Integrating qualitative and quantitative analyses of stance: A case study of English *that*/zero variation

TIMOTHY GADANIDIS ⓘ, ANGELIKA KISS ⓘ,
LEX KONNELLY ⓘ, KATHARINA PABST ⓘ,
LISA SCHLEGL ⓘ, POCHOLO UMBAL ⓘ AND
SALI A. TAGLIAMONTE ⓘ

University of Toronto, Canada

ABSTRACT

Previous work has shown that stance—the way speakers position themselves with respect to what they are talking about and who they are talking to—provides powerful insights into why speakers choose certain linguistic variants, beyond correlations with macro-social categories such as gender, ethnicity, and social class. However, as stancetaking moves are highly context-dependent, they have rarely been explored quantitatively, making the observed variable patterns difficult to generalize. This article seeks to contribute to this methodological gap by proposing a formal guide to coding stance and demonstrating how it can be operationalized quantitatively. Drawing on a corpus of eight individuals, self-recorded in three situations with varying levels of social distance, we apply this method to variation between English complementizers *that* and zero (i.e. no overt complementizer), providing a replicable and theoretically grounded protocol that incorporates both quantitative and qualitative analyses in a variationist sociolinguistic study. (Stance, complementizers, *that*, English)*

INTRODUCTION

Stance has been shown to be crucial to the study of social meaning in linguistic variation. Various frameworks for analyzing stance (e.g. Du Bois 2007; Kiesling 2009, 2011, 2016) have gained theoretical purchase in contemporary variationist research, accounting for variation where traditional macro-social categories alone cannot (e.g. Kiesling, Onuffer, & Hardware 2012; Podesva 2016; Levon 2016; Nycz 2018). Though stance has been operationalized in many ways, a common theme across studies is how stance can be informative in explaining the idiosyncratic patterns of individual speakers. However, our view of how stance influences variation between speakers across different local interactions on a larger scale remains limited, partly because stance is fundamentally dialogic and qualitative.

Our study engages with this growing body of work by developing a coding protocol that allows for a nuanced picture of both inter- and intra-speaker variation.

27

Specifically, this protocol provides a formal and replicable method to code for stance in quantitative variationist approaches. To illustrate this, we apply the protocol to the variation between English complementizer *that* and zero (i.e. no overt complementizer) using a corpus of eight individuals who recorded themselves in three interactional contexts, each varying in level of social distance between participant and interlocutor(s). Our main goal is to demonstrate the capabilities of this framework in accounting for intraspeaker variation. Our results also provide additional insight into how stance, in addition to grammatical factors, is correlated with the variable realization of *that* or zero, while sociodemographic factors are not. Consequently, we argue that complementizer use may not simply be grammatically constrained, but may also be a resource for performing social work.

STANCE FROM A QUALITATIVE AND QUANTITATIVE PERSPECTIVE

Qualitative perspectives on stance

Stance has been variously defined in the sociolinguistic and linguistic-anthropological literature, and also has a rich tradition in other linguistic subdisciplines (Biber & Finegan 1988, 1989; Ochs & Schieffelin 1989; Ochs 1992, 1996; Conrad & Biber 2000; Du Bois 2007; Jaffe 2009; Kiesling 2009; inter alia; for more detailed discussion see Jaffe 2009:3). Despite their differences, all of these definitions share some crucial core features: stance can be broadly conceptualized as a form of positionality, or attitudinal expression, that a speaker takes towards the content of their talk and their audience. It is this core meaning that characterizes the approach that we outline in the following sections.

Our study is not the first to propose a protocol for analyzing stance, and we are indebted to researchers who laid crucial groundwork for the methodology we propose. For instance, Ochs (1992:341) describes how one or more linguistic features may come to index particular stances, social acts, or social activities, which in turn helps constitute social meaning through the (real or imagined) association of these activities with certain social groups. That is, much of the social meaning behind variation is indirect, with the association between the two mediated by stance and ideology. Ochs (1996:410) further characterizes stance as being 'affective' or 'epistemic': affective stance refers to 'a mood, attitude, feeling, and disposition… vis-à-vis some focus of concern', while epistemic stance refers to 'knowledge or belief vis-à-vis some focus of concern'. Where Ochs was primarily concerned with the precise nature of stance itself, Du Bois's (2007) influential 'stance triangle' shown in Figure 1 provides a template for what happens when stancetaking occurs: according to Du Bois, we position ourselves as subjects with respect to the topic of our talk (the 'stance object'), and second, as co-stancetakers with respect to other subjects.

Kiesling (2011) identifies three primary axes of stance along which a speaker's position can vary: Affect, Alignment, and Investment. *Affect* (which is referred to as
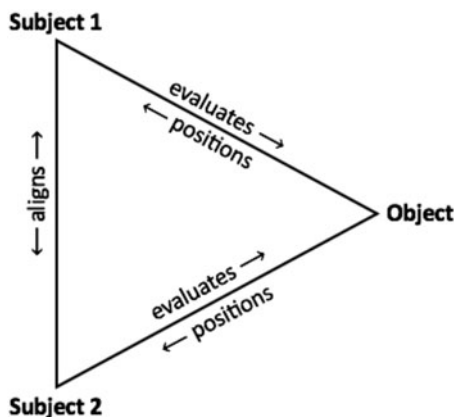
FIGURE 1. The stance triangle (reproduced with permission from Du Bois 2007:163).

Evaluation in Kiesling 2020), refers to the polarity or quality of the stance (positive or negative), and encompasses the speaker's assessment of the stance object. *Alignment* refers to whether a speaker aligns with or diverges from their interlocutor, both epistemically and interactionally. Finally, *Investment* refers to epistemic modality, or how strongly invested in their claims and opinions the speaker is (Kiesling 2011:5). These axes are not necessarily independent: depending on the discourse context, they can be interrelated or semiotically aligned with one another. While the stances in Kiesling's framework broadly fall under these Affect, Alignment, and Investment headings, they can also index more specific characteristics. For example, in Kiesling's (2004:282) analysis of the use of vocative *dude* in an American fraternity, he argues that the intersections of each axis convey the overall meaning of 'cool solidarity'. The frat members' use of *dude* aligned the speakers with each other and signaled some positive affect (in that it served to construct and maintain solidarity between interlocutors), but it also allowed the speakers to indicate the characteristically low investment leading to the 'cool' interpretation. *Dude* therefore does not necessarily add a distancing effect, but instead lowers the perceived enthusiasm or effort of the speaker. Kiesling (2016) expands the three initial axes to four by splitting Alignment into Solidarity and Hierarchy, in order to more comprehensively capture the relationships between interlocutors. *Solidarity* refers to whether the speaker is agreeing (converging) with the interlocutor or disagreeing with (diverging from) them; *Hierarchy* refers not only to existing hierarchical relationships between speakers (e.g. boss and worker, parent and child), but also to whether a speaker presents themself as an expert or a novice in relation to their interlocutor on the subject of talk. Kiesling (2020) highlights the importance of a multifaceted stance framework by examining how discourse markers *I mean* and *just sayin'* are used as stancetaking tools in signaling more

than just increased or decreased investment in one's utterance. While they both alter the strength of investment generally, these markers can also be used to indicate shifted evaluation of a stance object or to decrease the degree of disalignment with interlocutors. The importance of these and other specific and conventionalized grammatical structures in stancetaking has been recognized in pragmatic research as well. For example, Kärkkäinen (2012) reveals how *I thought* is systematically combined with different formulaic sequences to denote not only epistemic and evaluative stances about the content of one's talk, but affective stances as well.

Stance's flexibility as a methodological tool has led to its increased usage in research on interactional meaning, and this research has been carried out on language data across multiple modalities. Acts of stancetaking have been studied not only in spoken language, but also in online discourse: in a study of online comments on news articles by Langlotz & Locher (2012), the authors found that even without explicit emotional evaluation, various display strategies were used by commenters to take emotional stances about the subjects of news stories themselves. In addition, studies that use stance as an interpretive tool in understanding social and conversational meaning may utilize different subsets of the axes of stance mentioned above in their analyses. For example, Park (2013) used stance to interpret practices of vocal mimicry in Korean pop culture and explained that in reproducing a stylized version of the speech of another, the act of mimicry is inherently an act of stancetaking—regardless of the evaluative stance taken, the act emphasizes the positionality of the mimic relative to their subject.

Stance analyses across sociolinguistics and pragmatics have thus focused on different facets of stance itself, reflective of the fact that different elements of stancetaking may be more or less salient in any given interactional context.

Quantitative perspectives on stance

Several variationist analyses demonstrate how powerful stance can be as an analytic tool and how helpful it can be in explaining speakers' choice of one linguistic variant over another. Kiesling and colleagues (2012) analyzed the use of variable (-ing) and coronal stop deletion in a women's friendship group, and found that stance accounted for patterns of style-shifting while identity categories such as social class could not. Similarly, Podesva's (2016) study analyzing the use of (t,d) deletion and falsetto amongst Black and white speakers in Washington, DC found that the greatest use of these features among Black speakers occurred when they took evaluative stances about race and gentrification. Levon's (2016) case study of creaky voice by a forty-year-old married Orthodox Jewish man who engages in sexual and romantic relationships with other men showed that the speaker was creakiest in moments when he took deontic stances in relation to presupposed conflict between his religious affiliation and his intimate relationships. Holmes-Elliot & Levon (2017), proposing their own analysis of stance within the variationist framework, showed that an interactional stance approach was more

explanatory in accounting for /s/ variation in British television speech than the broad social categories of speaker class or speaker gender. Kiesling, Pavalanathan, Fitzpatrick, Han, & Eisenstein (2018) demonstrated how the dimensions of Affect, Alignment, and Investment could be used to evaluate stancetaking in Reddit posts, exposing the degree to which the strength of one of these dimensions in a given post was determined by the strength of that same dimension in a preceding post. Moreover, Nycz (2018) found that stance was a useful heuristic in accounting for speakers' use of second-dialect features: analysis of Canadian and US pronunciations of the (awT) vowel (e.g. *house)* showed that Canadians living in the United States produced the most US-like pronunciations when expressing distance from or ambivalence toward Canada, and the most Canadian-like pronunciations when expressing positive affect regarding or solidarity with Canada.

Lastly, Barnes's (2018) analysis of copula variation in a situation of language contact between Asturian and Spanish in urban Asturias is particularly noteworthy, as it outlines an approach to coding stance that can easily be applied to other datasets. Barnes found that the local Asturian variant *ye* 'he/-she/it is' indexes low epistemic commitment to the content of the speaker's utterance or mitigates the face threat of disagreeing with their interlocutor—stance properties that Barnes maps to the locally relevant features of an informal, laid-back, easy-going Asturian persona. Following Kiesling's (2011) framework, Barnes also provides a possible operationalization of stance that takes into account epistemic, affective, and alignment stances, as well as consideration of the particular speech activity. Like Barnes, our goal is to develop a replicable protocol for coding stance. However, our analysis differs in that unlike the variation between the third person copulas *es* and *ye* in Barnes's study, which are means of 'doing' local Asturian identity and its associated indexical features, we focus on applying our coding protocol for analyzing stancetaking to a linguistic feature that is not regarded as doing salient social work: variation between the complementizers *that* or zero.

Consistent coding for stance in a quantitative analysis of intra-speaker variation raises some methodological complications. First, stance is dialogic: stances can be taken up and responded to as stance objects in their own right and they can be constructed collaboratively between speakers. Determining speakers' stances requires contextual, ethnographic knowledge of the speakers and their interactions so that analysts can be reasonably certain of what speakers are responding to and how. To that end, we concur with Hunston's (2007:28) observation that 'interpreting the role of stance in discourse entails a deeper understanding of the discourse as a whole'. The present framework aims to meet this requirement by leveraging the unique knowledge that we have as the participants in the case study. This knowledge provides critical insight into the dynamics (and the interactional goals) of our own conversations, enabling us to identify linguistic and extralinguistic factors that must be accounted for in the development of a generalizable protocol for coding stance.[1]

ENGLISH COMPLEMENTIZERS

Variation between the English complementizer *that*, as in (1), and zero, as in (2), has been previously described as being primarily grammatical rather than socially constrained (Finegan & Biber 2001:258–59). All examples are taken from a corpus of self-recordings and include two supplementary pieces of information: the speaker's pseudonym and the social distance between interlocutors in the recording.

(1) So we know **THAT** there must be some kind of overlap between these two. (Quokka/high)
(2) I think **Ø** it sounds like a dream job. (Unicorn/medium)

The extensive literature on English complementizers (e.g. Warner 1982; Rissanen 1991; Poplack & Walker 2002; Tagliamonte & Smith 2005; Torres Cacoullos & Walker 2009) shows that a variety of factors influence complementizer choice. Linguistic material intervening between the matrix clause and complement clause increases the syntactic complexity of the phrase, thus increasing the likelihood of an overt complementizer in order to separate the two clauses (Rohdenburg 1998; Torres Cacoullos & Walker 2009). However, subject-verb collocations are frequently invariant for complementizer form. For example, Thompson & Mulac (1991) posit that *I think* and *I mean* have grammaticalized into epistemic markers, which frequently act as parentheticals and tend to appear with null complementizers (see also Kärkkäinen 2007 for *I guess*). Thus, factors associated with grammaticalized epistemic parentheticals also predict higher frequencies of null complementizers: present tense, first person subjects, and specific verbs such as *think* (Tagliamonte & Smith 2005; Torres Cacoullos & Walker 2009).

Previous research also suggests that overt complementizers, though grammatically conditioned, vary according to style. Finegan & Biber's (1985) analysis of the ARCHER corpus of Modern English demonstrates that more formal genres such as sermons and medical writings favour the use of *that*, while informal genres like personal letters favour zero. Storms (1966:262) characterizes overt *that* as 'less personal, less familiar, less warm, less friendly, less emotive', an observation that Rissanen (1991:277) echoes in comparing oral versus written registers. Given contemporary convictions that stance is the underlying precursor of stylistic variation (Kiesling 2009), we explore stylistic variation of the English complementizer with respect to stance. Specifically, we probe whether stance explains the patterning of this variable across and within speakers.

However, coding for stance is methodologically complex (see Hunston 2007). This study therefore develops a coding protocol, providing a formal guide to coding stance and demonstrating how it can be operationalized quantitatively. To highlight its utility, we implement a case study using a corpus of self-recordings across eight individuals in three situations with varying levels of social distance. Crucially, these recordings were created before it had been determined which linguistic variable would be analyzed, and before it had been decided that the analysis

would focus on stance—these decisions were made after exploring the collected data. The only knowledge that speakers (the authors) had about the study at the time that the recordings were made was that it would relate to stylistic variation in different contexts. While effects of speakers' expert knowledge as linguists cannot completely be ruled out, choosing the variable only after all data was collected aimed to mitigate such effects: all speakers were as naïve as possible to their use of variable patterns of English complementizer *that*/zero during their interactions. The goal of this case study provides a theoretically grounded and potentially replicable scheme through which to integrate quantitative and qualitative analyses in variationist study.

DATA AND METHOD

The data come from a corpus of self-recorded interactions collected by the seven co-authors of this article and one of their colleagues in a graduate seminar. Each self-recording was between sixty and 120 minutes long. In each case, interlocutors were informed beforehand that the recordings were being made and that they would be used for research purposes, but they were informed that only the speech of the researcher-participant would be analyzed. Some recordings involved only a single interlocutor, while in other recordings the speaker may have had an audience of dozens of people, such as while teaching a class or tutorial. Given that recording contexts varied, recording equipment also varied by speaker and context: some recordings were conducted using Zoom H2n recorders while some were conducted using smartphones. Regardless, the resultant audio quality was in all cases sufficient for analysis of syntactic phenomena like complementizers. Following Torres Cacoullos & Walker (2009:11), we extracted all complements in object position that could take either *that* or zero, including complements of verbs in (3), adjectival predicates in (4), and extrapositions in (5).

(3) I found **THAT** they sell like fake bologna, fake ham, and fake turkey. (Penguin/medium)
(4) I'm sure **Ø** they'll love that, yeah. (Unicorn/medium)
(5) It's good **THAT** you could make it short. (Platypus/low)

Following standard variationist methodology, we restricted the analysis to contexts where both *that* and zero could occur. Therefore, we excluded the following matrix subject-verb collocations: *I mean* in (6), which has fully grammaticalized and no longer accepts *that* (Tagliamonte & Smith 2005); *I guess* in (7), which exclusively occurred with zero in our data; and *I think* in (8), which was also categorically zero for most speakers. Note that only these specific collocations were excluded: uses of *guess* and *think* with non-first-person subjects were not excluded. The collocation *you know* did not occur in contexts where it could be followed by a complementizer, so it did not pose any problems for our analysis. We also excluded

TABLE 1. *Speaker sample.*

PSEUDONYM	GENDER	AGE	ETHNICITY	L1	N	% ZERO
Mole	woman	28	White Canadian	English	67	88
Fox	man	28	Filipino Canadian	Tagalog	87	79
Quokka	woman	24	White Canadian	English	151	74
Finch	woman	57	White Canadian	English	56	68
Unicorn	woman	30	White European	German	155	64
Platypus	woman	34	White European	Hungarian	35	57
Penguin	man	21	White Canadian	English	119	58
Orca	non-binary	27	White Canadian	English	52	50
Total					722	67

instances of emerging complementizer variant *like* (López-Couso & Méndez-Naya 2012:189).

(6) I mean Ø maybe there's a better way to do it. (Finch/high)
(7) I guess Ø it's normal. (Penguin/low)
(8) I think Ø that's a really really good point though. (Orca/low)

This procedure yielded 722 tokens which were coded for social, linguistic, and stance-related factors.

Social factors

In order to capture a wide range of stylistic performances, the researcher-participants recorded themselves in three different situations in which the social distance between interlocutors varied across high, medium, and low. Although there were some general guidelines for how social distance was to be operationalized—for example, low social distance was roughly defined as 'feeling comfortable discussing intimate topics or using profanity', while high social distance was roughly defined as 'closely-monitored communication in professional contexts'—the final decision regarding the categorization of social distance in each interaction was made by the speakers themselves based on their knowledge of their own behaviour and their relationships with interlocutors. For example, some of the speakers chose to record themselves with a colleague for their low social distance recording, while others classified a conversation with a colleague as their medium social distance recording dependent upon their unique relationship and level of social distance with that individual. To protect the anonymity of our interlocutors, each coder was only able to view their own data. All examples were approved by our interlocutors before they were included.

The speakers were between twenty-one and fifty-seven years old at the time of data collection and include two cisgender men, five cisgender women, and one non-binary speaker. Five are native speakers of English, while three learned a language other than English as their first language but consider themselves balanced bilinguals. The speaker sample is presented in Table 1.

Speakers varied along several social dimensions such as gender, age, ethnicity, and language background. Due to the small sample size, we cannot determine conclusively if the sociodemographic factors are influential. However, there is no reason to predict that the patterning of this variable is related to individual identity; indeed, most previous research has found no difference in complementizer use between different social groups (e.g. Finegan & Biber 2001:258–59; but see Staum 2005 for a discussion of potential regional differences). Therefore, there is no reason for us to expect that sociodemographic factors will have an effect on complementizer realization.[2]

Linguistic factors

The linguistic factors included features that have been previously reported to influence complementizer choice (Thompson & Mulac 1991; Tagliamonte & Smith 2005; Torres Cacoullos & Walker 2009) including matrix verb, matrix subject, matrix tense, and the presence of intervening material between verb and complement clause. The following subsections motivate the inclusion of each of these grammatical constraints in greater detail.

Thompson & Mulac (1991) argue that zero is favored by epistemic matrix verbs such as *think* and *bet* (Palmer 1979:21), whereas non-epistemic verbs such as *say* correlate with greater use of the overt complementizer *that* (Torres Cacoullos & Walker 2009). All tokens were therefore coded for matrix verb.

(9) So I think **Ø** we should just send it back just the way it is. (Finch/medium)
(10) He said **THAT** he feared for his life because the guy was reaching for a gun. (Penguin/medium)

Epistemic verb collocations typically occur with first person subjects, as in (9). As noted above, some matrix constructions with a first-person subject and an epistemic verb have grammaticalized to a parenthetical in speech with variable syntactic placement and more adverbial meaning (Thompson & Mulac 1991:238–39; Tagliamonte & Smith 2005; Kärkkäinen 2007). Consequently, matrix clauses with first person subjects as in (9) are more likely to favour zero complementizers (Thompson & Mulac 1991). When the matrix subject is not first person singular, the construction cannot function as a parenthetical, making *that* more likely, as in (10). Accordingly, matrix subject was coded as either 'I' or 'other'.

Based on Thompson & Mulac's (1991) hypotheses, Tagliamonte & Smith (2005) argue that the tense of the matrix verb may also be implicated such that verbs in present tense favour zero, as in (9), and past tense favour *that*, as in (10). We coded for the tense of the matrix verbs in (9) and (10) as either simple present or simple past, respectively. Other tense and aspect combinations were collapsed into a single additional category. Following previous findings (Thompson & Mulac 1991; Tagliamonte & Smith 2005), we hypothesize that the overt complementizer *that* will be disfavored by matrix verbs in the present tense.

Finally, we coded for the presence of intervening material between verb and complement clause. Previous research (Bolinger 1972; Thompson & Mulac 1991; Rohdenburg 1998) has argued that the presence of such material favours the use of *that* in order to explicitly demarcate the two clauses. Following Torres Cacoullos & Walker (2009), we distinguished between two types of intervening material: intervening verbal arguments (11) and other intervening material (12). Consistent with these previous findings, we predict that the overt complementizer *that* will be favoured in the presence of any kind of intervening material.

(11)   I told you **Ø** we should have picked up some chips. (Mole/low)
(12)   So imagine now **THAT** we're looking at just the- just the right side. (Quokka/high)

Coding protocol

Characterizing stance.   We take Kiesling's (2011, 2016, 2020) dimensions of stance as a point of departure because of their robust contribution to the stance literature, having formed the basis of a number of studies taking similar quantificational approaches to stance. At the same time, we seek to develop a replicable coding protocol grounded in pragmatic tests that maximizes consistency and objectivity across multiple coders.

We begin by determining the context of each token, including the nature or genre of the discourse in which it occurs (e.g. university lecture, family gathering, etc.), the interlocutors' relations to each other (colleagues, friends, instructor–student, etc.), and any other salient circumstances. Tokens containing a variable complementizer are small segments taken from this larger discourse, and are segmented according to sub-topics of a reasonable size. We model these subtopics as Questions Under Discussion (QUDs), and we assume that discourse can be modeled as a joint activity where participants raise and resolve QUDs explicitly or implicitly to achieve shared goals (see Roberts 2012). Assuming that the meaning of a question is a set of its answer propositions (Hamblin 1973), questions such as *Who came?* may have subquestions such as *Did Mary come?* and *Did John come?*, and so on—the answers to which provide partial answers to the superquestion (see Büring 2003:516).

Any utterance conveying a proposition implicitly raises a QUD, and when identifying it, multiple QUDs may seem acceptable. In order to understand the interlocutor's stance, coders need to hear everything that the speakers say about the current QUD of the discourse. The discourse segments analyzed by coders must be large enough that the speaker's stance can be determined clearly enough for coders to make a judgment. One such segment can possibly contain more than one instance of a variable. In such cases, we included the same stance token as two separate data points. Thus, the stance object is defined as the narrowest current QUD. We refer to the tokens (discourse segments) that cover the narrowest QUDs revealing the speaker's stance as *stance acts* (a term drawn from Du Bois 2007).

Building on Kiesling (2016, 2020), the four aspects of stance we use in our protocol are Evaluation, Investment, Alignment, and Hierarchy. These aspects of stance can be characterized by being affective (i.e. evaluative), epistemic (i.e. knowledge-related, following Ochs & Schieffelin 1989 and Ochs 1996), or both. They also align with Du Bois's stance triangle, in that Evaluation and Investment describe the stancetaker's relation to the stance object, and Alignment and Hierarchy, their relation to the other interlocutors.

Evaluation (which we take to be the same as Affect, the term used for this dimension by Kiesling 2016) describes how the speaker evaluates the stance object; it is inherently affective and has no epistemic attribute.

Investment comprises both affective and epistemic dimensions. Affective Investment characterizes the intensity of the speaker's emotions; epistemic Investment characterizes their degree of certainty. Together, Evaluation and Investment represent the speaker's relation to the stance object.

Alignment, too, can be characterized in terms of both affective and epistemic stance. The affective side of Alignment refers to whether and how much the speaker cares about the other interlocutors' face, manifesting in politeness strategies (Brown & Levinson 1987). On the epistemic side, Alignment refers to whether the speaker agrees or disagrees with their interlocutor's position.

Finally, Hierarchy expresses the speakers' differential knowledge concerning the stance object. We follow Heritage & Raymond (2005) and Heritage (2009) who claim that speakers distinguish between who has 'primary' or 'secondary' rights to assess a certain issue; they can present themselves as novices, as being at the same level as the addressee, or as experts in the current stance object when compared to the addressee (see also Isaacs & Clark 1987). In our protocol, Hierarchy is understood as having an epistemic dimension but no affective dimension. It is not to be confused with social hierarchy in Brown & Levinson's (1987) terms—this particular notion of hierarchy is encoded in the subjective side of Alignment as part of the speaker's evaluation of their interlocutor.

Table 2 summarizes the stance framework that we adopt in developing the protocol, connecting each aspect to the literature from which it is drawn. The questions provided in each cell guided our thinking in how to develop the tests outlined in the following paragraphs.

While previous studies have taken a more intuitive approach to coding stance based on varying amounts of surrounding utterance context (e.g. Barnes 2018), we aim to determine the stance object and speakers' orientations to it using tests that reveal pragmatic inferences, which we lay out in the following section.

Pragmatic tests.   The test question to uncover the value of Evaluation is whether one of the two utterances in (13) harmonizes with the stancetaker's contribution and can therefore be felicitously added to one of their utterances, at the discretion of the coder. These additional utterances express positive and negative stance explicitly. If

TABLE 2. *Stance framework informing the analysis.*

| | Ochs (1996) | | Du Bois (2007) |
	AFFECTIVE	EPISTEMIC	STANCE TRIANGLE
EVALUATION	How do I evaluate the stance object?	n/a	
			Relating to stance object
INVESTMENT	How much do I like/dislike the stance object?	How sure am I in what I say about the stance object?	
ALIGNMENT	How much do I care about the interlocutors' face?	Do I agree with my interlocutors?	
			Relating to co-stancetaker
HIERARCHY	n/a	Who is more knowledgeable with respect to the stance object?	

a speaker's utterance conveys any value of an aspect of stance, an explicit reformulation of the stance taken must always be acceptable.

(13)   a.   …and this is great / …which is great.
      b.   …and I'm not happy about this. / …which is terrible.

If (13a) is compatible with the speaker's utterance, Evaluation is coded as 'positive'; if (13b) is compatible, then as 'negative'. When there is no particular reason to choose either, it is coded as 'neutral'.[3]

(14)   Stance object: Internet access in Muskoka (Finch/medium)

Finch: Well you don't want to move to Muskoka then.
Interlocutor: (laughter) Yeah.
Finch: Because you know, I don't know how they deal with it. You know, and sometimes the wireless is just not available. Not not not.
Interlocutor: Yeah.
Finch: Cause it's just down. [**And I don't like it.** / **#And I'm happy about it.**]

For Investment, the test question is identical to the fundamental test question: 'Could the speaker have said the utterance in a less marked way/in more neutral terms?'. If the answer is 'yes', the value for Investment is non-neutral (i.e. 'high' or 'low'). In cases of neutrality, Investment is coded as medium. In (14), Finch could have conveyed the same information without emphasizing the fact that the wireless is not available. However, she chose to emphasize it by repeating *not*. As a less marked or more neutral version of the utterance could contain less emphasis, we conclude that the value for investment is 'high'. Conversely, uncertain utterances or ones that show a low degree of certainty, such as those containing *I don't know* or *maybe*, as in (15), are coded as low epistemic investment.

(15) Not the discussions of- **I don't know**, **maybe** we might have a fight with our partner, on the recordings, we- we might have delicate conversations about any number of things. (Finch/high)

For Alignment, different test questions apply to its affective and epistemic aspects. The test questions for affective Alignment, which concerns facework, are: 'Does the speaker care about the addressee's face?' or 'Does the speaker make any effort so that the addressee will not lose face?'. Epistemically, Alignment can also express a match or mismatch in propositional attitudes between speakers: that is, whether speakers agree or disagree on the truth of some proposition. The test question for epistemic Alignment is therefore 'Does the stancetaker agree with their interlocutor?'. The stance act in (16) shows a case where the stancetaker, Finch, aligns with her interlocutor along both dimensions.

(16) Stance object: Graduation photo of Finch's child (Finch/medium)

Interlocutor:	[Name] already has a graduation photo of himself in a frame? Way to go!
Finch:	Well, that's graduating from primary school.
Interlocutor:	From… yeah, from…
Finch:	Oh no!
Interlocutor:	Middle school, right? [Name of school]?
Finch:	Middle school. Sorry, middle school. Right.

In (16), Finch implicitly disagrees with her interlocutor but then corrects herself. According to Brown & Levinson (1987), overt disagreement is a face-threatening act, and hence epistemic disalignment may give rise to affective disalignment (Langlotz & Locher 2012). However, the two aspects are not inseparable, as one can disagree politely: by seeking agreement, Finch minimizes imposition on the addressee's negative face by avoiding the addressee's face-loss. In correcting herself, Finch avoids disagreement. This negative politeness strategy is supported by an explicit apology (*Sorry*). Thus, the stancetaker both makes an effort to save the addressee's face and agrees with the addressee. While this may be a reason for coding separately for the affective and epistemic side of Alignment and Investment, we did not separate them in coding or in the analysis; whether the affective/epistemic nature of stancetaking has an effect remains to be answered in future research. The possible values for Alignment are 'align', 'neutral', and 'disalign'.

For Hierarchy, coders evaluate whether the speaker is presenting themself as an expert by adding 'Believe me, I know this better than you' to one of their utterances within the same stance act. If this addition harmonizes with their overall contributions in the stance act, the value 'expert' is assigned to Hierarchy. If adding 'You know this better than I do' results in a better match, the value 'novice' is assigned. Experthood in our coding protocol is not to be confused with any actual or professional expertise, but rather a self-construction of experthood (see Isaacs & Clark

1987); for example, the only German participant of a conversation can present themselves as an expert in the topic of German beer without actually being an expert in German beer. Their experthood is to be understood as relative, a stance that the stancetaker takes based on their judgment of the fact that not everyone has equal access to the stance object. In this way, the speaker is the expert on German beer as part of their broader self-construction as an expert on all things German despite not being a beer drinker.

By default, assertions pronounced without any discourse markers that weaken their assertive force (e.g. *I believe*, *apparently*, *so I heard*, etc.) are taken to commit the speaker as an expert (Gunlogson 2008). Experthood can also reveal itself by referential choice. For example, speakers are sensitive to whether their discourse participants know the name of a certain referent (a person, a building, or an event). The proper name is safe to use when the addressee also knows that name (e.g. *CN Tower*), but when the addressee is a novice, the speaker may choose a definite description instead (i.e. *the tallest building in Toronto*). By choosing a referring expression that demonstrates knowledge about the referent, the speaker can present themself as an expert (Isaacs & Clark 1987).

Consider (17), where the stancetaker, Finch, presents herself as an expert.

(17)   Stance object: Situation types in Finch's linguistic research (Finch/medium)

Finch:	Okay. So that's the midpoint, we figured, because what the- we've all decided what our midpoint is, like we have to figure what our- our- like the most formal one is gonna be me teaching.
Interlocutor:	Yeah.
Finch:	And then the- the least formal is gonna be me interacting with [name] or the kids or something like that and then the midpoint was our hardest problem. [**Believe me, I know this better than you.** / **#But you know this better than I do.**]
Interlocutor:	Right.

Finch is a linguist explaining her research to a non-linguist colleague. Her superior knowledge is demonstrated by using unadorned assertions, and in addition, by using *we* and *our* as exclusive pronouns, referring not to the sum of Finch and her addressee, but to the sum of Finch and non-present linguist colleagues. There is thus a certain dissociation from the non-expert addressee conveyed by the exclusive reading of these personal pronouns. In this case, Finch can felicitously add 'Believe me, I know this better than you' to the end of her utterance.

By contrast, in (18), Platypus does not act as an expert, but rather as a novice, and therefore they cannot add this statement.

TABLE 3. *Coding protocol for stance.*

STYLE/STANCE FACTORS	LEVELS
Social distance to interlocutor	high, medium, low
Evaluation	positive, neutral, negative
Alignment	align, neutral, disalign
Hierarchy	novice, same level, expert
Investment	high, medium, low

(18)   Stance object: The Canadian school system (Platypus/low)

Interlocutor: And it is shifting now because now we don't have those kind of jobs. We're not in factories any more. We have so many other jobs that require so many other skills necessary. So they are making those shifts now. That's our problem.

Platypus: Yeah. And it's happening now? I mean, from- from when you went to school? And- between when you went to school and now? Is this shift happening recently? [**You know this better than I do.** / **#Believe me, I know this better than you.**]

Interlocutor: This shift just started.

Platypus knows that her interlocutor attended teacher's college, allowing the interlocutor to take on the role of an expert when discussing the Canadian school system. Platypus asks genuine questions, which are canonical ways of expressing lack of knowledge, as well as rising declaratives, which can present the addressee as more knowledgeable (Gunlogson 2008). Platypus thus presents herself as a novice: her questions exclude scenarios in which she can take an expert stance. In accordance with this, her stance can be reinforced by adding 'you know this better than I do'. In (18), then, the value of Hierarchy is coded as 'novice'. A summary of the coding protocol for stance including the four predictors outlined here is presented in Table 3.

RESULTS

The overall rate of *that* in our dataset is 42.6%. However, there is considerable inter-speaker variation, with the most frequent *that*-user, Orca, having a rate of 68%, and the least frequent *that*-user, Quokka, having a rate of 28.7%. The individual variation is visualized in Figure 2.

No clear patterns emerge for any demographic factor considered. Our small number of speakers makes significance tests based on sociodemographic factors impractical and, as we have noted above, we would not expect to find such correlations in any case. However, there is a significant effect of social distance,
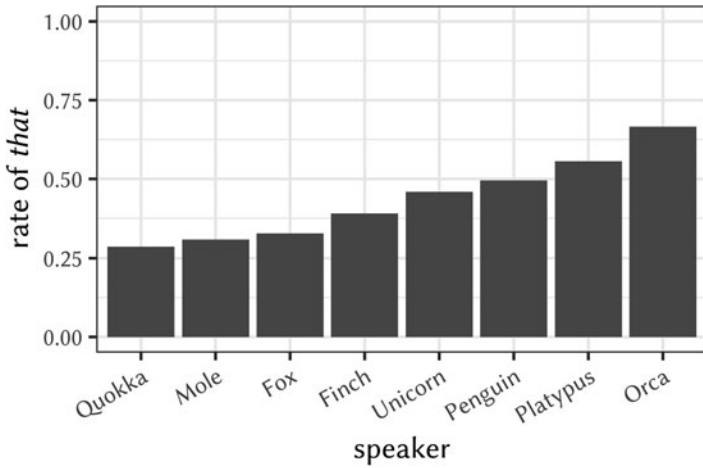
FIGURE 2. Individual variation in rate of *that*.
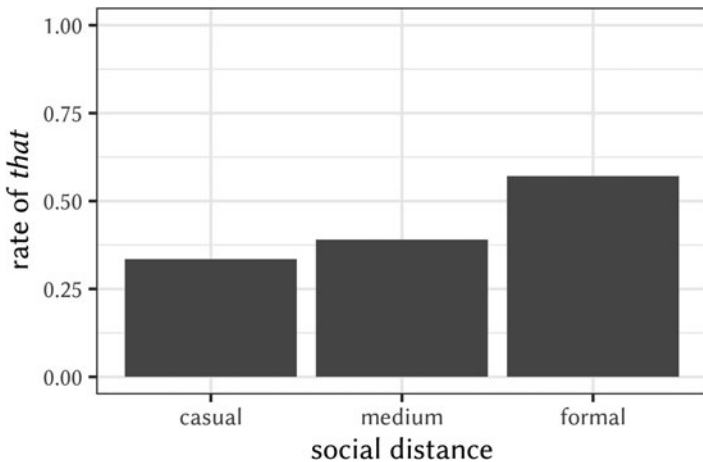
FIGURE 3. Effect of social distance in rate of *that*.

whereby greater social distance correlates with more *that*. This effect is visualized in Figure 3. This stepwise pattern supports previous observations about *that* which characterize it as a stylistic variable associated with more formal and less familiar genres (Storms 1966; Finegan & Biber 1985).

In terms of linguistic factors, there is a straightforward pattern by matrix subject: when the subject is not *I*, the rate of *that* is 54%, but when it is, the rate is only 27%. There are also distinct contrasts with respect to both types of intervening material.
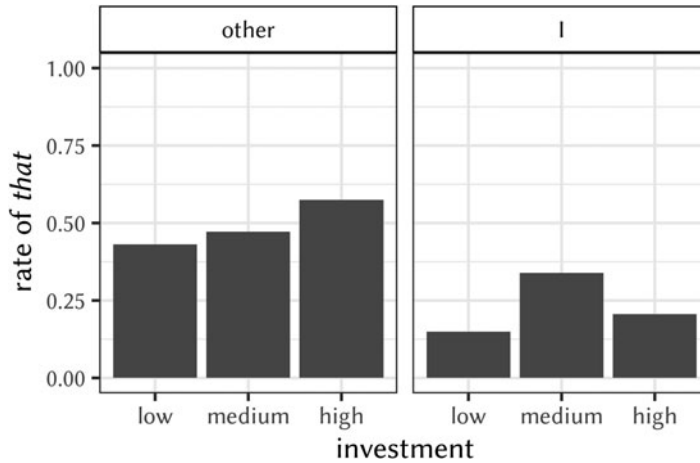
FIGURE 4. Apparent interaction between Investment and matrix subject.

When there are no intervening verbal arguments, the rate of *that* is 42%, compared to 68% when there are; and when there is no other intervening material, the rate of *that* is 39%, compared to 61% when there is. These are not visualized here for space; however, they are well in line with previous work on this variable (Tagliamonte & Smith 2005; Torres Cacoullos & Walker 2009).

Turning to our stance predictors, the only factor with a discernible pattern is Investment. Figure 4 shows an apparent interaction between matrix subject and investment. When the matrix subject is not *I*, the rates for both low and high Investment (and medium Investment, to a lesser degree) are much higher than when the matrix subject is *I*. Across both conditions, low Investment is consistently associated with the lowest rate of *that*.

To confirm the significance of these patterns, we fit a generalized linear mixed-effects model with a binomial linking function to the data. All models were fit using the glmer() function from the R (R Core Team 2018) package lme4 (Bates, Mächler, Bolker, & Walker 2015). We included the following predictors: Investment, matrix subject, intervening verbal material, other intervening material, and social distance. We also included a random intercept for speaker to account for interspeaker variation in the data. The output of the model is shown in Table 4.

Table 4 shows a significant effect for both intervening verbal arguments ($\beta = 1.35$, $p = 0.01$) and other intervening material ($\beta = 0.95$, $p < 0.001$): when there is any intervening material, *that* is significantly more likely. There is also a significant effect of social distance whereby *that* is more likely in more formal social contexts ($\beta = 0.27$, $p < 0.001$). Finally, there is an effect of Investment: *that* is less likely with low Investment ($\beta = -0.68$, $p = 0.046$). The interaction between Investment and the matrix subject approaches significance ($p = 0.06$).

43

TABLE 4. *Mixed-effects logistic regression model.*

	ESTIMATE	STANDARD ERROR	Z VALUE	PR (>|z|)	%	N
(Intercept)	0.72952	0.41635	1.752	0.07974		
Investment						
Medium Investment	*reference*				59%	132
High Investment	−0.09796	0.2516	−0.389	0.697028	55%	302
Low Investment	−0.68411	0.3432	−1.993	0.046229	71%	91
Matrix subject						
Other	*reference*				46%	364
I	−0.56771	0.39374	−1.442	0.149345	72%	263
Intervening verbal material						
Absent	*reference*				58%	602
Present	1.35172	0.52358	2.582	0.009831	32%	25
Other intervening material						
Absent	*reference*				61%	519
Present	0.94709	0.28039	3.378	0.000731	38%	108
Social distance						
Medium context	*reference*				61%	218
Casual context	0.08828	0.26398	0.334	0.738071	66%	218
Formal context	0.94328	0.26577	3.549	0.000386	42%	191
Interactions						
High Investment: Matrix subject *I*	−0.80923	0.49572	−1.632	0.102586		N/A
Low Investment: Matrix subject *I*	−1.28107	0.68123	−1.881	0.060035		N/A

Consistent with Figure 4, the trend is that the rate of *that* with low Investment is lower when the matrix subject is *I* (β = -1.28) than when it is not.

This model demonstrates that the operationalization of Investment has a significant impact on the distribution of *that* vs. zero in our data. While the results are intriguing, we refrain from making strong claims about how Investment interacts with this variable more generally due to the low number of speakers in our sample.

DISCUSSION

The theoretically grounded approach to stance taken in this article—qualitative in nature but applicable to large data sets—has the benefit of being able to incorporate aspects that are otherwise lost in analyses that are either strictly qualitative or quantitative. One of the most important contributions of this procedure is the ability to code stance using pragmatically informed test questions, drawing on the discourse hierarchy of QUDs.

This approach has several advantages. Previous variationist analyses that have incorporated stance (e.g. Holmes-Elliot & Levon 2017; Nycz 2018; inter alia)

have generally appraised stance within larger discourse units (i.e. sections of talk focused on a particular topic as part of a conversation). Our analysis goes one step further by contextualizing each token not only within its larger discourse context but also within specific stance acts themselves, which could be as small as a single utterance. In addition, our coding scheme attempts to frame the qualitative, intersubjective components of stancetaking as a set of social moves which can be captured quantitatively in a precise way. The development of this protocol follows in the footsteps of previous research by problematizing the idea that stance is an unquantifiable phenomenon. Specifically, we have attempted to provide greater understanding of the interactional process of stancetaking by examining how its subcomponents (stance dimensions) are instantiated in speech.

The pragmatic tests that we use here can be applied to code other data, and as a guide in the experimental investigation of different sources of pragmatic meaning such as morphology (Grandi & Körtvélyessy 2015), intonation (Prieto & Borràs-Comes 2018), discourse particles (Prieto & Roseano 2021), and gestures (Borràs-Comes, Kiagia, & Prieto 2019). Furthermore, the pragmatic reasoning behind our coding protocol is not restricted to the aspects of stance proposed in this article. It is also applicable to aspects of stance not considered here, such as Martin & White's (2005) 'graduation' or Levon's (2016) 'deontic stance'.

The coding protocol we outline has also proven capable of contributing new cursory insights into even such a well-studied variable as the English complementizer. This variable has offered a unique avenue through which to explore the effect of stance and the effectiveness of this coding protocol, partly because it has been considered in much previous research to be the subject of internal linguistic and processing constraints rather than a locus for social meaning. Our results demonstrate that rather than being a product solely of grammatical constraints, this variation is socially conditioned. In particular, complementizer choice is constrained by social distance, whereby overt *that* is less likely to be realized in less 'formal' speech contexts. Within the realm of stance, our analysis demonstrates that speaker investment, a subcomponent of interactional stance, constrains the realization of *that* vs. zero: speakers are more likely to employ *that* when their Investment is 'high'.

Much work on stance has focused on highly salient or socially significant features, such as *dude* (Kiesling 2004) or *güey* (Bucholtz 2009), and we might expect that more salient features are more likely to accrue social meaning. A previous finding that may be related is *that*'s putative status as prescriptively favoured or proper (Kroch & Small 1978); for Storms (1966:62), 'it is objective, factual, formal, official, sometimes tending to hostility', and Elsness (1984) similarly argued that *that* is much more common in written registers than in speech. This tendency surfaces in our data as well: *that* is favoured in formal contexts, zero in casual ones. It might be the case that when speakers feel strongly about the proposition they are introducing or discussing, formal variants can signal this elevated certainty. This parallels Labov's (2001:91) observation that speakers in what he calls 'soapbox style' tend to use more formal or standard variants. In fact, *that* may be prominent

in standard contexts BECAUSE it is linked to high investment in an indexical field (Eckert 2008). Because *that* is optional, its use may be perceived as more effortful, along the same lines as [ɪŋ] (Campbell-Kibler 2010) and released /t/ (Benor 2001, among others), and thus may be linked to increased care (Eckert 2008) and investment. Moreover, because *that* (complementizer) and *that* (demonstrative determiner) have the same phonetic and orthographical form, meanings associated with the demonstrative *that* (i.e. increased specificity) may also be relevant here. A matched-guide study contrasting *that* and zero could test these postulations directly. However, it may not be surprising that we see social work being done by leveraging *that*—these findings strongly support Kiesling's (2009) argument that stance is a core element underlying linguistic variation, as well as Eckert's (2019) proposal that all variation is associated with social meaning—even what we may think of as ubiquitous, stable variables below the level of consciousness.

Finally, our results raise the question of why we only found Investment to have a significant effect on *that*/zero variation, and did not find an effect for the other dimensions of stance. Of course, absence of evidence is not evidence of absence: we cannot rule out the possibility that these other dimensions do influence *that*/zero variation, and that we just did not detect it, for example due to our relatively small sample size and set of speakers. We thus refrain from speculating further about why we did not find effects for Evaluation, Alignment, and Hierarchy.

CONCLUSION

In this article, we introduced an innovative coding protocol informed by discourse semantics and pragmatics that contextualizes and correlates linguistic forms with speaker Evaluation, Investment, Alignment, Hierarchy, and other elements of interest within the literature on interactional sociolinguistics. Although the present study only unearthed an effect of Investment, the means of coding for stance outlined within this article are not only applicable to our data, but also have the potential to be applied to other data or corpora.

This protocol is not meant to replace qualitative methods, but to serve as an additional tool. The coding process is itself qualitative, drawing on pragmatic insights to capture theoretically important dimensions of stance. Critically, it enables the analyst to make judgments by using test questions with a closed set of answers that one can then quantify over. To test the applicability of the protocol in additional contexts, one promising direction for future work would be to take a corpus of spontaneous speech, extract excerpts, and code for a variable in the same manner. Future studies might also consider the utility of this protocol for capturing the nuances of social interactions, such as personal relationships between interlocutors or particularities of a given speech context, and whether it is capable of capturing phenomena like irony or linguistic ambiguity. It would be particularly interesting to determine to what extent the proposed method can be used across languages and cultures other than that considered here. Languages and cultures have their own norms of

interaction, but we expect that the pragmatic questions involved in the coding protocol will remain applicable.

Our findings converge with earlier literature arguing that the overt complementizer *that* is more objective or scientific-sounding than the null complementizer (Storms 1966), but our methods probe the variable patterning more deeply by providing an explanation of what kind of social work complementizers are doing in interaction and, potentially, how these social meanings developed. Complementizers—and potentially other functional elements—are not simply grammatically conditioned. Rather, they are resources for stylistic work in ways that traditional grammarians may never have envisaged. In sum, the success of our approach substantiates existing studies (e.g. Kiesling et al. 2012; Nycz 2018), demonstrating not only the possibility of including stance in quantitative analysis, but also the power of stance as an explanatory tool in understanding the complexity of linguistic variation.

NOTES

[1]As one reviewer observes, sociolinguistic studies normally analyze data from participants external to the research, and it is difficult to say with certainty the implications of relying on this direct knowledge for inter-coder reliability. While we acknowledge this limitation, future studies' employment of this protocol could shed light on this. Given the work of Kiesling et al. (2018), the potential for replicability seems promising nonetheless.

[2]Our speaker group consisted of individuals present in the graduate seminar from which the idea for this project originated, and is 'unbalanced' in terms of traditional social factors (e.g. age, gender, ethnicity, first language, etc.). We chose not to construct an artificially 'balanced' sample through addition of more speakers, as this would have been a misrepresentation of the speaker diversity in the city in which the researcher-participants all lived at the time. Furthermore, we contend that framing 'balanced' samples as a necessary requirement of quantitative sociolinguistic studies should be re-thought: the practice of privileging homogeneous 'balanced' samples has led to an overrepresentation of sociolinguistic studies and theories based on studies of white, cisgender, anglophone speakers.

[3]If the answers to the affective and epistemic questions contradict one another (for example, in the case of polite disagreement), coders are directed to choose the more salient aspect.

REFERENCES

Barnes, Sonia (2018). Copula variation in Asturian Spanish and the multidimensionality of stancetaking in interaction. *Journal of Sociolinguistics* 22:29–54.

Bates, Douglas; Martin Mächler; Ben Bolker; & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Benor, Sarah Bunin (2001). The learnèd /t/: Phonological variation in Orthodox Jewish English. *University of Pennsylvania Working Papers in Linguistics* 7(3), article 2. Online: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1462&context=pwpl.

Biber, Douglas, & Edward Finegan (1988). Adverbial stance types in English. *Discourse Processes* 11:1–34.

——— (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* 9:93–124.

Bolinger, Dwight (1972). *That's* that. The Hague: Mouton.

Borràs-Comes, Joan; Evangelia Kiagia; & Pilar Prieto (2019). Epistemic intonation and epistemic gesture are mutually co-expressive: Empirical results from two intonation-gesture matching tasks. *Journal of Pragmatics* 150:39–52.

Brown, Penelope, & Stephen C. Levinson (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Bucholtz, Mary (2009). From stance to style: Gender, interaction, and indexicality in Mexican immigrant youth slang. In Alexandra Jaffe (ed.), *Stance: Sociolinguistic perspectives*, 146–79. Oxford: Oxford University Press.

Büring, Daniel (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26:511–45.

Campbell-Kibler, Katherine (2010). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22:423–41.

Conrad, Susan, & Douglas Biber (2000). Adverbial marking of stance in speech and writing. In Susan Hunston & Geoff Thompson (eds.), *Evaluation in text*, 56–73. Oxford: Oxford University Press.

Du Bois, John W. (2007). The stance triangle. In Robert Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 139–83. Amsterdam: John Benjamins.

Eckert, Penelope (2008). Variation and the indexical field. *Journal of Sociolinguistics* 12:453–76.

——— (2019). The individual in the semiotic landscape. *Glossa: A Journal of General Linguistics* 4(1), article 14.

Elsness, Johan (1984). *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65:519–33.

Finegan, Edward, & Douglas Biber (1985). *That* and zero complementizers in Late Modern English: Exploring *ARCHER* from 1650–1990. In Bas Aarts & Charles F. Meyer (eds.), *The verb in contemporary English*, 241–57. Cambridge: Cambridge University Press.

———, & ——— (2001). Register variation and social dialect variation: The register axiom. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 235–67. Cambridge: Cambridge University Press.

Grandi, Nicola, & Lívia Körtvélyessy (2015). *The Edinburgh handbook of evaluative morphology*. Edinburgh: Edinburgh University Press.

Gunlogson, Christine (2008). A question of commitment. *Belgian Journal of Linguistics* 22:101–136.

Hamblin, Charles L. (1973). Questions in Montague English. *Foundations of Language* 10:41–53.

Heritage, John (2009). Conversation analysis as social theory. In Bryan S. Turner (ed.), *The new Blackwell companion to social theory*, 300–320. Oxford: Wiley-Blackwell.

———, & Geoffrey T. Raymond (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly* 68:15–38.

Holmes-Elliot, Sophie, & Erez Levon (2017). The substance of style: Gender, social class and interactional stance in /s/-fronting in southeast England. *Linguistics* 55:1045–72.

Hunston, Susan (2007). Using a corpus to investigate stance quantitatively and qualitatively. In Robert Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 27–48. Amsterdam: John Benjamins.

Isaacs, Ellen A., & Herbert H. Clark (1987). References in conversation between experts and novices. *Journal of Experimental Psychology* 116:26–37.

Jaffe, Alexandra (2009). The sociolinguistics of stance. In Alexandra Jaffe (ed.), *Stance: Sociolinguistic perspectives*, 3–28. Oxford: Oxford University Press.

Kärkkäinen, Elise (2007). The role of *I guess* in conversational stancetaking. In Robert Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 183–219. Amsterdam: John Benjamins.

——— (2012). *I thought it was very interesting*: Conversational formats for taking a stance. *Journal of Pragmatics* 44:2194–2210.

Kiesling, Scott F. (2004). Dude. *American Speech* 79:281–305.

——— (2009). Style as stance. In Alexandra Jaffe (ed.), *Stance: Sociolinguistic perspectives*, 171–94. Oxford: Oxford University Press.

——— (2011). Stance in context: Affect, alignment and investment in the analysis of stancetaking. Paper presented at iMean, Bristol, UK.

——— (2016). Stance: Theory and practice in sociolinguistics. Pittsburgh, PA: University of Pittsburgh, ms. Online: https://www.academia.edu/13000752/Stance_and_Stancetaking_Theory_and_Practice_in_Sociolinguistics.

——— (2020). Investment in stancetaking: *I mean* and *just sayin*'. *Language Sciences* 82:1–15.

———; Spencer Onuffer; & Alexander Hardware (2012). Operationalizing stance as an independent variable. Paper presented at New Ways of Analyzing Variation (NWAV) 41, Bloomington, IN.

———; Umashanthi Pavalanathan; Jim Fitzpatrick; Xiaochuang Han; & Jacob Eisenstein (2018). Interactional stancetaking in online forums. *Computational Linguistics* 44:683–718.

Kroch, Anthony, & Cathy Small (1978). Grammatical ideology and its effect on speech. In David Sankoff (ed.), *Linguistic variation: Models and methods*, 45–55. New York: Academic Press.

Labov, William (2001). *Principles of linguistic change, vol. 2: Social factors*. Hoboken, NJ: Wiley-Blackwell.

Langlotz, Andreas, & Miriam A. Locher (2012). Ways of communicating emotional stance in online disagreements. *Journal of Pragmatics* 44:1591–1606.

Levon, Erez (2016). Conflicted selves: Language, religion, and same-sex desire in Israel. In Erez Levon & Ronald Beline Mendes (eds.), *Language, sexuality, and power: Studies in intersectional linguistics*, 215–39. Oxford: Oxford University Press.

López-Couso, María José, & Belén Méndez-Naya (2012). On the use of *as if*, *as though*, and *like* in present-day English complementation structures. *Journal of English Linguistics* 40:172–95.

Martin, James R., & Peter R. R. White (2005). *The language of evaluation*. New York: Palgrave.

Nycz, Jennifer (2018). Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. *Language Variation and Change* 30:175–202.

Ochs, Elinor (1992). Indexing gender. In Alessandro Duranti & Charles Goodwin (eds.), *Rethinking context: Language as an interactive phenomenon*, 335–58. Cambridge: Cambridge University Press.

——— (1996). Linguistic resources for socializing humanity. In John J. Gumperz & Stephen C. Levinson (eds.), *Rethinking linguistic relativity*, 407–37. Cambridge: Cambridge University Press.

———, & Bambi Schieffelin (1989). Language has a heart. *Text* 9:7–25.

Palmer, Frank R. (1979). *Modality and the English modals*. New York: Longman.

Park, Joseph Sung-Yul (2013). Stance, style, and vocal mimicry. *Journal of Pragmatics* 53:84–95.

Podesva, Robert J. (2016). Stance as a window into the language-race connection: Evidence from African American and White speakers in Washington, DC. In H. Samy Alim, John R. Rickford, & Arnetha F. Ball (eds.), *Raciolinguistics: How language shapes our ideas about race*, 203–20. Oxford: Oxford University Press.

Poplack, Shana, & James A. Walker. (2002). A majority language in minority guise: The future of Quebec English. Paper presented at Canadian Linguistic Association, Dalhousie University, NS, Canada.

Prieto, Pilar, & Joan Borràs-Comes (2018). Question intonation contours as dynamic epistemic operators. *Natural Language and Linguistic Theory* 36:563–86.

———, ———, & Paolo Roseano (2021). The encoding of epistemic operations in two Romance languages: Intonation and pragmatic markers. *Journal of Pragmatics* 172:146–63.

R Core Team (2018). R: A language and environment for statistical computing. Online: http://www.R-project.org/.

Rissanen, Matti (1991). On the history of *that*/zero as object clause links in English. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, 272–89. London: Longman.

Roberts, Craige (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5(6):1–69.

Rohdenburg, Günter (1998). Clausal complementation and cognitive complexity in English. In Fritz-Wilhelm Neumann & Sabine Schülting (eds.), *Anglistentag Erfurt: Proceedings of the Conference of the German Association of University Teachers of English*, 101–112. Trier: Wissenschaftlicher.

Staum, Laura (2005). *When stylistic and social effects fail to converge: A variation study of complementizer choice*. Palo Alto, CA: Stanford University, ms.

Storms, Gotfrid (1966). *That*-clauses in Modern English. *English Studies* 47:249–70.

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

Thompson, Sandra A., & Anthony Mulac (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15:237–51.

Torres Cacoullos, Rena, & James A. Walker (2009). On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47:1–43.

Warner, Anthony (1982). *Complementation in Middle English and the methodology of historical syntax: A study of the Wyclifite sermons*. London: Croom Helm.

Address for correspondence:
Angelika Kiss
University of Toronto, Department of Linguistics
Sidney Smith Hall, 4th Floor, 100 St. George Street
Toronto, ON M5S 3G3, Canada
angelika.kiss@mail.utoronto.ca