Quasi-neutralization of stress contrasts in Spanish

Francisco Torreira [1], *Miquel Simonet* [2], *José I. Hualde* [3]

[1] Max Planck Institute, Nijmegen, The Netherlands

[2] University of Arizona, USA

[3] University of Illinois at Urbana-Champaign, USA

`Francisco.Torreira@mpi.nl, simonet@email.arizona.edu, jihuade@illinois.edu`

Abstract

We investigate the realization and discrimination of lexical stress contrasts in pitch-unaccented words in phrase-medial position in Spanish, a context in which intonational pitch accents are frequently absent. Results from production and perception experiments show that in this context durational and intensity cues to stress are produced by speakers and used by listeners above chance level. However, due to substantial amounts of phonetic overlap between stress categories in production, and of numerous errors in the identification of stress categories in perception, we suggest that, in the absence of intonational cues, Spanish speakers engaged in online language use must rely on contextual information in order to distinguish stress contrasts.

Index Terms: lexical stress, stress cues, phonological neutralization, Spanish.

1. Introduction

In Spanish, the correlates of lexical stress appear to be subtler than in other languages with lexically contrastive stress that have been investigated in this respect, such as English or Dutch, and also closely related languages such as Portuguese and Catalan. Spanish lacks systematic reduction of vowels in unstressed syllables [1], and durational differences between stressed and unstressed syllables in this language are relatively small compared to Portuguese [2, 3]. This results in lexical differences in stress placement often being difficult to perceive for learners of Spanish as a second language [4, 5].

The most robust cues to determine the position of lexical stress in a Spanish word are present when it carries an intonational pitch accent, in which case the pitch accent is associated to the lexically-stressed syllable, lending it phonetic prominence. When the word does not carry an intonational pitch accent, on the other hand, the perception of lexical stress may be jeopardized. Given the important role of pitch accents in disambiguating stress contrasts, the question arises whether lexical stress distinctions are maintained under such conditions, as has been reported for other stress languages such as Dutch and English [6, 7, among others]. To address this issue, [8] conducted a study where Spanish speakers were asked to produce parenthetical reporting clauses, which systematically exhibit a low flat pitch contour judged to lack pitch accents. Under this condition, minimal pairs involving paroxytone and oxytone verb forms (e.g. *deter**mi**no* vs. *determi**nó**,* stressed syllables in bold) were found to be distinguished in production by consistent differences in vowel duration, and less reliably by intensity.

Reporting clauses such as the ones used by [8] are rather rare in conversational Spanish. On the other hand, a very frequent context where words do not appear to carry pitch accents in Spanish is in phrase-medial position within broad-focus long intonational phrases (IPs). This is illustrated by the following examples, which are typically realized as single intonational phrases ending in rising continuation intonation:

a) [***Siem**pre que **mi**ro la **ho**ra*]$_{IP}$, …
    'Every time I look up the time, …'

b) [***Siem**pre que mi**ró** la **ho**ra*]$_{IP}$, …
    'Every time she looked up the time, …'

Our observations from the Nijmegen Corpus of Casual Spanish (NCCSP) [9] suggest that in such contexts stress contrasts might be lost, without duration and intensity cues compensating for the absence of a pitch accent. Audio examples of unaccented words in phrase-medial position extracted from the NCCSp can be accessed online at [10]. These examples contain the words *dejo* 'I leave' and *dejó* 'she left', which contrast only in the position of lexical stress.

In the present study, we investigate to what extent the lack of pitch accents in phrase-medial position in Spanish leads to a neutralization of stress contrasts. In the following sections, we present a production experiment and a perception experiment aimed at answering this question.

2. Experiment I: Production

In Experiment I, we investigate the extent to which Spanish speakers maintain lexical stress contrasts in words lacking a pitch accent in phrase-medial position. We elicit unaccented verbal forms contrasting in the position of stress (e.g. *tapo* vs. *tapó*) located in phrase-medial position, and examine their phonetic realization in terms of f0, duration, intensity, formant values, and two forms of consonantal lenition using regression modeling and a cross-validation procedure. We use verbal forms from the first conjugation (with infinitives ending in –ar) as targets words, because they provide a systematic contrast in stress placement (oxytones for 3rd p. sg, past tense, vs. paroxytones for, 1st p. sg, present tense).

2.1 Method

Nine native speakers of European Spanish (4 female, 5 male) were recorded in a quiet room with an Audix HT5 head-worn microphone connected to a Sound Devices MM-1 pre-amp, which was in turn connected to a Marantz PMD660 digital recorder. The signal was digitized at 44.1 kHz, 16-bit quantization. Subjects were seated in front of a computer screen and presented with strings of words consisting of a subject pronoun, a verbal tense (present or past), an interrogative pronoun, a target verb in the infinitive form, and

a noun phrase serving as a verbal object. The participants' task was to first examine the elements on the screen, and then produce a fluent wh-question. For instance, for the stimulus *PASADO/él: ¿Cómo/tocar/tu estrella?* 'PAST/he: How/touch/your star?', the expected response was *¿Cómo tocó tu estrella?* 'How did she touch your star?'. Wh-questions were chosen as carrier sentences because they are typically produced as one intonational phrase, both in spontaneous and laboratory speech. This method elicited a large number of phrase-medial verbal forms belonging to three minimal pairs: *toco* 'I touch' vs. *tocó* 'she touched'; *tapo* 'I cover' vs. *tapó* 'she covered'; *corto* 'I cut' vs. *cortó* 'she cut'.

The experimental materials included 60 target sentences (3 verbs * 2 tenses * 10 repetitions) and 60 distractors containing different subject pronouns and verbal tenses. From the 540 elicited tokens (60 target sentences * 9 speakers), we analyzed 456 that were produced as fluent broad-focus intonational phrases. The remaining utterances, exhibiting disfluencies, and less frequently, other phrasing and accentual patterns, were discarded. The majority of the analyzed utterances (74%) were produced with an initial rising accent (L*+>H* in SpToBI notation) on the question word (e.g. *cómo*) and a valley on the stressed syllable of the object (*estrella*) followed by a final rise (L* H%), with falling f0 throughout the phrase-medial verb (*toco* or *tocó*). The second most common intonational contour (24%) consisted of a low accent on the question word (L*), and a rise-fall on the object noun (L+H* L%), with flat f0 throughout the phrase-medial verb.

Acoustic analyses were performed as follows: as a first step, the two syllables of the target verb forms were segmented following standard criteria, with stop consonants starting and vowels ending at the stretches of silence attributable to oral stop closures. The following acoustic measurements were then taken with Praat, and registered as differentials between the first and the second syllable: (a) f0 at the midpoint of the vowel, in Hz; (b) syllable duration, in ms; (c) peak intensity within the syllable, in dB (d); first and second formant values (F1, F2) at the midpoint of the vowel, in Hz. According to several recent studies, Spanish /p t k/ consonants are often voiced and spirantized (realized as approximants, without a complete stop closure) in intervocalic position. These lenition phenomena may be conditioned by prosodic factors such as lexical stress or pitch accent [11, 12, 13]. For this reason, we also annotated (e) the presence of uninterrupted periodicity throughout the closure of each of the two stops in each utterance, and (f) whether vowel formants could be observed throughout the consonantal closure. In cases of spirantization, that is, without clear stop closures, segmentation of the syllables in the target words could not be performed, and the differentials mentioned above were left undefined.

2.2 Results

We first fitted a series of mixed-effects regression models with the f0, duration, F1, and F2 differentials between the first and second syllable of the target verbs as responses, stress pattern (paroxytone, oxytone) as the main predictor, verb type (*cortar*, *tapar, tocar*), and contour type as covariates, and speaker as a random factor. Positive differentials indicate that the first syllable has higher values than the second, while negative differentials indicate the opposite. If lexical stress in phrase-medial unaccented words is distinguished by speakers, some or all of these acoustic differentials should differ between the two stress patterns, with higher differentials for paroxytone words (i.e. with more phonetically prominent first syllables).

Consistent with our impressions that the phrase-medial target words did not carry a pitch accent, no f0 difference was found between the two stress patterns ($p = .46$). A small difference was observed for F1, with paroxytones having a slightly more open first vowel than oxytones ($\beta = 21.51$, $t = 4.1$, $p < .0001$), but no difference was observed for F2. Greater differences were observed for duration and intensity. In line with the possibility that a stress distinction is maintained in the absence of pitch accents, paroxytones tended to exhibit longer and more intense first vowels than oxytones (duration: $\beta = 22.3$, $t = 12.4$, $p < .0001$; intensity: $\beta = 1.77$, $t = 7.85$, $p < .0001$). Despite these statistical differences, however, we observed a considerable amount of overlap between the two stress patterns, in particular for the intensity differential. This can be appreciated in Figures 1 and 2, which show verb-normalized duration and intensity differentials as a function of stress pattern.

As mentioned above, we also examined consonant voicing and spirantization as possible cues to stress. Around a third (33%) of the consonants in the first syllable of the target word were fully voiced when the syllable was lexically stressed. This percentage was considerably higher (46.4%) in unstressed syllables. As for the consonant onset of the second syllable, it was voiced in 33% of tokens when this syllable was stressed vs. 53.7% when unstressed. Both differences were statistically significant in mixed-effects logistic regression models with consonant voicing as the response, stress pattern as the main predictor, verb type as a covariate, and speaker as a random factor (1st syllable: $\beta = 0.74$, $z = 3.12$, $p < .005$; 2nd syllable: $\beta =-0.75$, $z = -3.24$, $p < .005$).

Regarding spirantization, we observed that, in general, it was more frequent in the second consonant ($n = 43$; 9.4% of the data) than in the first one ($n = 10$; 2.1%). However, it was only in the first consonant that spirantization seemed to be affected by lexical stress, as suggested by a slight statistical trend in a regression model with spirantization as response, stress pattern as predictor, verb as a covariate, and speaker as a random factor ($\beta = -1.41$, $z = -1.77$, $p = .07$). In the second consonant, where spirantization was more frequent, no statistical difference was observed (8.3% in stressed syllables vs. 9.5% in unstressed syllables; $p = .63$).
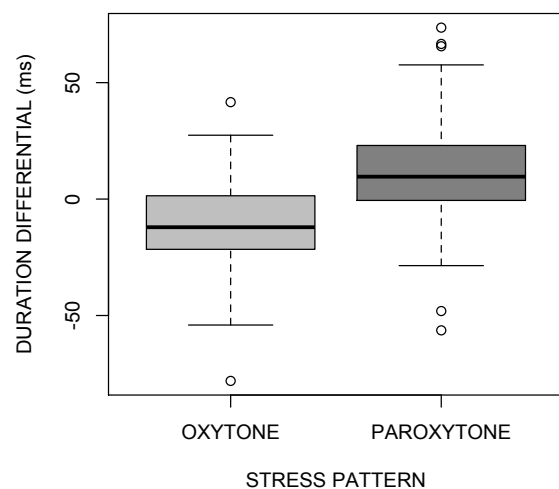

Fig. 1. *Verb-normalized duration differential between first and second syllables in our target words as a function of stress pattern.*
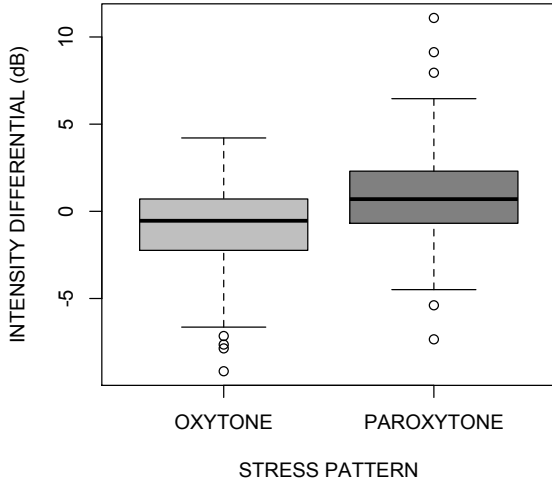
Fig. 2. *Verb-normalized peak intensity differential between first and second syllables as a function of stress pattern.*

To assess the strength of the different identified cues to stress position, we subjected our data to a leave-one-out cross-validation procedure. We simulated predicting stress patterns for unknown data in the following way: for each token in the dataset, we predicted its stress pattern with logistic regression models trained on the rest of the dataset. These models included different combinations of relevant features identified above (i.e. F1, intensity, and duration differentials, and consonantal voicing), plus speaker and verb type.

Table 1 shows percentages of correct classification obtained with five different models. Duration offered the best cue to the stress contrast, achieving 73.8% of correct classification, almost as much as a model containing all features (75.8%). Intensity provided a moderate gain over chance level (62.4%), whereas consonantal voicing and F1 only allowed for results slightly above chance level (55.2% and 51.8%).

We conclude from this analysis that duration and intensity, in this order, provide the best cues to lexical stress in Spanish unaccented words in phrase-medial position. It is also worth noting, however, that almost a quarter of the dataset could not be classified correctly even when all relevant cues where used. This suggests that in such cases there may be little acoustic information that listeners could use to distinguish stress contrasts. We address this issue in the following section.
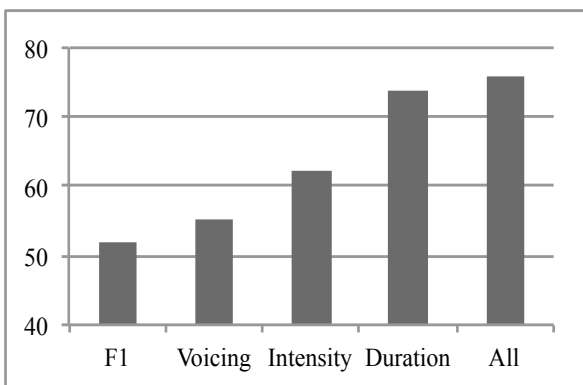

Fig. 3. *Percentages of correct stress pattern classification in Exp. I obtained with different acoustic features in a cross-validation procedure (see text for details).*

3. Experiment II: Perception

In Experiment II, we investigate the extent to which native Spanish listeners can distinguish different stress patterns in phrase-medial unaccented words. We also investigate what phonetic cues are used by listeners to discriminate between stress patterns.

3.1 Method

A random subset of 100 utterances was selected from the 456 utterances analyzed in Experiment I, and used as stimuli in Experiment II. We decided to use only a subset of 100 stimuli in order to avoid fatigue in the participants, which could lead to unreliable results. The experiment consisted of a two-alternative forced-choice task, in which listeners had to classify auditory stimuli in one of two groups according to the tense and person of the verb in the utterance, which, as explained in the previous sections, are distinguished only by the stress pattern of the verbal form (oxytone: 3rd p. sg., past vs. paroxytone: 1st p. sg, present).

Thirteen listeners, all of them native speakers of European Spanish, participated in the experiment. They wore closed headphones and sat in front of a desktop computer in a quiet office. In each trial, two written options appeared on the screen upon presentation of an auditory stimulus: *ÉL/PASADO* 'he/past', in the left part of the screen, and *YO/PRESENTE* 'I/present', in the right part of the screen. The message in the left part of the screen was congruent with an oxytone verb (e.g. ta**pó**), whereas the message in the right part of the screen was congruent with a paroxytone verb (e.g. **ta**po). Participants were asked to press the left or right-arrow key on a computer keyboard according to which message on the screen was congruent with each heard utterance.

3.2 Results

Participants correctly classified the stress pattern of the presented target verbs in 62.9% of the cases. Responses were correct significantly above chance level according to a mixed-effects logistic regression model with stress pattern as dependent variable, the participant's response as the only fixed predictor, and participant as a random factor ($\beta = 1.08$, $z = 9.2$, $p < .0001$). Note, however, that the automatic classification carried out in Experiment I outperformed participants in Experiment II by a considerable margin (75.8 % vs. 62.9%).

An analysis by participants revealed that all of them were able to identify the stress pattern of the verb in the stimuli slightly or moderately above chance level, with percentages of correct classification ranging from 55% to 70%. In an analysis by items ($n = 100$), we found that many of them tended to be correctly classified significantly above chance level (47 above 70%; 27 above 85%), but we also observed a considerable number of items below chance level (16 below 50%; 6 below 30%). Taken together, these results indicate that the stress pattern in the target words could be identified above chance level in most cases, but also that identification errors were common in the data.

We then investigated which cues participants followed when responding to the stimuli in the experiment. We subjected the data to a leave-one-out cross-validation procedure similar to the one employed in Experiment I. This time, we fitted logistic regression models with the participant's response (not the correct answer) as dependent variable, and different combinations of relevant acoustic cues as predictors (i.e. F1, intensity, and duration differentials, consonantal voicing), plus information about participant and verb type.

Figure 4 shows percentages of correct classification of participants' responses obtained with five different models. The model including all acoustic cues achieved 65.5% of correct classification. Interestingly, when only one cue was used, intensity offered the best performance, achieving 59.7% of correct classification. Duration followed closely, with 59.2%. Although in the production data from Experiment I duration clearly offered a better cue to the stress contrast than intensity, these perception data suggest that listeners use both intensity and duration cues in a similar degree when discriminating stress patterns. Regarding F1 and voicing, we found that, as in Experiment I, theses cues played a lesser role than intensity and duration, with percentages of 52.9% and 55.1% respectively.
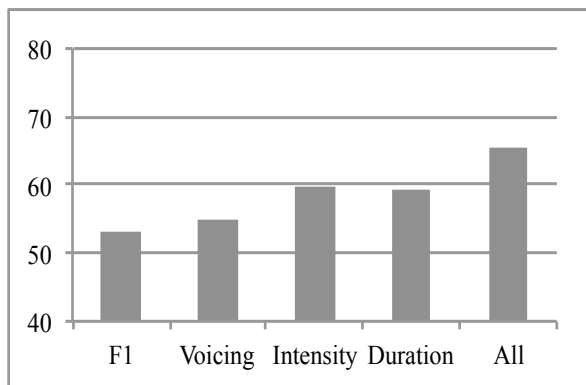

Fig. 4. *Percentages of correct classification of subjects' responses in Experiment II obtained with different acoustic features in a cross-validation procedure (see text for details).*

4. Discussion and conclusion

The previous two sections have presented two experiments aimed at investigating the production and perception of lexical stress contrasts in unaccented words in Spanish. More particularly, we have examined the case of unaccented words in phrase-medial position, a very, and possibly the most, frequent context of deaccenting in Spanish.

Our experiments have shown that, both in production and perception, a contrast in the position of lexical stress is maintained in spite of the lack of prominence-lending pitch cues. In our production experiment, lexically stressed syllables tended to be longer and more intense than their unstressed counterparts. Duration cues allowed for the correct classification of almost three quarters of the data in a cross-validation procedure simulating the discrimination of new unseen data, and they provided a considerably better cue than intensity cues. In perception, we found that all speakers could distinguish the position of lexical stress above chance level, and that, interestingly, their perception of the position of lexical stress was guided as much by intensity cues as by duration cues. Other phonetic cues, such as F1 and the presence of lenitory voicing in consonants also appeared to play a role in distinguishing lexical stress contrasts, but, both in production and perception, in a lesser degree than duration and intensity. Our results for phrase-medial unaccented words are therefore in line with those reported in [8] for parenthetical reporting clauses.

Despite the identified phonetic differences in Experiment I, and the listeners' performance above chance level in Experiment II, it should be noted nevertheless that we observed a considerable amount of phonetic overlap between stress patterns. In production, roughly a quarter of the data could not be classified correctly by a model containing several relevant features such as duration, intensity, voicing and F1; and, in the perception experiment, listeners made identification errors in a two-alternative forced-choice task in 37.1% of the trials. If we take into account that our speech materials were produced in a laboratory setting, and that the perception experiment explicitly required participants to choose between two alternative choices, we may wonder to what extent native Spanish speakers use stress-related phonetic contrasts such as the one examined in this study during online language use. In everyday conversation, it is likely that contextual information is necessary for discriminating between stress patterns, since phonetic cues to stress do not appear to be very robust even in highly controlled data such as ours.

The lack of across-the-board robust cues to stress in Spanish may be related to the low information load that the position of lexical stress has in this language. Except in verb forms, the location of stress in Spanish words is largely predictable: although lexical stress may fall in principle on any of the last three syllables of the word in theory (e.g. *lámpara* 'lamp', *mampara* 'screen', *Panamá*), there is an overwhelming tendency for consonant-final words to be stressed on the final syllable, and for vowel-final words to be stressed on the penultimate syllable, with more than 95% of all words in the Spanish lexicon following this rule. Regarding verbs, for which several kinds of minimal pairs exist, ambiguities are likely to be rare in discourse, since these minimal pairs involve verb forms in different tenses (e.g. present vs. past) and with different subjects (e.g. 1st sg. vs. 3rd p. sg.). Robustly conveying information on lexical stress pattern is therefore of little communicative value in Spanish, which may explain the significant amount of phonetic overlap between stress patterns observed in our data.

On the other hand, the fact remains that Spanish words are specified for the position of stress. Even if there are general rules of stress assignment, exceptions must be learned by children during language acquisition. In this regard, the considerable amount of phonetic neutralization that we find in phrase-medial position is not likely to pose a serious challenge to learners with continued exposure to the language, since the position of stress is often conveyed more robustly in other prosodic contexts thanks to the presence of intonational pitch cues [8]. In this sense, the location of lexical stress is not very different from other features that create contrasts between words, and that are sensitive to contextual and realizational factors. For instance, the fact that Spanish /ptk/ consonants are frequently voiced and spirantized in intervocalic position [11, 12, 13], giving rise to occasional ambiguities, does not prevent the /ptk/-/bdg/ opposition from being fully functional in the language. Nevertheless, the fact that Spanish stress contrasts are often lost under specific, but common, conditions, such as the context examined in this study, should be taken into account in descriptions of the prosody of this language.

5. Acknowledgements

6. References

[1] Nadeu, M. "The effects of lexical stress, intonational pitch accent and speech rate on vowel quality in Catalan and Spanish". Doctoral dissertation, Univ. of Illinois at Urbana-Champaign. 2013.

[2] Ferreira, L. "High Initial Tones and Plateaux in Spanish and Brazilian Portuguese Neutral declaratives: Consequences to the relevance of f0, duration and vowel quality as stress correlates". Doctoral dissertation, Univ. of Illinois at Urbana-Champaign. 2008.

[3] Barbosa, P., Eriksson, A. & Åkesson, J. 2003. "On the Robustness of some Acoustic Parameters for Signaling Word Stress across Styles in Brazilian Portuguese". Interspeech 2013, Lyon, 25-29 August.

[4] Saalfeld, A. 2009. "Stress in the beginning Spanish classroom: an instructional study". Doctoral dissertation, Univ. of Illinois at Urbana-Champaign. 2009.

[5] Ortega-Llebaria M., Hong, G. & Fan, J. "English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception". Journal of Phonetics, 41:186-197, 2013.

[6] Sluijter, A. M. C. & Heuven, V. J. van. "Spectral balance as an acoustic correlate of linguistic stress". J. Acoust. Soc. Am. 100:2471–2485, 1996.

[7] Beckman, M. & Edwards, J. "Articulatory evidence for differentiating stress categories" in P. Keating [Ed], Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III, 7-33, Cambridge University Press, 1994.

[8] Ortega-Llebaria, M. & Prieto, P. "Acoustic correlates of stress in Central Catalan and Castilian Spanish". Language and Speech, 54(1):73–97, 2010.

[9] Torreira, F. & Ernestus, M. "Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish". Phonetica, 69:124–148, 2012.

[10]http://corpus1.mpi.nl/ds/imdi_browser/?openhandle=1839/00-0000-0000-001B-8046-2 (to access a file, select it in the left panel and click on the URL link appearing in the right panel).

[11] Hualde, J. I., Simonet, M. & Nadeu, M. "Consonant lenition and phonological recategorization". Laboratory Phonology, 2:301-329, 2011.

[12] Kim, M. The phonetics of stress manifestation: Segmental variation, syllable constituency and rhythm. Doctoral dissertation, Stony Brook Univ. 2011.

[13] Torreira, F. & Ernestus, M. "Realization of voiceless stops and vowels in conversational French and Spanish". Laboratory Phonology, 2:331-353, 2011.