Research Article

The phonetics of information structure in Yoloxóchitl Mixtec

Christian DiCanio [a,*], Joshua Benn [a], Rey Castillo García [b]

[a] University at Buffalo, United States
[b] Secretaría de educación pública (SEP), Chilpancingo, Guerrero, Mexico

ABSTRACT

Research on speech prosody has shown that higher-level phonological constituents can be examined directly via their influence on low level phonetic processes (Beckman & Edwards, 1990; Fougeron & Keating, 1997). Despite the strong tradition of research in this area, the existing work has focused mainly on languages which lack lexical tone. This contributes to the view that prosodic structures show little influence on tone, i.e. a language may either have lexical tone or lexical/phrasal stress, the latter of which fits into the prosodic hierarchy. The current paper examines prosodic focus in Yoloxóchitl Mixtec, an endangered Otomanguean language spoken in Mexico. Using experimental data from ten speakers in the field, we investigated how sentence position, stress, and focus type influenced the realization of F_0 and duration in different tonal melodies. The findings show that the tonal F_0 space was expanded and raised on words produced with contrastive focus, less on words produced with narrow focus, and least on words produced under broad, sentential focus. Focus-related lengthening asymmetrically affected stressed syllables in the language more than unstressed syllables. In stressed syllables, this resulted in an increase in tonal hyperarticulation.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Research throughout the past several decades has shown that lexical tone targets can vary substantially (Andruski, 2006; Chang & Hsieh, 2012; DiCanio, 2012; Gandour, Tumtavitikul, & Satthamnuwong, 1999; Liu & Xu, 2005; Peng, 1997; Scholz, 2012; Xu, 1994; Xu, 1999; Xu & Xu, 2003; Zhang & Liu, 2011). Such variation, either due to coarticulation with adjacent tones and consonants or due to individual speaker differences, has an impact on both the accuracy and timecourse of tone perception (Francis, Ciocca, King Yu Wong, Ho Yin Leung, & Cheuk Yan Chu, 2006; Nixon, Chen, & Schiller, 2014; Peng, Zhang, Zheng, Minett, & Wang, 2012; Xu, 1994). Apart from these local phonological and socio-indexical sources of variation, an additional source of variation in tone production is the informational content of the word or unit to which the lexical tone is assigned. The flow of discourse among speakers requires that certain lexical items be brought to the attention of the listener while others be backgrounded (Baumann, 2006; Lambrecht, 1994). This has an impact on the production of lexical tone (Scholz, 2012; Xu, 1999) as well

as the degree to which segmental targets are hyperarticulated (de Jong & Zawaydeh, 2002; de Jong, 1995; Mücke & Grice, 2014).

The current study investigates how information structure influences tone production and the degree to which it is sensitive to stress. These topics are investigated in Yoloxóchitl Mixtec (YM, henceforth; ISO 639 code xty), an indigenous Oto-Manguean language of Southern Mexico (Castillo García, 2007). YM possesses both a complex lexical tone inventory and fixed lexical stress. The relationship between information structure and intonational pitch accents is well-established in non-tonal languages (for an overview, see Baumann (2006), Gussenhoven (2004), Jun (2005), Ladd (2008)), but substantially less is known about how information structure impacts lexical tone production. Moreover, work on non-tonal languages encompasses a typologically-diverse sample of languages, but the existing work on tonal languages is mostly limited to those lacking lexical stress (Kügler & Genzel, 2011; Liu & Xu, 2005; Scholz, 2012; Xu, 1999). YM is different in this regard. If the placement of nuclear pitch accents in non-tonal languages is sensitive to the prosodic hierarchy within the word (Gussenhoven, 2004; Jun, 2005), where do tone languages fit in?

We investigate the relationship between information structure and tone in YM through a speech production study carried out in the field with a population of ten native speakers. We examined how narrow (argument) focus and contrastive (or corrective) focus are realized via a naturalistic response task (c.f. Clopper & Tonhauser, 2013; Kügler & Genzel, 2011) and separately compare these results to tones produced under broad (sentential) focus via a repetition task. This experiment addresses both an empirical question and a theoretical one. First, how do tones vary in their realization in a complex tone language and with lexical stress? Do unstressed syllables vary more than stressed syllables? Second, how is the prosodic realization of information structure functionally constrained in a lexical tone language? Do tonal type (level, rise, etc) and tonal position within the tonal space (highest tone, lowest tone) matter?

1.1. Background: focus

Information structure refers to those components of the linguistic system that interlocutors use to negotiate shared knowledge of entities and states in discourse (Lambrecht, 1994). Languages utilize different strategies for expressing whether an entity is new/old, topical/focal, or recently identifiable (or not) in the discourse. Of primary importance to phonetic and phonological studies of information structure is the realization of *focus*. Focus refers to "the speaker's assessment of the relative predictability or unpredictability of the relations between propositions and their elements in a given discourse situation." (Lambrecht, 1994, p. 6). In utterances produced with broad focus, the entire sentence or predicate conveys pragmatically unpredictable information. In utterances produced with narrow focus, a single argument or state is pragmatically unpredictable.[1] This single constituent may then be linguistically-marked as distinct from others in the utterance. Languages frequently utilize one of three possible tactics for marking focus: morphosyntactic marking, focus particles, and prosodic marking (Ladd, 2008). Additionally, languages like Northern Sotho may mark only pragmatically *predictable* (non-focal) information via backgrounding and pronominalization, leaving focus entirely unmarked (Zerbian, 2007).[2]

In many languages which mark narrow focus with suprasegmentals, the focus domain (c.f. Lambrecht, 1994) may be marked with an intonational pitch accent. This accent is aligned to the most prominent syllable in the phrase via the *focus-to-accent* (FTA) principle (Gussenhoven, 1983a). In this way, utterance-level prosodic distinctions are directly sensitive to stress and the heads of prosodic constituents. Intonational pitch accents "arrange themselves according to the demands of the metrical structure" (Ladd, 2008, p. 268). A simple corollary of this view is the idea that metrically-weak prosodic constituents within words will be less affected by FTA than metrically-strong ones. It is this particular corollary that we investigate in the current study.

There is some debate over the extent to which different focal domains and types are distinguished by speakers/listeners. Bishop (2013) provides an overview of this debate. In terms of focal domain, speakers of English, Dutch, and German distinguish broad and narrow focus in speech production with prosodic features (Baumann, Grice, & Steindamm, 2006; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Gussenhoven, 1983b; Xu & Xu, 2005). However, while listeners may be able to successfully discriminate between focal domains using prosodic cues (Breen, Fedorenko, Wagner, & Gibson, 2010), they are less reliable at using these cues to identify the the context which elicited them (Birch & Clifton, 1995). In terms of focal types (i.e. narrow vs. contrastive focus), there is some debate regarding the extent to which these are phonologically categorical or pragmatically unique (c.f. Büring (2007), Katz & Selkirk (2011) and the references therein). Broad, sentential focus and narrow, nominal focus are distinguished morphosyntactically in YM. In the latter context, the NP is pre-verbal and ex-situ while in the former, the NP is in situ (post-verbal). Yet, both contrastive and non-contrastive (narrow) focus occur ex-situ in the same syntactic position. For the purposes of the current paper, we explore whether this distinction is prosodically marked and compare it with the in situ context.

The general goal of morphosyntactic marking of focus cross-linguistically is to align the constituent with the edge of a prosodic domain (Féry, 2013). This need not involve any particular type of prosodic marking, but in many cases, it does.[3] While a language may be described as marking focus by constituent dislocation, e.g. Italian (Lambrecht, 1994), such dislocation does not preclude prosodic marking at the same time. For instance, speakers of Bilbao Spanish or Central Catalán may front a constituent with narrow focus while simultaneously producing it with a pitch accent (Vanrell & Fernández Soriano, 2013). This same type of "double marking" is found with speakers of Balearic Catalán and Castillian Spanish when producing contrastive focus (ibid). In Zulu, focus is realized both through lengthening of the penultimate vowel of the focused word and via dislocation into the post-verbal position (Cheng & Downing, 2012). Finally, a focused constituent that is pre-posed in English may also be produced with a particular intonational pitch accent, e.g. '**BROCCOLI** *I hate,* **PEARS** *I love.*' (see Prince, 1981). The use of morphosyntax to mark information structure does not preclude prosodic marking. In fact, aligning a constituent to a phrase boundary may be used as a strategy to increase the unit's prosodic prominence (Féry, 2013).

1.2. Background: prosodic marking of focus in tone languages

The idea that there are multiple, simultaneous strategies for marking focus is pertinent to understanding prosodic focus marking in tonal languages. From the standpoint of the functional load hypothesis (Berinstein, 1979), one predicts that languages which use tone to mark lexical or morphological contrasts (lexical tone languages) would avoid the use of pitch to mark pragmatic distinctions like focus. However, numerous studies have shown that tone languages can use pitch to mark

---

[1] The interaction of predictability and focus is specifically explored in Turnbull (2017), to which the reader is referred.

[2] While there is little work on this question, languages appear prima facie to treat constituents with narrow focus as marked (either prosodically or morphosyntactically) and un-focused constituents as un-marked.

[3] See Féry (2013) for additional examples where languages do not mark focus prosodically.