

New Information in Naturalistic Data Is Also Signalled by Pitch Movement: An Analysis from Monolingual English/Spanish and Bilingual Spanish Speakers

Stefan Th. GRIES & Viola G. MIGLIO

University of California, USA
stgries@gmail.com
miglio@spanport.ucsb.edu

ABSTRACT

In communication, speakers and listeners need ways to highlight certain information and relegate other information to the background. They also need to keep track of what information they (think they) have already communicated to the listener, and of the listeners' (supposed) knowledge of topics and referents. This knowledge and its layout in the utterance is commonly referred to as information structure, i.e., the degree to which propositions and referents are given or new. All languages have 'chosen' different ways to encode such information structure, for instance by modifying the pitch or intensity of the vocal signal or the order of words in a sentence. In this study, we assess whether the use of pitch to signal new information holds in typologically different languages such as English and Spanish by analyzing three population group monolingual California English speakers, bilingual speakers of English and Spanish from California (Chicano Spanish), and monolingual Mexican Spanish speakers from Mexico City.

Our study goes beyond previous work in several respects. First, most current work is based on sentences just read or elicited in response to highly standardized and often somewhat artificial stimuli whose generalizability to more naturalistic settings may be questionable. We opted instead to use semi-directed interviews whose more naturalistic setting provides data with a higher degree of authenticity. Second, in order to deal with the resulting higher degree of noise in the data as well as the inherent multifactoriality of the data, we are using state-of-the-art statistical methods to explore our data, namely generalized linear mixed-effects modeling, to accommodate speaker- and lexically-specific variability. Despite the noisy data, we find that contour tones including H+L or L+H sequences signal new information, and that items encoding new information also exhibit proportionally longer stressed vowels, than those encoding given information. We also find cross-dialectal variation between monolingual Mexican Spanish speakers on the one hand and monolingual English speakers and Chicanos on the other: Mexican Spanish speakers modify pitch contours less than monolingual English speakers, whereas the English patterns affect even the Spanish pronunciation of early bilinguals. Our findings, therefore, corroborate Gussenhoven's theory (2002) that some aspects of intonation are shared cross-linguistically (longer vowel length & higher pitch for new info), whereas others are encoded language-specifically and vary even across dialects (pitch excursion & the packaging of information structure).

Keywords: information structure, intonation, pitch movement/contours, bilingual vs. monolingual speakers, mixed-effects modeling

INDEX: 1. Introduction. 1.1. General introduction. 1.2. Goals and characteristics of the present paper. 2. Methods. 2.1. Subjects. 2.2. Speech samples and their annotation. 2.3. Statistical evaluation. 3. Results. 3.1. Overall regression results and main effects. 3.2. Interaction effects. 4. Discussion and concluding remarks. 4.1. Acoustic correlates. 4.2. Interaction between information structure & dialectal/linguistic variation. 4.3. Gender and intonation. 4.4. Implications and where to go from here. References.

1. INTRODUCTION

It is probably uncontroversial to assume that communication between individuals is perhaps the primary means to which language is put to use. To that end, speakers and listeners require ways to refer to – as well as direct their attention to – propositions, referents, actions, state of affairs and ways to express their attitudes and emotions to towards all these. Directing attention to referents etc. also requires speakers to keep track of and manipulate knowledge/inferability of topics and attentional states, towards of propositions, referents, etc. This knowledge/attentional state, i.e. the degree to which propositions, referents, etc. are (assumed by the speaker to be) given or new or various intermediate states such as inferable. (e.g., not mentioned before directly but inferable from the discourse or the linguistic or non-linguistic context of the utterance) is commonly referred as information structure. All languages have ways to encode information structure, for instance by modifying the pitch or intensity of the vocal signal. Since some of the meanings attributed to these physical ways of encoding information are also shared by other animals' means of communication, some correlates are deduced to be universal: For instance what Ohala (1983) and Gussenhoven (2002) call the frequency code, i.e. that higher pitch signals non-dominance because it correlates with smaller production organs, and vice versa threatening calls are usually delivered through the lowest possible frequencies (indicating bigger body size). The other two factors that can have a universal 'interpretation' are the effort code, i.e. more effort in the movement of the larynx can avoid undershooting of targets and therefore more muscular effort implies imparting importance to the information delivered; and the production code whereby physical correlates of intonation must interact with breathing, and therefore beginnings of utterances are usually more energetic than their ends because they must correspond with exhalation phases. These physical factors can however also be grammaticalized in human language, and this gives rise to an interaction between universal interpretations of these acoustic correlates, and language-specific ones.

Moreover, in human language, there may be different, purely grammatical ways of encoding such information for hearers (cf. Haviland & Clark 1974, Chafe 1976, Prince 1981), including lexical, morphological, syntactic, or word-ordering means (Féry 2007, Gussenhoven 2007), as well as the intonational means that are at the core of the present paper. Intonation, i.e., the non-lexical variation of spoken tones as manifested in a multi-layered complex of modulation of pitch (F0), intensity, and vowel duration, is manipulated by speakers and their speech communities, and, as

part of the phonological system of a language, can become grammaticalized in the sense of developing conventionalized (intonational-)form-function pairings (Gussenhoven 2002). One such function is concerned with signaling information structure. Crucially, the indicators of different information-structural states typically also have functions unrelated to the packaging of information in an utterance (Féry 2007:162), such as the marking of sociolinguistically relevant information (Warren & Daly 2000, Daly & Warren 2001; Clopper & Smiljanic 2011) and are subject to constraints imposed on them by their physical expression (cf., e.g., Ohala 1983, Cruttenden 1997, Gussenhoven 2007). This means that the different functions of information-structural devices and their encoding give rise to potentially complex interactions with other grammatical or semantic components and their physical correlates, whose interpretation is language-/dialect-/variety-specific (cf. Gussenhoven 2002, 2007; Arvaniti & Garding 2007): different varieties can attribute a different semantic interpretation to the same tunes (sequences of H(igh) and L(ow) pitch on the different syllables of the intonation units) regardless of information-structural packaging. For instance, in peninsular Spanish, Italian, and in English, a neutral statement ends in a L tone indicating the finality of the utterance and hence the boundary of the intonation unit (Ladd 1996, Martínez Celdrán & Fernández Planas. 2003:185, D'Imperio et al. 2005), but in Mexican Spanish a neutral statement is more likely to end in a circumflex tone (Butragueño 2004).

1.1. GENERAL INTRODUCTION

A further complexity is caused by the scope of information structure, which is necessarily laid out in a multi-word, multi-sentence domain, since the speaker and listener can only keep track of whether an item belongs to given or new information over a textual chunk comprised of at least several utterances. This means that processing information from a read text or spoken discourse requires a complex cognitive engagement of memory and domain-general attention strategies in order to extract meaning from a string of separate word units while they are being assembled into larger, multi-word constituents, and while the listener keeps track of the most important components of the conversation. This complex effort in processing linguistic information has been termed unification, an activity central to the language faculty recruiting frontal lobe structures, such as the left inferior frontal gyrus (Hagoort 2005).

Despite the complex interactions to which the different language-specific, grammatical and pragmatic functions of intonation give rise, some studies have underlined the cognitive importance of the prosodic markings of information structure for sentence processing (Cowles et al. 2007, Wang et al. 2009, van Leeuwen et al. 2014). The above-mentioned process of unification has been shown to be particularly sensitive to the encoding of information structure: ERP studies showed for instance

that an N400 effect is obtained when an unexpected word is found in a reading task after a focusing device, such as clefting in English (Cowles et al. 2007, Wang et al. 2009), but also in studies of auditory processing of language (van Leeuwen et al. 2014:65 and references cited therein). These neurolinguistic studies show that whatever is mentioned in previous discourse/textual context creates expectations as to the information structural status of a specific item that follows, and that more processing resources (as measured through ERPs) are required if there is a mismatch between the salience of the information presented and the expected way in which it is supposed to be encoded through pitch manipulation (van Leeuwen et al. 2014).

There is a considerable amount of literature on intonation in English (starting from Pierrehumbert 1980, Pierrehumbert & Hirschberg 1990, Ladd 1996, Gussenhoven 2004, Arvaniti & Girding 2007 and literature cited therein), but most of it concerns the meaning of intonational tunes and the realization of different types of focalizations; less has been published on acoustic correlates of information structural packaging. Moreover, there are different layers of emphasis that can be influenced by pitch: words both in English and Spanish have lexical stress, but an added level of prominence is provided by phrasal emphasis, or 'pitch accents,' i.e. pitch modifications on the phrasal or intonational unit level (Ladd 1996, Gussenhoven 2004). While pitch accents may be used as focusing devices or to mark prosodic boundaries, we are only interested in their information-structural use. Baumann (2005) found that, while a H pitch accent correlates with new information and deaccentuation (L) with given information (as in Pierrehumbert & Hirschberg 1990), neither completely new nor completely given items, a status which arguably covers most items in discourse, were marked by contour tones, i.e. sequences of H+L pitch movements¹. In this paper, therefore, we focus on the role that is played by these ever so common contour tones and we provide further evidence for the use of pitch movement as an acoustic correlate of information structure in spoken discourse. We address typological questions by focusing on the distinctions between two languages that are supposed to privilege different means to set off new information from given information: English privileges pitch changes (Reinhart 1981, Pierrehumbert and Hirschberg 1990, Cruttenden 1997), whereas Spanish is supposed to prefer the manipulation of syntactic structures and word order (Zubizarreta 1998, Zubizarreta & Nava 2011). We also analyzed the speech of a group of Spanish-English early bilinguals speaking Spanish in order to assess the effects of bilingualism on the encoding of information structure in the Spanish of these speakers.

2.1. GOALS AND CHARACTERISTICS OF THE PRESENT PAPER

The present paper has two main goals. First and as already mentioned briefly above, we are exploring (i) how speakers of languages that are known to mark information structure differently: monolingual English (argued to use pitch movement) and

monolingual Spanish (argued to use syntax and constituent order) and (ii) how bilingual speakers' Spanish compares to the monolingual speakers' (lack of) use of pitch movement. That is, we focus on how intonation – i.e. the suprasegmental melody of language and its acoustic correlates – affects the encoding of the information-structural status of items in discourse. Specifically, the main hypotheses we explore are the following:

1. New information is generally signaled by pitch movement on the relevant word;
2. monolingual English speakers use pitch excursion more than monolingual Spanish speakers to signal new information (the latter may not do it at all);
3. balanced early bilingual speakers speaking Spanish may be influenced by English and use pitch excursion to signal new information more than their monolingual Spanish counterparts.

Second, as mentioned briefly above, we are also trying to advance the study of intonational correlates of information structure in two methodological ways: (i) by using much more naturalistic data than most prior work has, and (ii) by using more statistically sophisticated methods than has been customary in this area of research. With regard to these two methodological goals, it is necessary to bear in mind that a vast majority of studies in this area use constructed stimuli or passages, typically in reading or auditory tasks. Specifically, information-structural states can be simulated and/or targeted with manipulations of pitch and syntactic structure (this has been done in many existing studies on various languages)², as exemplified also by Daly & Warren (2001:88) or Röhr & Baumann (2010). However, such experimental designs expose speakers to overall unrepresentative stimuli – unrepresentative in the sense that the range of stimuli/situations that speakers/subjects are exposed to are by design (i) limited in various ways compared to the richness of naturalistic situations and (ii) characterized by (typically balanced) probability distributions that do not represent the typically skewed and Zipfian distributions of natural data.

Given these considerations, we decided to use a corpus of semi-directed interviews of different dialects of English and Spanish collected by the Phonetics Lab of the Spanish and Portuguese Department at the University of California, Santa Barbara (see Section 2 for details). A frequent counterargument to the use of (such more) naturalistic speech is that it is supposed to provide less robust data sets (Butragueño 2004, Clopper & Smiljanic 2011). However, not only can the same be true of the supposedly less noisy experimental conditions – see for instance, the constructed sentences read by Röhr & Baumann's participants, which produced rather noisy data (2010:4) – but we are using statistical methods that are well-suited to handle the kinds of interrelated and potentially noisy data that arise from (more) naturalistic samples. This in turn allows us to work with speech samples that are more attuned to regular language use (again, see Section 2 for details) as well as cover data from a larger

variety of speakers than some previous work has been able to include (see, e.g., Sluijter & van Heuven 1996's analysis of a mere six speakers).

The remainder of the paper is structured as follows: Section 2 explains how our data were gathered, the characteristics of the participants, and how the data were annotated and analyzed both acoustically and statistically. Section 3 explains the results obtained by our statistical multifactorial analysis of the factors correlated with pitch movement. In the final section, Section 4, we discuss the results, the conclusions, and future research developments.

2. METHODS

In this section, we discuss how our data were gathered, prepared for analysis, and then analyzed using corpus-linguistic and statistical tools. Specifically, Section 2.1 outlines how the materials for analysis were gathered, Section 2.2 outlines the type of speech samples obtained, and Section 2.3 the statistical analysis with which we explore them.

2.1. SUBJECTS

Data from three different groups of subjects were culled for the present study. The three groups were all composed of 10 subjects each³; all subjects were students at a major university of the area where they were interviewed; five female and five male participants were interviewed per group, with age ranges between 20 and 25, and similar linguistic, socio-economic, cultural, and ethnic background within each group. They were asked to provide a minimum of biographical data that would ensure the correct ascription of the speaker to the relevant group, while maintaining the anonymous character of the data gathered in the interviews. Such biographical data allowed the researchers to establish whether the students were monolingual or bilingual (language spoken at home, languages spoken by the parents/caregivers, place of birth and number of years of residence in California or Mexico City respectively). The speakers were split into three rather homogenous groups: a group of monolingual Southern California English speakers and one of monolingual Spanish speakers, who had never resided abroad for a period of more than 4 weeks, and were born and raised either in Southern California, or in Mexico City by monolingual parents/caregivers of English and Spanish, respectively. The third group was also composed of undergraduate university students born and raised in Southern California, but raised in Spanish-speaking households and encountering English either since birth because both Spanish and English were spoken in the household, or as soon as they entered the US school system, in any case before their 8 years of age. Bilingual subjects were fluent both in English and Spanish at the time of recording.

Recordings of monolingual English and bilingual Spanish-English speakers from Southern California were made in the phonetics lab of the Spanish and Portuguese

Dept. at UCSB, using a Gretch-Ken Industries professional sound booth (anechoic chamber with NIC rating of 34) with a Shure SM86 condenser vocal microphone, connected directly to an M-Audio Fast Track Pro interface, feeding into a computer with the program Audacity (<http://audacity.sourceforge.net>). Recordings in Mexico City were carried out in a silent room at U.N.A.M. university's main campus in Mexico City using a portable MacBook computer, and an M-Audio Microtrack 24/96 professional digital recorder with a dual electret microphone, and GarageBand software.

The participants reported having no known hearing or speech impediments, and they were all asked in writing to agree to the recordings according to Human Subjects handling both in the U.S.A. and abroad. Participation was entirely voluntary and unpaid. No distinction was detected in the extent to which individual participants or the different groups engaged in the tasks they were requested to perform.

2.2. SPEECH SAMPLES AND THEIR ANNOTATION

Speech samples culled from participating speakers were obtained with semi-directed interviews lasting between 10 and 20 minutes each. The participants were given free rein to tell anecdotes after receiving the same set of prompts. These included items such as 'Tell me about the scariest moment of your life,' 'Tell me what you remember about the first day of school/university,' 'Tell me the plot of your favorite movie,' 'Who was your favorite teacher in high school and why?,' 'What was your favorite subject in high school and why?' etc. This allowed participants to speak in fully fledged utterances without interruptions – unless these were self-imposed pauses – in as naturalistic a way possible according to their own speech patterns and rhythms while being recorded.

Although information status can be broken down into a more complex hierarchy than just the distinction of new vs. given, to simplify matters and to make sure we obtained a sufficient number of tokens from the naturalistic interviews, in this study, we applied only this binary distinction to nouns in declarative sentences – questions and other types of syntactic frames where pitch could be used for different semantic purposes were excluded from the sample (e.g. narrow or contrastive focus). The speech samples were analyzed manually using PRAAT software (Boersma & Weenink 2014), pitch was normalized visually where the program provided spurious values due to creakiness, or excluded where creakiness impeded measurements. The resulting 1043 data points were then annotated with regard to the following variables, which had proven useful in a pilot study (Miglio, Gries, & Harris 2014):

- PITCHMOVEMENT, the binary dependent variable: no (the annotated word exhibits no pitch movement/excursion over the word through the rater's visual and aural perception) vs. yes (the annotated word exhibits pitch movement/excursion);

- **SPEAKERTYPE:** monoengl (for utterances by monolingual English speakers) vs. monospan (for utterances by monolingual Spanish speakers) vs. bispan (for Spanish utterances by bilingual speakers of Spanish and English);
- **GIVENNESS:** no (the referent of the word whose pitch movement was annotated was mentioned in the discourse for the first time) vs. yes (the referent of the word whose pitch movement was annotated was mentioned before in the discourse);
- **PHRASEFINALITY:** no (the annotated syllable is not in a phrase-final position) vs. yes (the annotated syllable is in a phrase-final position);
- **SEX:** the sex of the speaker, female vs. male;
- **DURATION:** the natural log of the duration of the stressed vowel in milliseconds;
- **INTENSITY:** the average intensity of the stressed vowel in decibels.

In addition, we also noted the specific speaker from whose speech the token was sampled as well as the specific word whose PitchMovement level was studied in order to include those as random effects in the regression model.

2.3. STATISTICAL EVALUATION

We then explored the degree to which the above predictors can predict whether speakers will employ pitch movement in their utterances by using generalized linear mixed-effects modeling (GLMEM). This kind of model has several attractive characteristics for the present study. First, it allows the researcher to study several predictors' effects as well as their interactions at the same time. That is to say, one avoids the potential risk of monofactorial studies – studies in which only one predictor is studied at a time – namely that (i) the studied predictor may be significant but only because it is correlated with another one or (ii) the studied predictor might not have the same (significant) effect in all parts of the data (e.g., GIVENNESS may not have the same effect on PITCHMOVEMENT for all speaker types).

A second big advantage is that this kind of modeling approach allows to ensure that statistical assumptions of standard regression modeling are not violated. In our data, as in most linguistic data sets in fact, every speaker contributes more than one data point, which means that the assumption that all data points are completely independent of each other is violated. The GLMEM approach, on the other hand, allows us to include in the analysis individual speakers' preferences to (not) use pitch movement in the analysis, as well as account for the possibility that particular lexical items are more likely to come with a (dis)preference for pitch movement.

We undertook a model selection process in which we first fit a regression model that in Miglio, Gries, & Harris's (2014) pilot study proved useful to distinguish uses of pitch movement in a part of the present data. In that model, we modeled PITCHMOVEMENT as a function of SPEAKERTYPE, GIVENNESS,

PHRASEFINALITY, DURATION, and the interaction SPEAKERTYPE:GIVENNESS, with varying intercepts for both speakers and lexical items. We then considered adding potential two-way interactions of fixed effects (using an exploratory significance level of $p=0.1$) and varying slopes for GIVENNESS to the regression model to achieve the best possible model fit while at the same time following Occam's razor.

3. RESULTS

3.1. OVERALL RESULTS AND MAIN EFFECTS

In this section, we discuss the results of the model selection process. That process was concluded quickly because only one additional predictor had to be added to the final model of Miglio, Gries, & Harris (2014), the interaction GIVENNESS:PHRASEFINALITY – no other fixed-effect predictors nor the varying slopes for GIVENNESS improved the model significantly.

The final model makes for an intermediately good fit: it achieves a classification accuracy of 72.1%, which, compared to the baseline of correct random choices of 51.2% is highly significantly better (pbinomial test<10-40); this degree of accuracy yielded a C-value of 0.78. However, the extreme variability of observational data also results in comparatively low amounts of explained variability: R2marginal (i.e. the 'correlation coefficient' that quantifies the effect of the fixed-effect predictors) is a mere 0.164; R2conditional (i.e. the 'correlation coefficient' that quantifies the effect of both fixed-effect predictors and the random effects) is also just 0.249. Overdispersion and collinearity did not pose any problems: poverdispersion>0.98 and all VIF<2.85.

Predictor	<i>b</i>	<i>se</i>	<i>z</i>	<i>p</i>
Intercept	-3.905	0.9691	-4.03	<0.0001
SPEAKERTYPE: <i>bilingual</i> vs. <i>monolingual</i>	0.0334	0.1937	0.172	0.8632
SPEAKERTYPE: <i>monoengl</i> vs. <i>monospan</i>	0.9636	0.3444	2.798	0.0051
GIVENNESS: <i>given</i> → <i>new</i>	1.1714	0.1955	5.993	<0.0001
PHRASEFINALITY: <i>no</i> → <i>yes</i>	1.1539	0.2204	5.236	<0.0001
SEX: <i>female</i> → <i>male</i>	-0.5089	0.2193	-2.321	0.0203
Duration (logged)	0.6269	0.2122	2.955	0.0031
SPEAKERTYPE (<i>bilingual</i> vs. <i>monolingual</i>) : GIVENNESS (<i>given</i> → <i>new</i>)	-0.3082	0.2045	-1.507	0.1318
SPEAKERTYPE (<i>monoengl</i> vs. <i>monospan</i>) : GIVENNESS (<i>given</i> → <i>new</i>)	-0.7686	0.3485	-2.205	0.0274
GIVENNESS: (<i>given</i> → <i>new</i>) : PHRASEFINALITY: <i>no</i> → <i>yes</i>	-0.5042	0.2928	-1.722	0.085

Table 1. Coefficients of the final mixed-effects regression model

As Figure 1 indicates, there is a clear effect such that the longer the duration of the stressed vowel of the word analyzed, the higher the predicted probability of pitch movement, an effect that is attested across all speaker types and givenness levels.

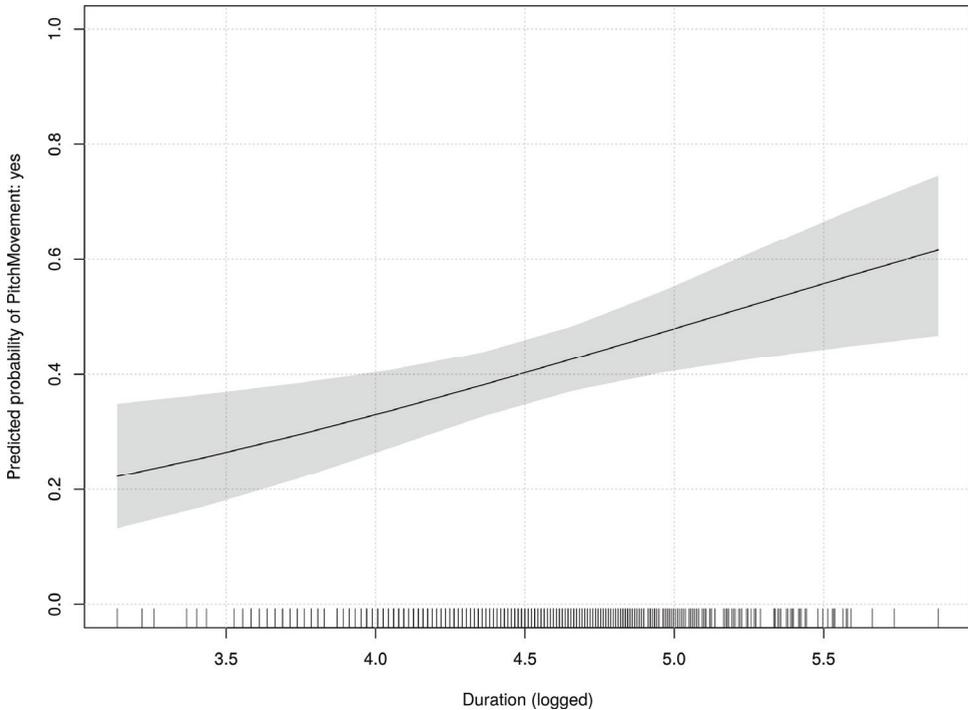


Figure 1. The effect of DURATION (logged) on the predicted probability of pitch movement (regression line with 95%-confidence band)

Figure 2 shows that female speakers make more use of pitch movement than men, again regardless of speaker types and givenness levels.

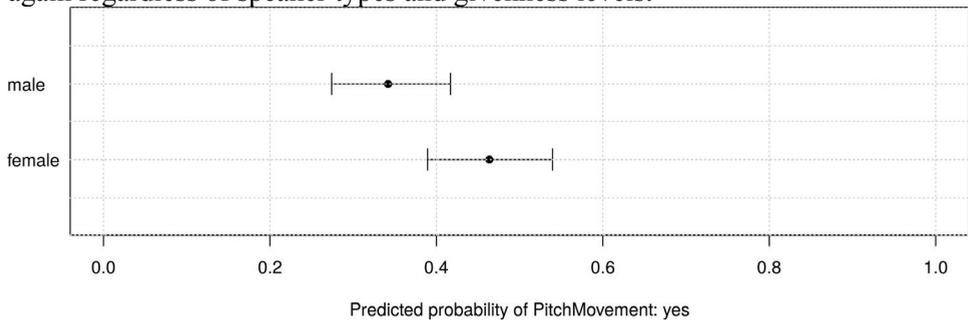


Figure 2. The effect of SEX on the predicted probability of pitch movement (with 95%-confidence intervals)

Figure 3 reflects the overall effect that GIVENNESS has on pitch movement – new information is more marked with pitch movement than given information – but this effect is qualified in an interaction, which is why we revisit it again below.

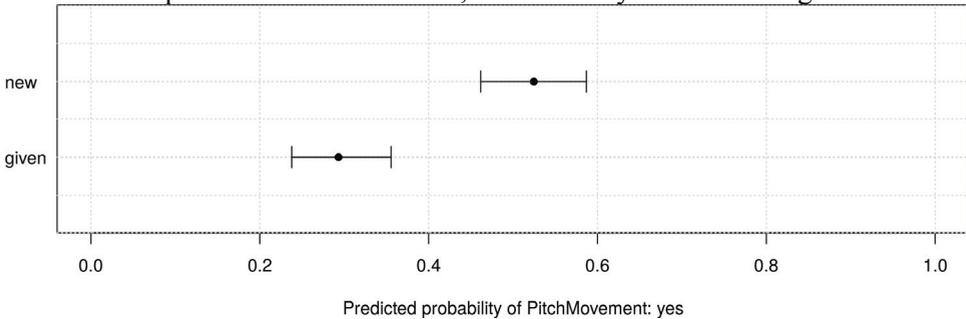


Figure 3. The effect of GIVENNESS on the predicted probability of pitch movement (with 95%-confidence intervals)

A similar situation arises with the effect of PHRASEFINALITY in Figure 4: Its overall effect is that utterance-final phrases exhibit more pitch movement than non-final ones, but PHRASEFINALITY participates in a significant interaction with GIVENNESS and will thus be analyzed in more detail below.

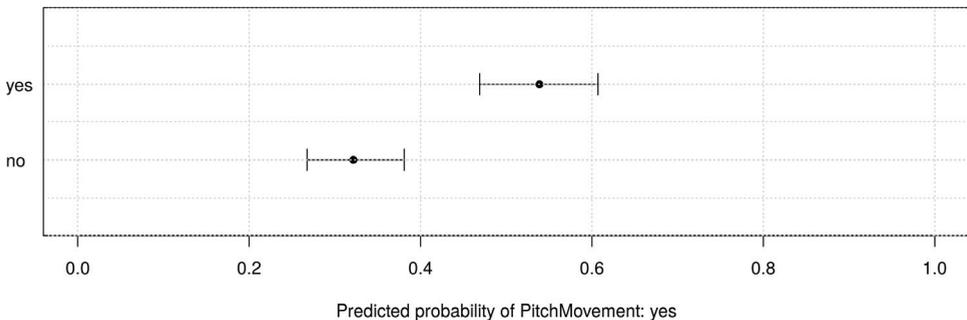


Figure 4. The effect of PHRASEFINALITY on the predicted probability of pitch movement (with 95%-confidence intervals)

The final main effect to be discussed briefly is that of SPEAKERTYPE in Figure 5: As the planned contrasts in Table 1 indicate, the main findings are (i) that the speaker types form a cline from monolingual English speakers via bilingual

Spanish/English speakers to monolingual Spanish speakers and (ii) that the bilingual speakers do not differ from the two kinds of monolingual speakers combined, but the monolingual English speakers use pitch movement significantly less than the monolingual Spanish speakers. However, this effect, too, will have to be revisited given the significant interaction with GIVENNESS that it participates in.

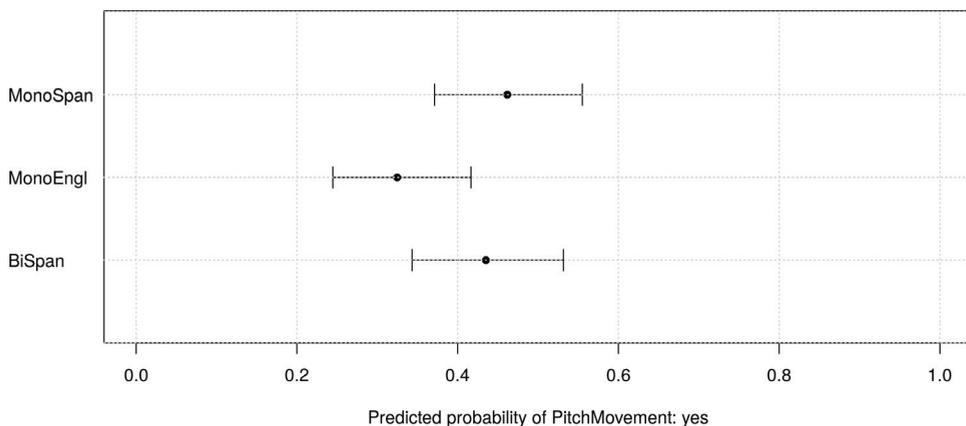


Figure 5. The effect of *SPEAKERTYPE* on the predicted probability of pitch movement (with 95%-confidence intervals)

3.2. INTERACTION EFFECTS

In addition to the above main effects, we also obtained two significant two-way interactions in the data, which qualify three of the above main-effects results.

Figure 6 represents the first of these two relevant interactions to be discussed here: *SPEAKERTYPE:GIVENNESS*, which qualifies the main effects of the predictors involved in it; both panels show the same results but perspectivized differently. While we saw above how *GIVENNESS* (given → new) results in a strong overall increase in the probability of pitch movement, we now see that this effect is different for the different speaker types. The left panel shows clearly that, for monolingual Spanish speakers at the top, the contrast of *GIVENNESS* has the least effect (resulting in a just about significant but still small adjusted pitch-movement probability difference of 10.7%). However, for both the monolingual and the bilingual speakers of English, the difference that *GIVENNESS* makes is much more pronounced: For the monolingual English speakers in the middle, *GIVENNESS* results in a highly significant pitch-movement probability difference of nearly 26%; for the bilingual Spanish speakers at the bottom, the difference is an even greater (and more significant) difference of 30.5%.

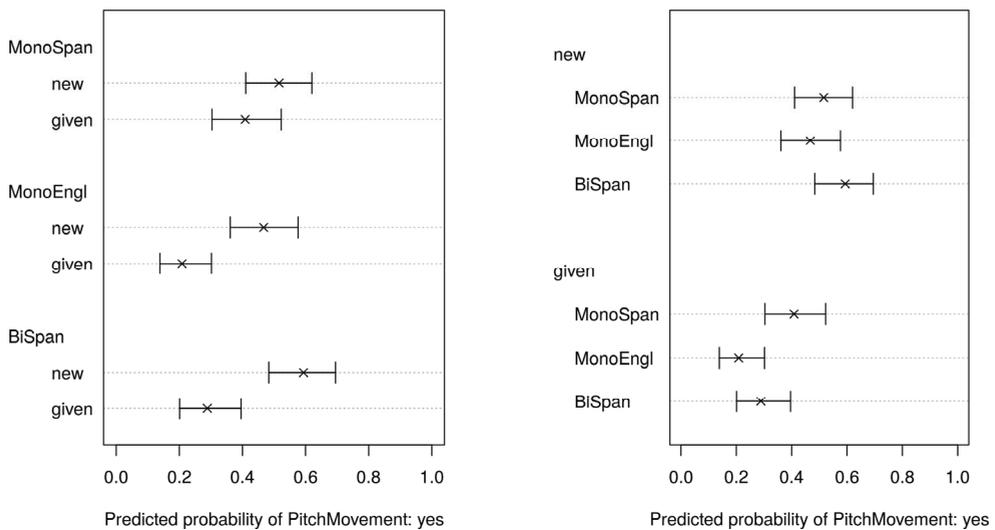


Figure 6. The effect of *SPEAKERTYPE:GIVENNESS* on the predicted probability of pitch movement (with 95%-confidence intervals)

The right panel, on the other hand, makes it very obvious that the above results for *SPEAKERTYPE* shown in Figure 5 still hold, but only for when the information embodied by the word analyzed is given: at the bottom of the right panel we again find the cline from monolingual English speakers via bilingual Spanish/English speakers to monolingual Spanish speakers – but the top of the right panel shows that, with new information, the bilingual speakers now do not fall between the two monolingual speaker groups anymore because of the big difference in pitch movement in response to *GIVENNESS*.

Finally, Figure 7 represents the final predictor, the interaction *PHRASEFINALITY:GIVENNESS*. Again, we saw above how both *GIVENNESS* (given → new) and *PHRASEFINALITY* (no → yes) result in a strong overall increase in the probability of pitch movement, but now we also find that neither effect is equally strong but connected to where the word under scrutiny occurs (phrase-finally or non-phrase-finally) or to whether the referent of the word in question is given or new: When the word being analyzed is phrase-final, pitch movement is more likely overall, but the different levels of *GIVENNESS* make a very significant but smaller difference (16.6%); however, when the word is not phrase-final, pitch movement is less likely overall, but the different levels of *GIVENNESS* make a highly significant much larger difference (25.2%).

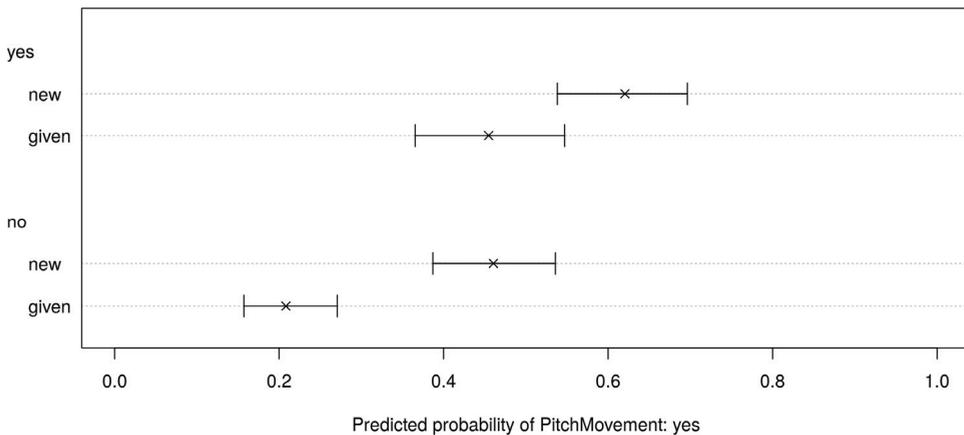


Figure 7. The effect of *PHRASEFINALITY:GIVENNESS* on the predicted probability of pitch movement (with 95%-confidence intervals)

As a result of the analysis, we also obtained the regression model's adjustments for the lexical items whose pitch levels we measured as well as for all the speakers in our data. Space does not permit a more systematic exploration of these, but it is instructive to note that the adjustments made for the speaker are larger than those for the lexical items, which makes sense given that one would not expect words to have default pitch movement characteristics associated with them whereas it is easily conceivable that speakers differ more consistently in their use of pitch excursion. In addition, it is this aspect of the model that allows us to model each speaker's baseline tendency to use pitch movement separately and, therefore, get results for all other predictors that are not tainted by the fact that all data points of a speaker may be characterized by idiosyncrasies.

On a final and more methodological note, in addition to the mixed-effects model discussed above, we also fit a standard binary logistic regression (BLR) model in order to compare both the classification accuracies and the coefficients of the model predictors. Figure 8 reveals the dangers of not using the right kind of regression modeling. All coefficients of both models are shown on the x-axis and the percentage to which the BLR model misestimates the coefficients is shown on the y-axis. On average, the coefficients of the binary logistic regression are off by 12.4%, but one coefficient – for one contrast of the main effect of *SPEAKERTYPE* – is off by more than 50%. Correspondingly, the overall classification accuracy of the BLR is about 10% worse than that of the more sophisticated mixed-effects model, which would also generalize better when applied to new speakers' data. Finally, the standard BLR approach would also suggest to include another predictor in the model, *SPEAKERTYPE:DURATION*, whereas the mixed-effects model recognizes that this effect is better considered to consist of speaker idiosyncrasies rather than what

speakers of the different types share. Thus, the mixed-effects modeling approach not only makes the predictors it flags as significant more precise and generalizable, it also protects researchers against falsely accepting effects as significant.

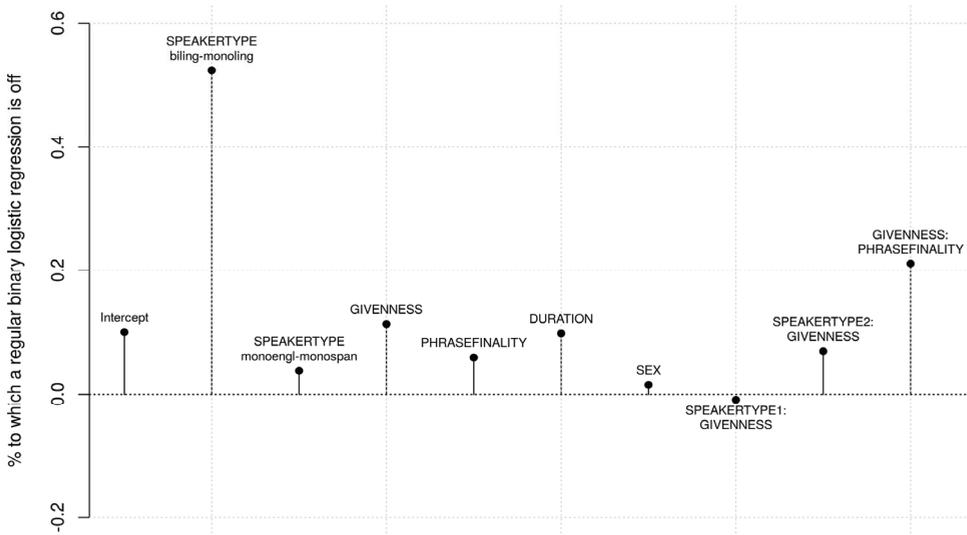


Figure 8. The difference between a generalized linear mixed-effects model and a traditional binary logistic regression

4. DISCUSSION AND CONCLUDING REMARKS

Given the results laid out in the previous section, the discussion will focus on three different aspects of the analysis: one related to the acoustic correlates of intonation in marking information structure (Section 4.1), one on the interaction between dialectal/linguistic variation and the encoding of information structure (Section 4.2), and finally one on gender and intonation (Section 4.3).

4.1. ACOUSTIC CORRELATES

In our data, as manifested Figure 1, we found a clear correlation between stressed vowel duration and pitch movement, i.e. the predicted probability that there would be a raising or lowering of pitch on longer vowels, across speaker types and regardless of information structure. This result confirms that segments with a particular structure, in this case longer vowels, are more likely targets for contour tones, i.e. complex tones made up of two separate targets, either a H+L sequence or L+H sequence. This is unsurprising, since from both an articulatory point of view and a perceptual point of view, longer duration provides more time to produce separate movements of the

muscles and cartilages of the larynx in order to modify F0, as well as more time to perceive them as separate targets. Thus, this finding confirms what previous literature has remarked about the feasibility of certain segments to bear tones; especially for complex tones such as those exhibiting more than one target (i.e. those with pitch movement), longer vowels are better suited than shorter ones, since "contour tone bearing ability is [...] crucially dependent on duration" (Zhang 2001:33).

One of our initial hypotheses was that new information would be marked by a complex pitch movement (based on the high frequency of H*+L sequences found in Baumann's (2005) study). As mentioned in the introduction, while English tends to use pitch modification to signal new information (Cruttenden 1997, Vallduví 1992), Spanish is supposed to use its more flexible word order for the same purpose (Suñer 1982, Zubizarreta 1998, Zubizarreta & Nava 2011 and literature cited therein). However, as we can already see from Figure 3, which embodies a main effect found in the overall data, we noticed that in naturalistic speech a contour tone in fact is more likely to mark new information than given/old information for all three speaker groups, i.e. regardless of language or dialect spoken. This is remarkable in and of itself, since at least some of the Romance languages, the ones that Vallduví terms as 'non-plastic' (1992) such as Spanish and Italian, are supposed to manipulate word order rather than use pitch modulations for this purpose. Yet even the final main effect (of SPEAKERTYPE) seems to contradict the predictions found in the literature⁴, since monolingual Mexican Spanish speakers are shown to be most likely to use pitch movement, compared to bilingual and monolingual English speakers.

As mentioned in the methodology section, we controlled for pitch accents unrelated to information structure by eliminating utterances with narrow and contrastive focus from the data. Another area where pitch accents are likely to appear unrelated to information structure is at the end of the utterance, since they can mark boundary tones in different languages such as English and German (Baumann 2005:3). In fact, we do find an interaction between the prosodic packaging of information structure in our data and phrase-final position, as shown in Figure 7. When the word analyzed is in phrase-final position, it is more likely to show pitch movement overall, and the fact that this did not also interact with SPEAKERTYPE shows that this effect is found for all three speaker groups, i.e. regardless of language spoken. This is, in a sense, not surprising, given that both in English and Spanish the main pitch accent in an utterance (also called phrasal or nuclear stress) can fall on the rightmost content word, i.e. towards the right boundary of the sentence; in Spanish this is strictly enforced, whereas English has more positions where nuclear stress can fall (Zubizarreta & Nava 2011:652). However, as mentioned in the Section 3 above, what is important here is that there is an interaction with the expression of information structure: We see in fact that pitch movement is a much more important resource to mark information as new (as opposed to given) when the word is not in utterance-final position, where the givenness difference amounts to a 25.2% difference of predicted probability of pitch movement, as opposed to the smaller

corresponding difference due to GIVENNESS (16.6%) when the word is in final position. Since pitch movement is also used across languages to mark boundary tones, this is an important finding because our naturalistic data show that pitch movement is a discriminating factor especially when the word is not found in utterance-final position. This also confirms that our study corroborates hypotheses that different acoustic correlates of intonation are used for different purposes in language (Gussenhoven 2002, Gordon and Nafi 2012): we show that pitch movement is a correlate of information-structural marking at least across the languages and varieties we analyzed here.

4.2. INTERACTION BETWEEN INFORMATION STRUCTURE AND DIALECTAL/LINGUISTIC VARIATION

As we have seen above, the role of GIVENNESS for PITCHMOVEMENT is qualified by interactions, as shown for instance in Figure 6. The left panel specifically corroborates the existing literature in showing that while monolingual Spanish speakers do use pitch movement, they exhibit the smallest difference in the likelihood of using pitch movement to distinguish given and new information in comparison with the other two groups. For monolingual Spanish speakers, in fact, that difference only accounts for approximately 11% predicted probability, while for the other two groups (English monolinguals and English-Spanish bilinguals), the distinction made through pitch movement accounts for 26% and 30.5% respectively. This is compatible with the view of Spanish – as embodied by the monolingual population – as a language that has other mechanisms at its disposal, such as word order, to foreground new information and therefore uses pitch sparingly for this purpose. English, on the other hand, has a more fixed word order, and therefore uses pitch movement more in order to distinguish between new and given information in spoken discourse. The right-hand panel in Figure 6 is also revealing in so far as it still clearly shows the main effect whereby new information is characterized by pitch movement for all SPEAKERTYPE groups (top part of right panel); moreover, for given information (bottom part of right panel) we find the bilinguals neatly nested between the monolingual English and the monolingual Spanish speakers as expected in our third hypothesis above – showing that English bilingualism does influence Chicano linguistic behavior even when they are speaking Spanish. However, in using pitch movement to encode new information, the bilingual group is no longer wedged between monolingual Spanish and monolingual English groups (top part of right panel), but goes 'over the top' in exploiting intonation for information structural purposes, showing that Chicanos use pitch movement more than either monolingual group in signaling new information. This shows that intonational packaging of information structure is a language-specific area of linguistic behavior, and one that is not easily mastered natively even by early bilinguals. Such difficulties are also corroborated by L2 studies: Zubizarreta & Nava (2011:667), for instance, find that in

grammatically comparable contexts⁵, native speakers of Spanish find it hard to acquire English pitch modulations that encode information structure in broad focus contexts, but not necessarily those that encode contrastive focus.

Although Chicano Spanish is a peculiar variety of Spanish in the sense that it is a contact variety spoken by early bilinguals, our findings confirm that research cannot avoid distinguishing among different dialectal varieties, especially for languages that have many millions of speakers scattered across vast surfaces of the globe, such as English and Spanish. This is the kind of criticism that Arvaniti & Garding (2007:5) level at many intonation studies, namely that a language such as English is "treated as a homogenous language when in fact in most cases the research involved speakers of quite distinct varieties."⁶ The same could certainly be said for Spanish (cf. Prieto & Roseano's 2010 careful distinction of different varieties of Spanish), where research exploring the acoustic correlates of intonation is generally relatively scarce; as far as we know, in fact, no study has been published on acoustic correlates of intonation and information structure in Spanish before this one. However, what has been published on general intonation in different dialects of Spanish points to considerable distinctions in the semantic interpretation of prosodic cues depending on the dialect analyzed (Butragueño 2004, and the articles collected in Prieto & Roseano 2010). We do not wish to maintain, therefore, that Mexican Spanish is representative of the use of intonation in encoding information structure for all or even just for any other non-contact variety of Spanish, such as, say, Iberian Spanish. We chose Mexican Spanish because the bilingual Chicano speakers from California are most likely to speak a dialect of Spanish closely related to Mexican Spanish (Parodi 2011).

4.3. GENDER AND INTONATION

Finally, our naturalistic speech also provides new data to corroborate previous findings in the literature related to gender and pitch movement, as shown in Figure 2. Females, regardless of language spoken and of information structure packaging, are always more likely than males to use pitch modulations. This finding is in tune with Daly & Warren's findings (2001:92, also Warren and Daly 2000) that women use more dynamic pitch than males in their New Zealand English study. They found that this was especially true of their story-telling task⁷ (rather than the read sentences), and this may well be why we also find it in our data, since semi-directed interview can be considered akin to a story-telling task, where participants relate anecdotes from their past. There are still relatively few studies on acoustic correlates of discourse and gender identity (see Clopper & Smiljanic 2011:238 and literature cited therein) that go beyond evolutionary observations (Ohala 1983, Gussenhoven 2002). Our findings corroborate what has often been considered a stereotype, which has nonetheless been hard to substantiate with actual data: i.e. that women's speech exhibits more swooping pitch changes, a truism no doubt related to women expressing their emotions more

patently than men. Some early studies had failed to produce actual data proving that there was any truth in the stereotype (Henton 1989, 1995), whereas Daly & Warren (2001:85) did find more pitch dynamism in (New Zealand) female speech compared to the prosodically 'flatter' speech of men, and the experiments discussed by Gussenhoven (2002, section 3.1) also show that there is some widespread expectation of wider pitch ranges in female than in male speech. Whatever the sociolinguistic interpretation of a more dynamic pitch use in female speech may be, we do find that women are more likely to use contour tones, i.e. pitch movement, regardless of language variety or information structural concerns. Our study, therefore, also contributes new data for the study of prosody as a sociolinguistic marker of gender identity. Impressionistically, however, we can say that from the simple coding of the data, many women are simply more engaged story-tellers, carefully evaluating the character of the information they communicate and imparting the value they themselves attribute to it using pitch dynamism as a performative device to alert the listener of the importance of the various parts of the utterance. This evaluation of female pitch dynamism, while admittedly impressionistic, seems, however, to adjust well to some findings discussed by Gussenhoven for a Bantu language (2002, section 3.1), where a compressed pitch range indicates withdrawal of information.

4.4. IMPLICATIONS AND WHERE TO GO FROM HERE

There is still a lot to be done in researching intonation: both regarding acoustic correlates of information structure, and regarding the interpretation of different tunes in various languages and dialects, particularly in Spanish. What has been published on general intonation in different dialects of Spanish points to considerable distinctions in the semantic interpretation of prosodic cues depending on the dialect analyzed (Butragueño 2004, and the articles collected in Prieto & Roseano 2010), which makes instrumental studies of the prosodic characteristics of different Spanish dialects such as ours all the more timely.

With the use of sophisticated statistical modeling such as that used in this paper, it is possible to use naturalistic data, rather than ad-hoc read sentences or artificial stimuli in order to study intonation and its various acoustic correlates. This type of study has wide-reaching implications not only for phonetics and phonology, but also for the effects of language dominance and education on the speech of early bilinguals, on the sociolinguistic analysis of gender identity, on different textual & discourse registers, as well as performativity in language.

The study of bilinguals in this paper yielded interesting and unexpected conclusions as to the use of pitch movement in the Spanish of Chicano speakers: an analysis of their English is what we will carry out next to compare their use of intonation between both languages and compare their English intonation to their monolingual counterparts.

The analysis of further dialects of Spanish in relation to information structural packaging also promises to yield interesting results and we have a corpus of central Iberian Spanish semi-directed interviews that we intend to analyze for this purpose.

Finally, a further study of correlates of stress and intonation, such as pitch range, intensity, and duration in different dialects of English and Spanish would certainly provide much needed materials and analysis to improve our understanding of universal and language-specific phonetic features overall.

NOTES

1. Pitch contours or 'tunes,' i.e. sequences of H+L, L+H relative pitch frequencies, are referred to in this paper as 'pitch movement.' Flat tones, H or L, are sometimes subsumed under the 'lack of pitch movement.'
2. Dutch: van Leeuwen (2014); Dutch and Italian: Swerts et al. (2002); German: Röhr & Baumann (2005); for intonation in Spanish see Butragueño 2004, Herrera & Butragueño (2003), and Prieto & Roseano (2010).
3. One monolingual English student had to be excluded because of technical problems with the recording.
4. However, some authors do talk about these tendencies in non-categorical ways: "Germanic languages and, to a lesser extent, Romance languages use pitch accents to mark focused parts of sentences" Gussenhoven (2002, section 3.3).
5. They explored the use of prosody in wide focus clauses in signalling the distinction between sentences that distinguish between topic and comment and eventive, topicless clauses in native Spanish speakers learning English as L2.
6. It should also be pointed out that most studies in intonation use a small number of speakers from which to analyze data, despite the fact that inter-speaker variability is well known to be problematic. Studies such as Arvaniti & Garding for English (2007) use 13 speakers, the studies they mention in their article vary from an undefined number, to two, to five speakers (2007:5); Sluijter & van Heuven (1996) use 6 speakers, Clopper & Smiljanic (2011) use 10 speakers – all considerably smaller numbers of speakers than those analyzed in our study, which takes data from 29 speakers (10 for monolingual Spanish, 10 for bilingual Spanish, and 9 for monolingual southern California English).
7. Although they call it story-telling in Table 1 (p. 92), it is really a story-reading task as explained in the section on materials on page 90.

REFERENCES

- Arvaniti, Amalia & Gina Garding (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (eds), *Papers in Laboratory Phonology 9*. Berlin & New York: Mouton de Gruyter, 547-576.
- Baumann, Stefan (2005). Degrees of givenness and their prosodic marking. *The International Symposium on Discourse and Prosody as a Complex Interface*. Aix-en-Provence. http://phonetik.phil-fak.uni-koeln.de/fileadmin/home/ritters/Allgemeine_Dateien/baumann-aix-idp05.pdf. (5 Oct 2014).
- Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.86, retrieved 28 September 2014 from <<http://www.praat.org/>>.
- Bolinger, Dwight (1978). Intonation across Languages. In Joseph H. Greenberg (ed.), *Universals of Human Language 2 (Phonology)*. Stanford University Press, pp. 471-524.
- Butragueño, Pedro Martín (2004). Configuraciones circunflejas en la entonación del español mexicano. In *Revista de Filología Española* LXXXIV(2): 347-373.
- Chafe, Wallace L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and topic*. New York, Academic Press, pp. 27-55.
- Clopper, Cynthia G. & Rajka Smiljanic (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39(2): 237-245.
- Cowles, H. Wind., Robert Kluender, Marta Kutas, & Masha Polinsky (2007). Violations of information structure: an electrophysiological study of answers to wh- questions. *Brain and Language* 102(3): 228-242.
- Cruttenden, Alan (1997). *Intonation*. Cambridge: Cambridge University Press.
- Daly, Nicola & Paul Warren (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics* 5(1): 85-96.
- D'Imperio, Mariapaola, Gorka Elordieta, Sonia Frota, Pilar Prieto & Marina Vigário (2005). Intonational Phrasing in Romance: The Role of Syntactic and Prosodic Structure. In Sonia Frota, Marina Vigário & Maria João Freitas (eds.). *Prosodies: With Special Reference to Iberian Languages*. Berlin: Walter de Gruyter, pp. 59-97.
- Fernández Planas, Ana María & Eugenio Martínez Celdrán (2003). El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español. In *Estudios de Fonética Experimental* 12: 165-200.
- Féry, Caroline (2007). Information structural notions and the fallacy of invariant correlates. In Caroline Féry, Gisbert Fanselow, & Manfred Krifka (eds.). *Interdisciplinary studies on information structure*. Vol. 6. Potsdam: Universitätsverlag, pp. 185-204.

- Gordon, Matthew & Latifa Nafi (2012). Acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics* 40(5): 706-724.
- Gries, Stefan Th. (2013). *Statistics for linguistics with R: a practical introduction*. 2nd rev. & ext. ed. Berlin & New York: De Gruyter Mouton.
- Gussenhoven, Carlos (2002). Intonation and Interpretation: Phonetics and Phonology. *Speech Prosody 2002: Proceedings of the 1st International Conference on Speech Prosody*. Aix-en-Provence: ProSig and Université de Provence Laboratoire Parole et Langue: pp. 47-57.
- Gussenhoven, Carlos (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos (2007). Notions and subnotions in information structure. In Caroline Féry, Gisbert Fanselow, & Manfred Krifka (eds.). *Interdisciplinary studies on information structure*. Vol. 6. Potsdam: Universitätsverlag, pp. 185-204.
- Hagoort, Paul (2005). On Broca, brain and binding: A new framework. *Trends in Cognitive Sciences* 9(9): 416-423.
- Halliday, Michael A.K. (1967). Notes on transitivity and theme in English. Part II. *Journal of Linguistics* 3(2): 199-244.
- Haviland, Susan E. & Herbert H. Clark (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13(5): 512-521.
- Henton, Caroline G. (1989). Fact and fiction in the description of female and male speech. *Language and Communication* 9(4): 299-311.
- Henton, Caroline G. (1995). Pitch dynamism in female and male speech. *Language and Communication* 15(1): 43-61.
- Herrera, Esther & Pedro Martín Butragueño. (eds.). 2003. La tonía: dimensiones fonéticas y fonológicas. México City: El Colegio de México.
- Krifka, Manfred (2007). Basic notions of information structure. In Caroline Féry, Gisbert Fanselow, & Manfred Krifka (eds.). *Interdisciplinary studies on information structure*. Potsdam: Universitätsverlag: pp. 13-56.
- Ladd, Roger (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lambrecht, Knud (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Martínez Celdrán, Eugenio & Ana M^a Fernández Planas (2003). El tono fundamental y la duración: dos aspectos de la taxonomía prosódica de las modalidades de habla (enunciativa e interrogativa) del español. *Estudios de Fonética Experimental* vol. XII: 165-200.
- Miglio, Viola G., Stefan Th. Gries, & Michael Harris (2014). Mexican & Chicano intonation: Differences related to information structure. 6th Annual PSSLT Conference: Pronunciation in Second Language and Teaching 2014. September 5-6, 2014, University of California, Santa Barbara.

- Parodi, Claudia (2011). El otro México: español chicano, koineización y diglosia en Los Ángeles, California. *Realismo en el análisis de corpus orales: Primer Coloquio de Cambio y Variación Lingüística*, 217-243. R. B. Villanueva and P. M. Butragueño. Mexico, D.F.: El Colegio de México.
- Pierrehumbert, Janet (1980). *Phonetics and phonology of English intonation*. PhD dissertation, MIT.
- Pierrehumbert, Janet & Julia Hirschberg (1990). In Phillip Cohen, Jerry Morgan, & Martha Pollack (Eds.). *Intentions in communication*. Cambridge, MA: MIT Press: pp. 271-311.
- Prieto, Pilar & Paolo Roseano (eds.) (2010). *Transcription of Intonation of the Spanish Language*. Lincom Europa: München.
- Prince, Ellen F. (1981). Toward a taxonomy of Given-New information. In Peter Cole (ed.), *Radical Pragmatics*. New York: Academic Press, pp. 223-256.
- Reinhart, T. (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27 (1): 53-94.
- Röhr, C.T. & Stefan Baumann (2010). Prosodic marking of information status in German. *Proc. Speech Prosody Chicago*, 100019: 1-4.
- Sluijter, A.M.C., & van Heuven, V.J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the 4th international conference on spoken language processing*: 630-633.
- Suñer, Margarita (1982). *Syntax and Semantics of Spanish Presentational Sentence-Types*. Georgetown University Press, Washington, DC.
- Swerts, Marc, Emiel Kraemer & Cinzia Avesani (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4): 629-654.
- Vallduví, Enric (1992). *The informational component*. Garland Publishers: New York & London.
- van Leeuwen, Tessa M., Monique J. A. Lamer, Karl Magnus Petersson, Carlos Gussenhoven, Toni Rietveld, Benedikt Poser, & Peter Hagoort (2014). Phonological markers of information structure: An fMRI study. *Neuropsychologia* 58 (May): 64-74.
- Wang, L., Hagoort, P., & Y. F. Yang (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, 1282(July 28): 50-56.
- Warren, P. & N. Daly (2000). Sex as a factor in rises in New Zealand English. In Janet Holmes (ed.) *Gendered speech in social context: Perspectives from gown and town*. Wellington: Victoria University Press, 99-115.
- Zubizarreta, María Luisa (1998). *Focus, Prosody, and Word Order*. Cambridge, MA: MIT Press.
- Zubizarreta, María Luisa & Emily Nava (2011). Encoding discourse-based meaning: Prosody vs. syntax. Implications for second language acquisition. *Lingua* 121(4): 652-669.