence via code-switching. *International Journal of Bilingualism* 15(3). 241–267. DOI: 10.1177/1367006910371025

Torres Cacoullos, Rena & Catherine E. Travis. 2014. Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63. 19–34. DOI: 10.1016/j.pragma.2013.08.003.

Torres Cacoullos, Rena & Catherine E. Travis. Forthcoming. Assessing inter-linguistic (dis)similarity via intra-linguistic variability for subject expression. In Ana M. Carvalho, Rafael

Orozco & Naomi Lapidus Shin (eds.). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown: Georgetown University Press.
Torres Cacoullos, Rena & James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47(1). 1–43. DOI: 10.1515/LING.2009.001
Tottie, Gunnel. 1991. Lexical diffusion in syntactic change: Frequency as a determinant of linguistic conservatism in the development of negation in English. In Dieter Kastovsky (ed.), *Historical English syntax*, 439–467. Berlin: Mouton de Gruyter.
Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalisation. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511554469.003
Travis, Catherine E. 2005. *Discourse markers in Colombian Spanish: A study in polysemy* (Cognitive Linguistics Research). Berlin & New York: Mouton de Gruyter.
Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. DOI: 10.1017/S0954394507070081
Travis, Catherine E. & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. DOI: 10.1515/cog-2012-0022
Van Bogaert, Julie. 2010. A constructional taxonomy of *I think* and related expressions: Accounting for the variability of complement-taking mental predicates. *English Language and Linguistics* 14(3). 399–437. DOI: 10.1017/S1360674310000134
Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1). 29–58. DOI: 10.1017/S0022226700007441

Appendix

Transcription Conventions (Du Bois et al. 1993)

.	final intonation contour	=	lengthening
,	continuing intonation contour	[ ]	speech overlap
?	appeal intonation contour	[2 2]	used to distinguish adjacent overlaps
…	medium pause (0.5–0.7 secs)	!	booster: emphatic speech
..	short pause (about 0.5 secs)	X	one syllable of unclear speech
--	truncated intonation contour	@	one syllable of laughter
–	truncated word	@word	word uttered while laughing
(H)	in-breath	(( ))	researcher's comment
%	glottal stop (TSK) click	<VOX VOX>	speech produced with marked voice quality

Corresponding author's address

Catherine E. Travis
School of Literature, Languages and Linguistics
Australian National University
Baldessin Precinct Building (110)
Room W3.17
Canberra, ACT
0200 Australia

Catherine.Travis@anu.edu.au