The functions of prosody in the reported speech of rural Southwestern Nicaraguans

Las funciones de la prosodia en el discurso referido de nicaragüenses rurales del suroeste

Jeff Michno
Furman University, South Carolina, EE.UU.
jeff.michno@furman.edu

Original recibido: 30/10/2020
Dictamen enviado: 22/12/2020
Aceptado: 01/02/2021

Abstract

This paper contributes new evidence of the evidential and stance-taking functions of prosody in reported speech. Analysis of direct reported speech in an understudied vernacular Spanish suggests that prosody: (1) marks the boundaries of reported speech, (2) marks boundaries between multiple story-world voices, and (3) serves as a stance-taking device. Pitch average and, to a lesser extent, pitch range, were found to play a role in contextualizing reported speech. Pitch shifts tended to accompany reported speech as a unit rather than centering on a flanking intonational phrase boundary, although post-hoc analysis suggests an effect for prosodic breaks preceding reported speech. A mixed-methods analysis answers a call to assess the simultaneous functions of prosody in marking reported speech and in stance-taking. Pitch changes consistently marked dialogic reported speech, while changes in voice quality, rate, and rhythm variably accompanied salient linguistic markers in the reported speech of in- and out-group characters.

Keywords: evidentiality, identity, personal narratives, pitch, stance.

Resumen

Este trabajo presenta nueva evidencia de las funciones de la prosodia en el discurso referido para expresar evidencialidad y posturas afectivas. Los análisis de discurso referido directo en una variedad de español vernácula poco estudiada sugieren que la prosodia: (1) delimita las fronteras del discurso referido, (2) delimita las fronteras entre múltiples voces dentro del discurso referido, y (3) sirve como herramienta para expresar posturas afectivas. Se encontró que el promedio entonativo y, en menor medida, el rango de entonación, jugaron un papel determinante en la contextualización del

discurso referido. Los cambios entonativos tendieron a acompañar al discurso referido como una unidad en vez de centrarse en una frontera entonativa, aunque el análisis post-hoc sugiere un efecto para las pausas que preceden al discurso referido. Un análisis de métodos mixtos responde a una llamada a evaluar las funciones simultáneas de la prosodia en delimitar el discurso referido y expresar posturas afectivas. Los cambios de entonación delimitaron sistemáticamente el discurso referido dialógico, mientras que los cambios en la calidad y en el ritmo del discurso acompañaron variablemente a los marcadores lingüísticos que destacaron en el discurso referido según la identidad del personaje representado.

Palabras clave: *entonación, evidencialidad, identidad, narrativas personales, postura afectiva.*

Introduction

The prosodic structure of reported speech, or reported discourse, has been claimed both to index the speaker's stance and to distinguish the reported speech from the speaker's own. The rate of occurrence of prosodically-marked reported speech, however, has varied from study to study and has been shown to fluctuate according to the subtype of reported speech: *indirect*, *direct*, and *free direct*.

(1) a. They said that they weren't going to eat their vegetables. Indirect
b. They said / were like / were all, *"we ain't gonna eat that!"* Direct
c. I asked them to eat them, but, *"we ain't gonna eat that!"* Free Direct

Scholars have often viewed prosodic changes as a redundant marker of the reported speech segment, noting that other lexico-syntactic markers, such as *verba dicendi* (e.g., *they said*) or "new quotatives" such as *they were all / were like*, typically occur, as in (1a) and (1b). Instances of free direct speech, however, in which no lexico-syntactic markers are present, as in (1c), provide an ideal context for testing this claim. Presumably, if prosody does indeed serve to mark reported speech, free direct speech should yield higher rates of accompanying prosodic changes given the lack of other contextualization cues (Maldonado, 1999). Recent findings support this prediction and suggest that marked prosody may even be obligatory in such contexts, functioning as the sole evidential marker (Estellés-Arguedas, 2015).

Dialogic direct reported speech, in which multiple story-world voices are reported by the same speaker, offers a particularly rich context to assess both the evidential and stance-taking functions of prosody. How do speakers mark the

reported speech of different story-world characters while also distinguishing it from their own? Instances of dialogic *free* direct reported speech should prove particularly helpful in addressing this question, given the complete lack of lexico-syntactic cues among multiple sequential voices. Nonetheless, analysis of this type of data has been limited, presumably due to their low frequency of occurrence in comparison to other reported speech forms.

Among scholars who have argued that prosody functions to contextualize reported speech amid conversational speech, there has been conflicting evidence regarding the locus of pitch changes. While some studies suggest that pitch shifts are localized at the intonational phrase (IP) boundaries immediately prior to or following reported speech (Jansen, Gregory and Brenier, 2001; Oliveira and Cunha, 2004), others point out that pitch patterns often occur on a more global level, spreading across several intonational phrases comprising a reported speech utterance (Couper-Kuhlen, 1998). While prosodic marking has been identified across languages, variation has also been reported, with primary evidence coming from English, as well as from German, Russian, Japanese, and Spanish, among others. Scholars signal a need to continue cross-linguistic research that includes understudied vernacular speech and to consider prosodic features in addition to pitch, such as voice quality and rhythm.

The present study builds upon current knowledge of the role of prosody in direct reported speech by analyzing conversations in rural vernacular Nicaraguan Spanish, thus expanding cross-linguistic evidence with a highly understudied variety and an underrepresented group of speakers. The study assesses the role of both pitch level (mean) and pitch range in contextualizing reported speech to test diverging claims regarding the locus of pitch changes –whether they occur prior to, after, or across the entire reported speech segment. Simultaneously, it considers the impact of prosodic breaks (i.e., pauses) reported in some studies. The study also adds to recent findings regarding the polysemic evidential and stance-related functions of reported speech by supplementing a quantitative analysis with a qualitative inspection of dialogic free direct reported speech. It responds to a call to consider not only pitch but other prosodic features, including voice quality, intensity, speech rate, and rhythm, in speech representing in- and out-group voices.

In the following section, I provide background literature on reported speech, describing its types, corresponding lexico-syntactic and prosodic cues, and functions. After detailing the trajectory of research on the functions of prosody in reported speech, I briefly describe the Southwestern Nicaraguan community where data

were collected to provide sufficient sociolinguistic context for interpreting key lexical, segmental, and prosodic features of analysis. I end the section by presenting the study's research questions. In the following section, I describe the data and the methods used in their collection and analysis. Next, I share the study's results in order of the research questions, first presenting quantitative analyses of the occurrence and locus of pitch changes, followed by a qualitative analysis of stance-related functions of prosodic marking. I concurrently interpret and discuss the significance of the findings. Finally, I conclude with a summary of the study's implications, limitations, and suggestions for future research.

Background

Klewitz and Couper-Kuhlen (1999) remark that speakers typically mark reported speech (i.e., speech or thoughts of others or of their own from another time), often using lexico-syntactic resources such as deictics or a reportative phrase (e.g., *she said*). In the absence of such marking, however, identification of reported speech can be challenging for the hearer. Research into prosody has identified patterns that may help to disambiguate such cases. Klewitz and Couper-Kuhlen (1999) observe that reported speech in conversation is usually accompanied by a change in pitch register or range. The researchers interpret these prosodic shifts as "accomplishing something, namely marking… delimiting one's own territory or speech from someone else's" (p. 469). They note that changes in volume, speech rate, voice quality, and rhythm may also occur, but that pitch change nearly always accompanies these. Pitch is the primary prosodic feature of focus in the present study, while other acoustic correlates provide secondary evidence. Furthermore, pitch change has been found to accompany all forms of oral reporting, including *direct* (2a and 2b) and *indirect* (2c) forms (e.g., Günthner, 1997).

(2) a. *"No, I can't go with you."*
b. She said, *"No, I can't go with you."*
c. She said that she couldn't go with me.

Instances of direct reported speech such as (2a) demonstrate the value of marking spoken reported speech in some way to distinguish it from the speaker's own conversational speech. Lack of lexico-syntactic contextualization cues could make it problematic for hearers to determine the story-world voice of the reported speech (Goffman's (1981) *author* and/or *principal*), which is being brought to life by the real-world speaker or *animator*. Prosodic cues, such as pitch shifts, can aid

in this distinction. Some researchers (e.g., Jansen, Gregory and Brenier, 2001; Oliveira and Cunha, 2004) have provided evidence that pitch shifts are localized at the intonational phrase boundaries immediately prior to or following reported speech. Couper-Kuhlen (1998), on the other hand, observes that pitch patterns often occur on a more global level, spreading across several intonational phrases comprising a reported speech utterance. At the same time, Klewitz and Couper-Kuhlen (1999) caution that not all reported speech is prosodically marked, and that such marking may be a stylistic choice of the speaker.

While the aforementioned generalizations are drawn from English data, similar prosodic patterns have been discovered in other languages, such as Russian (Bolden, 2004) and German (Günthner, 1999). Not all languages appear to follow suit, however; Maier, (2014), for example, reported more ambiguous prosodic marking according to the directness of reported speech in Japanese. Coordinated efforts to compare reported speech cross-linguistically date at least as far back as Coulmas (1986). Much of the early work, however, relied on elicited speech, while more recent research analyzes naturally-occurring conversations. Cabedo (2007), for example, analyzed colloquial Iberian Spanish conversations and discovered, in line with Klewitz and Couper-Kuhlen's (1999) findings for English, that pitch patterns were inconsistent.

More recent research has focused on prosodic patterns corresponding to the different subtypes of direct reported speech. In a study of colloquial Iberian Spanish conversations from Valencia, Spain, Estellés-Arguedas (2015) found that prosody marked 100 % of direct reported speech that lacked a *verbum dicendi* (or other introductory mark) and 97.5 % of dialogic reported speech. The author interpreted this near-universal prosodic marking as evidence that prosody functions as an evidential marker in reported discourse. In examples of dialogic reported speech (i.e., multiple story-world voices in the same reported speech segment) the study found that changes in prosody (pitch) accompanied only one of the story-world voices, with additional voices realized in conversational-level pitch. The present study presents some evidence that diverges from this trend, wherein additional pitch rise separates the speech of adjacent story-world voices, all of which lack *verba dicendi* or other lexico-syntactic cues.

Reported speech and stance

As Günthner (1999) points out, reported speech is often imbued with the evaluation of the narrator, embodying Bakhtin's (1981) notion of polyphony, or layering, of voices –whereby the narrator not only produces the speech of a

real or imagined figure, but simultaneously comments on it (e.g., through the sociophonetic features and content of the speech). This observation has motivated research into the role of reported speech in stance-taking (Goffman, 1981; Goodwin, 2007) and identity-construction (Auer, 2007). The term *'reported speech*' itself has been contested regarding its authenticity. Tannen (1986) has argued that, even when tied to actual referents, the label can be misleading. Citing conversational data and the general limitations of human memory, she suggests that much reported speech is not in fact reported, but rather an approximation or embellishment, and therefore deserves an alternative label: *constructed dialogue*. These observations inform the present study's qualitative analysis of reported speech segments, which may include both real and hypothetical story-world characters, situations, and speech. They also allow for simultaneous consideration of the performative nature of reported speech (and its corresponding prosodic contours) alongside the role that prosodic features might play in marking the boundaries between conversational and reported speech segments, particularly in the absence of other linguistic markers.

Present study

Data for this study were collected in a rural Spanish-speaking community in Southwestern Nicaragua that has experienced intense culture, language, and dialect contact over the past twenty years. A growing tourism and property development industry has brought a steady stream of English- and Spanish-speaking tourists and workers to the formerly isolated coastal community. Commentary from 28 sociolinguistic interviews and casual conversations with locals during three visits over a 6-year period indicate that the growing tourist and expatriate presence is having social, economic, and linguistic impacts on the local community. Metalinguistic commentary suggests a sensitivity by locals to certain linguistic features associated with Nicaraguan Spanish and, in particular, the local vernacular. For example, the second-person singular pronoun *vos* 'you' (in contrast to *tú* and *usted*) and /s/-reduction (aspiration, deletion, or glottalization of [s]),[1] were frequently reported by locals as stigmatized features. In local speech during interviews, *vos* was primarily confined to personal narratives and the other two variants, *tú* and *usted,* were frequently used as pronouns of address with the

[1] See Lipski (1994, 2008) for a detailed description of Nicaraguan Spanish, including a discussion of its high rates of *vos* usage and /s/-reduction. See also Christiansen (2014) and Chappell (2015) for recent studies of pronoun usage and /s/-reduction, respectively, in Nicaraguan Spanish.

interviewer. *Vos* was often used to refer to locals or provide a local perspective (e.g., as a voice in reported speech) while *tú* was used to address or represent non-locals in reported speech. The pronouns often covaried with other sociophonetic features associated with local and non-local varieties. These observations suggest that locals use pronouns and other salient linguistic markers as semiotic resources in identity construction (Bucholtz, 2005; Eckert, 2008).

The present study contributes to the growing body of knowledge regarding the functions of prosody in reported speech by analyzing naturally-occurring speech in an underrepresented variety: rural Nicaraguan vernacular Spanish. The study utilizes modern acoustic technologies to move beyond impressionistic analyses of acoustic correlates, examining pitch level (mean) and range across intonational phrase and reported speech boundaries, while also assessing the effect of pauses prior to reported speech, to contribute new quantitative evidence regarding the location(s) of pitch changes. In addition, the study contributes new insights into the overlapping evidential and stance-taking functions of prosody in reported speech through qualitative analysis of naturally-occurring free direct reported speech.[2] The analysis answers a call to consider the polysemic nature of reported speech by attending to multiple prosodic features in addition to intonation, such as speech quality, intensity, and rhythm, and by considering the overlapping functions they might serve in naturally-occurring reported speech. Specifically, the study aims to answer the following research questions:

1. Does prosody aid in contextualizing the reported speech of rural Nicaraguan Spanish speakers (i.e., Is there a change in pitch level and/or pitch range at the boundaries between reported speech and adjacent conversational segments)?
2. If so, is the change in pitch localized at the IP boundary preceding or following the reported speech, or is it applied to the reported speech segment as a whole?
3. Is prosody used both to mark reported speech (in an evidential capacity) and to take a stance vis-à-vis the reported speech?
    a. Is there an effect for second-person singular pronoun use on pitch in reported speech (i.e., Does pitch vary according to use of stigmatized in-group *vos* versus prestigious out-group *tú*?)

[2] Similar to Estellés-Arguedas (2015) and Estellés Arguedas and Albelda Marco (2014), the present study embraces a broad view of evidentials as language that, principally, points to the information source, regardless of whether or not that language is considered part of the grammar. This view aligns with Chafe and Nichols (1986), Biber and Finnegan (1988), and Diewald and Smirnova (2010), while diverging from Aikhenvald (2004).

b. Do pitch and other prosodic changes accompany dialogic free direct reported speech containing in-group linguistic markers (*vos* pronoun and /s/-reduction) versus out-group markers (*tú* pronoun and [s]-realization)?

Methodology

Data collection and analysis

This study utilizes data taken from sociolinguistic interviews between the researcher and Southwestern Nicaraguans living in the same rural community. To investigate the role(s) of prosody in the reported speech of participants, I identified and analyzed 20 instances of direct reported speech produced by 7 male speakers during personal narratives.[3] I focus here on male speakers for three reasons: (1) I encountered a greater quantity of reported speech in interviews with males; (2) I follow Jansen *et al.*'s (2001) approach of analyzing male and female speakers separately,[4] and (3) I aim to complement Jansen *et al.*'s (2001) study, which was limited to the analysis of female productions due to small male sample size. The 7 male speakers analyzed in this study ranged in age from 19 to 51 (19, 21, 25, 31, 45, 46, 51) and varied widely with respect to highest education level achieved: literacy training only (1), primary school (3), secondary school (1), some university (2). They were second-hand acquaintances of the researcher recruited through a local contact. The interviews, which lasted from 30-45 minutes, proceeded in a spontaneous fashion loosely guided by a list of core questions about life in the community and changes in lifestyles over time.

The present analysis is restricted to direct reported speech, as no tokens of indirect reported speech were discovered in the interview transcripts, mirroring low-frequency counts in other studies (e.g., Estellés-Arguedas, 2015). I annotated and transcribed each of the reported speech segments using PRAAT (version 6.0.33). In addition, I analyzed the intonational phrases (IP) immediately preceding and following each segment of reported speech. Following the Spanish ToBI labeling scheme (a prosodic annotation tool),[5] I ensured that

[3] In line with Ochs and Capps (1996), the present study views personal narratives as enactments of "actual or possible life events" (p. 19). The first 20 instances of reported speech identified in the personal narratives of interviewees were selected for this analysis.

[4] The authors recommend this approach because speech perception/production in the frequency domain is logarithmic, speaking to general physiological differences (i.e., vocal tract length) between males and females that bear on speech frequency.

[5] http://prosodia.upf.edu/sp_tobi/en/

each intonational phrase was bounded by a prosodic break index of 3 or 4. For instances of reported speech spanning across multiple intonational phrases, I coded for the first and last IP as well as the entire segment of reported speech.[6] This allowed inspection of more localized pitch effects across adjacent IP boundaries (at each edge of the reported speech) in addition to more globalized pitch movements over the entire stretch of reported speech. This enabled testing of conflicting claims by researchers (e.g., Couper-Kuhlen, 1998; Jansen *et al.*, 2001; Oliveira and Cunha, 2004) regarding the locus of pitch level and pitch range shifts and patterns. In tandem, I coded for pauses (break index 4+) at the reported speech-flanking IP boundaries. Next, I manually identified the maximum (f0-max) and minimum pitch (f0-min) values (in Hertz) of each intonational phrase, and recorded the difference as *pitch range*. This was done by hand to avoid errors by PRAAT's autocorrelation pitch tracker function due to perturbation by creaky voice and background noise. It also served to determine the upper and lower limits for calculating the *pitch level* (mean f0) in Hertz of each intonational phrase. With these boundaries set, I measured the *pitch level* using the autocorrelation pitch tracker function. I used one-way ANOVAs to test for significant differences between the mean pitch level and pitch range values of the preceding and following IPs with respect to the reported speech segment. Where effects were found, I utilized Tukey HSD post-hoc tests to identify the locus of effects at one or both boundaries.

Regarding prosody and speech style, given the literature linking reported speech to stance-taking and identity-construction, and the suspected local usage of pronouns in these capacities, each instance of reported speech was coded according to the pronoun it contained: *vos*, *tú*, *usted*, and *none*. As a secondary consideration, I coded for voice qualities such as breathy, creaky, and rhythmic voice, and /s/-realization. As /s/-reduction was frequently identified by locals as a stigmatized feature of Nicaraguan Spanish, its status as a stereotype makes it a readily available sociophonetic resource (e.g., for use in reported speech and identity-construction (Auer, 2007)).[7]

6 Reported speech segments averaged 2.75 phonological phrases in length.

7 Given this paper's primary focus on pitch and due to space limitations, the qualitative analysis does not include a detailed acoustic description of speech features such as voice quality, speech rate, and intensity, and /s/-realization utilizing acoustic technologies. These features were coded by the primary researcher impressionistically and verified using Praat acoustic software. Future research primarily focused on these co-occurring features would benefit from more detailed acoustic descriptions.

Results and Discussion

To determine the role of prosody in contextualizing reported speech, I considered both *pitch level reset* and *pitch range reset*. Again, *pitch level* refers to mean f0, and *pitch range*, to the difference between f0-max and f0-min of the given segment. *Reset* was calculated as the difference in pitch level and pitch range values between the reported speech segment and each of its flanking intonational phrases.

Global pitch patterns

First, I follow Couper-Kuhlen's (1998) suggestion to consider pitch reset patterns on a global level by treating each instance of reported speech as a unit –even if it was comprised of multiple intonational phrases. Table 1 shows the mean pitch level values and standard deviations for the reported speech (entire RS utterances) and flanking intonational phrases.

Table 1. Mean pitch level (Hz) by intonational phrase

	Preceding IP	Reported Speech	Following IP
Avg.(SD)	130(19)	171(38)	120(16)

The mean pitch level values suggest a difference across the factor groups. A one-way ANOVA shows that the difference across intonational phrases is highly significant, $F(2, 57) = 20.9$, $p < .001$. A Tukey HSD post-hoc test shows that there are highly significant differences between the reported speech segment and the preceding IP ($t(19) = 41.4$, $p < .001$) and the reported speech segment and the following IP ($t(19) = -50.8$, $p < .001$), but not between the flanking IPs themselves ($t(19) = -9.4$, $p = 0.51$). The results are clearly displayed in the boxplot in Figure 1.

Turning to *pitch range* reset, Table 2 shows the mean pitch range values and standard deviations for the reported speech (entire segment) and flanking intonational phrases.

Table 2. Mean pitch range (Hz) by intonational phrase

	Preceding IP	Reported Speech	Following IP
Avg(SD)	45(27)	92(49)	31(15)

Once again, the mean values suggest a difference across the factor groups, although the high SD values indicate high variability. An almost identical pattern emerges: a one-way ANOVA shows that the difference is highly significant, $F(2, 57) = 17.6$, $p < .001$. A Tukey HSD post-hoc test shows that there are highly significant differences between the reported speech segment and the preceding IP ($t(19) = 46.6$, $p < .001$) and the reported speech segment and the following

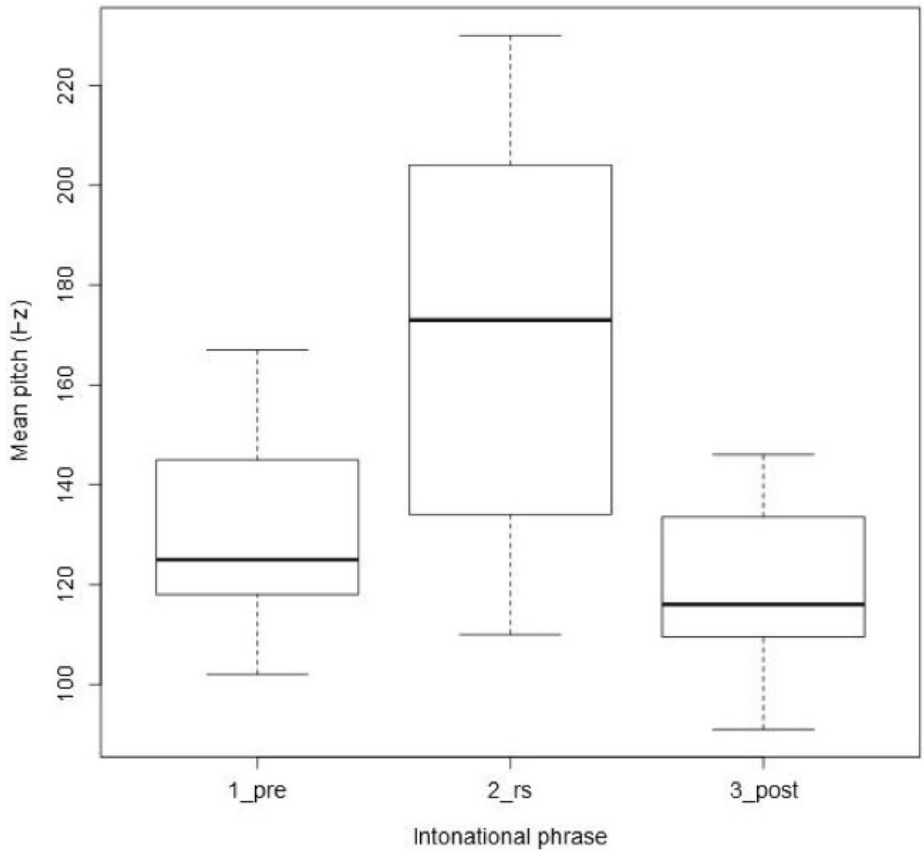

Figure 1. Mean pitch (Hz) by intonational phrase

IP ($t(19) = -60.1$, $p < .001$), but not between the flanking IPs themselves ($t(19) = -13.5$, $p = 0.42$). The results appear in the boxplot in Figure 2.

One consistency across the pitch level and pitch range results, in addition to the significant main effects, is a tendency for greater variability within the reported speech IP. This may reflect the performative nature of story-world voices, with speakers utilizing broad ranges of phonetic attributes such as pitch, intensity, and voice quality to animate a broad range of characters. In response to Research Question 1, these results suggest that both pitch level and pitch range play a role in contextualizing reported speech.

Localized pitch shifts

A secondary approach was taken to test disparate claims regarding the specific locus of pitch level and range reset with respect to the reported speech and its

boundaries. While Jansen *et al.* (2001) found a main effect located at the IP boundary *preceding* the reported speech, Oliveira and Cunha (2004) found the effect to lie at the IP boundary *following* the reported speech. This section narrows the focus to only the first and last IP of each reported speech segment. In this way, it aims to identify local shifts at the level of an IP boundary, which can then be compared to the more global shifts explored in the section *Global pitch patterns.*

For this analysis, four intonational phrases were considered. In addition to the two IPs flanking the reported speech, the reported speech itself was broken into two IPs–the first and last IPs of the reported speech utterance. Similar to the previous analysis, a one-way ANOVA was run to identify any group effects, one for *pitch level* and another for *pitch range*. The results closely mirrored those in the section *Global pitch patterns*, with a significant main effect for IP in both cases. Tukey HSD post-hoc tests showed significant differences across both preceding

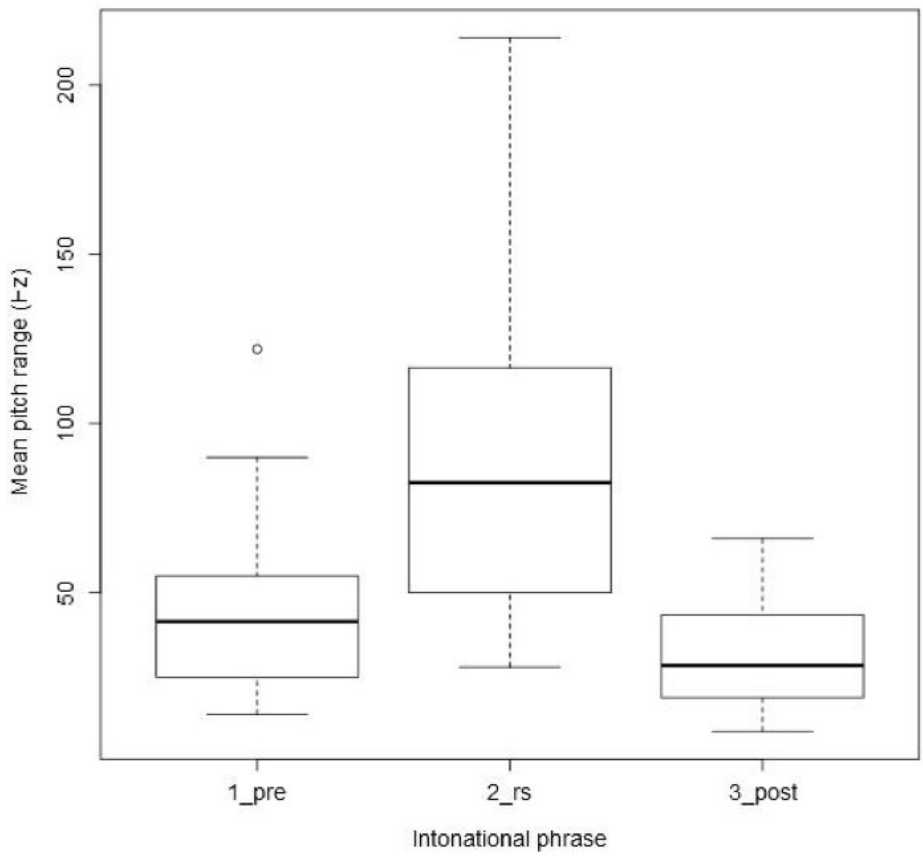

Figure 2. Mean pitch range (Hz) by intonational phrase

and following reported speech IP thresholds, $p < .05$. There were no significant differences between the two internal reported speech IPs or between the two flanking IPs. The results are displayed in boxplots in Figures 3 and 4. The first IP of the reported speech segment is labeled *rsin*, and the last, *rsout*.

These results support Couper-Kuhlen's (1998) suggestion that pitch reset patterns may be applied at a more global level across the entire reported speech utterance, rather than centering on the preceding or following IP boundary. The similarity in pitch throughout the reported speech utterance, and its contrast with both flanking IPs, contrasts with findings by Jansen *et al.* (2001) that pitch shifts centered on the preceding boundary, and by Oliveira and Cunha (2004) that they centered on the following boundary.

As a post-hoc analysis based on considerations of Jansen *et al.* (2001), I tested whether the presence of an intonational phrase break of Spanish ToBI index 4 or

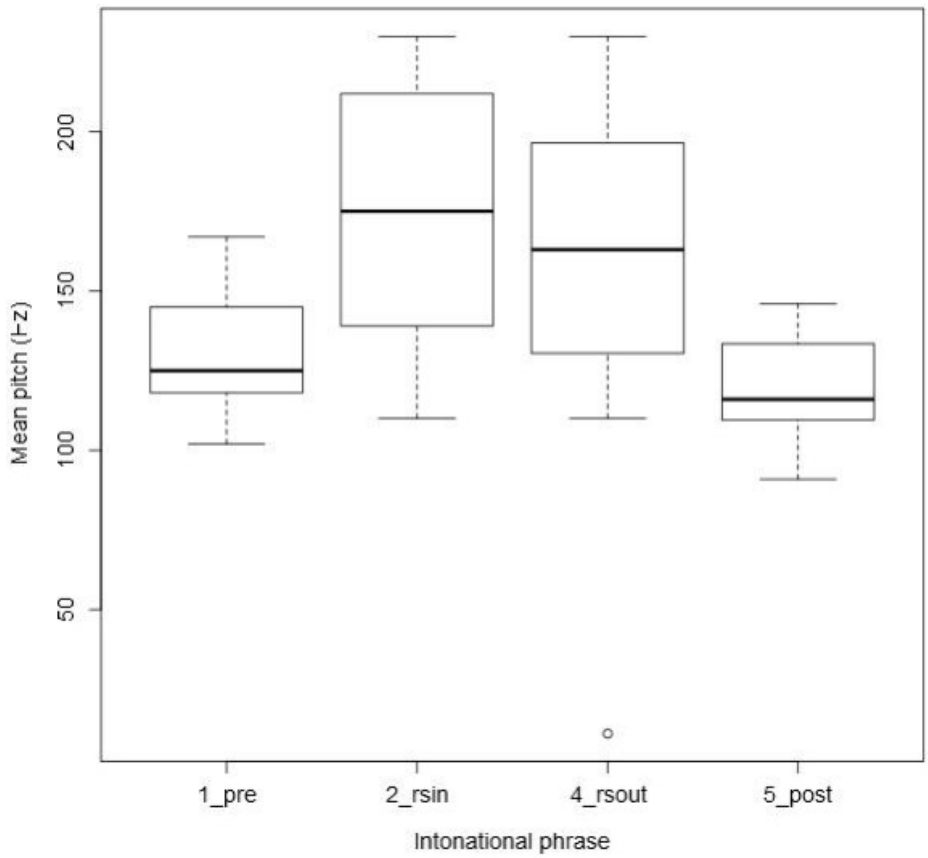

Figure 3. Mean pitch (Hz) as a function of first and last reported speech intonational phrases

greater (pause) would have an effect on pitch. While the phrase break *after* the reported speech showed no effect, the *preceding break* showed a highly-significant main effect ($F(1,57) = 13.4$, $p < .001$). The result is displayed in Figure 5, in which reported speech without a preceding intonational phrase break is labeled *prebreak: no*.

The line plot in Figure 5 shows that pitch was significantly higher in reported speech that lacked a preceding pause than in reported speech that did have a preceding pause (intonational phrase break of Spanish ToBI index 4 or greater). These results suggest that the pause before reported speech may serve as a salient phonetic cue in distinguishing between the speaker's voice and the story-world voice. The observation corroborates claims by Jansen *et al.* (2001) regarding the significance of the preceding IP boundary and motivates a follow-up study focused centrally on the role of prosodic breaks in pitch movements.

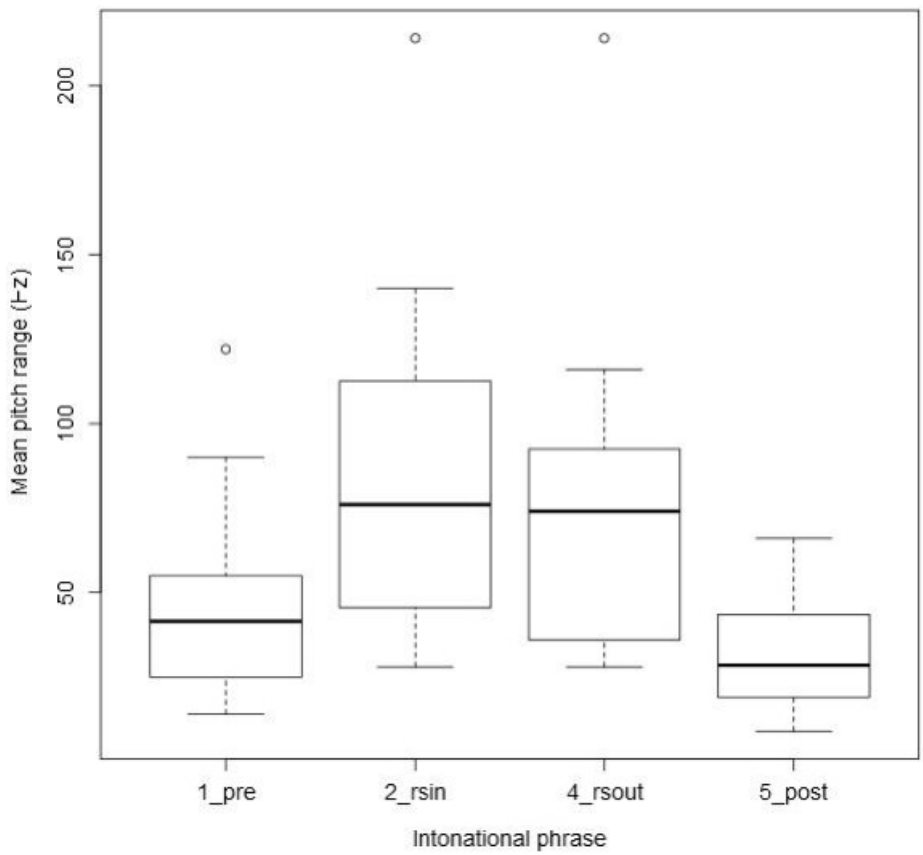

Figure 4. Mean pitch range (Hz) as a function of first and last reported speech intonational phrases

Prosody and stance

Finally, to consider the use of prosody as a stance-taking device, a two-part analysis was employed. The first part quantitatively assesses the prosodic effects of pronoun usage in reported speech. The analysis includes 9 instances of reported speech containing the pronoun *tú*, 5 with *vos*, 2 with *usted*, and 4 with no pronoun. A 3 (*IP*) X 4 (*pronoun*) factorial ANOVA was run, first with *pitch level*, then with *pitch range*, as the dependent variable. The results for *pitch level* confirm a strong main effect for *IP* ($F(2, 57) = 25.6$, $p < .001$) as well as main effect for *pronoun* ($F(3, 57) = 3.8$, $p < 0.01$), and no interaction between the two. The results for *pitch range*, on the other hand, confirm a strong main effect for *IP* ($F(2, 57) = 18.5$, $p < .001$), but show no significant effect for pronoun. The results for *pitch level* are displayed in Figure 6.

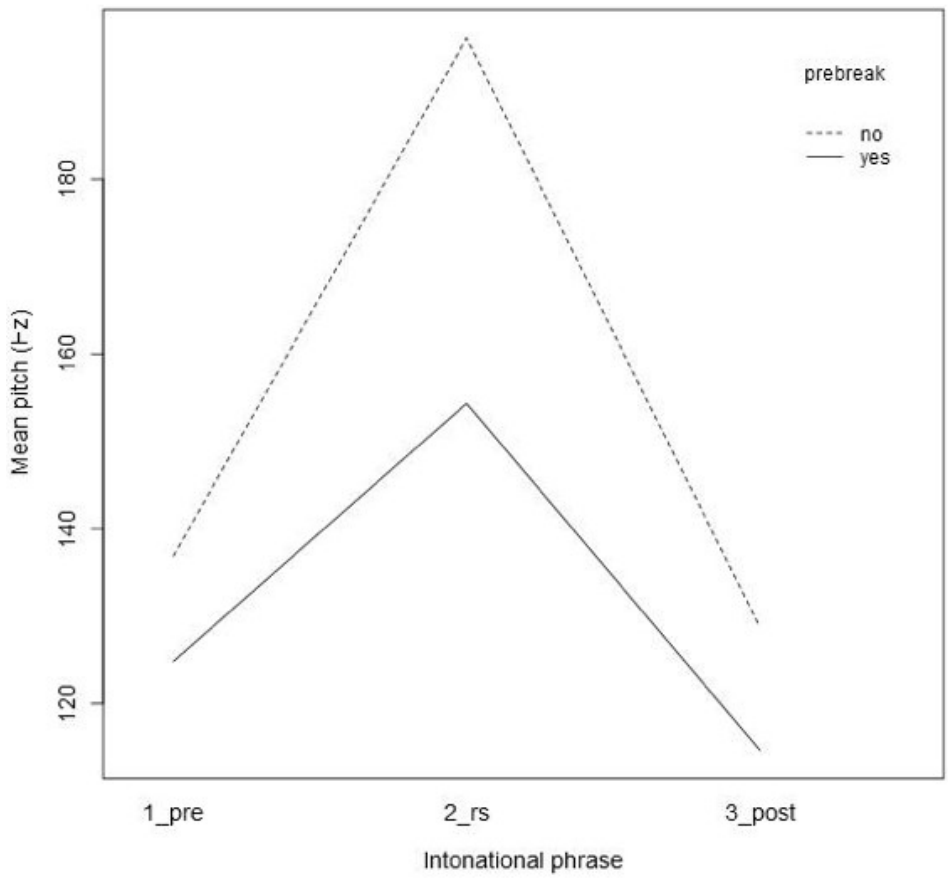

Figure 5. Mean pitch (Hz) by intonational phrase and reported speech preceding break

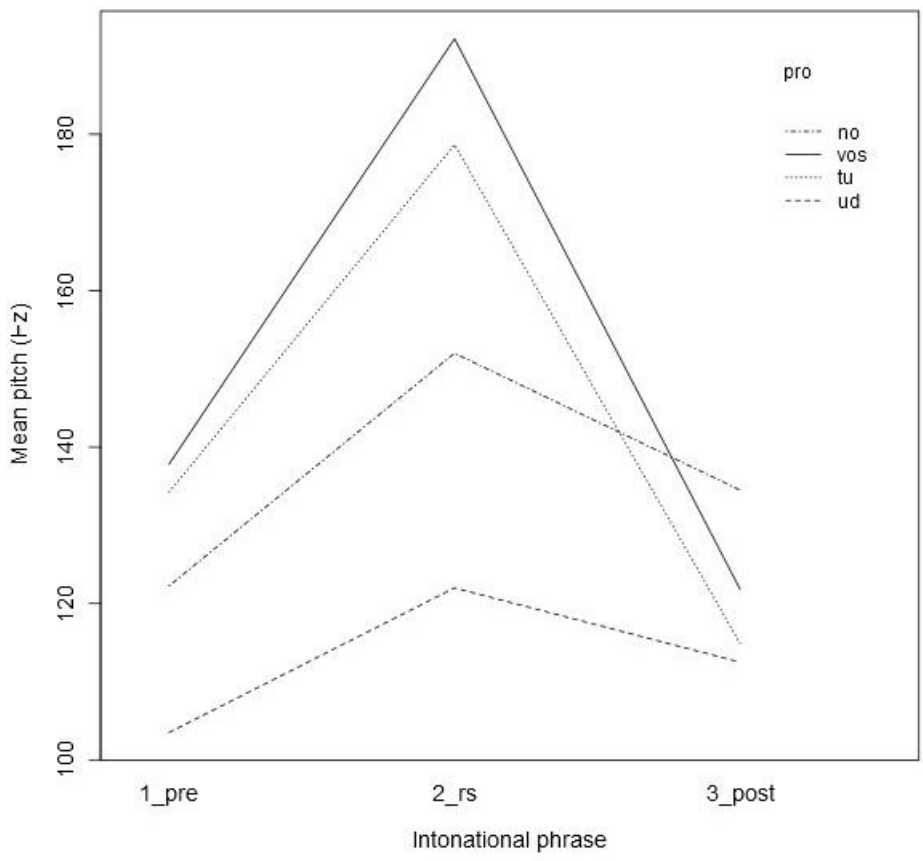

Figure 6. Mean pitch (Hz) by intonational phrase and pronoun

Figure 6 shows a clear effect on pitch level for *IP*, as expected, but it also shows a main effect for *pronouns*. Most notably, *tú* and *vos* appear near the top of the mean pitch level, with sharp slopes both into and out of the reported speech. *Usted*, on the other hand, lies near the bottom, its low slopes reflecting much smaller shifts in pitch level in and out of reported speech. A tentative explanation is that *usted* is the unmarked variant, while *tú* and *vos* are marked in terms of potential social meaning embodied by the pronouns. As mentioned in the introduction, participant usage of the pronouns during interviews hinted at in-/out-group membership along *vos*/*tú* lines, and metalinguistic commentary supported this interpretation: locals often referred to *vos* as Nicaraguan and *tú* as a form used by outsiders. As such, it could be that locals are using the pronouns in reported speech to take a stance vis-à-vis local and outsider story-world voices as well as with real-world recipients.

The second part of the analysis of stance in reported speech provides a qualitative look at prosody as it is used moment-to-moment in the following transcripts to represent multiple voices. The speaker produces both direct and free direct dialogical reported speech, at times alternating between two story-world voices alongside his own voice without any lexico-syntactic cues.

In the first transcript, the speaker, a 19-year-old first generation university student, comments on why he would not use the prototypical Nicaraguan second-person pronoun *vos* to address a recently-made acquaintance at university.[8]

Transcript 1:[9]

1		si yo comienzo
		if I begin
2	f_1:	hh ↑<<*f, breathy, rhythmic*>entOnce(s) vO(s) como te llamÁ(s)?>
		hh well what's your(vos) name?
3		↓es como más vulgar como::(-)
		it's like more vulgar like
4		má::s incorrecto, pues no, no sé
		more inappropriate, like, I don't know

Following the introductory mark in line 1 *si yo comienzo* 'If I begin,' an increase in pitch signals a switch to a story-world voice, in this case, the speaker's own imagined voice addressing a new acquaintance.[10] The increase in pitch, however, is not the only prosodic cue corresponding to the reported speech segment; it is accompanied by an increase in intensity and a breathy, rhythmic quality. During his hypothetical use of the stigmatized local form *vos*, which he subsequently describes as vulgar and inappropriate, the speaker dynamically adjusts multiple components of the reported speech prosody, before returning to his

8 The symbol 'f' is utilized in transcripts to index a story-world figure (i.e., reported speech). Superscripts distinguish between multiple figures in the same transcript (e.g., f_1, f_2).

9 Transcription conventions are based on GAT, Gesprächsanalytisches Trankriptionssystem (Selting *et al.*, 1998).

10 It could be argued, alternatively, that the change in pitch corresponds to the transition to an interrogative in line 2, given the intonational differences between Spanish interrogatives and declaratives described in the literature (e.g., Navarro Tomás, 1944; Canellada and Madsen, 1987; Quilis, 1993). Indeed, the side-by-side analysis of declarative and interrogative modalities in the reported speech of this study somewhat limits pitch pattern generalizability. Nonetheless, data such as those utilized in Transcript 2 and in the quantitative analysis support the view that pitch changes accompany shifts from speakers' conversational voices to both declarative and interrogative reported speech segments.

own conversational voice and prosodic contours in lines 3 and 4. Tellingly, the reported speech also contains another local speech feature reported as stigmatized, /s/-reduction, which appears three times in quick succession as fully deleted coda-s. The combination of stigmatized markers of local speech (*vos* and /s/-reduction) with a loud, breathy, and rhythmic speech quality, which contrasts with the preceding and following conversational speech, suggests the performative nature of the reported speech segment. In effect, the speaker is producing, while commenting on or evaluating, stereotypical local speech, embodying Bakhtin's (1981) notion of polyphony, or layering, of voices. While an increase in pitch is present, it co-occurs with several other prosodic, segmental, and lexical features.

In the next transcript excerpt, the same speaker offers an example of how he would address a recent acquaintance on the public bus. This time, he opts for the *tú* pronoun and corresponding verbal forms often attributed by locals to out-group members and other varieties of Spanish. The speaker also produces dialogical free direct reported speech containing two story-world voices alongside his own conversational voice.

Transcript 2:

1	f_1:	↑Oye (.) tú no eres e:::l chico que estaba la otra vez?
		↑*Hey aren't you(tú) the guy that was here the other time?*
2	f_2:	↑<<*all, falsetto*>Oh sí yo soy.>
		↑*Oh yes I am.*
3	f_1:	↓Oye tú eres buena onda.
		↓*Hey you're(tú) nice/cool/etc.*
4		↓como cosas así.
		↓*like things like that.*

Again, the consistency of pitch changes across voices is apparent, signaling breaks between conversational and reported speech *and* between consecutive free direct speech by different story-world voices. It is worth noting that the reported speech in line 1 is a rising declarative, which could be argued to explain the initial pitch rise from the speaker's conversational voice (preceding the transcript). The transition to the next story-world voice (f_2) in line 2, however, corresponds to an additional rise in pitch. Furthermore, the pitch lowers from line 2 to line 3, and once again to line 4, each of these lines containing a falling declarative produced by a different voice (transitioning from f_2 to f_1 to the speaker himself). These pitch

shifts, therefore, suggest that pitch primarily serves here to differentiate between each of the two story-world voices and the speaker's own.

In this segment, the speaker utilizes multiple prosodic elements in creating the story-world voices. The loud, breathy, rhythmic speech from his former imagined self in Transcript 1 has been replaced with smooth speech similar to his own conversational speech. It contains fully-realized coda-s, which differed from his typical aspiration in conversational speech. The second character's speech corresponds to an additional rise in pitch and increase in speech rate, giving it a falsetto quality. The additional rise in pitch supports observations that a change in speech rate rarely occurs alone (Klewitz and Couper-Kuhlen, 1999). At the same time, it contrasts with findings by Estellés-Arguedas (2015) that when multiple voices appear in dialogic reported speech, only one is prosodically marked. It could be that the order of voices with respect to the speaker's own conversational voice bears on whether one or more voices are marked. In this case, for example, the first reported speech segment represents the speaker's own voice. His increase in pitch suggests its role in demarcating the reported speech boundary. The subsequent pitch rise, then, is utilized to mark the free reported speech of another story-world voice. Were the order of the voices switched, perhaps the marking would have differed, aligning with observations by Estellés-Arguedas (2015) that, in her corpus of Iberian colloquial Spanish, the speaker's own voice was never marked in dialogic reported speech. Instead, this example refutes the view that, in dialogic direct reported discourse, prosody functions to mark only story-world voices that do *not* belong to the speaker. This piece of evidence suggests that prosody functions to demarcate both (1) conversational from story-world voices and (2) multiple adjacent story-world voices –regardless of whether or not one of those voices represents the speaker. It supports the claim that in dialogic direct reported speech prosody functions in an evidential capacity (Estellés-Arguedas, 2015). Concurrently, the transcript excerpts demonstrate how clusters of features, such as intensity, voice quality, and rhythm, can be attributed to a given story-world voice, mirroring co-occurring linguistic features involved in social style formation (Gumperz and Cook-Gumperz, 1982; Auer, 2007), and showing how they can be used in tandem with salient linguistic markers to take stances on types of speech and speakers.

Conclusion

These findings suggest that prosody, in particular pitch level and, to a slightly lesser extent, pitch range, play a role in contextualizing reported speech in rural

southwestern Nicaraguan vernacular Spanish. The study also suggests that the pitch shifts are applied to the reported speech as a unit rather than centering on one flanking IP boundary or the other, although a post-hoc analysis motivates further research into the role of prosodic breaks immediately prior to reported speech. In addition, observations of an effect for second-person singular pronoun use on pitch in reported speech suggests that prosody may concurrently function in a stance-taking capacity, in light of the strong association between pronominal variants *vos* and *tú* and in- and out-group speech, respectively. While the sample size is limited, the pattern motivates follow-up research into the pronouns' effects on other prosodic elements, such as intensity, voice quality, and rhythm. A qualitative analysis of multiple voices in dialogic free direct reported speech provides additional evidence that pitch changes function to demarcate both story-world from conversational speech, and the speech of multiple story-world characters, even if the character represents the speaker him/herself. Combined, these findings provide evidence of the polysemic nature of prosody in reported speech, suggesting that it is used both as an evidential to mark the speech, and as a performative device in taking a stance vis-à-vis the reported speech. While this study is limited to a particular group of male speakers, it identifies robust patterns that motivate follow-up research among female speakers in the same community and across age and gender groups in other communities. In addition to bolstering the quantitative findings, additional data could provide more qualitative evidence of pitch changes –and other prosodic attributes– occurring in dialogic free direct reported speech.

References

Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.

Auer, P. (Ed.). (2007). *Style and social identities: Alternative approaches to linguistic heterogeneity.* Berlin: Mouton de Gruyter.

Bakhtin, M. M. (1981). "Discourse in the novel." In M. Holquist (Ed.), *The dialogic imagination* (pp. 259-422). Austin: University of Texas Press.

Biber, D. & Finegan, E. (1988). "Adverbial stance types in English." *Discourse Processes, 11*, pp. 1-34.

Bolden, G. (2004). "The quote and beyond: Defining boundaries of reported speech in conversational Russian." *Journal of Pragmatics, 36*(6), pp. 1071-1118.

Bucholtz, M. (2005). "Identity and interaction: A sociocultural linguistic approach." *Discourse Studies*, *7*(4-5), pp. 585-614.

Cabedo, A. (2007). "Caracterización prosódica del estilo directo de habla en la conversación coloquial." *Estudios de Lingüística Universidad de Alicante, 21*, pp. 1-13.

Canellada, M. J. y Madsen, J.K. (1987). *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.

Chafe, W. & Nichols, J. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, N. J.: Ablex.

Chappell, W. (2015). "Formality strategies in Managua, Nicaragua: A local vs. global approach." *Spanish in Context*, *12*(2), pp. 221-254.

Christiansen, A. (2014). "El *vos* es el dialecto que inventamos nosotros, la forma correcta es el *tú*." *Borealis: An International Journal of Hispanic Linguistics*, *3*(2), pp. 259-297.

Coulmas, F. (Ed.). (1986). *Reported speech: Some general issues. Direct and indirect speech.* Berlin: Mouton de Gruyter.

Couper-Kuhlen, E. (1998). "Coherent voicing: On prosody in conversational reported speech." *Pragmatics and Beyond New Series*, pp. 11-34.

Diewald, G. & Smirnova, E. (Eds.). (2010). *Linguistic realization of evidentiality in European languages.* Berlin/New York: De Gruyter Mouton.

Eckert, P. (2008). "Variation and the indexical field." *Journal of Sociolinguistics, 12*(4), pp. 453-476.

Estellés-Arguedas, M. (2015). "Expressing evidentiality through prosody? Prosodic voicing in reported speech in Spanish colloquial conversations." *Journal of Pragmatics, 85*, pp. 138-154.

Estellés Arguedas, M. & Albelda Marco, M. (2014). "Evidentials, politeness and prosody in Spanish: A corpus analysis." *Journal of Politeness Research, 10*(1), pp. 29-62.

Goffman, E. (1981). *Forms of talk.* Philadelphia: University of Pennsylvania Press.

Goodwin, C. (2007). "Interactive footing." In E. Holt & R. Clift (Eds.), *Reporting talk: Reported speech in interaction* (pp. 16-46). Cambridge: Cambridge University Press.

Gumperz, J. J. & Cook-Gumperz, J. (1982). "Introduction: Language and the communication of social identity." *Language and Social Identity*, pp. 1-21.

Günthner, S. (1997). "The contextualization of affect in reported dialogues." In S. Niemeier & R. Dirven (Eds.), *The language of emotions. Conceptualization, expression, and theoretical foundation* (pp. 247-276). Amsterdam: Benjamins.

Günthner, S. (1999). "Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech." *Journal of Pragmatics, 31*(5), pp. 685-708.

Jansen, W., Gregory, M. L. and Brenier, J. M. (2001). "Prosodic correlates of directly reported speech: Evidence from conversational speech." In *ISCA*

Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding.

Klewitz, G. & Couper-Kuhlen, E. (1999). "Quote-unquote? The role of prosody in the contextualization of reported speech sequences." *Journal of Pragmatics, 9*(4), pp. 459-485.

Lipski, J. M. (1994). *Latin American Spanish*. London: Longman.

Lipski, J. M. (2008). *Varieties of Spanish in the United States*. Georgetown University Press.

Maier, E. (2014). "Japanese reported speech: Towards an account of perspective shift as mixed quotation." In E. McCready, K. Yabushita & K. Yoshimoto (Eds.), *Formal approaches to semantics and pragmatics: Japanese and beyond* (pp. 135-154). Netherlands: Springer.

Maldonado, C. (1999). "Discurso directo y discurso indirecto." En I. Bosque y V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (vol. 3), (pp. 3549-3596). Madrid: Espasa Calpe.

Navarro Tomás, T. (1944). *Manual de entonación española.* New York: Hispanic Institute in the United States (1974. 4th ed. Madrid: Ediciones Guadarrama).

Ochs, E. & Capps, L. (1996). "Narrating the self." *Annual Review of Anthropology, 25*(1), pp. 19-43.

Oliveira Jr, M. & Cunha, D. A. (2004). "Prosody as marker of direct reported speech boundary." In *Speech Prosody 2004*, International Conference.

Quilis, A. (1993) *Tratado de fonología y fonética españolas.* Madrid: Arco.

Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kuhlen, E., Günthner, S., Meier, C., Quasthoff, U., Schlobinski, P. & Uhmann, S. (1998). "Gesprächsanalytisches transkriptionssystem (GAT)." *Linguistische Berichte, 34*(173), pp. 91-122.

Tannen, D. (1986). "Introducing constructed dialogue in Greek and American conversational and literary narrative." *Direct and indirect speech, 3*, pp. 11-32.

Appendix – Transcription Conventions*

(.)	micro-pause
(-), (--), (---)	short, middle or long pauses of 0.25 - 0.75 seconds, up to ca. 1 second
:,::,:::	lengthening, according to its duration

* Transcription conventions are based on GAT, Gesprächsanalytisches Trankriptionssystem (Selting *et al.*, 1998).

ACcent	primary or main accent
↓	pitch step down on the following syllable
↑	pitch step up on the following syllable
<<f>>	forte, loud
<<all>>	allegro, fast
h, .hh, .hhh	breathing in, according to its duration
( )	unintelligible passage, according to its duration