# 3 Theoretical background and literature review

In this chapter, I lay out the theoretical groundwork necessary for the analytical chapters. An empirical investigation into the prosody of Huari Spanish and Quechua necessitates not only the establishment of theoretical and analytical terminology, but also an understanding of the range and limits of prosodic variability.

## 3.1 Introduction

In the following I will introduce and develop fundamental concepts (sections 3.2 and 3.3), putting particular emphasis on the range of typological variation observed in the areas of stress, accent, and prosodic phrasing (section 3.4), and the relation between tones, prosodic structure, and segmental material (section 3.5). We will see that this variation goes beyond the descriptive coverage of labels such as "culminative stress", "delimitative phrasing", or "segmental anchoring". Such a detailed exploration of the ranges of prosodic variability is relevant because only from a comprehensive viewpoint is it possible to characterize the prosodic phenomena we will observe in both Spanish and Quechua, and to bring them into an appropriate perspective towards each other. Not doing so would run the risk of studying Conchucos Quechua, about whose prosody and intonation very little is known, with Spanish as an analytical foil, simply because so much more is already known about the prosody of Spanish varieties. With such a widened typological perspective, on the other hand, it will hopefully be better possible to analyze both local languages, Quechua and Spanish, on their own terms and to locate them, with their individual variability and areas of overlap and divergence, on the variation space mapped out by what is known about prosodic typology. Over the course of this chapter, the discussion of the relevant prosodic concepts therefore will often be based not only on what is already known about Spanish, but also about typologically more distant languages as comparison. Quechua will also feature wherever possible, but unfortunately, for the most part to mark gaps in our knowledge.

Finally, the phenomena observed in the areas of pitch peak and accent distribution, as well as of (scaling-based) cues for prosodic phrasing in the analyses of Huari Quechua and Spanish (in particular sections 5.1.2, 5.1.3, 5.2, 6.2.3.3, and 6.4), make it necessary to include the issues of possible recursiveness in prosody and the prosodic cueing of information structure into the analysis. The theoretical groundwork for these later analytical decisions is also laid in this chapter, in sections 3.6 and 3.7, respectively. Their analysis is intended to contribute to suggestions about how the prosodic variation space can be conceptualized cross-linguistically.

Open Access. © 2024 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
https://doi.org/10.1515/9783111304595-003

From the second half of the 20th century onwards, it has been increasingly recognized that speech sounds exhibit systematic behaviour that can only be described by going beyond individual segments. The study of these aspects of speech is therefore sometimes called *suprasegmental* phonetics and phonology. Linguistic sound systems organize individual speech sounds into larger units via the concerted manipulation of the parameters[19] pitch, loudness, spectral quality of vowels and consonants, and duration, and they allow for the signaling of meanings independent of those conveyed by the segmental string in particular by the manipulation of pitch. All aspects of this system concerned with the grouping and composition of units are together called *prosody*. The systematic manipulation of pitch for postlexical linguistic functions[20] in particular is called *intonation* (cf. Ladd 2008: 5–7). Prosody and intonation are intimately connected, since prosody defines the domains on which an intonational event must occur in order to fulfill a specific function, and the two terms are often used nearly interchangeably. The restriction to postlexical functions distinguishes intonation from the use of pitch to distinguish lexical contrasts in tonal languages.

[19] The names given here are strictly speaking those of the perceptual/psychoacoustic correlates of physical properties of the acoustic signal. Pitch is the perceptual correlate of fundamental frequency, or F0, loudness that of intensity (energy amplitude), vowel quality that of formant frequencies above (overtones of) F0, consonant quality a mixture of that, aperiodic signals in the spectrum and timing measures such as voice onset time (VOT), and duration that of quantity (Ladefoged 2005; Fastl & Zwicker 2007; Ladefoged & Johnson 2011). The relationship between acoustic and psychoacoustic correlates is not straightforward in human perception, see e.g. Traunmüller 1990, 2017; Dawson et al. 2017; Fahey & Diehl 1996; Kuang & Liberman 2016; Ladd et al. 2013. For practical purposes in intonation research, the labels are often used interchangeably (Ladd 2008: 5).

[20] Ladd (2008) stresses the distinction between *linguistic* and *paralinguistic* uses of pitch and other prosodic parameters. Paralinguistic uses signal emotional and attitudinal states of the speaker, and are characterized by a gradual and iconic relationship between what is signaled and the signal itself: paralinguistically, loudness signals arousal or emotional involvement; the louder, the more emotionally involved. Linguistic uses, on the other hand, are supposed to convey aspects of meaning that are discrete and categorical and be expressed by formal means that are also essentially discrete. It has long been recognized that intonation straddles the boundary between these domains (Bolinger 1978: 475 famously calling it a “half-tamed savage”), in particular because pitch variation is by nature gradual and continuous, instead of discrete and categorical. The idea of *paralinguistics* is furthermore problematic because many meanings thought to belong to it, such as propositional speaker attitudes, are not only expressed by quite discrete morphosyntactic means in several languages (e.g. German discourse particles, morphology for expressing *mirativity*, etc.), but can also be shown to be semantically well-defined, and expressable by equally well-defined intonational cues (Fliessbach 2023). Ladd (2008: 34–42) maintains the distinction conceptually, but allows for the integration of gradual form into phonology.

## 3.2 Prosodic units and their hierarchy

Prosody research has proposed the following hierarchy of units apart from segments: segments group together to form onsets, nuclei and codas of syllables (σ). Nucleus and coda together are called the rhyme. A rhythmic unit, the mora (μ), is often positioned below the syllable, but is better seen as orthogonal to it. One or several syllables together form feet (F or Σ), which in turn form prosodic or phonological words (ω). Prosodic words are not isomorphic with lexical or morphological words, but it is a reasonable heuristic to assume that they broadly map onto each other. In Spanish, some morphological words such as adverbs formed by attaching *–mente* to an adjectival root (e.g *rápidamente* "quickly", *nuevamente* "newly, again") are often produced as two prosodic words, as can be seen from them realizing two stresses, and even pitch accents, one on the stressed syllable of the root, and another on the penult of *–mente*. On the other hand, clitics do not affect stress and pitch accent placement on the words they attach to (e.g. *te lo traduzco* "I translate it for you", *cocinárselo* "to cook it for oneself"). Thus on the criterion of being the domain of stress assignment, they form a single prosodic word on more than a single morphological one. For Quechua, this is largely unexplored. Given that Quechua is agglutinating and allowing for quite long polysyllabic words consisting of a root plus a number of suffixes, observations about multiple stresses or accents on a single word in several varieties (e.g. in Parker 1976; Adelaar 1977; Hintz 2006) indicate that some morphological words are produced as more than one prosodic word. Non-isomorphy between the morphological word and the prosodic word as domain of stress assignment is assumed by Stewart (1984: 207–209) for Conchucos Quechua,[21] and by Levinsohn (1976: 20) for Inga Quechua (Quechua IIB, Southern Colombia), with the domain of stress assignment sometimes larger, sometimes smaller, than the morphological word.

Above the prosodic word, a number of units can be roughly classed into two phrasal groups: a smaller phrasal unit has been identified under the name of Accentual Phrase (AP), Minor Phrase, phonological phrase (PhP or ϕ), or intermediate phrase (iP). A larger unit has been called Major Phrase or Intonational Phrase (IP or ι), sometimes said to be equal to or yet below, the utterance (U or υ). Their attestation differs substantially between languages as well as between descriptive and theoretical approaches. Table 2, taken from Frota (2012: 257), gathers proposed hierarchies of units into three groups based on different approaches in the litera-

[21] Conchucos Quechua includes Huari Quechua. When giving information about a different Quechua variety, I indicate the classificational group the variety belongs to (cf. section 2.2) as well as its geographic region.

**Table 2:** Different prosodic hierarchies proposed in the literature, from Frota (2012: 257).

| a. Rule-based | b. Intonation-based | c. Prominence-based |
|---|---|---|
| Intonational Phrase (IP) | IP | Nuclear Accent |
| Phonological / Major Phrase | Intermediate Phrase | |
| Clitic Phrase / Minor Phrase / Prosodic Word Group | Accentual Phrase | Accent |
| Prosodic Word (PW) | (PW) | Stress |
| Foot | Foot | Full Vowel |
| Syllable | Syllable | Syllable |
| Mora | Mora | |

ture. Early proposals in the rule-based generative tradition (Selkirk 1984; Nespor & Vogel 2007) assume that the prosodic hierarchy is universal across all languages. Later approaches (Selkirk 1996, 2011; Ito & Mester 2007, 2012; Féry 2017) maintain this claim but reduce the inventory by making all domains recursive in a departure from earlier views.

### 3.2.1 Different kinds of evidence for the existence of prosodic units in languages

Empirically speaking, a prosodic unit should only be said to exist in a language if there is tangible evidence for it, i.e. if phonological or phonetic processes can be shown to make reference to it, or if it is perceptually or articulatorily robust. In many languages, prosodic domains constrain how segments are distributed. This is most frequently seen with syllables, which impose restrictions on segment distribution in nearly all languages, based on preferences that enhance contrast and rhythmicity (Vennemann 1988). Thus, e.g. the sequence /sp/ is not ilicit *per se* but must always straddle a boundary between two syllables in Spanish, because codas may not contain more than one consonant and in onsets the two segments do not create a sufficiently steep sonority gradient. But such restrictions can also act on larger domains: either an aspirated or a glottalized consonant, never both, may optionally occur only once per word  (on the initial stop in the root and never in the suffixes) in Cuzco Quechua (Quechua IIC, southern Peru; Quechua I varieties do not have these consonants), unlike their simple counterparts, which are not constrained in this way (cf. (6); Parker 1997: 2; Cusihuamán 2001: 34). This restriction of once-per-domain is an example of a *culminative* property by which the relevance of a word-level unit can be argued, here for Cuzco Quechua.

(6) Distribution of glottalized and aspirated consonants in Cuzco Quechua: culminative but optional
a. t’anta “bread”
b. thanta “old, used up”
c. tayta “father, old man”
d. *tant’a
e. *t’antha

Hyman (2006: 229) lists other properties besides culminativity that can be used to define prosodic word domains in a language. Phonological processes specific to a domain can e.g. affect the segmental makeup by allowing assimilation processes only within a domain, but blocking it at its edges. For s-aspiration in Spanish, the domain of application differs between varieties:

(7) /s/-aspiration in Spanish varieties (adapted from Strycharczuk & Kohlberger 2016)

| | Andalusian / Honduras Spanish (Hualde 1991; Kaisse 1999) | Rio Negro Argentinian Spanish (Kaisse 1999) | Chinato Spanish (Hualde 1991) | Buenos Aires Spanish (Kaisse 1999) |
|---|---|---|---|---|
| *dieces* “tens” | [die.seh] | [die.seh] | [die. ðeh] | [die.ses] |
| *desigual* “unequal” | [de.hi.gual] | [de.si.gual] | [de.ði.gual] | [de.si.gual] |
| *dos palas* “two shovels” | [doh.pa.lah] | [doh.palah] | [doh.pa.lah] | [doh.pa.las] |
| *dos alas* “two wings” | [do.ha.lah] | [do.ha.lah] | [do.ða.lah] | [do.sa.las] |

Data such as in (7) is often interpreted to show that in all varieties, the minimal condition for aspiration of /s/ is for /s/ to be in coda position. However, what counts as a coda position is influenced by another process, usually called “resyllabification”, which causes postvocalic consonants to be produced as an onset when followed by a vowel, even if a word boundary intervenes. The syllable boundaries, symbolized by the dot (.), that are given in (7), are intended to be those after this “resyllabification” has occurred. To explain (7), Hualde (1991) and Kaisse (1999) propose that aspiration and resyllabification occur in different orders between the varieties (see there for details). In Buenos Aires Spanish, but not in the other varieties, aspiration does not occur before a pause (Kaisse 1999: 206–207), i.e. before a phrase boundary (Strycharczuk & Kohlberger (2016: 2)). Thus, /s/-aspiration interacts with the boundaries of up to three different domains: syllables, prosodic words and (phonological) phrases. Resyllabification also does not occur across pauses, but its complex workings lead Cardinaletti & Repetti (2009) to propose a new prosodic constituent, the "phrasal syllable level", as its domain. In the next section we will see what the pitfalls of proposing prosodic constituents based on individually observed phenomena can be.

### 3.2.2 Universal aspects of the prosodic hierarchy

Assuming universality of the units of the prosodic hierarchy would at this point mean postulating the phrasal syllable level also for all other languages, even if no processes ever take it as their domain apart from Spanish resyllabification. However, on the criterion of demonstrably being the domain of at least one phonological process in all languages, it turns out that few if any of the proposed units are really universal (cf. also Grijzenhout & Kabak 2009). Even the syllable, maybe the most universally accepted of these units, is perhaps not the domain of any phonological processes in at least one language, Gokana (Niger-Congo, cf. Hyman 2011, 2015 for a discussion). The prosodic word has been argued not to be universal based on several languages, including Vietnamese and Limbu (Sino-Tibetan, cf. Bickel et al. 2009; Schiering et al. 2010). In contrast, Himmelmann et al. (2018) present quite robust evidence that a unit corresponding to the intonational phrase can be identified in perception consistently even in languages the listeners are unfamiliar with. Its length also seems to average at 1.5 seconds in some data on English, French, and German reported on by Jun (2005d: 443), suggesting that there is at least some amount of overlap among observing linguists regarding what constitutes an IP crosslinguistically.

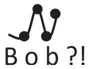
Arguments made about segmental prosodic processes are often made based on symbolic, categorical data such as given in (7), but increasingly, instrumental (acoustic and articulatory) findings are also brought to bear on them. They have resulted in similar observations of gradual variability in segmental realizations depending on prosodic domains across a variety of languages and are in general known as the prosodic strengthening of domain edges. The two edges of domains do not show the same effects: very broadly speaking, domain-initial strengthening often seems to make consonants at the beginning of larger prosodic domains such as utterances or IPs more consonant-like, judging from articulatory measurements such as increased linguopalatal contact and acoustic measures such as increased voice onset time (Fougeron & Keating 1997; Onaka 2003; Keating et al. 2004; Keating 2006; Cho & Keating 2009), while vowels enhance those features that make them more contrastive against other vowels (Georgeton & Fougeron 2014). On the other hand, final lengthening, as the name suggests, is an effect of increased duration at the end of phrasal domains that can affect both individual segments as well as syllables (Beckman & Edwards 1990; Rao 2007, 2010; Fletcher 2010; Petrone et al. 2017). The durational measurements are sometimes able to distinguish between positions defined with respect not just to a single, but several, prosodic domains (e.g. Strycharczuk & Kohlberger (2016: 7–9) on /s/-realization in Peninsular Spanish). Although both of these phenomena have been observed across a variety of languages and are thought to be "phonetic" markers of prosodic structure rather than (language-specifically) "phonological" ones (cf. Vaissière 1983; Keating 2006; Cho & Keating 2009: 466), individual studies also show differences both in terms of the precise nature and strength of the effects observed as well as the domain at which they occur between languages, and also between different information structural conditions (cf. Cruttenden 1997: 33; Fletcher 2010: 529–532). Differences between individual speakers should also not be discounted. For example, Strycharczuk & Kohlberger (2016: 9) note several degrees of durational domain-sensitivity in /s/-realization among their speakers, ranging from differential realization for each of the categories to total insensitivity across them.

## 3.3 Intonation in the autosegmental-metrical model

For the purposes of this work the most relevant phenomena identifying domains of the prosodic hierarchy are tonal ones. In the model of intonational phonology adopted here, the autosegmental-metrical (AM) model of intonation (Pierrehumbert 1980; Ladd 2008), tones are represented on a tonal tier that is autonomous from the segmental tier,[22] and both are independently assigned to a specific level and position in the prosodic structure. This autonomy of tones and segments and their mediation via the prosodic structure (called "tune-text-association") is vital for making sense of how tonal contours relate to segmental strings of different length and morphosyntactic complexity:

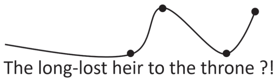
(8) The "incredulity" contour on utterances of different length and complexity

a.

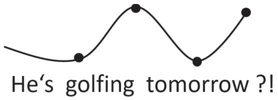


[22] „Autosegmental" does not only mean that tones and segments are on separate tiers, although this was one of the original applications of autosegmental phonology (Goldsmith 1976). Different classes of segmental features are also taken to be located on separate tiers (e.g. McCarthy 1986; Hellmuth 2013; Venditti et al. 2008: 458–459).

b.



c.



In (8), the “incredulity contour”[23] (Ward & Hirschberg 1985; Hirschberg & Ward 1992), represented by the schematic rise-fall-rise above the text, is produced on three different utterances. In (8)a, it is realized on the monosyllabic name *Bob*, appropriate to a context where Bob has just been suggested as the answer to a pending question by someone else but prior to that was deemed by the speaker not to be a likely candidate for the question (e.g. who might cook a complex meal for six in the evening if the only culinary action the speaker has ever witnessed Bob performing was to bake a frozen pizza, and to burn it). In (8)b, it is produced on the complex noun phrase *the long-lost heir to the throne*, felicitous in a context e.g. where the speaker has just been told that Princess Peach, the royal scion, has made a public appearance and the speaker up to that point had believed the princess's whereabouts to be unknown. In (8)c, the same contour is found on the intransitive sentence *he's golfing tomorrow*, e.g. in a context where the speaker knows that an important parliamentary debate is taking place the next day and has just been told that the president will be golfing at the time. The domain for such a contour – not just this one, but all comparable ones – is taken to be the Intonational Phrase (Hirschberg & Ward 1992: 242). This is one aspect why AM assumes tune-text-association to happen via the prosodic structure: the important point is that whether the contour is felicitous depends not on the morphosyntactic makeup of the utterance, nor its length (although it must certainly consist of at least one sufficiently sonorous sound), but only upon the appropriate context conditions and that it be realized on one intonational phrase, in the correct form.

The correct form, however, depends on more than just proportionally adapting the tonal movement to the length of the text, spreading the rises and falls evenly across it: in both (8)b and (8)c, the high and low points in the contour are again

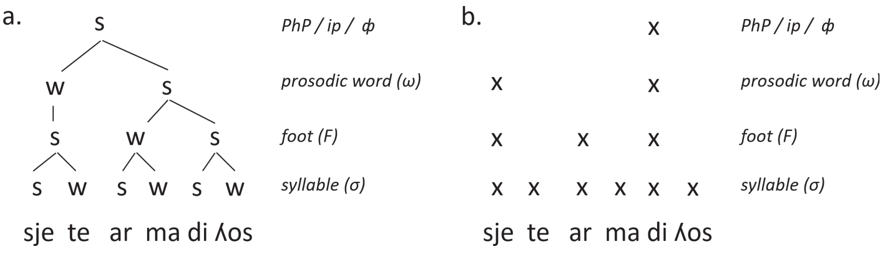
[23] Also discussed under various labels in Liberman & Sag (1974); Liberman (1975); Marneffe & Tonhauser (2019) and elsewhere. Descriptions of its meaning are multifaceted. Hirschberg & Ward (1992) find evidence that increased pitch excursion shifts listeners' perception towards an “incredulity” reading of ambiguous sentences, while perception is shifted towards an “uncertainty” reading when pitch excursion is comparatively decreased. The same contour has also been found to evoke more negative scalar implicatures in listeners when used as an indirect answer to a polar question than a neutral declarative contour (Marneffe & Tonhauser 2019). I will use “incredulity” as a shorthand for the meanings associated with it.

located relative to specific positions in the prosodic structure. AM assumes just two tonal primitives, a high (H) and a low (L) tone. The rise-fall-rise contour under discussion here is usually taken to be made up of a sequence of four tones, LHLH. These are responsible for the pitch movement in the tune: only where a tone is specified is pitch actively manipulated; between tonally specified locations, any tonal movement is due to interpolation (Pierrehumbert 1980: 52). The tones in the phonological representation are not quite directly reflected in the actual pitch contour: they cause tonal targets to exist in the phonetic implementation, which is responsible for realizing the tonal contour together with the segmental string in speech production by, amongst other things, assigning pitch values relative to the overall pitch range of the utterance: high tonal targets, due to high tones, get assigned relatively high pitch values, low tonal targets, due to low tones, get assigned relatively low pitch values. In the LHLH contour, the position of the final H tone is determined simply by the right edge of the intonation phrase: it is as rightmost in the phrase as it can be. Such tones are called boundary tones. For the adequate specification of the other tones' position in the contour, the edges of prosodic units are not enough. That specification must make reference to the metrical structure of the utterance.

### 3.3.1 Stress and metrical structure

Metrical structure (Liberman & Prince 1977; Hayes 1995; Ladd 2008; Calhoun 2010b) assigns strength relations (prominence) throughout the prosodic structure. At each level of prosodic structure, all the constituents of the level below that are dominated by a constituent at that level, are assigned a strength relation such that only one of them is strong (s), and the others are weak (w).

(9) Metrical structure for the phrase *siete armadillos* "seven armadillos"



In (9), the metrical structure for the phrase *siete armadillos* "seven armadillos" is given in branching tree notation (a) and grid notation (b). The two notations are effectively equivalent for our purposes. In both, prominence is built from the ground up. Crucially, it is relative: at the level of the foot, the two syllables [ar] and [di] in *armadillos* are still equal in strength, but at the level of the prosodic word, the foot which contains [ar] is weak relative to the strong one containing [di], which is how the fact that [di] is the stressed syllable in *armadillos* is represented. Strong nodes at one level must always be founded on a strong node at the level below, so that strength propagates upwards. In this way, structures such as (9) can represent three facts at one. Firstly, that [sje] and [di] are the stressed syllables in their respective words. Secondly, that [di] is the strongest syllable in the phrase. And finally, that *armadillos* is stronger than *siete* in it. Even though it might seem so from this example, metrical structures are not maximally binary branching, but n-ary branching in principle, in order to allow the assignment of exactly one strong position for level x-1 at each constituent of level x dominating it (Nespor & Vogel 2007: 7). The metrical structure also assigns prominence at levels higher than the iP/PhP: in the utterance *Juan encontró siete armadillos* "Juan met seven armadillos", it would also put *siete armadillos* in a strength relation with the rest of the utterance. How it would do that however would depend on how the utterance is phrased and on its information structure (cf. section 3.7).

### 3.3.2 Word stress crosslinguistically

The strongest position at the word level is particularly important. It is usually called "word stress", "lexical stress" or simply "stress", although this is a somewhat confusing terminology, since "stress" also signifies the particular way in which many (European) languages configure their prosody in order to mark this position both phonologically and phonetically. This typically includes increased duration and intensity on stressed syllables in comparison to unstressed ones (Fry 1955 for English, Ortega-Llebaria & Prieto 2007, 2011 for Spanish, Gordon & Roettger 2017 for a broad crosslinguistic overview), but the evidence for such claims has to overcome a considerable number of methodological pitfalls (cf. Roettger & Gordon 2017). The most relevant of those is to separate stress from pitch accent. For Spanish in particular, Ortega-Llebaria & Prieto (2007, 2011) find that stressed syllables are longer and louder than unstressed ones even when not pitch accented. Acoustic cues to prominence are statistically robust in many languages, but in individual instances, they can often be missing or misleading. Yet experimental subjects securely identify prominent positions, both at the word level and above also under less than ideal cues and even against cues, up to a point (Terken & Hermes 2000; Bishop 2012; Cole et al. 2019, cf. also the works cited in Calhoun 2010b: 4–5), and the expectation of a prominent position to occur increases attention to detail even when cues are absent (Zheng & Pierrehumbert 2010). All of this indicates that metrical structure should primarily be thought of in relation to this generation of expectations about prominence (Reich & Rohrmeier 2014), which can then be exploited for interpretative information structural effects (Ladd 2008; Calhoun 2010b; Bishop 2012), rather than with regards to its acoustic correlates.

Language-specific stress-related phonological phenomena include e.g. the historical alternation in Spanish in the morphological paradigms of many words whereby /o/ and /e/ occur when a syllable is unstressed, but the diphthongized or pre-glided counterparts /we/ and /je/, respectively, occur when it is stressed, e.g. *apostár – apuésto* "to bet – I bet", *tenér – tiéne* "to have – s/he has" (accent indicates stressed syllable). Historically, there is also a tendency to gradually reduce material following the stressed syllable (syncope), resulting in increasingly few word forms where the stressed position is followed by more than one further syllable. This is a particular instance of the much more general observation that stressed syllables in many languages are most resistant to reduction processes, both historically and in production, relatable also to the fact that more peripheral and sonorous vowels tend to occur in stressed position, while unstressed positions often have a more restricted set of less sonorous vowels in many languages (Crosswhite 2004). For an overview of stress-related phonological processes in various Romance languages, see Meinschaefer (in press). In contrast, there are languages that do not share in any or most of the properties associated with "stress" but still have a comparable unique syllabic position at the word level. The stereotypical example is Japanese, where about 45% of the words in the lexicon (Kubozono 2008: 167) have such a unique lexically specified position which can be anywhere in the word, but it is not marked by increased duration or intensity, or any other phonetic or phonological process except a characteristic pitch movement (Beckman 1986; Venditti et al. 2008). Japanese is usually said to have a lexical (pitch) accent, which is somewhat confusing terminologically, because *pitch accent* is what the postlexically assigned tones linked to prominent syllables are called in AM. Leaving aside the issue of the phonetic and phonological correlate for the moment, word stress has two defining properties that directly derive from the nature of the metrical structure, *culminativity* and *obligatoriness* (cf. Liberman & Prince 1977: 263; Hayes 1995: 29; Hyman 2014: 60). Culminativity is the property that only exactly one syllable can be stressed (= assigned highest prominence) in a word; as we have already seen, this is a property of metrical structure that holds at each level. Obligatoriness means that every lexical content word in a language must be stressable. This is a corollary of every utterance having a metrical structure that assigns strength relations at each level. Having more than half of the lexicon consist of words that are not accented is clearly a different matter than having a closed and relatively small set of function words that cannot be stressed: these can often taken to be clitics and therefore never form prosodic words on their own (see Hualde 2007, 2009 for stress on function words in Spanish). Japanese then does not have word stress under this definition quite independently of its acoustic correlates, and in the analysis of our own data we will see that Huari Quechua shares some, but not all of the properties that lead to this conclusion. That does not mean that Japanese does not have a metrical structure, it just means that the position of its lexical pitch accents does not necessarily have something to do with it. This is borne out by the fact that accented words are not in any sense more prominent in Japanese than unaccented ones. High vowels bearing the H tone of the lexical pitch accent are often reduced (Venditti et al. 2008: 480–481), which would be unexpected in a stress system where the stressed syllable is the most prominent at the word level. The Japanese lexical pitch accent is simply a lexical specification that has the additional property of being culminative at the word level, similar to the distribution of aspirated and glottalized plosives in Cuzco Quechua (cf. (6), see also Hyman 2006: 238–239). Yet Japanese, and other languages that do not have word stress, clearly do have prominence above the prosodic word, so that one word is more prominent than the others in a phrase (cf. Venditti et al. 2008 on Japanese, Roettger 2017: 135 on Tashlhiyt Berber). One theoretical solution to account for this would be to assume that the relevant prosodic domain, the prosodic word, is not headed (Gussenhoven 2018: 398 on Ambonese Malay), which effectively means that metrical structure is not assigned at its level. Another option could be to assume that it is assigned a metrical structure, but this finds no, or only very little, expression in the language under discussion. The last option might be less abstract than it sounds: Hyman (2014: 57–58, 61) develops properties and functions that stress systems may have and shows that typologically, languages do not just cluster around having either all or none of these, but seem to occupy many of the available positions in between.

(10) Properties of a highly stress-oriented phonological system (Hyman 2014: 57–58)
  a. Stress location is not reducible to simple first or last syllable (which could simply represent a boundary phenomenon).
  b. Stressed syllables show positional prominence effects:
    i. Consonant, vowel, and tone contrasts are greater on stressed syllables.
    ii. Segments are strengthened in stressed syllables (e.g. Cs become aspirated or geminated, Vs become lengthened, diphthongized)
  c. Unstressed syllables show positional non-prominence effects:
    i. Consonant, vowel, and tone contrasts are fewer on unstressed syllables.
    ii. Segments are weakened in unstressed syllables (e.g. Cs become lenited, Vs become reduced).

d. Stress shows cyclic effects (including non-echo secondary stresses).
e. Stress shows rhythmic effects lexically/post-lexically (cf. the English ‘rhythm rule’).
f. Lexical stresses interact at the post-lexical level, e.g. compounding/phrasal stress.
g. Lexical stress provides the designated terminal elements for the assignment of intonational tones (‘pitch-accents’).
h. Other arguments that every syllable is in a metrical constituent which can be globally referenced.

The properties in (10) all apply in the case of English, but as Hyman (2014) points out, other languages “care” far less about word stress than English. Spanish could be said to be less interested in word stress at least with regards to properties (10) b, c, and probably e. Thus it is also quite conceivable that a language might rank very low on these dimensions, with a stress system that serves only very few of these functions and leaves barely any trace, so to speak. Some languages have also been argued to have a stratified lexicon, with one set of lexical items exhibiting one configuration of properties, and a second set another (Hyman 2006: 228). We will return to a discussion about languages without stress in section 3.5.3. How the position of word stress is determined in different languages is subject to a vast research body (cf. van der Hulst 1999b; Goedemans et al. 2010 for surveys). Some stress systems seem very simple in their regularity, e.g. stress always occurs on the initial syllable of a word in Hungarian (if its expression is not blocked by postlexical processes, Siptár & Törkenczy 2007: 21–23). Other languages require much more complex metrical algorithms that can differ at least in whether they are quantity-sensitive (sensitive to syllable weight, e.g. as expressed in moras), whether feet in it are left- or right-headed (trochaic or iambic), whether stress is oriented towards the left or the right edge of the word, and whether they systematically ignore constituents at one of the edges (extrametricality, cf. Hayes 1995: especially 54–61, 71–74). Together with additional morphophonological constraints e.g. on how certain classes of affixes interact with stress, this can then lead to systems that at first sight seem quite opaque, such as that of English, but they also allow for the exploitation of stress contrasts for minimal pairs (e.g. English *óbject* (noun) vs. *objéct* (verb), German *Ténor* “thrust of an idea” vs. *Tenór* “tenor voice”, Spanish *páso* “step / pass.1SG” vs. *pasó* “pass.3SG.INDEF”), which is of course impossible when stress is always assigned to the same position in a word. There are also proposals that the primary word stress in many languages is simply specified in the lexicon, but interacting with the metrical structure (cf. van der Hulst 1999a: 73–75). In Spanish, word stress always falls on either of the three final syllables in a content word, unless it is followed by two clitic pronouns (e.g. *cuéntamelo* “tell it to me”), but within this three-syllable window, the difference between stress on the antepenult (proparoxytonic), on the penult (paroxytonic) or on the ultima/final syllable (oxytonic) can encode lexical and grammatical contrasts, as seen above. Among these three options, paroxytonic words are by far the most common overall (almost 75%), oxytonic words are only more common in the subgroup of words ending in a consonant other than /s/, and proparoxytonic words make up just a little less than 6% of a corpus of the 4289 most frequent Spanish words, according to Eddington (2000: 96). Stress assignment in Spanish has been analyzed algorithmically (Harris 1983, 1987; Roca 1997, 1999, 2006; Lipski 1997, among many others) as well as via an exemplar-based lexical approach (Eddington 2000).

Stress assignment can be subject to considerable variation within a language (cf. e.g Behnstedt & Woidich 1985a, 1985b on substantial regular differences in regional varieties of Egyptian Arabic). Lexical items with low use frequency seem often to show more variability in their stress placement, which would indicate that stress position is at least partly lexicalized, but the relationship between the two seems to be quite indirect in many cases, at least in English (cf. Tokar 2017: 19, 21–22). Since stress derivation is not the focus of this study, we will simply treat stress position as a given property of prosodic words in Spanish, part of the knowledge of speakers, and also mostly ignore the level of feet in representations henceforth. For now, we resolve to call the culminative, obligatory highest prominence assigned by metrical structure at the level of the prosodic word "word stress" or "lexical stress", without necessarily implying the concomitant bundle of phonological and phonetic properties attached to stress in many European languages. "Stress accent", on the other hand, will be used when necessary to distinguish it from other kinds of lexical accent like the Japanese one. For the treatment of intonation, word stress is crucial in many languages: the position specified by it provides the other anchoring site, besides constituent edges, for tones in AM. Those tones assigned to positions designated as prominent by the metrical structure are called pitch accents. Before returning to the discussion about how intonational contours are related to prosodic landmarks crosslinguistically and in AM, we will look at what is known about metrical structure and word stress in Quechua.

### 3.3.3 What is (not) known about stress in Quechua

For Quechua, the facts about word stress are far less clear than in many European languages and making efforts at resolving them is one of the contributions of the present work. In overviews based almost entirely on impressionistic data, word stress is said to regularly fall on the penult, or on the initial syllable, in some of the Quechua I varieties (Cerrón-Palomino 1987: 128–129; Adelaar & Muysken 2004: 207–

208; Wetzels & Meira 2010: 352–353). Focusing on those central Peruvian Quechua I varieties closest to Huari Quechua, for the Quechua varieties spoken in the department of Ancash, Parker (1976: 57–60) describes a complex, partly weight-sensitive system: broadly speaking, stress (which he equates with *intensidad* "intensity") regularly falls on the penult, as in other varieties. In exclamatives, it seems to be final, but he tentatively proposes that this is due to a particular exclamative intonation (cf. also Cusihuamán 2001: 79–81, who similary describes the difference between declaratives and exlamatives in terms of intonation for Cuzco Quechua). In particular for the varieties spoken in the Callejón de Huaylas, the following details are given: in phrase-final words, a non-final heavy syllable ((C)VC or (C)V:) receives stress; or the initial syllable, if no heavy syllable is present. In words in a non-final position in the phrase, the initial syllable is stressed, or, in slow speech, one of the heavy syllables, if present, preferentially the prefinal one (Parker 1976: 58–59). Parker (1976: 59–60) also explicitly remarks on the inadequacy of applying notions of stress from Spanish to the Quechua he describes, because he observes a disjunction of phonetics cues: in a word consisting of three light ((C)V) syllables in final position in the phrase, the highest intensity is heard initially, but the highest pitch on the penult. A particularly exceptional case is also described: syllable structure does not allow closed syllables with a long vowel ((C)V:C) in this variety; when they occur morphologically, the vowel is shortened. For some exceptional lexical items, they do however surface word-finally, and Parker (1976: 56) states that they then also cause irregular oxytonic stress. For Conchucos Quechua, Stewart (1984) paints a picture similarly full of complex exceptions and optionality. In general, she also describes word stress to fall on the penult. Unlike Parker, she treats only syllables containing a long vowel as heavy, at least in all words that do not consist of three syllables (Stewart 1984: 195). Quantity-sensitivity acts variably, and differently on words of different length: in words of more than three syllables, a final heavy syllable following a light penult can but need not attract stress, but in bisyllabic ones, it normally does not (Stewart 1984: 199, 201–203). No explanation for this optionality is offered. For trisyllabic words, the situation becomes more complicated. Only in them, closed syllables ((C)VC) also count as heavy. They attract stress in a positional hierarchy: if the penult is heavy, it is stressed. If it is not, then the initial syllable is stressed if heavy. If that is also not the case, then the final syllable may receive stress if heavy (Stewart 1984: 205–206). In all words, when both final syllables are heavy, then the penult is stressed. Words where all three syllables are light, such as *yaku-ta* (water-OBJ) "water (obj.)" or *hacha-man* (tree-DIR) "towards the tree", are variably stressed either on the initial syllable or the penult (Stewart 1984: 198–199). This last case sounds similar to what Parker describes for such trisyllabic words (see above), but unlike Parker, Stewart does not differentiate between intensity and pitch. It is imaginable that some of the unexplained optionality, based presumably on what she as analyzing linguist perceives as prominent, might be resolved when considering phonetic cues individually.

Unlike the previous studies, Hintz (2006) uses instrumental acoustic data in her study on stress in South Conchucos Quechua. She does not find evidence for weight-sensitivity, even though the varieties studied by her and Stewart (1984) overlap.[24] Based on data from four speakers, she finds that stress is word-initial in spontaneous connected speech, with a secondary stress on the penult. In isolated words and under conditions of "emphasis", the situation reverses, and primary stress is now located on the penult, with secondary stress on the initial syllable. Her perceptions are partially confirmed by judgments by three of the same speakers, who were asked to identify prominent positions on a subset of the recorded words. Both in her own judgment and that of the native speakers, variation occurred in general, but in particular in questions and exclamatives, where stress was found to have moved to the final syllable of the word (Hintz 2006: 488). Identical word forms containing the suffix *–shun* (1.PL.INCL.FUT), such as *aywakushun* "we will go/let's go" were found to have different stress patterns depending on whether the speaker intended to express agreement or a request (Hintz 2006: 490). Vowels in a voiceless environment are occasionally observed to be voiceless, with a large majority of them (>80%) occurring word-finally (Hintz 2006: 498, cf. Delforge 2011 for a similar phenomenon in Cuzco Quechua). To explain part of the remaining variation in stress perception, Hintz (2006: 499–500) proposes to view such devoiced syllables as "optionally extrametrical", i.e. optionally not entering into the stress assignment algorithm, giving a number of suffixes where she has found this to be the case. Interestingly, Stewart (1984: 209) also describes extrametricality, but directly relates it to certain morphemes, which do not enter into the stress assignment algorithm, again explicitly optionally. The list of candidate syllables men-

[24] "South Conchucos Quechua" is a label for Quechua I varieties spoken in the Southern part of the Conchucos valley, including Huari province and town, by about 250.000 speakers according to Hintz (2006: 478) Her speakers are from Huaripampa, a small community outside of the town of San Marcos, about an hour's drive by car away from the town of Huari. "Conchucos Quechua", the label used by Stewart (1984), seems to be more comprehensive, but she does not indicate where precisely her data is from. In Stewart (1987: 5–8), she explains however that "Conchucos Quechua" is spoken in an area bounded to the north by the town of Pomabamba, and to the south by that of San Marcos, by about 200.000 speakers, which is largely coterminous to the area described by Hintz (2006) to fall under the label "South Conchucos Quechua". A large part of the data in Stewart (1987) comes from a community close to Pomabamba, which is a car drive of 3–5 hours away from the town of Huari. If the description in Stewart (1987) also applies to Stewart (1984), then her data is thus from the northern edge of the region in which "South Conchucos Quechua" is spoken, Hintz' data is from its southern edge, and the data on which the present study is based is from a more centrally southern part of it. It is not known how much prosodic variation exists within this area.

tioned by the two authors mostly does not overlap, only *–ta* (OBJ) and *–qa* (TOP) are mentioned by both. Primary, secondary and unstressed syllables were found to be significantly different according to the cues of F0, intensity, and duration in Hintz' analysis. However, only for one speaker was there actually a difference between all three conditions, via pitch height; both pitch height and intensity were different in the data of all speakers between stressed and unstressed syllables; duration was the least reliable cue (Hintz 2006: 505–506). This last observation is somewhat expected: as vowel length is phonemic in Quechua I varieties, it is perhaps less likely to be used additionally as a cue to stress.

As already pointed out by Parker (1976) and also by Hintz (2006), many of these observations would benefit from a treatment that separates phonetic cues, intonational phenomena, and word stress. In our own Huari Quechua data, we found a situation where our non-native perceptions were very heterogeneous. In multiple recordings of the same lexical items, impressions of highest prominence did not agree between recordings and hearers; in many words, prominence could be heard on different syllables depending on which cue we focused on. We therefore decided to refrain from relying too much on our own perceptions which are apparently biased from exposure to languages with other prosodic systems. In a study based on the speech of 2 speakers, Buchholz & Reich (2018) investigated whether acoustic cues (pitch height, pitch range, duration and intensity), both individually and taken together, served to make syllable positions stand out from their environment. We did not find consistent evidence across words of different length that either the penult or the initial syllable in the word was cued to stand out relative to the others, except for an indication that the penult seemed to have slightly higher intensity values overall (Buchholz & Reich 2018: 147, 155). CVC-syllables in the word penult also were found to be somewhat lengthened in this position, but not CV-syllables in the same position (Buchholz & Reich 2018: 153–155). On the other hand, we found that pitch height formed a distinctive pattern on phrases, operationalized as material between pauses, such that in general a gradual rise from the beginning of the phrase was observed, until a more abrupt fall taking place across the last two syllables of the phrase (Buchholz & Reich 2018: 151–153).

There are several indications that what has been described as word stress might partially consist of postlexical intonational phenomena. First, from a functional perspective it seems inefficient to have a system of stress assignment that is as complicated as described above yet does not fulfill a distinctive function: all descriptions agree that stress in Quechua is not lexically distinctive. Secondly, Parker (1976) and Cusihuamán (2001) state that the "stress shift" in exclamatives is likely an intonational phenomenon; the same argument could also be made for the "stress shifts" due to pragmatic conditions observed by Hintz (2006), namely that these are possibly due to different pitch accent and boundary tone configurations at a phrasal level. Boundary tones adjacent to a phrase edge might also better explain why voiceless syllables do not seem prominent, because such syllables are insufficient landing sites for tones. Thirdly, Stewart (1984: 207–209) gives evidence that the domain of stress assignment is not isomorphic with the morphological word (cf. (11)).

(11) Conchucos Quechua data points that necessitate a domain of stress assignment larger or smaller than the morphological word, according to Stewart (1984: 208)

a. áchikày mikùkuskínàa  
achikay miku-ku-ski-naa  
*Name eat-MID-ITER-PST*  
"The Achikay [wicked old woman] ate it all up"

b. hákàakunáta  
hakaa-kuna-ta  
*guinea.pig-PL-OBJ*  
"the guinea pigs (obj.)"

c. kèedanantsíkpaqkáqta  
keeda-na-ntsik-paq ka-q-ta  
*stay-NMLZ-1.INCL-BEN COP-AG-OBJ*  
"that which is to stay for us"

d. taríntsikpístsu  
tari-ntsik-pis-tsu  
*find-1.INCL-ADD-NEG*  
"we don't find it at all"

All the examples in (11), where the acute accent (´) marks primary and the grave accent (`) secondary stress, are attested but incompatible, according to Stewart (1984: 208), with her stress algorithm unless the domain of stress assignment is not the individual (morphological) word. In (11)a, the words *achikay*, the name of a wicked woman from folklore, and *mikukuskinaa* "s/he ate", are both full content words, yet if they were assigned stress independently according to Stewart's system, then the initial syllable of *mikukuskinaa* would have to have primary stress. In contrast, (11)b-d lead Stewart (1984: 208) to conclude that the morphemes –*kuna* (noun plural), *ka*- (copula),[25] and "frequently" –*pis* (additive meaning, "also" or "even")

[25] Stewart (1984) treats *kaq* as a "definitivizer" and apparently as a dependent morpheme. As indicated in the glosses, I treat it as a combination of the copula verb *ka*- and the agentive nominalization –*q*, and as also morphologically independent because it can occur on its own (which

form their own domains of stress assignment. If we take a postlexical perspective, both types of cases can be seen as indicative of correlates of phrasing or postlexical pitch accenting rather than of lexical stress.

Finally, formulations such as that "stress can be distributed over several long syllables" (Adelaar & Muysken 2004: 207) also indicate that what has been described as stress in much of the impressionistically based literature on Quechua is probably to be understood as a broad label collecting a number of suprasegmental phenomena, rather than culminative word stress as defined in the previous section.[26] The results in Buchholz & Reich (2018) about an identifiable pitch trend (different from declination) across a phrase-like unit can also be seen as indication that at least pitch movement is sensitive less to stress as a word-level phenomenon, but to a unit above the prosodic word.

Hyman (2006: 246–247) cautions against interpreting any prosodic differences observed in an unfamiliar language via the lens of stress accent familiar to the analyzing linguist from European languages. Citing the example of Indonesian, he proposes the heuristic that "if word-stress is so hard to find, perhaps it is not there at all". Indonesian has received a number of conflicting stress-based accounts, yet van Zanten et al. (2003); van Zanten & Goedemans (2009) showed that what these accounts had taken to be correlates of word stress seems most likely to be pitch accenting independent of word stress, instead seeking proximity to a phrasal edge. Gordon (2014: 111–112) even estimates that perhaps a majority of word stress-based accounts especially of lesser-studied languages actually reflect such systems of phrasal, not lexical, prominence realized via pitch events that seek to occur close to phrasal edges. In this vein, as an alternative hypothesis to the complex and optionality-heavy stress-based accounts reviewed here, I will consider the possibility that Huari Quechua has no word stress, or at least only "cares" very little about it, following the characterization by Hyman (2014). In chapter 6, I will present data in evidence for this hypothesis and also provide an analysis of the intonational phenomena of Huari Quechua that only marginally needs to make reference to a lexical stress position.

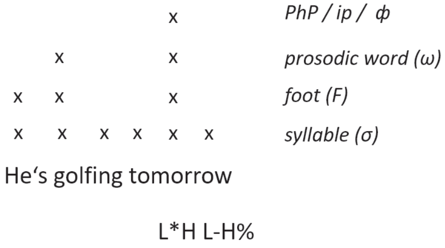
–*kuna* and –*pis* cannot). For an in-depth treatment of the functions of *ka-q* and its derived forms, see Bendezú Araujo (2021).

[26] See also Wetzels & Meira (2010: 314–315), who state that the definitions for phenomena like stress or pitch accent in descriptions of suprasegmental phenomena in South American indigenous languages in general are often vague. They make it clear that much more research is needed before reliable generalizations can be made.

### 3.3.4 Pitch accents and arriving at phonological tones from pitch contours

Returning to the discussion of how to relate the pitch contour and the text to the tones along the example of the rise-fall-rise contour, we can now see that the position of the first rise in (8)a-c is clearly linked to the stressed syllable in the words *Bob*, *heir*, and *golfing*, respectively. Specifically, it is linked to the stressed syllable of that word which is most prominent in the entire IP. In some languages, among them English and German, it is only the most prominent word in a phrase whose stressed syllable has to be pitch accented. Because of this link to a prominent position, such pitch accents are sometimes called "prominence-lending" (e.g. Welby 2006: 364). They have also sometimes been confused with a direct correlate of stress, but this should be kept apart: syllables that are not pitch accented but stressed are still often longer and louder than unstressed syllables. In languages with stress, a stressed syllable is a necessary condition for prominence-lending pitch accents to occur, but not a sufficient one. In other languages, this does not have to be the case. Continuing with our example, we can see that it is the strongest position in the phrase that receives the pitch accent because in *he's golfing tomorrow*, it could also occur on another word than where it does in (8)c, e.g. on *tomorrow* (see (12)).

(12) Metrical structure and rise-fall-rise contour on *he's golfing tomorrow* with highest prominence on *tomorrow*



That would, however, indicate another context:[27] e.g. one in which the parliamentary debate is still taking place the next day, but the speaker has just asked when the

[27] Note that in principle, the same thing could be done with 8b, e.g. moving the highest prominence and also the pitch accent to the stressed syllable on *throne*. However, that would imply a context in which long-lost heirs to different things are contrasted with each other. Stereotypically, long-lost heirs are often those to thrones ("throne" is by far the most frequent collocate to "heir to" in a four-word window to the right of "heir" in the >13 billion-word web corpus English Web 2015 as searched via Sketch Engine (Jakubíček et al. 2013; Kilgarriff et al. 2004)), so such a context is simply not highly likely from our knowledge of the world.

president will be going golfing next and has been told it will be tomorrow. In this context, (8)c would clearly be odder than (12). However, the other way around, (12) in the context given for (8)c would arguably be more acceptable, because there is a general preference for locating the highest prominence rightmost (cf. Ladd 2008: 252). This hints at the complex influence that metrical structure, pitch accenting and context have on the interpretative categories of focus and contrast, which will be treated in more detail in section 3.7. Note that the "incredulity" aspect of the meaning conveyed (and made plausible by the context) does not change, just the location at which alternatives to (8)c and (12) would have to differ from them, respectively, in order to be less incompatible with the speaker's expectations. That it can change, however, is evidence that it does not suffice to simply say that "the position of the first rise is linked to the stressed syllable" in a word, as we just did above. Instead, the linking relationship has to be established between the individual tones making up the rise, the L and the H. If the rise is timed so that the pitch trough preceding it (caused by the presence of the first L tone) extends throughout the stressed syllable of the most prominent word, then the "incredulity" reading obtains, while if most of the rise, sometimes including its peak, takes place on the stressed syllable, then a different reading obtains which is also contrastive but without the additional meaning of "incredulity", cf. (a) and b) in Figure 4.

That this difference in relative timing between text and tune creates a difference in perception that for the majority of speakers seems to be close to categorical[28] was first established by Pierrehumbert & Steele (1989) in a categorical perception and imitation task. Providing contexts that make one or the other realization more felicitous (Pierrehumbert & Steele 1989: 185) and thus demonstrating that

[28] Note that of their five test subjects, one did not reproduce any difference in contours between the two conditions on average. Pierrehumbert & Steele (1989: 190) readily ascribe this to the L*H pitch accent not occurring in the speaker's tonal inventory. It's since been amply demonstrated that intonation can vary considerably also between experimental tasks, speaking styles and various social categories, even within speakers from a locally relatively restricted area, and also show effects of interlocutor accomodation (cf. Face 2003; Henriksen 2013; Romera & Elordieta 2013; Huttenlauch et al. 2016; Huttenlauch et al. 2018; Martín Butragueño & Mendoza 2018; Vanrell & Fernández Soriano 2018, to cite only a few recent works on Spanish). Furthermore, categorical perception experiments have been less successful in other instances (see also Gerrits & Schouten 2004 for a critical assessment of the paradigm). Overall the conclusion seems to solidify that for intonation, the relation between form and meaning is often distributional and probabilistic, rather than strictly categorical: realizations of two meaning categories often show a bimodal yet also clearly overlapping distribution (meaning that statistically, differences with clear trends do emerge, but two randomly chosen instances from each of the categories are relatively likely to be similar or even to counteract the trend); they can also differ substantially in their internal homogeneity, all of which seems to firmly locate variability at the heart of what intonation is (cf. Ladd 2008: 150–154, 2014; Cangemi & Grice 2016; Roettger 2017).

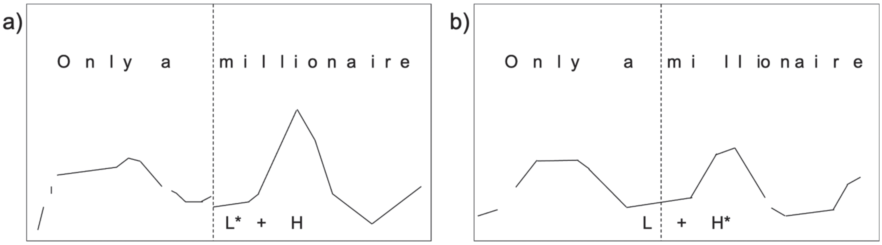


**Figure 4:** Two pitch track contours of *only a millionaire*, with the “incredulity” reading (a, L*H L- H%) and with a contrastive reading without “incredulity” (b, LH* L- H%), adapted[29] from Pierrehumbert & Steele (1989: 182).

the contours in Figure 4 can be intonational minimal pairs, they propose to encode their difference phonologically via specifying which of the two tones in the first rise is linked to the stressed syllable in *millionaire*. This link between the TBU of the stressed syllable and one of the tones of the pitch accent is called association, and it is marked in the phonological annotation by the asterisk after that tone (see Figure 4). Note that even in the contour in Figure 4 b), where the association holds between the stressed syllable and the H tone (LH*), the pitch peak only occurs quite late in [mil]. Association of a starred tone with a stressed syllable does not necessarily mean that the pitch peak always occurs within that syllable, but that the timing of its tonal event can be reliably defined relative to its position (Arvaniti et al. 2000, 2006; Ladd 2006, 2008). How the other types of tones in the contour are related to the segmental string will be treated in section 3.5. It is a reasonably good initial heuristic in AM-based intonation analysis to take pitch turning points (peaks, i.e. local maxima, and shoulders, points which form the left or right edges of high pitch plateaux, for high tones; troughs, local minima, and elbows, points which form the left or right edges of low stretches of pitch, for low tones) as indication for tonal targets. Yet there are many factors that impede this simple equation (cf. Ladd 2008: 134–138; Barnes et al. 2010). For one, it is well known that voiceless obstruents tend to locally raise F0 while voiced obstruents will lower it (so-called *consonantal microprosody* even though its effects can be quite large, cf. Ohala 1978; van Santen & Hirschberg 1994; Hanson 2009; Kirby & Ladd 2016; Ladd & Schmid 2018). These have to be discounted when trying to deduce the presence of tones from the pitch signal. Then, as we have seen, quite subtle but inarguable pragmatic contrasts can affect pitch in also quite subtle but perceptible ways. Phonological context can also

[29] It is problematic to obtain open access reproduction rights for pitch tracks originally published under more restrictive copyright schemes. I therefore use schematized adaptations throughout in these cases.

play an important role, e.g. when the proximity of other tones or prosodic boundaries causes tonal targets to undershoot (*truncation*, which can vary between languages and prosodic positions, cf. Rathcke 2013, 2016; Cho & Flemming 2015). Thus, inferring tones from the pitch signal can only be done with some certainty once the relevant landmarks in the prosodic structure and the tonal inventory of the language are known. Both the relevant landmarks as well as the tonal inventory can vary considerably between languages.

## 3.4 Intonation systems typologically

In the following, we will explore how languages vary in their intonation systems. The goal is to arrive at a variation space mapped out by what is known about prosodic typology. This will help us avoid a treatment of Huari Quechua and Huari Spanish only in terms of what is known about other Spanish varieties or other European languages.

### 3.4.1 Domains of tone assignment

The very broad typological division of languages into "tone" and "intonation" languages[30] is partially based on the domain of tone assignment. Tone languages assign tones at the level of the syllable or the prosodic word, but more crucially, different combinations of segmental strings together with tonal contours can encode distinctive lexical meanings in them. In intonation languages, with which we are here mostly concerned, tones never carry lexical meaning; instead they can convey meanings that are orthogonal to the lexical meaning. In this sense, they are always *postlexical* in these languages. However, while this distinction is useful in itself, it would be misleading to characterize languages as only belonging to either category. See e.g. the studies in Downing & Rialland (2017) on how intonation interacts with tone in African tone languages. Languages that are not tonal according to this

[30] This division is used here only descriptively. The prosodies of many languages pose problems for this dichotomic typology, in particular the so-called "lexical pitch accent languages" among which Japanese, Basque and Turkish have sometimes been counted; but also many African tone languages, or languages like Egyptian Arabic or West Greenlandic, which assign tones to a unit that looks like a word but without conveying lexical meaning. Hyman (2014) proposes a prosodic typology that does not "pigeon-hole" languages, but instead allows to class them along several dimensions via a number of criteria that do not necessarily have to cluster. See also Hyman (2009, 2014, 2018); Beckman & Venditti (2010) for extensive discussion.

definition (and their descriptions) are also sometimes described as assigning tones also at the level of the prosodic word (Egyptian Arabic according to Hellmuth 2005, 2007; West Greenlandic according to Arnhold 2014). It is not entirely clear, however, whether the unit identified as the prosodic word in these descriptions is really distinct from a small phrasal domain such as the AP or PhP.[31] The tonal movements resulting from the placing of (combinations of) tones at prosodic constituent edges help to identify these constituents. However, this identification is not just such a straightforward matter. In the following section we will briefly consider how pitch cues can signal prosodic constituency beyond a delimitative and culminative placement of phonological tones.

### 3.4.2 Tonal scaling and other gradual cues for prosodic constituency

Besides marking the edges of prosodic units with tones, thus delimiting these units and making the domain they belong to identifiable, languages also use further tonal cues for the delimitation of prosodic units. They are related to tonal scaling, the height of a tonal event in relation to another or a reference point. Each human speaker has a typical vocal range for speech, the interval on the acoustic spectrum on which they are most comfortable speaking and at which they have the greatest control over pitch modulations, differing according to gender and individual. Within this range, pitch register, the global height of the pitch of an utterance, can still differ according to social context or emotional state of the speaker (Ladd 2008). Pitch registers can differ either in pitch level, if both the lowest and the highest points of a contour are lowered or heightened together, or in pitch span, if the lowest and highest points are moved closer towards or further away from each other. Within a given register, speakers have a lower and an upper baseline, in relation to which individual tonal events are scaled; this local scaling is called pitch excursion. The baselines normally fall gradually over the course of an utterance, this is called declination and likely due to decreasing subglottal pressure over an exhalation (Pompino-Marschall 2009: 246–247). It has been shown for some languages (Pierrehumbert 1980; Liberman & Pierrehumbert 1984 for American English, Prieto et al. 1996 for Mexican Spanish) that at the very end of

[31] In particular, Arnhold (2014: 221–222) argues for the prosodic word as the domain of tone assignment mainly on the grounds of it being identical to a morphological or syntactic word. However, since individual words can also be realized without tones (Arnhold 2014: 222), the tones could also be analysed as belonging to a phrasal category that is usually, but not always, isomorphic with a morphosyntactic word (which can also be quite long in West Greenlandic). Solving this argument will most likely always depend on one's theoretical stakes.

some utterances, pitch events take place at a lower local level than projected from the declining lower baseline; this has been called *final lowering* and is one of the ways in which pitch scaling can be used to cue the edges of prosodic units. Pitch reset, the return of the pitch level to the initial baseline, is the second. Pitch reset often occurs between larger chunks of speech and is proposed by Himmelmann et al. (2018: 239–240) to be a universal cue, together with pauses, for what they call the "phonetic IP", a universally present speech unit which serves as the basis for the language-specific development of "phonological IPs". The third is systematic manipulation of the pitch level according to prosodic constituency. We will return to this last issue in section 3.6. The suggestion by Himmelmann et al. (2018) that there is both a "phonetic" and a "phonological" IP in some languages is targeted at the open question whether the different types of prosodic cues, such as segmental processes, phonetic cues such as initial strengthening and final lengthening, tonal cues based on the location of phonological tones, and tonal cues based on pitch scaling and pauses, actually align in signaling the *same* prosodic units. This is an ongoing discussion. It is evident that the consequent assignment of segmental phonological processes to prosodic domains has led to a proliferation of them. Prosodic domains also seem to differ in the way they manifest: e.g., unlike other prosodic domains, there is quite robust articulatory evidence for the syllable even though it is perhaps not the domain of phonological processes in all languages (Hyman 2015). Pauses, sometimes seen as the quintessential signal that a part of speech has ended, have been shown to also occur in hesitations or as indicators of increased cognitive load, instead of just at the end of larger prosodic groups and not even always there (Cruttenden 1997: 30–32; Frota et al. 2007; Seifart et al. 2018). Presumably, such points will all have to be considered in order to eventually resolve the issue of how many prosodic hierarchies there are and how they are related. However, since solving this problem is far beyond the scope of this work, I will here follow Frota (2012: 257–258) in her assessment that there is good evidence for a convergence at least between the segmental processes and the tonal cues to prosodic structure, as well as pauses, at least on average. I will take this as an optimistic methodological heuristic, undergoing constant re-evaluation. It is also quite evident that languages differ in the specific weight they assign these various cues. In the following section we will leave these murky waters and consider the differing types of tonal inventories across languages.

### 3.4.3 Tonal inventories

Research based on the AM model, by now the dominant approach, has produced intonational analyses of a considerable number of languages and varieties: see Jun

(2005c, 2014a) and the works therein for a typologically broad overview, Frota & Prieto (2015) and the works therein for an overview over intonational variation in Romance languages, and Prieto & Roseano (2009–2013, 2010) as well as the works therein and in particular Sosa (1999, 2003); Beckman et al. (2002); Estebas Vilaplana & Prieto (2008); Hualde & Prieto (2015) on intonation in Spanish varieties. Jun (2005a) classes languages into whether they assign tones to prominent positions and edges of prosodic domains, or to edges only, and to my knowledge it still holds that no language has been described as not assigning tones to the edge(s) of at least one prosodic domain. However, languages differ considerably with regards to the paradigmatic tonal inventory that they employ according to available descriptions. AM only assumes two underlying tones, a high (H) and a low (L) one, but for many languages it has been argued that bi- or even tritonal (e.g. Gabriel et al. 2010 for Argentinian and García 2016 for Peruvian Amazonian Spanish) complexes are assigned to a single position under certain conditions. As an aid for facilitating comparison between data, a transcription system has been developed based on the AM description of American English by Pierrehumbert (1980), called Tones and Break Indices (ToBI, Beckman & Hirschberg 1994; Beckman et al. 2005). ToBI conventions for a wide range of languages have been developed that all use a related set of symbols. An asterisk following a tone (e.g. H*, L*, a "starred" tone) signifies that the tone is a pitch accent and associated with a tone-bearing unit (TBU) in a prominent position (a stressed or accented syllable). TBUs are usually taken to be either the rhymes of syllables, or moras, depending on the language (cf. Gussenhoven 2004, 2018). Languages sometimes make more fine-grained differentiations than that, e.g. only allowing consonants of sufficient sonority as TBUs (cf. the situation in Tashlhiyt Berber according to Roettger 2017). In bi-or tritonal pitch accents the tones are either connected with a plus sign (+) or simply written together (the convention followed here), with the tone associated with the prominent position marked by the asterisk (e.g. LH*/L+H*, L*H/L*+H, the tones not followed by * are also called "unstarred"). Unstarred tones preceding the starred tone in a more than monotonal pitch accent are called leading tones, those following are called trailing tones. Boundary tones of the smaller phrasal level (ip or AP) are marked by a minus sign (e.g. H-, L-), those of the higher phrasal level (IP) with a percent sign (e.g. H%, L%). Placing the minus or the percent sign after the tone symbol indicates a boundary tone at the right edge, placing it before it one at the left edge of a phrase (e.g. -H, %H vs. H-, H%). The specific status of starred and unstarred tones in pitch accents as well as boundary tones and their phonetics and phonology will be discussed below, in section 3.5. There are further symbol conventions used in some language-specific ToBI schemes, the most frequent of which are ! for downstep and ¡ for upstep (e.g. !H, ¡H), intended to mark a tone whose excursion is markedly lower (downstepped) or higher (upstepped) relative to its surroundings.

SpToBI, the ToBI system developed for Spanish (Beckman et al. 2002; Face & Prieto 2007; Estebas Vilaplana & Prieto 2008; Aguilar et al. 2009; Hualde & Prieto 2015) also specifically marks delayed rising peaks, i.e. pitch accents occurring in prenuclear or prefinal position in which the peak is realized after the stressed syllable but the H tone is taken to associated with it, using the "smaller than" symbol between the L and the H tone (L<H* / L+<H*). Whether tonal scaling (down/upstep) is best encoded paradigmatically at the level of individual tones in all cases and the independent status of the delayed rising pitch accent in Spanish are both somewhat contested issues. The former will be discussed in section 3.6, while we discuss the latter in the next section.

### 3.4.4 Delayed peaks in Spanish and the notion of the nuclear accent

Regarding the issue of rising pitch accents with a delayed (posttonic) peak, it is uncontroversial that in many peninsular varieties of Spanish, prenuclear/prefinal rises are generally[32] delayed in the way described, but in nuclear / final position in the utterance, they are not (see Hualde 2002: 103 for a list, since then further enlarged, of works concurring with this finding). Other varieties, e.g. Cuzco Spanish, do not show this behaviour as strongly (O'Rourke 2005). In those that do, however, a pitch accent that is not delayed on a prefinal word causes the word to be interpreted as being narrowly focused; equally, oxytonic words (stressed on the final syllable) do not usually delay the peak into the following word (cf. Hualde 2002: 103–105; Face & Prieto 2007). A rather neat unifying explanation is proposed by Hualde (2002: 106), based on findings by Nibert (1999) that intermediate phrase boundaries follow the non-delayed accents in prefinal position: the delayed realization of the peak due to the H tone of the pitch accent is blocked by the presence of the boundary tone realized on the last syllable of the word, whether the word is in ip-final or IP-final position. A concomitant assumption is that focused constituents are always directly followed by the right edge of an ip or IP. This is the view espoused by Gabriel (2007), based on further empirical evidence and formulated in optimality-theoretic (OT) terms.[33] We will also follow this view here and

[32] There are some curious exceptions: Prieto & Torreira (2007: 475) state that the prenuclear peak can regularly occur late in the accented syllable, instead of in the postaccentual one, on a first accent in a phrase "when the first accent belongs to an utterance-initial phrase which contains two accents".

[33] Even though he proposes that ip- or IP-boundaries follow focused constituents, Gabriel (2007: 191) formulates doubts that it is really the presence of the ip boundary tone which causes the peak to be realized earlier, based on the observation that proparoxytones in narrow focus also realize

can then also define nuclear accents, with Ladd (2008: 134), as the final and only obligatory accent in the intermediate phrase. We already saw that in English or German, the nuclear accent, the pitch accent assigned to the strongest prominence in the phrase, is often the only one. This is mostly not the case in Spanish, where words in prenuclear position are also regularly pitch accented (Hualde & Prieto 2015: 358). However, using the reasoning and evidence given above, we will take the nuclear accent in the Spanish case to be not only the rightmost one (which it is also in English, all following ones being deaccented, Pierrehumbert 1980: 37), but also on the final word in an ip.[34] That is to say, an ip-boundary is inserted after the nuclear accent in Spanish if it is not already final in the IP. The special status of the nuclear accent is also corroborated by the fact that in many languages, fewer pitch accents are available in prenuclear position than in nuclear position (cf. Ladd 2008: 286).[35] This is reflected in the Spanish ToBI system, which only recognizes two different prenuclear pitch accents (L*H and L<H* (which we here take to be a variant of LH*)) in prenuclear position, but five in nuclear position (cf. Hualde & Prieto 2015). We will return to the issue of focus and nuclear accent further below (section 3.7.3).

the peak within the stressed syllable. However, Prieto et al. (1995: 438–439) find that at least for one of their two Mexican Spanish speakers, the presence of a following phrase boundary clearly reduces peak delay also on proparoxytones, showing that tonal crowding effects can persist also at distances larger than single syllables. Their other speaker does not show this effect, highlighting the importance of considering individual variation in these discussions.

[34] This is not the same as saying that *focus* is always rightmost in a *sentence* or in an *utterance* in Spanish, the categorical claim by Zubizarreta (1998) which has since then been amply shown to be unsubstantiated (Face 2001; Hualde 2002; Gabriel 2007; Muntendam 2010; Hoot 2012, 2016; Uth 2014; Vanrell & Fernández Soriano 2018; Dufter & Gabriel 2016 and the works cited therein). The nuclear accent is here defined prosodically as a statement about pitch accent location and metrical structure in an intermediate phrase. In the conception espoused here, this only relates probabilistically to the interpretative category of focus (Calhoun 2010b). The optimality-theoretic constraints that relate focus to prominence and phrasing are also formulated as violable in Gabriel (2007: 235, 278).

[35] As an example for a case where this does not hold as strongly, cf. Gussenhoven (2005: 121) who states that standard Dutch has a large choice of prenuclear pitch accents. However, in a sequence of several prenuclear accents within an IP, they tend to be of the same type, thus still largely conforming to the assertion by Ladd (2008: 286) that the type of prenuclear accent within a single "tune is – or may be – a single linguistic choice".

### 3.4.5 Differing degrees of combinatorial freedom in the tonal make-up of prosodic constituents

Intonational descriptions of languages also differ with regards to the (number of) phrasal domains at which tones are assigned and whether edge tones occur at the right or the left edge of a prosodic constituent. For example, German as analysed in Grice et al. (2005) has (monotonal) tones assigned to the right edges of intermediate phrases (iP; L-, H-, !H-) and to the left (%H) as well as right edges of intonational phrases (IP; L%, H%, ¡H%, not all of which are attested in free combination with the ip tones), as well as six monotonal and bitonal pitch accents (L*, H*, LH*, L*H, HL*, H!H*) available for assignment at prominent positions. Unangan (Eastern Aleut), according to Taff (1999), makes do with a single monotonal pitch accent (H*), three monotonal (H-, L-, ¡H-) and one bitonal (LH-) ip-boundary tones and two monotonal (L% and H%) IP-boundary tones; all of the boundary tones occur at the right edge of their respective phrases. Tokyo Japanese assigns two tones (%LH-) at the left and one tone (L%) at the right edge of the Accentual Phrase (AP), a single bitonal pitch accent (H*L) on lexically accented syllables and provides a choice between four (H%, LH%, HL%, HLH%) boundary tone combinations (mostly) at the right edge of the IP, according to Venditti (2005); Venditti et al. (2008); Igarashi (2015).

"Peninsular" Spanish, in the description by Hualde & Prieto (2015), has seven monotonal and bitonal pitch accents (L*H, L<H* (these two only prenuclear), L*, H*, LH*, HL*, L¡H*), three monotonal boundary tones at the right edge of the ip (L-, H-, !H-), and six monotonal and bitonal boundary tones at the right edge of the IP (L%, H%, !H%, LH%, L!H%, HL%). In the only available descriptions of the (declarative) intonation of a Quechua variety in AM terms, O'Rourke (2009) tentatively proposes that Cuzco Quechua assigns an –L tone at the left edge and optionally a H- or L- tone at the right edge of the ip, and has an inventory of two pitch accents, L*H, and LH*, the first of which occurs phrase-medially and the second phrase-finally. O'Rourke (2005: 182–201), discussing interrogative intonation, does not add a further pitch accent. IP-finally, L% is attested most often, also in questions, but H% does occasionally occur. As stated before, the intonation of both Huari Quechua and Spanish is as yet unexplored. Thus the range of variation in tonal inventories is quite large. A further relevant aspect pertains to the combinatorial freedom of these tones in the different languages. While in Spanish and German, there are several options to choose from at each point in the prosodic structure that a tone could be assigned to, in Japanese, an AP will always begin with a rise, either because of the AP-initial %LH- tones or because of a combination of AP-initial L plus the high tone of the lexical pitch accent H*L if the first or second syllable in the AP is lexically accented, in which case the AP-initial H- is superseded by the lexical H (Venditti 2005). In addition, the only pitch accent is H*L. That is to say, there is less combina-

torial freedom in the Japanese system compared to the Spanish or German one (cf. Igarashi 2015: 559–560). The Japanese case is comparable to other languages like French (Jun & Fougeron 2000, 2002) or Korean (Jun 1993, 2005b) that have also been described as having an AP with quite fixed initial and final boundary tones, and to the Cuzco Quechua case in the description by O'Rourke (2009). Quite obviously, less optionality at a given structural point means fewer possibilities for exploiting these options to encode meaning differences, but on the other hand, it means that the unit that is delimited by these tones (here, the AP) is made more easily recognizable, because its edges are signaled by a less variable set of cues. Paradigmatic variability at a given point in the prosodic structure therefore stands in a tradeoff relationship with the delimitation of larger prosodic units.

From the point of view of information transmission, this is a tradeoff between having many paradigmatically variable cues at each point, each signaling some differential information, on the one hand, and having cues converge on only a few units that are signaled so that listeners may adjust their expectations to these units, and also recognize when deviations from this convergent pattern are exceptionally exploited for the coding of meaning, on the other. Redundant coding is part of what constitutes recognizable structure in languages, and it is especially necessary in the acoustic signal for human communication to be robust under noisy conditions (Shannon 1948). Paradigmatically variable prosodic cueing thus is capable of encoding more information at each structural point, but it runs a greater risk of information loss. By having a greater combinatorial freedom for tones, such cueing systems can also encode information syntagmatically at the level of the tones, but we will also see that it is less easy to encode information syntagmatically via the sequence of the larger units that are delimited by these tones because the make-up of the tonal cues that signal their edges is more varied. Paradigmatically fixed cueing, on the other hand, while being less free at each individual structural point, can invest more resources in ensuring lossless transmission and it is freer to signal information syntagmatically at the level of the units delimited by the tones, e.g. by varying the size of the few units that are robustly cued in relation to the morphosyntactic units that are contained in them (*phrasing*).

### 3.4.6 "Phrasing" and "Accenting" languages: An appropriate typology?

In fact, Korean, Japanese and French have occasionally been called "phrasing" languages (cf. Igarashi 2015, also called "edge-prominence languages", in the typology developed by Jun 2005d, 2014b), because they optimize the (accentual) phrase in this way and use phrasing to a large degree for the encoding of information structure. Languages like English, German, or Spanish, on the other hand, have been called “accenting” languages (“head-prominence languages” in Jun 2005d, 2014b), because information structure and other pragmatic meanings are mainly encoded via the placing of tonal accents and choice of tones. Venditti et al. (1996) show that indeed, very similar information structural effects are conveyed via (de-)phrasing in Japanese and Korean, on the one hand, and by (de-)accenting in English, on the other, which supports the point that these are divergent coding strategies but with similar purposes (cf. Ladd 2008: 279–280). While deaccentuation in “accenting” languages means that pitch accents are not realized on constituents where they would otherwise be expected, often following a focus, dephrasing in Japanese or Korean signifies that AP-tones are not realized following a focus, so that the last AP in an IP is the one beginning with the focused constituent. This is also called “prosodic subordination” in Venditti et al. (2008), which is an apt term also for the phenomenon observed in the “accenting” languages. They also note that in Japanese, “full” dephrasing, i.e. the total deletion of AP tones and lexical pitch accents does not always take place and that often instead, tonal movements on focused constituents are scaled higher, followed by a substantial reduction in the local excursion of subsequent movements, with the difference to “full” dephrasing being gradual (Venditti et al. 2008: 484–485).

The same is probably true for “accenting” languages, with a gradual continuum between “deaccentuation” and “compression” (Kügler & Féry 2017; Vanrell & Fernández Soriano 2018). In general, the dichotomic typology between “phrasing” and “accenting” languages is probably misleading, because on the one hand, it depends to a certain degree on the descriptive approach chosen: French e.g. has also been described in a more “accenting” framework (Post 2000, 2002), and under closer inspection, individual languages seem to nearly always occupy an intermediate position between the two theoretical endpoints (cf. Igarashi 2015: 561). It has also been argued that phrasing always plays a role for the encoding of information structural categories in nearly all languages, and that only in some, accenting is an additional optional means for it (Féry 2013), but Kügler & Calhoun (2020) point out that there are counterexamples to the universal validity of this claim. Furthermore, the dimensions of freedom of combinatoriality and “phrasing”-“accenting” must clearly be kept apart: Egyptian Arabic realizes pitch accents on the stressed syllables of prosodic words, and is thus an “accenting” language. However, unlike in German or Spanish, there is no space for accent choice: the only available pitch accent is LH*. Additionally, this pitch accent usually occurs on each prosodic word, no matter its information structural status (Hellmuth 2005, 2007; Chahal & Hellmuth 2014). That means that it also has few options to encode pragmatic meaning via accent choice (cf. Igarashi 2015: 561–562), and it tonally optimizes a recurring prosodic unit, like the “phrasing” languages. Madrid Spanish shares some, but not all of these characteristics: it is usually described, unlike German or English, as also realizing a pitch accent on each prosodic word, at least in prenuclear position (Hualde & Prieto 2015: 358), while postfocally, deaccentuation or compression does often take place (Hualde 2002; Face & Prieto 2007; Gabriel 2007; Torreira et al. 2014; Vanrell & Fernández Soriano 2018). In addition, as mentioned above, there are only two pitch accents available for words in prenuclear position, compared to the choice between five for the nuclear position (the last accent in a phrase, cf. Hualde & Prieto 2015). Thus Spanish seems to use tone distribution both to optimize words to a certain degree (by the recurrent realization of rising accents prenuclearly), while the placement of the nuclear accent in rightmost position in a phrase also serves as a delimiting cue for that phrasal category (cf. Kügler & Calhoun 2020 for a similar view), and it also maintains quite a large bit of combinatorial freedom through its tonal inventory. We can see some of the typological dimensions discussed exemplified in the utterances from Tokyo Japanese, Egyptian Arabic, Lima Spanish and German (Figure 5, Figure 6, Figure 7, Figure 8).

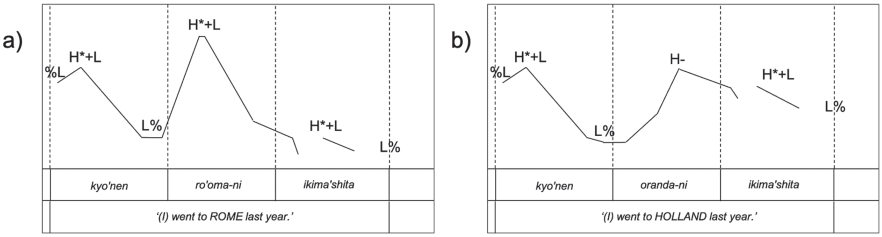


**Figure 5:** Two declarative Tokyo Japanese utterances with dephrasing, *kyo'nen ro'oma-ni/oranda-ni ikima'shita* "(I) went to Rome/Holland last year", adapted from Venditti et al. (2008: 485). The accent (') marks the location at which the fall from the lexical pitch accent takes place. Both utterances cue focus on the location, which in a) bears a lexical pitch accent (*ro'oma* "Rome") but in b) doesn't (*oranda* "Holland"). Note that the AP-initial H- does not get expressed separately if the mora bearing the pitch accent is initial or peninitial in the AP.

They all involve instances of "prosodic subordination" – dephrasing or deaccentuation – in which the highest prominence in the IP occurs on a prefinal constituent and this is cued, to differing degrees, by increased pitch scaling on that constituent, and tonal compression up to deletion on following ones. This asymmetry in scaling effects an interpretation of focus on the most prominent constituent, in accordance with the elicitation contexts in which the other elements were given but the element corresponding to the most prominent one in the utterances was asked for.[36] Of all

[36] Cf. Venditti et al. (2008: 484–486) for elicitation circumstances of the Japanese examples, Féry & Kügler (2008: 683–684) for those of the German one and Hellmuth (2006: 270–273) for those of

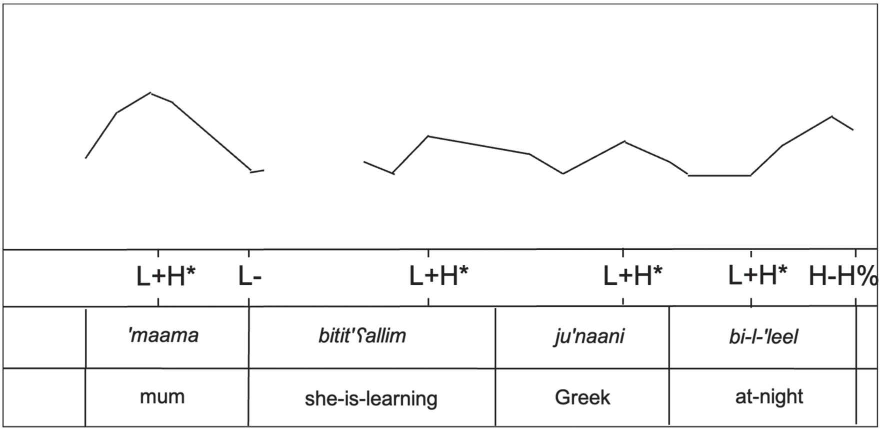


**Figure 6:** A declarative realization of *maama bititʕallim junaani bi-l-leel* “mum is learning Greek in the evening” by an Egyptian Arabic speaker, adapted from Chahal & Hellmuth (2014: 399). The accent (') is placed before the stressed syllable. The context is intended to yield a contrastive interpretation on *maama* “mum”, with *ju'naani* “Greek” given, yet there is no deaccentuation, only some compression.

the examples, it is obvious that the Egyptian Arabic (Figure 6) one shows the least effect of this context, with the contrastively focused *maama* clearly receiving the highest scaled pitch accent, but the following given ones still robustly accented with the only available LH* pitch accent. We can also see a gradient, rather than a categorical difference between the Spanish and German examples: before the focused constituent *dem Rammler* in the German example (Figure 8), tonal movement is compressed but the constituent *der Hammel* is still pitch accented (the degree of pitch compression/scaling reduction here differs depending on whether the prefocal element is given or not, Féry & Kügler 2008), while following it, the verb is completely deaccented. In the second Spanish example (b) in Figure 7), the two prefocal pitch accents on *cuatro* and *policias* are clearly not compressed, both in comparison to the prefocal constituent in the German example, and to the postfocal pitch accents in the first Spanish example (a) in Figure 7). Those are compressed, but at least the first two, on *policias* and *arrestaron*, are still undoubtedly present and not deaccented. Full deaccentuation most likely takes place on the final word *sospechosos* in the second Spanish utterance. Note also the presence of the low

the Egyptian Arabic one. The Spanish examples are spoken by Raúl Bendezú Araujo and recorded by the author. They were purposely recorded to elicit differing intonation contours dependent on question context on the same sentence. This is of course not a very natural procedure, since in normal conversation, (null) pronominalization and elision of given constituents would likely result in different segmental strings between the two context conditions.

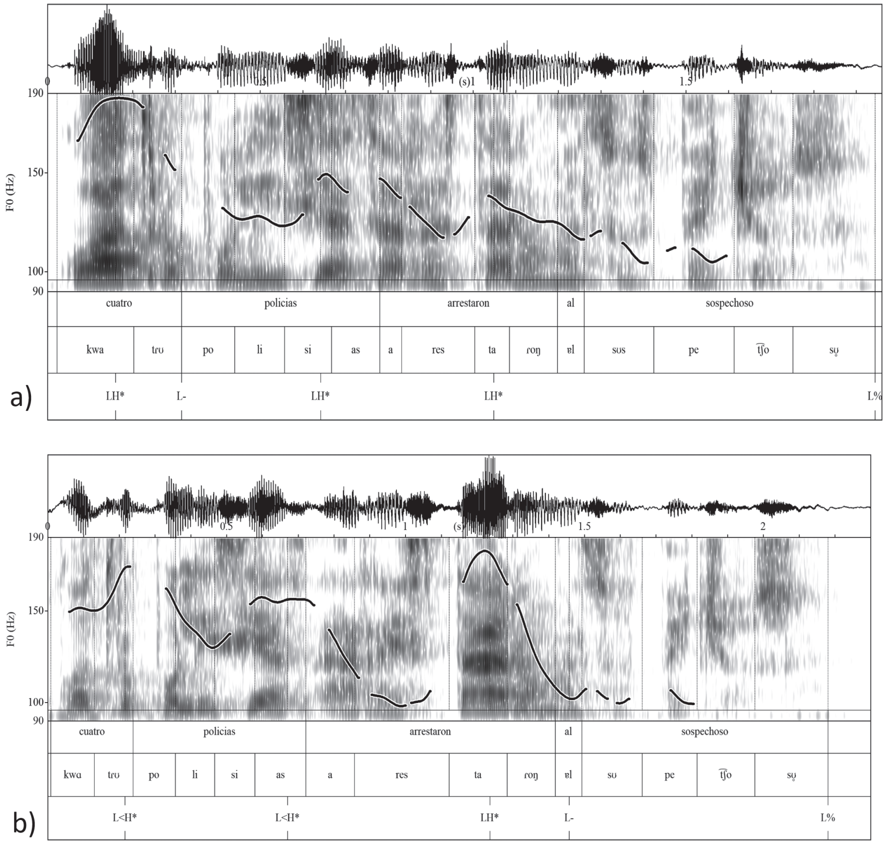


**Figure 7:** Two elicited declarative realizations of *cuatro policias arrestaron al sospechoso* “four police arrested the suspect” by a Spanish speaker from Lima. a) is an answer to the question “how many police arrested the suspect?” while b) answers “what did four police do to the suspect?”.

ip-boundary tone (L-/Lɸ) directly following the focused constituents in both the German and Spanish examples, causing pitch to steeply drop after reaching the accent peak, and resulting in an extended final low stretch in the German and the second Spanish example. This is where we can make out the most striking difference to the Japanese examples (Figure 5). In both Japanese examples, the tonal movement on the AP bearing highest prominence is scaled high, which is comparable to what happens in the examples from the other languages. However, only in the first one, where the highest prominence is on the lexically pitch accented *ro'oma* “Rome”, pitch then also drops to an overall low level, like in the German and Spanish examples – this is here caused by the increased scaling (relative lowering) on the L tone of the H*L lexical pitch accent. The lexical pitch accent on the verb

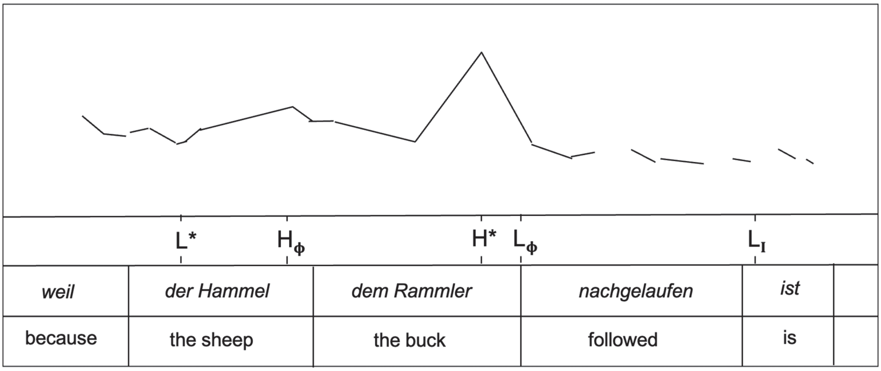


**Figure 8:** Declarative realization of *weil der Hammel dem Rammler nachgelaufen ist* „because the sheep ran after the buck" by a German speaker, adapted from Féry (2017: 155). Focus is cued on *Rammler "buck"*, note the difference between prefocal compression and postfocal deaccentuation.

*ikima'shita* is either very reduced or deleted. In the second one, such a steep drop cannot be observed. Instead, pitch actually stays quite high after reaching the high tonal targer of the AP-initial LH- boundary tone, and only gradually sinks down to the IP-final L%, with very slight additional movement caused by the pitch accent on the verb, which is strongly compressed. Here we can see that dephrasing in this case really results in the absence of the tones belonging to the following phrase, and that the high pitch following it is not cueing prominence (cf. Venditti et al. 2008: 484–486).

Another important difference is the directedness of phrasing: in Japanese, the focused or most prominent constituent begins a new AP that then extends until the end of the ip or IP, whereas the analyses for German and Spanish place this constituent at the (right) end of an (intermediate) phrase which might have begun considerably earlier (further to the left). Note that these differences in directedness or prosodic headedness are probably again tendencies, rather than absolutes, as Beckman & Pierrehumbert (1986: 285) already point out. A further similarity could be seen in the tendency shown in Japanese, Spanish, and German to make the tonal movement on the focused constituent the last one in the IP before the final boundary tone, or at least the last one of comparable pitch scaling. Thus there are real differences in the realization of how prominence asymmetries are encoded intonationally in these four languages (Egyptian Arabic, Japanese, Spanish, and German), but it also seems that not all of them are actually categorical, and that in order to adequately analyze them, it is also necessary to pay considerable attention to gradient factors, such as pitch scaling. The description in O'Rourke (2005, 2009) suggests that Cuzco Quechua shares attributes with "accenting" languages in that it is said to realize pitch accents at prominent positions. However, it also shares some with “phrasing” languages in that it has both initial and final ip-level boundary tones, although in “phrasing” languages, the relevant domain is usually taken to be the AP. The relative scarcity of different pitch accents is also more reminiscent of “phrasing” languages. No evidence is found for dephrasing or deaccentuation. For Huari Quechua, comparable findings have not been made. My own analysis will show that prominence asymmetries are in fact encoded via deaccentuation/dephrasing, and also argue that both the Spanish and the Quechua data from Huari defies easy categorization as either “phrasing” or “accenting”.

### 3.4.7 Restrictions on combinatorial freedom

Let us now return to a consideration of tonal inventories. Even in languages with a large inventory, combinatorial freedom seems to be actually more restricted than might be assumed. In this section, I will present evidence that argues that there are strong constraints on reducing the number of contours at a phrasal level, and that this aids both in the identification of contours and in consolidating this level as the domain at which intonational meaning is conveyed. The intonational finite-state grammar proposed by Pierrehumbert (1980) (see Figure 9b) in principle allows for any combination of pitch accent plus ip-boundary tone / phrase accent plus IP-boundary tone (in this order), as chosen from the inventory at each point, to be realized in American English. At each node in the sequence, there is a free choice between each of the options available, leading to 8 x 2 x 2 = 32 possible combinations of pitch accent plus ip and IP boundary tones for American English (leaving out the initial boundary tone), compared to 2 x 5 available combinations for Tokyo Japanese, according to the inventory in Figure 9a.

Empirical reality paints a somewhat different picture. Dainora (2001, 2002, 2006) shows, using a corpus of radio speech in American English comprising about 1200 IPs, that out of 100 possible sequences of prenuclear pitch accent plus nuclear pitch accent plus ip boundary tone plus IP boundary tone, only 44 are attested; that out of 20 nuclear sequences[37] – nuclear pitch accent plus ip and IP boundary tone – only 18 are attested and four of them (H* L- L%, H*L- H%, LH* L- H%, LH* L- L%, in order of decreasing frequency) account for nearly 80% of all attested nuclear sequences (Dainora 2006: 112–113). The identity of a pitch accent plus an

[37] The pitch accents H*L and HL*, contained in Pierrehumbert’s original inventory, are dropped in Dainora’s investigation because they were originally introduced only to trigger downstep and are thus not taken to constitute independent categories, see Dainora (2001: 99). Downstepped H tones are also collapsed together with non-downstepped ones, because Dainora (2001: 46–69) finds that they come from the same distribution.

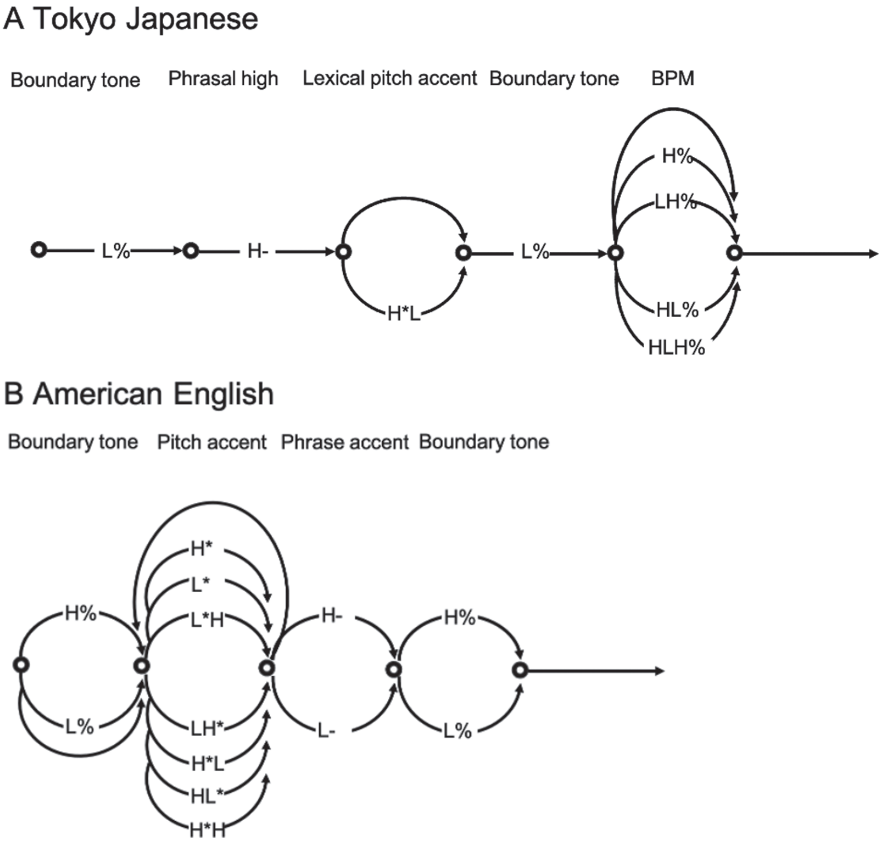


**Figure 9:** Comparative tonal finite-state grammars of Tokyo Japanese (A, above) and American English (B, below, originally from Pierrehumbert 1980: 29), adapted from Igarashi (2015: 560). The arrow pointing to the left in b) indicates that the pitch accent node can be repeated.

ip boundary tone is also a considerable predictor for the identity of an IP boundary tone, so much so that "in some cases the nature of the boundary tone is almost predetermined by other parts of the phrase" (Dainora 2002: 4). A further finding is that adjacent sequences of H tones are quite rare, irrespective of whether they are pitch accent or boundary tones: less than 16% of all attested contours contain such sequences (cf. table 1 in Dainora 2006: 113).[38] A similar observation, albeit

[38] To bring this into perspective: even without assuming anything about what part of the nuclear contour a tone T belongs to, two tones in a sequence will have a probability of 0.5 * 0.5 = 0.25 of being HH, which is already more than the attested 16%. If we take the five pitch accents Dainora assumes as fundamental, atomic units, then 12/20 = 60% of all possible nuclear contours (pitch

without data on the frequencies of individual contours, is made concerning Bengali by Hayes & Lahiri (1991). According to their analysis, Bengali has in nuclear position a pitch accent (L*, H*, LH*) followed by an optional iP-boundary tone (a PhP in their terminology, either $H_P$ or $L_P$), plus an optional IP-boundary tone ($H_I$ or $L_I$) and one obligatory IP-boundary tone ($H_I$ or $L_I$).[39] Free combination would yield 54 possible tone sequences, but only eight of those are attested, each with a distinct intonational meaning (Hayes & Lahiri 1991: 72). Quite comparably, the summary of attested realizations in Peninsular Spanish of *bebe la limonada* "s/he is drinking the lemonade" in Hualde & Prieto (2015: 389), modestly said to list "some possible intonations" of that sentence, but based on a very comprehensive overview over the intonation research on Spanish in at least the last two decades by two of the leading experts in the field, only gives ten such combinations of prenuclear plus nuclear pitch accent and IP-boundary tone as well as an optional iP-boundary tone (see Table 3).

**Table 3:** Attested intonation contours of *bebe la limonada* "S/he is drinking the lemonade" in Peninsular Spanish, adapted from Hualde & Prieto (2015: 389).

| | **_be_**_be_ | _la limo_**_na_**_da_ | **Function** |
|---|---|---|---|
| a. | L<H* | LH* L% | *Statement or command* |
| b. | L<H* | L* L% | *Statement or command* |
| c. | LH* L- | L* L% | *Statement or command with emphasis on first word* |
| d. | L<H* H- | LH* L% | *Statement or command with emphasis on second word. First word is topic.* |
| e. | L<H* | LH* L!H% | *Statement of the obvious (see also echo-question expressing surprise)* |
| f. | L*H | L* H% | *Information-seeking question* |
| g. | L<H* | LH* HL% | *Confirmation-seeking question* |
| h. | L<H* | L¡H* L% | *Echo question (surprise etc.)* |
| i. | LH* | H* H% | *Quiz question* |
| j. | L<H* | HL* L% | *Insistent explanation / Insistent request* |

This should be compared against the inventory, as stated above (cf. section 3.4.5), of 2 prenuclear and 5 nuclear pitch accents, six IP-boundary tones and three ip-boundary tones. The SpToBI-tradition has not followed the convention adhered to in the descriptions of American English or German that an IP-boundary tone must always be preceded by an ip-boundary tone (cf. Figure 9 and Pierrehumbert

accent + phrase accent + boundary tone) contain *at least* one sequence of an H followed by another H (some, like LH* H- H% or H*H H- H%, contain two or even three). From this expectation, the attested 16% clearly constitute a marked deviation.

[39] This is the same as saying that IP boundary tones are either monotonal or bitonal, but generally obligatory, as done e.g. in analyses of Spanish.

1980; Beckman & Pierrehumbert 1986 for English, Grice et al. 2005: 68 for German). Instead, the ip boundary tone is often taken to occur somewhere before the nuclear pitch accent, but it is effectively optional (see Table 3). This means that the possible combinations must include four options for ip-boundary tones, the three tonal options (L-, H-, !H-, cf. section 3.4.5) plus the one that no tone occurs. Calculated like this, the attested combinations in Table 3 are only ten out of a possible 2 x 4 x 5 x 6 = 240 (taking the nuclear pitch accent inventory as not including L*H and L<H*, which are only attested as prenuclear from Table 3). Counting only possible nuclear contours would still yield 30 possible combinations, of which only eight are attested. It is doubtlessly the case that even with the intensive research done on the intonation of "Peninsular" Spanish, the information presented in Table 3 is far from definite. It is quite likely that with even more empirical investigations, some additional contours not listed there might turn up, and it is certainly the case that the specification of the form-meaning correspondences given could bear considerable refinement (cf. Fliessbach 2023). These considerations notwithstanding, the discrepancy between what is theoretically possible and what is attested in terms of tonal combinations seems to be equally striking for this variety of Spanish as it is for American English or Bengali.

One explanation for this discrepancy is provided in the case of Bengali by Hayes & Lahiri (1991: 72–75) who point out that nearly all unattested contours are those that include a sequence of like tones (L-L or H-H) and therefore propose that the obligatory contour principle (OCP) is effective in Bengali "at the level of the entire tune". The OCP is a prohibition against sequences of like tones based on observations first made in Leben (1973). It is discussed and named as such in Goldsmith (1976). There and in Odden (1986), it is established that even though evidence for a principle like the OCP seems abundant in most of the African tone languages studied there, it cannot be said to hold universally and that asymmetries exist in how the OCP is applied at different levels of linguistic structure, also in terms of its directedness. The OCP has since been extended to non-tonal aspects of linguistic sound structures (McCarthy 1986; Frisch et al. 2004), and is functionally motivated in a broader context as a tendency to avoid sequences of perceptually overly similar elements (cf. Boersma 1998, especially ch. 18; Flemming 2002, 2004). Boersma (1998: 416) also describes a tendency against the repetition of similar articulatory gestures, which he takes to be part of the articulatory functional motivation for the OCP. The language-specific aspect of the OCP is also acknowledged in Hayes & Lahiri (1991: 74), and it also shows in the examples we have discussed: while in Bengali, any sequence of like tones seems to be prohibited, the data from American English seem to suggest a robust but violable tendency to avoid sequences of high tones, yet sequences of low tones are quite frequent.

For the Spanish data from Table 3, the same argument cannot be as easily made: six of the prenuclear + nuclear contours contain sequences of like tones (b, c, d, g, i, and j), of those, four contain sequences of high tones (d, g, i, and j); of the nuclear contours themselves, three (b/c, g, and i) have sequences of like tones, two of those (g and i) consisting of high tones. Can we thus say that Peninsular Spanish entirely ignores the OCP, and if yes, what else might account for the discrepancy between possible and attested tone sequences? First of all, it is important to recognize that for all three languages we are discussing here, the facts we have seen suggest that at least at some level, it is the tonal sequence as a whole, without an identification of individual tones as belonging to pitch accents or boundary tones, that is important as a domain to which constraints or processes might apply. That assumption is inherent in calling upon the OCP to explain the (relative) absence of sequences of like tones, no matter whether they are pitch accents or boundary tones, both in American English and in Bengali, for which this assumption is explicitly acknowledged in Hayes & Lahiri (1991: 74). It is also inherent in the practice, followed in the descriptions of all three languages here as well as several others[40] of assigning intonational meanings to tonal tunes consisting of a combination of at least nuclear pitch accent plus boundary tone(s), the nuclear contour. Especially for Spanish, it has even been proposed that for the identity of the tune as a whole, it is somewhat irrelevant which of the tones it is made up of form the nuclear pitch accent, as long as they stay in the same order (Torreira & Grice 2018). The domain for such a holistic tune, linked to some pragmatic meaning, is sometimes taken to be the IP (Hayes & Lahiri 1991: 52). This finds its parallels on the meaning side by statements to the effect that an IP encodes "an informational unit or sense unit" (Heusinger 2007: 280, referring to concepts developed in Halliday 1967 and Selkirk 1984, respectively), that it represents one "informational chunk" which is processed as such in production and perception, similarly across languages (Himmelmann et al. 2018: 236 and the works cited therein), or that it conveys exactly "one new idea" (Chafe 1994: 108–119). Actually, Chafe (1994: 57–58) proposes the „one new idea constraint" to hold for what he calls „intonation units", which he states are more similar to the intermediate phrase than the intonational phrase in the description of Beckman & Pierrehumbert (1986). Since we here assume with Ladd (2008: 134) that it is the intermediate phrase, and not the intonational phrase, which defines

[40] Cf. Grice et al. 2005: 72 for German, Prieto et al. 2015: 45–46 for Catalan, Gili-Fivela et al. 2015: 191–193 for several varieties of Italian, Frota et al. 2015: 278–279 for European and Brazilian Portuguese, Arvaniti & Baltazani 2005: 95 for Greek. The practice is not applied to Standard Dutch in Gussenhoven 2005: 137, who also claims that there are no restrictions prohibiting any of the possible combinations of nuclear pitch accents plus boundary tones (however, of the 24 possible resulting combinations, he only attests to 18, stating that the others are probably rare).

the nuclear accent and the nuclear contour, Chafe's idea is still compatible. Thus, the relevance of a holistic tune whose domain is the ip/IP seems relatively well grounded. If considered from the perspective of the (nuclear) tune, something like the OCP or rather its somewhat more abstract correlate of similarity avoidance in order to create perceptual contrasts (Flemming 2008) can probably still be made use of to explain the discrepancy between possible and attested tone combinations also in Spanish.

While we cannot provide a definitive answer here, the evidence seen so far suggests that only a relatively limited number of tunes at the level of the nuclear contour can be usefully associated with meanings in a language. Evidence that unfamiliar tunes can slow down processing (Braun et al. 2011) also suggests a way that tune frequency at this level can be brought into relation with notions like markedness. Potentially, it would make sense that a small number of acoustically least eventful contours are used most frequently, in a broad range of situations covering both unmarked meanings and more marked meanings when they are retrievable from context. This opens up a space for more eventful contours, acoustically and perceptually more prominent, to signal cases in which it is particularly necessary to convey a marked meaning even against expectations built up from a conversational context. They presumably need to be both acoustically salient and of relatively low frequency, so that the mere fact of their occurrence is sufficient to captivate attention to such an extent that a change of contextual expectations might be effected. At the same time, if they are to convey quite specific meanings, they need to be sufficiently recognizable, which means they must be spaced sufficiently far apart from each other in terms of their dynamic acoustic properties. This as well as their low frequency (by itself an obstacle to identification) presumably also contributes to setting an upper limit on the number of types that are practically differentiable, each with its separate meaning. We will revisit some of these issues in section 3.7.

## 3.5 Relation between tones and the segmental string: association and alignment

In the previous section, we considered aspects of intonation at the level of the (nuclear) contour. In this section, we will instead look at the level of individual tones, specifically the phonology and phonetics of how they relate to the segmental material, and how this differs both crosslinguistically and between types of tones.

Since Pierrehumbert (1980), intonational tones have been classed into different categories: the tones associated with metrically strong positions are pitch accents (T*) and those located at the edges of prosodic constituents are boundary tones (T%). In the rise-fall-rise contour L*H L- H%, that accounts for the starred L and the final H%, but leaves out the trailing H of the pitch accent and the L-. The latter was originally called a "phrase accent" in Pierrehumbert (1980) and was then reanalyzed as an intermediate boundary tone in Beckman & Pierrehumbert (1986: 288). The former, unstarred tone in a bitonal pitch accent is either called a leading or trailing tone, as already stated above. Even though we have just seen that there are aspects of tonal behaviour, especially in relation to their pragmatic function, that suggest it is useful to consider tonal contours in a somewhat holistic fashion, this does not mean that all of the tones in such a contour-forming tone sequence show the same properties with respect to how they relate to the text and the prosodic structure. In this regard, the "phrase accents" and unstarred tones in multitonal pitch accent have often been seen to behave differently from the starred tones and boundary tones. For the starred tone, metrical structure provides an association site (Pierrehumbert 1980: 32–33). For the IP-boundary tone, metrical structure is irrelevant but the edge of the IP provides a "straightforward" orientation (Pierrehumbert 1980: 32). Thus these two types of tones are seen as oriented by the prosodic and metrical structure, independent of each other and of their tonal environment. The phrase accent, on the other hand, "is found near the end of the word with the nuclear stress even when this is not a metrically strong position" (Pierrehumbert 1980: 32). Intonational descriptions of Spanish do not make use of the phrase accent at all, and use the ip boundary tone in a less restricted way, as we have already seen in 3.4.7. Regarding unstarred tones, Pierrehumbert concludes after investigation of a small corpus produced specifically for this purpose and in reference to the trailing H in a L*H pitch accent that it "is located at a given time interval after the L*, regardless of the stress pattern on the material following the accented syllable" (Pierrehumbert 1980: 77).

Even without saying anything about phonetic alignment (the timing of a tonal event relative to the segmental material), in these descriptions these two are therefore different from the starred tones and boundary tones because they aren't orientated directly with reference to the metrical or prosodic structure but only indirectly by reference to other tonal events. From this description alone it would be hard to argue that either of the phrase accent or unstarred tone can be associated to a specified TBU, as the starred tone is. At the same time, they also show "a certain amount of variation" in their placement (Pierrehumbert 1980: 32), mirroring their less directly determined phonological status in phonetic terms. We have also already seen that tonal alignment can affect the (perception) of intonational meaning, in the discussion of the L*H L- H% vs. LH* L- H% contour of American English. Meaning differences have also been related to alignment for example in

German (Kohler 1987, 2005). Thus, a relation clearly seems to exist between alignment behaviour and phonological categories of tones (which may then also relate to meaning differences). This relation has since been further refined. In the following sections, we will first take a closer look at the phonetic alignment of pitch accents taken to be associated with prominent positions, and then at how the absence of such robust alignment can be seen as contributing evidence for the absence of such positions at the word level in some languages. The variation in degree of anchoring both across languages and types of tones is highly relevant for the present work because it also represents one of the dimensions of variation between Huari Spanish and Quechua.

### 3.5.1 Segmental anchoring in pitch accents

On the one hand, a number of findings on the relatively solid temporal alignment of the pitch turning points of pitch accents have led to the formulation of the "segmental anchoring hypothesis", according to which they are aligned "with specifiable points in the segmental string" (Ladd 2006: 20). For Greek prenuclear rising accents, Arvaniti et al. (1998, 2000) showed that the elbow of the low tonal target consistently aligned just before (5 ms on average) the consonantal onset of the stressed syllable, while the peak of the high tonal target aligned just after (mean of 17 ms) the onset of the vowel in the posttonic syllable. That is to say, the distance between the low and the high tonal target is dependent on the length of the stressed syllable plus the length of the onset consonant in the posttonic, but not upon that of the following vowel, and both tones can be said to be stably aligned in relation to the stressed syllable and independently of each other, unlike in the case of the English L*H. This leads Arvaniti et al. (2000) to propose that both tones in this bitonal configuration are associated with the stressed syllable, and starred, i.e. L*H*.

On the basis of similar findings, in particular because both the distance between valley and peak and the distance from stressed syllable onset to peak correlate with stressed syllable duration, O'Rourke (2005: 105–108) also proposes the association of both tones (symbolized also as L*H*) with the stressed syllable for nuclear and prenuclear pitch accents in both Cuzco and Lima Spanish. However, she also finds on the one hand that for her Lima data, no syllable boundary serves as an anchor for the nuclear peaks, but since the peak always occurs within the bounds of the stressed syllable, association with it is nonetheless assumed (O'Rourke 2005: 108). On the other hand, prenuclear peaks seem also quite strongly attracted by the right boundary of the word containing the stressed syllable, more so for Lima than for Cuzco, which concurs with more prenuclear peaks in Lima being realized after the stressed syllable than in Cuzco[41] (O'Rourke 2005: 79–84, 105–106). For the first prenuclear pitch accent in a phrase in Madrid Spanish, Prieto & Torreira (2007) find that the H peak position varies dependent on the syllable structure of the stressed syllable, but again because it is in general realized within the stressed syllable, they assume association of the H tone with it. For Cuzco Quechua, O'Rourke (2009) finds that peaks in final words are realized significantly earlier than in prefinal words, with both however still occurring mostly within the tonic syllable. She therefore proposes an association of the low tone with the stressed syllable (taken to be the word penult), symbolized as L*H, for prefinal words, and of the high tone, symbolized as LH*, for final words. Another possible analysis she considers is to assign LH* throughout and to attribute the differing alignment to the presence or absence of an incoming phrasal boundary tone (O'Rourke 2009: 309, note 16).

Mücke et al. (2009) investigate peak alignment in German rising contrastive pitch accents, comparing Northern (Düsseldorf) and Southern (Vienna) varieties, prenuclear and nuclear position, syllable structure (open vs. closed syllables) and acoustic vs. articulatory measurements. The picture that emerges from the acoustic measurements is that while both dialectal background and syllable structure have measurable effects (Southern varieties and closed syllables have later peaks), these are also (as in O'Rourke's findings on Lima and Cuzco Spanish) subject to considerable speaker variation (n=2 each for both varieties), resulting in syllable structure being a significant factor only for nuclear accents overall, and dialectal background for neither. Yet the alignment difference between nuclear and prenuclear position was found to be significant and substantial for all speakers, leading Mücke et al. (2009: 336–337) to call the dialectal differences gradient and those between accent status discrete and symbolical. Articulatorily, their findings show that latencies between peaks and articulatory anchors (opening and closing gestures) are smaller than those measured acoustically between peaks and segmental anchors, but do not differ much in variability (Mücke et al. 2009: 336).

[41] Note that considerable individual differences between Cuzco Spanish speakers with regard to prenuclear peak placement lead O'Rourke to group them into 4 different patterns. The first pattern (A) is essentially the same as that exhibited by Lima speakers, with all prenuclear peaks realized posttonically, while in the last (D) almost all prenuclear peaks are realized within the stressed syllable. Patterns B &C are in between, with prenuclear late peaks more frequent on subjects than on verbs in the SVO sentences O'Rourke elicited. Pattern D was produced by the largest group (A: n=3, B: n=4, C: n=3, D: n=5). Interestingly, the grouping of speakers according to these patterns could not be correlated to their status as either Spanish monolinguals, early Quechua-Spanish bilinguals (<5 years) or late Quechua-Spanish bilinguals (having learned Spanish only after entering school), but as a whole, Cuzco speakers thus behaved clearly differently from Lima speakers (for details, see O'Rourke 2005: 79–86).

In summary, it seems that languages, speakers and individual tones in pitch accents differ with respect to the degree that they have invariable segmental anchoring behaviour. It is perhaps worth emphasizing the relatively large role of individual variation found by the studies discussed in this paragraph especially when considering that with the exception of O'Rourke (2005), none had more than 5 experimental subjects. Analysis of data from 72 speakers in total of German (n= 35) and Neapolitan (n= 17) and Pisa Italian (n= 20) has shown that this alignment variation is so far-ranging that Niebuhr et al. (2011) suggest speakers might even follow different global strategies, with one group broadly seeking to align f0 turning points with segmental landmarks, and the other seeking to realize contour shapes specific to pitch accents. Another summary finding seems to be that there is both evidence for independence between the individual tones making up bitonal pitch accents (their individual segmental anchoring) and for considering their movement as a single unit (the fact that their alignments are never totally independent of each other, cf. Ladd 2006: 27–28). Different bitonal pitch accents in different languages tend more to one or the other side of this continuum: in the English L*H, with the "fixed temporal distance" at which the H follows the L*, the two tones are less independent of each other than in the Greek L*H*. This difference recalls one that Grice (1995) has proposed to differentiate *within* English bitonal pitch accents: those with trailing tones (T*T), with the trailing tone occurring at a given distance after the starred one, are "melodic units", in which the unstarred tone is not directly associated, while those with leading tones (TT*) are "tone sequences", in which both are somewhat independently associated (Grice 1995: 216–219). Investigations into the coordination of articulatory gestures (e.g. Katsika et al. 2014; Tilsen 2016, 2019) are likely to provide more fine-grained and adequate explanations also of the effects of larger prosodic domains on peak alignment (cf. also Ladd 2006: 33–35), but this cannot be dealt with here. In sum, tones belonging to pitch accents have shown themselves to be aligned relative to a TBU assigned to a prominent position in the metrical structure, indicating association with that TBU. Where such independent alignment cannot be found, independent association is also in doubt.

### 3.5.2 Secondary association and tonal spreading

After having introduced segmental anchoring in 3.5.1, we can now turn to two mechanisms by which phonological tones can relate to the segmental string without independently associating, secondary association and tonal spreading. They allow to treat contours as similar that share the same number and type of turning points but differ in the way high or low stretches are extended across material, and to differentiate between points of similar tonal height according to whether pitch is actively manipulated there due to a relevant prosodic position, or simply maintained due to a specification from elsewhere. An optimality-theoretic approach to these issues is also introduced. This is crucial for the later analysis in sections 5.3 and 6.3, because it will allow for a principled generalization across superficially different contours containing such stretches in both Huari Quechua and Spanish, and to demonstrate how the different intonational variants we will encounter are relatable to each other via stepwise changes in constraint rankings.

Grice et al. (2000) study the "Eastern European Question Contour" (EEQT), a polar question contour consisting of a final rise-fall (LHL) in several related (varieties of) Eastern European languages, Standard and Transylvanian Hungarian, Standard and Cypriot Greek, and Standard and Transylvanian Romanian. They argue that in all varieties, the phonological representation is L* H- L%, i.e. the H tone is a phrase accent not associated with the nuclear accent and thus not "prominence-lending" in either of the varieties, but the position of its peak seems to vary discretely, rather than gradually, between the varieties. In a configuration in which nuclear stress is on the prefinal word in an IP, producing a low tonal target on the nuclear stressed syllable followed by a low stretch throughout the nuclear word, the H peak is aligned on the penult of the final word in Standard Hungarian (unless that is also the initial, i.e. stressed syllable, in which case the peak moves to the final syllable) and on the penult or final syllable in Cypriot Greek, independent of which syllable bears stress in the final word, and on the postnuclear stressed syllable (in the final word) in Standard Greek and Romanian. The Transylvanian varieties show the same behaviour as their standard counterparts, except that instead of the low stretch following the L*, pitch directly rises to form a plateau extending to the position at which the peak occurs in the standard varieties. In order to analyze this behaviour, Grice et al. (2000) take recourse to a mechanism originally proposed in Pierrehumbert & Beckman (1988) for Japanese, *secondary association.* Pierrehumbert & Beckman take the AP-initial H- tone in Japanese to associate *primarily* to the left edge of the AP, but *secondarily* to the second mora of the AP. For Japanese, this (secondary) association is reflected in the findings by Ishihara (2006: 72) that the AP-initial peak in unaccented words is quite stably aligned just at the beginning of the third mora, although Venditti (2005: 181) states that its alignment can also vary considerably. Returning to the EEQT, Grice et al. (2000) propose to analyze the H- phrase accent as primarily being associated with the right phrase edge, but secondarily with the position at which the peak is found. In the Transylvanian varieties, it is secondarily associated both with the nuclear stressed syllable *and* that additional position, via tone copying, not spreading (i.e., the tone does not associate with each intervening syllable; see Table 4 for a summary of the secondary association sites).

**Table 4:** Secondary association of H- phrase accent in the Eastern European Question Tune (EEQT), adapted from Grice et al. (2000: 158, 161).

| | ***nuclear accent in non-final word*** | ***nuclear accent in final word*** |
|---|---|---|
| **Standard Hungarian** | penult | penult |
| **Standard Greek** | postnuclear stress | final |
| **Cypriot Greek** | penult/final | final |
| **Standard Romanian** | postnuclear stress | final |
| **Transylvanian Romanian** | nuclear syllable *and* postnuclear stress | nuclear syllable |
| **Transylvanian Hungarian** | nuclear syllable *and* penult | penult |

Gussenhoven (2000a, 2000b, 2002b, 2004) develops a slightly different intonational phonology based on similar considerations. He takes phrase accents to be boundary tones, and adopts the notion of phonological alignment from prosodic morphology (cf. McCarthy & Prince 1993, 2001[1993]). Prosodic constituents (among which tones are counted) can align with the edges of other prosodic constituents. The relative ranking of optimality-theoretic phonological alignment constraints for the tones in a given input determines both their sequence and proximity to each other in the output.[42] Because his model is couched in OT-terms, constraints are formulated together with their conflicting counterparts. For alignment, this leads easily to situations in which tones are aligned in opposing directions (*multiple* alignment), and without any higher-ranking constraints intervening, the results are long pitch stretches, such as the plateaux in the Transylvanian varieties of the EEQT, in which the intervening TBUs are not associated to the tone (which would be spreading). The endpoints of such stretches do not have to have the same phonological status. Gussenhoven only allows association with TBUs, but not with edges of prosodic constituents. Association acts in two directions via two families of constraints, those that associate classes of tones with TBUs (e.g. H → TBU) and those that associate classes of TBUs with tones (e.g. σ ← T). In this way, a tone available at a position can associate with that position because it is a tone that has to associate, or because it is a position that has to associate, or it does not associate at all (if the constraints that would cause association in this case are outranked by a constraint militating against association). That means that in principle, tones in his model can be only aligned but not associated, and still be realized, if no constraint deleting unassociated tones is high-ranked. Secondary association, in Grice et al's terms, can then result out of a combination of multiple alignment, with the tone associating with a relevant TBU at both edges, or only at one of them. Additionally, an extended pitch stretch could also not be associated at either edge in Gussenhoven's model. Importantly here,

[42] For a more detailed introduction into Gussenhoven's OT model of intonation, see section 5.3.

the status of a tone as associated with a TBU or “only” aligned with some prosodic constituent is taken by Gussenhoven (2018: 406–407) to be reflected in its phonetic alignment behaviour: if a tone’s peak or valley alignment can be shown to be indifferent to phonological structure, this is taken as indication for a lack of association.

With this additional specification, the Gussenhoven model is in principle capable of distinguishing between more alignment-association-scenarios than the Grice et al. (2000) model: in the study on German mentioned above, Mücke et al. (2009: 335–336) identify a second low tonal target in their data as evidenced by a low elbow following the nuclear (but not the prenuclear) pitch peak at a fairly invariable distance (some 150 ms later), independent of the segmental material. They take the presence of this additional low tonal target in the nuclear condition to be the reason for the significantly earlier peak alignment of the nuclear pitch accents compared to the prenuclear ones, comparable with the analysis for postnuclear deaccentuation in German by Féry (2017) and also with the analysis of nuclear phrasing for Spanish espoused here. However, they analyse it as a phrase accent secondarily associated with the nuclear stressed syllable. This gives this low tone phonologically the same status as e.g. the Japanese AP-initial H, or the EEQT high phrase accent tone. However, the latter Grice et al. (2000) have shown to be sensitive to discretely variable aspects of structure in its dialectal variation (see Table 4), and the former is also much more sensitive to structure as evident from its temporal alignment (cf. Ishihara 2006). This difference in sensitivity to structure between the German postnuclear low tone, on the one hand, and the EEQT high phrase accent and the Japanese AP-initial H, on the other, could be seen as an argument against assigning them the same status. Indeed, it should be at least possible to keep them apart in their representation: in the Gussenhoven-style model, the German low tone found by Mücke et al. (2009) would certainly be analyzed as as an additional leftward alignment of the upcoming low IP-boundary tone without association. This can be seen from analyses of comparable low phrase accents in Dutch by van de Ven & Gussenhoven (2011), in West Germanic varieties by Peters et al. (2015), and in English according to the analysis by Gussenhoven (2018) of the findings reported on in Barnes et al. (2010).[43] Because in principle it allows for a more fine-grained representation (which does not mean that it is necessarily the more

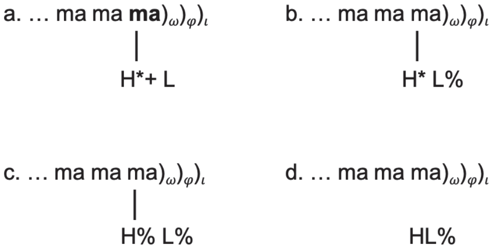
[43] Barnes et al. (2010: 989) themselves take the turning point to be “epiphenomenal, the result of a constraint on the phonetic implementation of deaccenting in the postnuclear region: The fall from the level of the H* maximum to the level of the eventual L- at the phrase boundary (i.e., E2) must be accomplished sufficiently swiftly to avoid creating the percept of a pitch accent (L* or downstepped H*, for example) on a lexically stressed syllable within the postnuclear region.” Arguably, the unblocked leftward alignment of the L%, resulting in a flat postnuclear stretch is precisely what would constitute deaccentuation here.

adequate representation, further studies will have to decide that), I am adopting the Gussenhoven-style model here together with its heuristic about the association status of a tone based on its alignment behaviour.

### 3.5.3 Representational possibilities for the association and alignment of tones in languages without word stress

We can now extend the discussion on variability in tonal anchoring to languages without word stress. Since we are pursuing the hypothesis that Huari Quechua has no word stress at all or that it is almost irrelevant in its phonological system, this section will provide criteria from alignment and association that can be brought to bear on the matter. Following the Gussenhoven-style approach, a phrase-final HL sequence can be represented in at least four different ways:

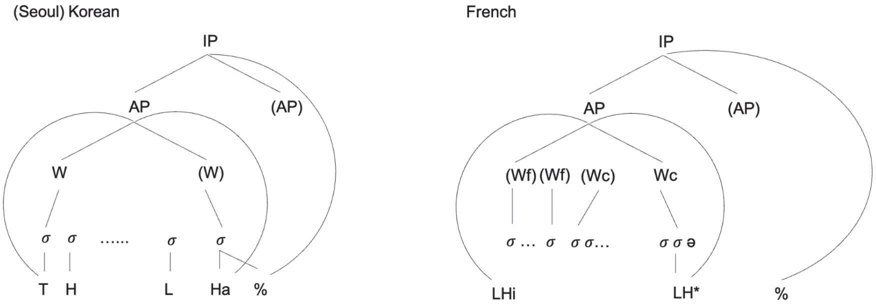
(13) Representational possibilities for a phrase-final HL sequence, adapted from Maskikit-Essed & Gussenhoven (2016: 355)


a. … ma ma **ma**$)_\omega)_\varphi)_\iota$
H*+ L
b. … ma ma ma$)_\omega)_\varphi)_\iota$
H* L%
c. … ma ma ma$)_\omega)_\varphi)_\iota$
H% L%
d. … ma ma ma$)_\omega)_\varphi)_\iota$
HL%

In (13)a, the starred H tone is associated with the syllable bearing word stress, and the L tone is trailing (which means that it is left-aligned with the right edge of the H). In this phrase-final configuration, it is unclear whether the model would have different phonetic alignment predictions for H*+L and H* L% here (L% being minimally right-aligned with the right edge of ɸ/ip or ι/IP). This would also depend on the constraints on tonal crowding. Rathcke (2016) demonstrates that falls in pitch accents and those due to boundary tones in German and Russian are treated differently in terms of compensation strategies under time pressure (truncation, compression, tonal re-alignment), but also that these strategies are language-specific to a certain degree.

#### 3.5.3.1 French

(13)b, in which no word stress exists but the H tone forms a pitch accent associated with a postlexically determined position (the final syllable in the phrase), is exemplified by French, according to Maskikit-Essed & Gussenhoven (2016: 356). In (13)c, two boundary tones simply align rightmost and associate with the rightmost available TBUs. In (13)d, finally, tones are only aligned with the right edge of the phrase, but remain unassociated. Maskikit-Essed & Gussenhoven argue that (13)b is different from (13)c, the latter of which they say is exemplified by Korean AP boundary tones, because in French, whether a postlexically assigned phrase-final position is eligible for accentuation can be exceptionally determined by the morphophonological makeup of the phrase (see Figure 10 for the intonational structure of (Seoul) Korean and French): without citing sources, Maskikit-Essed & Gussenhoven (2016: 356) state that in French *que sais-je?* “what do I know?”, the AP-final peak cannot be aligned on and the corresponding H tone not associate with the pronominal clitic *je*, but that this is possible for the pronominal clitic *le* in *prends-le* “take it”.



IP = Intonation Phrase, marked by a boundary tone %.
AP = Accentual Phrase, marked by THLHa tone pattern.
T = H if AP initial segment is aspirated or tense C, /h/, or /s/; L, otherwise.
W = phonological word; σ = syllable.

IP Intonation Phrase
AP Accentual Phrase
Wf function Word
Wc content Word
σ Syllable
% IP boundary tone

**Figure 10:** Intonational structure of (Seoul) Korean, left, and French, right, adapted from Jun (2007: 151), and Jun & Fougeron (2002: 152), respectively.

This is a somewhat controversial analysis. According to Welby (2006: 347), the general view is that the AP-final H will associate with the last full vowel, which excludes schwa, the vowel in both *le* and *je* (if one is realized at all). Welby & Loevenbruck (2006: 114) find that in contrast to this generalisation, the final peak can occasionally align on an AP-final schwa, but do not relate this to any lexical or morphosyntactic difference. Féry (2017: 182) distinguishes between two kinds of AP-final schwa in French: a postlexically inserted schwa, which cannot act as TBU for the final H tone, and an “underlying” schwa, which can. She explicitly gives examples including both *le* and *je* for the latter category. In spite of this, there is some agreement that the AP-final accent (LH*) in French is a pitch accent, indeed associated with what should perhaps be called the final licit TBU in the phrase, while the AP-initial accent (LHi or LH-) is a phrase accent and not associated with a

TBU (Jun & Fougeron 2000, 2002; Welby 2006; Welby & Loevenbruck 2006; D'Imperio et al. 2007). This is argued to be so because while the former is stably temporally aligned (allowing only for a kind of discrete variation in alignment with either the AP-final vowel or the penultimate one if the final one is a schwa) and the syllable on which it is realized is longer and louder than its neighbours, the latter is optional, varying in its alignment within the first three syllables of the AP and not accompanied by more loudness and longer duration (Jun & Fougeron 2002: 159).

Welby (2006) is able to further differentiate this picture based on measurements of phonetic alignment: in the initial rise, the L is stably aligned at the left edge of the first content word and thus taken to be associated both with that word's left edge and optionally secondarily with the left edge of the AP[44] (Welby 2006: 364–366). In Gussenhoven's model, association is only possible with TBUs, but not with constituents (Gussenhoven 2004: 155, 2018: 408–409). Yet the stable alignment of the initial elbow with the left edge of the first content word in Welby's data indicates association. It could be modeled with constraints that align it with both edges (with that aligning it with the content word edge ranked higher) and a constraint that stipulates that it must associate with whichever TBU is available in its location (L → TBU). The AP-final H is consistently aligned in the final syllable, accompanied by syllable lengthening and also interpreted as the starred tone of a pitch accent, albeit one that is not "prominence-lending" (Welby 2006: 364–365). Thus the two outside tones are stably aligned with segmental landmarks, even across different speech rates and taken to be associated. The two middle tones are different: the initial H has highly variable alignment not just within one syllable but extending over at least the initial two syllables of the first content word and the final L varies equally in its alignment. Both are taken not to be associated, yet "edge-seeking", i.e. showing alignment tendencies towards their respective AP-edges (in terms of the Gussenhoven model), but they are also not realized at a relatively constant distance to their initial L and the final H, respectively, varying also with speech rate (Welby 2006: 366–367). Thus, the French AP seems to consist of tones belonging to at least two of the configurations above. The position of all tones is postlexically determined, so (13)a is not appropriate. However, according to Welby's (2006) analysis, both the AP-final H* and the AP-initial L should be classed in category (13)b because they are clearly associated with a postlexically determined position whose eligibility is itself subject to (morpho-)phonological constraints. Both the middle tones, initial H and final L, should most appropriately be classed in category (13) d, since they are not associated (independent of whether they form part of bitonal

[44] Because it forms a low stretch if the content word is preceded by function words that in most cases extends to the beginning of the AP.

accents with the external tones). Category c) is perhaps best occupied by the Japanese AP-initial H- in cases of unaccented words, because it is associated (according to the findings of stable alignment by Ishihara 2006), yet its position can simply be described as aligning as leftmost as possible and associating where the leftmost free TBU (mora) is available, resulting in association with the second mora of the AP, because the initial one is occupied by the L tone. For Korean, which Maskikit-Essed & Gussenhoven (2016: 355) take to exemplify this category, it has been found that at least the AP-initial peak is also sensitive in its alignment towards the morphosemantic makeup of the words phrased in it (Kim 2013: 79–84) and so the AP-initial H can therefore also be thought of as exemplifying category (13)b. Note that the difference between the French AP-final H*, which is analyzed as a pitch accent because it also affects duration and intensity, and the AP-initial L, which does not, is actually not expressible in the typology in (13). This shows that in a language arguably without word stress (cf. also Féry 2017: 180–184), accents can also differ along this dimension.

#### 3.5.3.2 Tashlhiyt Berber

A further relevant case is that of Tashlhiyt Berber as investigated by Roettger (2017). Using a statistical assessment of durational and intensity differences between syllables as a proxy for prominence in a production study, Roettger (2017: 48–58) finds no consistent asymmetries that could be interpreted as evidence for word stress, and in particular no evidence that word stress is final as claimed before in the literature. His findings on peak alignment are highly interesting because they relate variability in peak placement with “phonetic enhancement” of tonal events, i.e. when syllables on which tones occur are also longer, louder and produced with more peripheral vowels (Roettger 2017: 47, 135, 144). In Tashlhiyt Berber, polar questions (both neutral and echo) and statements with a contrastive interpretation are realized with a very similar rise-fall (LHL) tonal movement. In questions, this is realized on the IP-final word, while in contrastive statements it is realized on the word which is contrasted against an alternative in the context. In cases where the contrasted word is phrase-final, these two positions obviously fall together. Yet the utterance modalities also differ in terms of their peak alignment in two additional ways:[45] on the one hand, quasi-discretely, such that peaks in statements are bimodally distributed with one mode towards the end of the penult and the other some

[45] They also differ in terms of global pitch level over the utterance and local pitch excursion on the pitch peak, with scaling in both cases such that polar questions > echo-questions > (contrastive) statements, a difference that also proves robust in perception (Roettger 2017: 76–79, 93). This reflects a crosslinguistically frequent tendency (cf. the works cited in Roettger 2017: 65).

way into the final syllable of the word bearing the pitch peak, while in questions, peaks are unimodally distributed with a majority (but by no means all) located inside the final syllable (Roettger 2017: 80–81). On the other, they also differ gradually in that even only within the tokens that realize the peak on the final syllable, statements align the peak earlier than questions. In a perception experiment only the former, but not the latter, difference significantly affected ratings as either statement or question, and with a considerable amount of variability both between and within test subjects (Roettger 2017: 92–95). The syllable on which the peak was realized was phonetically enhanced in so far as it was louder and longer, and also sounded more "prominent", although this was not systematically tested (Grice et al. 2015: 249, 254; Roettger 2017: 85, 135). In addition to utterance modality, peak placement is also affected by sonority differences between the syllables making up the word. Tashlhiyt has few vowels (/i a u/) but many and diverse consonants, all of which, including voiceless stops, are allowed as syllable nuclei, resulting in a large number of vowelless words in the lexicon and comparatively long vowelless strings in actual speech (Roettger 2017: 37–38). Grice et al. (2015) found in a study on bisyllabic words that peak placement was attracted to the syllable containing the more sonorous nucleus, showing a preference both for vocalic nuclei over consonantal ones as well as for more over less sonorant consonants, with liquids > nasals > voiced obstruents > voiceless obstruents (thus, e.g. /tu.gl̩/ "she hanged" and /tr̩.ks̩/ "she hid" were more likely to realize the peak on the penult, and /tn̩.za/ "it was sold" and /ʜf̩.dˤʁ̩/ "I learnt" more likely to realize it on the final syllable, cf. Grice et al. 2015: 248–251, 263). It also showed a preference for heavier (with a coda) over lighter syllables and a general preference for being rightmost; all these factors (including utterance modality) seem to be independent of each other and probabilistic rather than categorical: only when several of them came together in the same direction were counts of 100% reached for peak placement in one of the two syllables, but not by all speakers, and with within-speaker variability present even in the exact same condition on the same word (cf. Grice et al. 2015: 251; Roettger 2017: 86–88). The situation is exacerbated further when words are considered that contain no sonorants, with either voiced (e.g. /tb̩.dg̩/ "she was wet") or unvoiced obstruents in the syllable nuclei (e.g. /tk̩.ʃf̩/ "it dried"): in those cases, the peak can either a) not be realized at all, or b) it is realized on the word preceding the non-sonorant target word, or c) it is realized on a schwa-like vowel inserted somewhere in the word (Grice et al. 2015: 254–255; Roettger 2017: 98–101). These three alternatives were found regardless of utterance modality, but the presence of an inserted schwa in the target word and the realization of the peak on the preceding word (b) never occurred together in a word consisting only of voiceless obstruents, i.e. when a schwa was inserted in such a word, it always carried the tonal peak (Roettger 2017: 100). Roettger proposes that there are two types of schwa in Tashlhiyt: one is purely the result of articulatory timing, when two constricting gestures are timed far enough apart that a transitional period occurs ("gestural underlap") during which airflow can pass unhindered, resulting in a vowel-like sound (a "transitional vocoid") as long as the vocal folds are vibrating (Roettger 2017: 106–108). Such an account is of course insufficient for words consisting only of voiceless segments, since no phonation will occur. For those, he proposes a true epenthetic vowel, a schwa that is postlexically inserted in voiceless environments precisely in order to act as a TBU (Roettger 2017: 125–128). Its distribution is not categorically determined, and seemingly affected by a variety of heterogeneous factors, including sociolinguistic ones, that do not correlate with peak alignment variability (Roettger 2017: 113–114, 124). The difference in status between the two schwas accounts for the fact that in voiceless words, a schwa never cooccurs with the peak realized on the preceding word: because this is the postlexical epenthetic vowel inserted to serve as TBU, whereas in words with voiced obstruents, the transitional vocoid can occur but not serve as TBU because it is invisible to the phonology (cf. Roettger 2017: 126–128).

Roettger (2017) does not offer a full phonological analysis of Tashlhiyt intonation, because only certain aspects of its system were considered. However, he takes the discrete variability in stable alignment to different syllables depending on the factors outlined and the "phonetic enhancement" of the syllables the peak is aligned with as evidence that the H tone is not just aligned, but associated with the syllable it occurs on (Roettger 2017: 144). Note that this is not the same as saying that these syllables bear word stress, which is clearly not the case (since there are realizations of the same lexical item by the same speaker under the same experimental conditions, in one of which the peak is located on the penult, while in the other it is on the final syllable, and that syllable is phonetically enhanced in both cases); thus the Tashlhiyt case is further evidence that the determination of a culminative position at the lexical level and the acoustic correlates of increased duration and intensity, usually associated with stress accent, are orthogonal to each other (we have already seen that Japanese is in a sense the inverted case). It also seems that this behaviour is conditioned by the type of word the tonal movement is realized on and/or the utterance type: in a study on wh-words (question words) in Tashlhiyt, Bruggeman et al. (2017: 20–21) found that essentially the same rise-fall (LHL) contour is realized on them as on the final word in polar questions and the contrasted word in contrastive statements, but that while the peak is always aligned within the question word, no evidence for systematic alignment with any syllable nor for phonetic enhancement of the syllable realizing the peak could be found, making an analysis in terms of association with a TBU unlikely. For both the question words in Bruggeman et al. (2017) and the tone-bearing words in Roettger's (2017) study, secondary association to a higher prosodic constituent is also proposed: for the question words and the contrasted words in statements, this association is with the prosodic word, for polar questions, it is with the IP (Roettger 2017: 137–140). Tones are also aligned with the right edge of the domain they are associated with.

Despite the rather complex situation in Tashlhiyt and several unsolved puzzles,[46] it seems that parallels can actually be drawn to the French case. In both languages, an H tone (the AP-final one in French) seeks to occur rightmost in its domain and will associate with an available legitimate TBU. Assuming the account of schwa by Féry (2017), both languages have two different types of schwa, only one of which can serve as a TBU. While in French, that is essentially the only additional constraint able to prevent a rightmost placement of the H tone, in Tashlhiyt the segmental environment is much more adverse to tonal realization, resulting in cases in which the rightmost available TBU is actually placed before the target word. What a legitimate TBU is is thus affected by different factors in the two languages, but the generalization that the tone seems to almost occur as rightmost as possible, everything else being equal, makes a good case for the adequacy of the alignment constraints as proposed in the Gussenhoven model (cf. Grice et al. 2015: 260 for this argument, but Roettger 2017: 143–144 for a more sceptical view). The Tashlhiyt H tone in polar questions and contrastive statements would thus be classed, like the French AP-final H* pitch accent, under category (13)b, while its counterpart in wh-words would fall in category (13)d.

#### 3.5.3.3 Ambonese Malay

Let us make a final comparison with another language without word stress, Ambonese Malay as discussed in Maskikit-Essed & Gussenhoven (2016). Using durational and peak alignment measurements on target words from a task eliciting small read dialogues, they make the case that Ambonese Malay also has no word stress. In a comparison with data from a similar task made with Dutch speakers, where stress is uncontroversial, they find that peak alignment in Ambonese Malay does not correlate strongly with any segmental position in the word, especially not one in the penult, which has been proposed to bear stress in previous studies based on impressionistic analyses; the peak effectively varies relatively freely in its placement on the last two syllables (Maskikit-Essed & Gussenhoven 2016: 364–366). This leads them to propose that the H tone is aligned rightmost, but remains unassociated, thus belonging to category (13)d. Note that even though we have already mentioned above that the Gussenhoven model does not allow for association with higher prosodic constituents, Maskikit-Essed & Gussenhoven (2016: 382) state that

[46] The most theoretically challenging one of which is perhaps that the Tashlhiyt data do not in any way support a deterministic relation between form and function in intonation, an issue which is discussed in Roettger (2017: 145–148).

"words are referred to as domains within which the rising-falling pitch movement is placed", because the peak always occurs within the target word and because its distance from the word edge correlates with the duration of the final two syllables as well as of the word itself, more so than in the measurements on Dutch. If "reference to a domain" is not to be proposed as a third, as yet somewhat vaguely defined, means of tune-text linking (besides association and alignment), then this must be taken as a covert admission of association to higher-level prosodic constituents. In fact, this would not be a surprising finding if Grice et al. (2000: 148) turn out to be right in suggesting that (secondary) association to a TBU rather than a higher domain is typologically more frequent.

Intonation in Ambonese Malay seems also not to signal prominence in contrastive discourse contexts in terms of pitch scaling or any other means[47] (cf. Maskikit-Essed & Gussenhoven 2016: 372–374, this is also in line with results from prominence perception experiments on the related language Papuan Malay by Riesberg et al. 2018), and as far as their data allows them to say, the only functional differentiation intonation makes in that languages is one between declaratives, with a final low boundary tone, and non-declaratives (polar questions and continuation rises), with a final high boundary tone (Maskikit-Essed & Gussenhoven 2016: 377–380).

We have thus seen that also in languages without word stress, there is considerable variation in intonational systems in terms of association and alignment of different tones and their relation to various functions such as the signaling of prominence. It seems that the best measurable indicators for association are stable phonetic alignment with a consistently identifiable position, on the one hand, and "phonetic enhancement", on the other. However, phonetic enhancement is certainly not a necessary condition for association, as at least Japanese shows. In addition, neither is evidence for word stress: only when they indicate patterns that consistently point to a unique position in each lexical word can that connection be made. In absence of such a pattern, a postlexical pitch accent hypothesis is at least as likely, and more so when peak placement is discretely variable, as in Tashlhiyt Berber and French. Table 5 attempts to give an overview over the interrelation between association, phonological and phonetic alignment to lexically/morphologically specified positions in a word or postlexically determined ones in a larger prosodic unit, phonetic enhancement and word stress or lexical accent via tones from various languages discussed in this section. Two cells contain no data. For a type of tone to be entered there, it would need to exhibit continuously variable alignment that spans

[47] The results in Maskikit-Essed & Gussenhoven (2016: 372–374) show some variation between the speakers, in that for 2 out of the 4 speakers, the difference between focus conditions was actually significant, but with very small effect sizes (0.23 semitones on total average). In general, their small study size certainly leaves room for doubt until further results corroborate their findings.

Table 5: Attested possibilities for tonal association and alignment from some intonational systems.

| | | ***phonetic enhancement*** | ***no phonetic enhancement*** |
|---|---|---|---|
| **Association to a TBU** – stable phonetic alignment with a segmental or syllabic landmark | lex./morph. specified position | starred tone(s) of pitch accent on a stressed syllable (Germanic and Romance, Cuzco Quechua (?)) | starred tone of pitch accent on a lexical accent position (e.g. Japanese, Turkish (Levi 2005)), H phrase tone in EEQT in Standard Greek (sec. assoc. to final stressed syl according to Grice et al. 2000) |
| | postlex. determined position | Tashlhiyt H in rise-falls (Roettger 2017); AP-final H in French (Welby 2006) | Japanese AP-initial H (secondarily associated to a mora, according to Pierrehumbert & Beckman 1988, closely aligned according to Ishihara 2006), H phrase tone in EEQT in Standard Hungarian (sec. assoc. according to Grice et al. 2000); AP-initial L in French (Welby 2006); AP-initial H in Korean (Kim 2013); some phrase-final boundary tones (?) |
| **Phonological alignment to a constituent edge** (including another tone) – variable phonetic alignment across landmarks | lex./morph. specified position | – | some leading or trailing tones of a bitonal pitch accent on a stressed syllable (Germanic / Romance) |
| | postlex. determined position | – | some phrase-final boundary tones (?); H phrase tone in EEQT in Standard Hungarian according to Gussenhoven-style analysis; H tone in Ambonese Malay (Maskikit-Essed & Gussenhoven 2016), H phrase tone in EEQT in Cypriot Greek, AP-initial H in French, AP-final L in French (Jun & Fougeron 2002; Welby 2006), H tone in Tashlhiyt Berber wh-words (Bruggeman et al. 2017) |

at least two syllables, and at the same time there would have to be evidence that whatever syllable the tone is realized on is phonetically enhanced. That is conceptually not impossible, but perhaps not overly likely. Spanish in general contributes to the data resumed in the table, but Quechua mostly represents a large gap in our knowledge in this respect. I have tentatively included Cuzco Quechua among those languages where phonetic enhancement occurs on the stressed and pitch accented syllable, but this has not been conclusively shown (O'Rourke 2009: 298–299).

For Huari Quechua Quechua and Spanish I will present an analysis in section 6.1.6 that would allow us to locate them on this table and weigh in on the question of word stress in Huari Quechua. Since vowel length is contrastive in this Quechua variety (unlike in Cuzco Quechua), it should be less likely that duration will also be exploited strongly to provide phonetic enhancement for pitch accented positions. This seems to be corroborated by the results in Hintz (2006: 507), but see the findings in section 6.1.6 for a qualification.

## 3.6 Recursive prosodic structure

In this section, we will revisit the hierarchy of prosodic constituents initially introduced in section 3.2. We will consider arguments and evidence from several languages that the prosodic units above the prosodic word can have a recursive structure. In section 5.2 on complex Huari Spanish utterances, I will argue that the data presented there is also best analyzed as representing a recursive prosodic structure. The discussion about recursive prosody and cues to it ties in both with that about tonal scaling and which role it plays in conveying phonological tones and prosodic constituents from section 3.4.2, as well as that about the role of prosody in the cueing of information structure, which we will come to in section 3.7.

### 3.6.1 The Strict Layer Hypothesis and prosody-syntax mapping

In the early days of the development of the prosodic hierarchy, the following principles were thought to universally hold for it, besides the hierarchy and its levels being themselves universal:

(14) Architectural principles of the prosodic hierarchy (Nespor & Vogel 2007: 7)
*Principle 1.* A given nonterminal unit of the prosodic hierarchy, $X_i$, is composed of one or more units of the immediately lower category, $X_{i-1}$.
*Principle 2*. A unit of a given level of the hierarchy is exhaustively contained in the superordinate unit of which it is a part.
*Principle 3*. The hierarchical structures of prosodic phonology are n-ary branching.
*Principle 4*. The relative prominence relation defined for sister nodes is such that one node is assigned the value strong (s) and all the other nodes are assigned the value weak (w).

Principles 1 and 2 have also been formulated by Selkirk (1984: 26–27) under the heading *Strict Layer Hypothesis* (SLH). It claims that the prosodic hierarchy is strictly ordered and non-recursive: constituents of level $n_i$ exhaustively dominate all constituents of level $n_{i-1}$ in their domain and are themselves exhaustively dominated by a constituent of level $n_{i+1}$, with the index on n being prohibited to skip a number, so that structure (15)a is well-formed, while structures (15)b-d are not (deviant elements in bold[48]):

(15) Possible and impossible prosodic structures according to the SLH

a. $[_{\phi}[_{C}[_{\omega}[_{\Sigma}\ \sigma\ \sigma]_{\Sigma}\ [_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}]_{C}]_{\phi}$

b. $*[_{\omega}[_{\Sigma}\ \sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}\ \boldsymbol{\sigma}]_{\omega}$

c. $*\boldsymbol{[_{\omega}}[_{\omega}[_{\Sigma}\ \sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}\boldsymbol{]_{\omega}}$

d. $*\boldsymbol{[_{\phi}}[_{IP}[_{\phi}[_{C}[_{\omega}[_{\Sigma}\ \sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}]_{C}]_{\phi}]_{IP}\boldsymbol{]_{\phi}}$

While structure (15)a conforms to all stipulations of the SLH, (15)b does not, since it contains a syllable (σ) that, while being contained inside a phonological word (ω), is not contained within its next-higher constituent level, the foot (Σ). The level of the foot has been skipped. (15)c, on the other hand, is disallowed because it is recursive: a phonological word is contained within another phonological word. (15) d, finally, is out because in it a constituent of the level of the intonational phrase (IP) is dominated by a lower-level constituent, that of a phonological phrase (ϕ). Claims for the attestation of structures like (15)b-d have all been variously made in the literature, but we will be mainly concerned with (non)-recursive structure here. Principle 3 has to be understood at least to some extent in the context of principles 1 and 2: clearly, if there was a restriction to binary branching of constituents (as there still is, for example, in Liberman & Prince 1977), it would be impossible to maintain principles 1 and 2 with the same set of prosodic levels and without questionable results, such as having to posit two different prosodic domains for *white rabbit* ((16)a) and *white fluffy rabbit* ((16)b), while the problem does not arise with n-ary branching ((16)c).

(16) Differences between only binary and n-ary branching

a. $[_{C}[_{\omega}\text{white}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}$

b. $[_{\phi}[_{C}[_{\omega}\text{white}]_{\omega}]_{C}\ [_{C}[_{\omega}\text{fluffy}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}]_{\phi}$

c. $[_{\phi}[_{C}\ [_{\omega}\text{young}]_{\omega}\ [_{\omega}\text{white}]_{\omega}\ [_{\omega}\text{fluffy}]_{\omega}\ [_{\omega}\text{little}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}]_{\phi}$

There is another important consideration influencing the positing of principle 3, however: the flatter structures created in this way are to reflect the finding that while there is a certain degree of correspondence between morphosyntactic structure and prosodic phrasing, the constituents of the prosodic hierarchy are

[48] In order to hopefully improve legibility I use double marking of brackets in this section, using the same symbols for the prosodic constituents as described in section 3.2 for subscription. Thus, $[_{\phi}\text{XYZ}]_{\phi}$ marks one phonological phrase.

decidedly not isomorphic with morphosyntactic ones (Nespor & Vogel 2007: 1–2). This is even held to be true for the phonological phrase, which is at the same time considered to be the most important interface category with the morphosyntactic derivation (Nespor & Vogel 2007: xx–xxi). In some subsequent work, this stipulation of non-isomorphism between syntax and prosody has been moved away from to a certain degree and instead, a close correspondence between the two levels has been espoused as the norm that can occasionally be deviated from. This is reflected for example in how the following optimality-theoretic constraints are formulated:

(17) MATCH-constraints (Selkirk 2011: 439)
i. Match clause
A clause in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ι, in phonological representation.
ii. Match phrase
A phrase in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ϕ, in phonological representation.
iii. Match word
A word in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ω, in phonological representation.

Here, "matching" means that both the left and right edges of the prosodic constituent coincide with those of the morphosyntactic one. The idea of close mapping has taken such hold that for instance in Féry (2017: 36), the prosodic levels above the foot are introduced with the 'corresponding' syntactic units as an indicator of their size; the prosodic word "corresponds roughly to a grammatical word", the prosodic phrase to a syntactic phrase (NP, VP, PP, AP etc.), the intonational phrase to a clause and the utterance to a paragraph. It is somewhat unfortunate for those wishing to assess this correspondence claim that 'paragraph' is a notion not regularly discussed in syntax, and that even 'clause', which is much more ubiquitous, lacks a definition relevant for prosody-syntax interactions in "current syntactic theory", as Myrberg (2013: 94) admits. Furthermore, it is clear that 'corresponds roughly' must here in fact often mean 'does not correspond at all' when considering that utterances in actual conversation often consist of syntactic fragments only a few words long; nevertheless, they are utterances and intonational phrases: it is well-known in prosody research that recordings of words pronounced in isolation should not be used to draw conclusions about word-level prominences precisely because they are then always pronounced within their own intonational phrase, cf. Jun & Fletcher (2014: 495); Féry (2017: 179). Evidence from ellipsis phenomena also underscores the point that a close mapping to morphosyntax can be only one of several competing constraints affecting prosodic phrasing.[49] This insight is captured in the nature of these constraints as violable, as usual for OT. More recent studies have also reformulated the SLH in a family of OT-constraints. Selkirk (1996: 189–190) proposes the following four constraints, where $C_i$ is a prosodic constituent of level *i*, PWd is a prosodic (phonological) word, Ft is a foot:

(18) SLH as violable OT-constraints (Selkirk 1996: 189–190)
(i) *LAYEREDNESS* No $C_i$ dominates a $C_j$, j > i, e.g. “No σ dominates a Ft.”
(ii) *HEADEDNESS* Any $C_i$ must dominate a $C_{i-1}$ (except if $C_i$ = σ), e.g. “A PWd must dominate a Ft.”
(iii) *EXHAUSTIVITY* No $C_i$ immediately dominates a constituent $C_j$ , j < i-1, e.g. “No PWd immediately dominates a σ.”

[49] Consider the following short dialogues. In i), A’s monosyllabic utterance would be uttered with question intonation, i.e. IP-level boundary tones. B’s answers are equally full IPs but of varying syntactic status. That B can felicitously answer with a-d in i) but only a in ii) makes an account of A’s question as an elliptical version of a full question sentence doubtful because of question-answer congruence (but cf. Gretsch 2003 for the same observation and a syntactic account). The “elliptical” question by A encoded by the question intonation is clearly underspecified in a way which a “full” syntactic version can never be, either as a wh-question (Where are the keys? – Here / # Yup) or a polar question (Do you have the keys? – Yup / Here / # My pockets). In iii), it is clear that syntactic accounts can hardly explain the unfelicitousness of B’s c. compared to a. A’s question is here specified by the context (with a strong bias for expecting the answer to be epistemically accessible to the interlocutors but not the speaker) in a different way to how a “full” syntactic question would be specified (Who has the keys? - You [declarative intonation]). The constraints for prosodic well-formedness at the level of the IP are therefore primarily context-dependent and pragmatic, such that information which is sufficiently encoded to provide a context update given the discourse context can form an IP, but not less information (or more, according to the “exactly one idea”-proposal by Chafe 1994).

i. Context: couple leaving their flat together
A: Keys? [question intonation]
B: a. Here. b. Yup. c. Got them. d. My pockets [declarative intonation]

ii. Context: the couple have just closed the door behind them after leaving the flat and often forget to turn off the lights when leaving
A: Lights? [question intonation]
B: a. Yup. b. # Here. [declarative intonation]

iii. Context: three flatmates sharing one pair of keys leaving their flat together
A: Keys? [question intonation]
B. a. Me. b. You got them. c. # You. [declarative intonation] d. You [insistent intonation]

Here and elsewhere, the hash (#) before examples is used to indicate that the example is well-formed (grammatical) in the language, but infelicitous/pragmatically odd in the given context. The asterisk (*) in the same position indicates that the example is not well-formed (ungrammatical), independent of context, according to what is assumed about the grammar of the language in question.

(iv) *NONRECURSIVITY* No $C_i$ dominates $C_j$ , j = i, e.g. “No Ft dominates a Ft.”

For Selkirk (1996: 191), *EXHAUSTIVITY* and *NONRECURSIVITY* are violable constraints, so that level skipping as in (15)b and recursive structures as in (15)c can be allowed in some languages if other constraints are ranked higher. The first two constraints however, *LAYEREDNESS* and *HEADEDNESS*, are inviolable universally (i.e. in all natural languages) in her conception.[50] This is also maintained in Myrberg (2013), but she proposes to further complement *EXHAUSTIVITY* and *NONRECURSIVITY* with a slightly different constraint, *EQUALSISTERS*:

(19) *EQUALSISTERS* (Myrberg 2013: 75)
Sister nodes in prosodic structure are instantiations of the same prosodic category.

*EQUALSISTERS* is argued by Myrberg (2013: 78) to be able to explain preference patterns for prosodic structures in Stockholm Swedish which *EXHAUSTIVITY* and *NONRECURSIVITY* cannot differentiate between (the latter but not the former being violated uniformly in all of them). She argues that these structures are recursive (with an intonational phrase containing one or two intonational phrases) based on the distribution of IP-initial and IP-final as well as lexical and focal pitch accents and their downtrend behaviour (Myrberg 2013: 98–100, 108–110), i.e. on observable empirical grounds.

### 3.6.2 Empirical evidence for recursive prosodic structure from systematic pitch scaling

While some of the arguments for recursive prosodic structure in the literature are clearly motivated by maintaining a tight mapping between syntax and prosody and are principally theoretical, there is another line of arguments for it that is based on empirical findings and can be evaluated independently of assuming such a tight mapping. They mostly come from findings about pitch scaling, which is an area that AM is still

[50] Féry (2015) discusses examples from non-extraposed German relative clauses that she calls “prosodic monsters” and which she argues to also violate *Layeredness* (i.e. structures like (15)d) by containing an IP within a phonological phrase. An important point she wishes to make is that syntax and prosody interact as equals: syntactic and prosodic constraints must be in the same ranking hierarchy in order to account for the acceptability differences between the examples she gives. However, her analysis of the prosodic phrasing is based only on a close mapping between syntactic and prosodic categories and she does not provide any measurable intonational evidence against alternative analyses with less complex prosodic structure and a looser mapping.

struggling to gain a solid theoretical grasp on (cf. section 3.4.2). We will review some of those that argue for a recursive category at the top of the prosodic hierarchy (IP) in the following. Terminologically, “downstep” will be used to label a process analyzed as phonological and resulting in a lowering of pitch level between two successive units; “declination” is the phonetically (physiologically) determined lowering of pitch over longer stretches, and “downtrend” is used to describe either of these without committing to an underlying cause. “Upstep” is the counterpart of downstep in the other direction.

#### 3.6.2.1 Stockholm Swedish

Myrberg (2013) argues for prosodic recursion based on Stockholm Swedish data such as the utterance shown in Figure 11. Swedish has a lexical pitch accent so that content words belong to one of two accent classes, called accent I and accent II. IPs are delimited at their left edge by a so-called “initiality accent” located on the first prosodic word, L*H on an accent I word, and H*LH on an accent II word. At their right edge, they are delimited by a high or low boundary tone (H% or L%). The most prominent phonological phrase (PP here) in an IP gets assigned a “focal accent” on the rightmost prosodic word which is almost of the same form as the “initiality accent”. The “initiality accent” is argued not to be the same as the “focal accent” because unlike the latter, its position is only determined by the position in the IP and not at all by information structural criteria, appearing also on backgrounded and given words. Its final H is also aligned later and more variably than that of the “focal accent” (Myrberg 2013: 85–87, cf. also Horne et al. 2001; Roll et al. 2009). In utterances like the one shown in Figure 11 from a corpus of read utterances by three speakers (n=27, cf. Myrberg 2013: 93–94), the tonal movement realizing the initiality accent is produced twice (on *andra* and *Anna*), and that of the focal accent as well (on *utklädda*, which has a secondary stress on the second syllable leading to the association of the L tone as L*, and on *med*). The focal accent pitch movements are both followed by clear low valleys indicating an L% tone at the right boundary of the IP. Thus judging from the presence of these initial and final pitch movements, it is clear that these examples consist of two IPs, with each extending over one of the main clauses making up the coordinated syntactic structure (see the lower part of Figure 11). These utterances however all (Myrberg 2013: 98) also display a steady downtrend in the relative scaling of their local pitch maxima and minima, as shown by the dashed lines in the pitch track in the upper part of Figure 11. The fact that this downtrend continues across the utterance and that no or only partial pitch reset takes place at the boundary between the two IPs is taken to be evidence for the presence of a third, larger, IP which fully contains the two smaller ones.

Pitch reset normally occurs at the boundary between IPs in Swedish (Myrberg 2013: 108–109), and is also taken as a universal cue for IPs by Himmelmann et al.

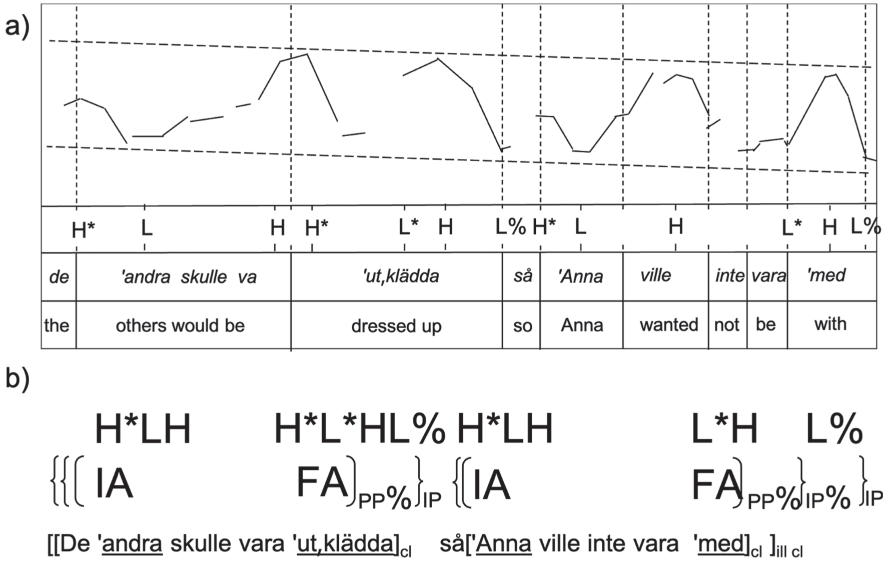


**Figure 11:** Pitch track of a Stockholm Swedish utterance, *de andra skulle vara utklädda så Anna ville inte vara med* “the others were getting dressed up, so Anna didn’t want to join”, with dashed lines indicating downtrend of the pitch maxima and minima over the time course (a), and its proposed phrasing as two coordinated minimal IPs in a maximal IP (b), adapted from Myrberg (2013: 97, 99).

(2018). That it does not occur in the Swedish examples is thus evidence that as a whole they consist of a constituent in whose domain reset is suspended, i.e. an IP.

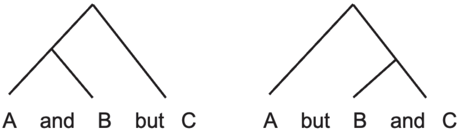
#### 3.6.2.2 British English

A very similar argumentation has been brought forward by Ladd (1986, 1988, 1990, 1993, 2008: 294–297) based on evidence from somewhat more complex English data. Ladd (1988) presents read speech data from speakers of British English in two conditions, the “A and B but C” condition and the “A but B and C” condition, exemplified by (20)a and (20)b, respectively. Their hierarchical structure[51] is given in (20)c, intended to show that the clauses connected by *and* are sister nodes under a higher constituent, while the clause preceded by *but* is a sister node at the level of

[51] Note that Ladd (1988: 531) argues for the hierarchical asymetry in clause status and boundary strength based on semantic and pragmatic considerations, not on syntactic ones: “The most natural interpretation of these sentences appears to be one in which the but opposes one proposition to the two conjoined by and [. . .]”. The experimental materials from all the studies in this section can be interpreted in this way, suggesting that hierarchical pitch scaling at least partially serves the function of cueing discourse cohesion and coherence.

that constituent. This leads to the intuitive understanding of the boundary before *but* as stronger than the one before *and* in these constructions, which Ladd (1988: 531) hypothesizes is reflected in pitch scaling.

(20) “A but B and C” and “A and B but C” sentences from Ladd (1988: 532)
a. Allen is a stronger campaigner, and Ryan has more popular policies, but Warren has a lot more money
b. Ryan has a lot more money, but Warren is a stronger campaigner, and Allen has more popular policies
c.



When read by Ladd’s test subjects, the sentences were consistently produced with several pitch accents on the prominent words and a final falling boundary movement at the end of each clause, indicating that each clause is produced as an IP (Ladd 1988: 532). In two experiments, the first one involving examples with three accented words per clause (4 speakers x 18 sentences x 2 conditions = 144 utterances), the second with four accented words per clause (same number of utterances), peak measures taken for the accents showed a downtrend across each IP corresponding to an individual clause, but with partial, not full, reset between the IPs, indicating them all to be part of a larger constituent in which full reset is suspended (Ladd 1988: 535, 539). Average peak measurements on the first accent following a *but*-boundary were higher than those on the corresponding accent in the other condition, when it followed an *and*-boundary, and this difference in boundary strength was also mostly supported by durational measurements at the boundaries (Ladd 1988: 535–536, 539). These differences were shown to be statistically significant in the second experiment for all speakers (significant interaction in ANOVA of sentence type x clause type x accent and/or sentence type x clause type), and for three in the first (Ladd 1988: 535, 539). A third experiment served as a control to ensure that it is not a local effect of *but* which is responsible for the observed scaling differences, but indeed the hierarchical structure.

#### 3.6.2.3 Northern German

Truckenbrodt & Féry (2015) (see also Féry & Truckenbrodt 2005) essentially replicate Ladd’s (1988) study with German data, using similar experimental conditions in a reading task with 5 northern German speakers as subjects.

(21) Example sentences for the three experimental conditions from Truckenbrodt & Féry (2015: 25–27), underlined syllables are expected to be accented

a. AX condition (Ladd’s A but B and C condition): *A während [B und C]*$_{\text{X}}$
*Context*: Warum meint Anna, dass Handwerker teurere Autos haben als Musiker?
‘Why does Anna think that craftsmen have more expensive cars than musicians?’
*Sentence:* Weil der Maler einen Jaguar hat, während [die Sängerin einen Lada besitzt, und der Geiger einen Wartburg fährt]
‘Because the painter has a Jaguar, while [the singer owns a Lada and the violinist drives a Wartburg]’

b. XC condition (Ladd’s A and B but C condition): *[A und B]*$_{\text{X}}$ *während C*
*Context:* Warum meint Anna, dass Musiker teurere Autos haben als Sportler?
‘Why does Anna think that musicians have more expensive cars than sportsmen?’
*Sentence:* [Weil die Sängerin einen Jaguar hat, und der Geiger einen Daimler besitzt], während der Ringer einen Lada fährt
‘[Because the singer has a Jaguar and the violinist owns a Daimler], while the wrestler drives a Lada’

c. No-X condition (control condition)
*Context:* Warum meint Anna, dass ihre Nachbarn teure Autos haben?
‘Why does Anna think that her neighbours have expensive cars?’
*Sentence:* Weil Möller und Hummel einen Jaguar haben, Meyer und Lerner einen Daimler besitzen, und Wollmann und Lehmann einen BMW fahren
‘Because Möller and Hummel have a Jaguar, Meyer and Lerner own a Daimler, and Wollmann and Lehmann drive a BMW’

As the examples in (21) show, their experimental sentences all create an argumentative contrast between two propositions one of which, X, is itself a coordination of two propositions. Each proposition is expressed through an individual clause, the two coordinated ones separated by *und* “and”, and the contrasting one separated by *während* “while”. In the control condition, no argumentative contrast exists. As in Ladd (1988), each clause is realized with several pitch accents (here, one prenuclear one and a nuclear one on the final accented word in the IP, both L*H) and a separate final boundary tone (H% at the end of the prefinal and L% at the end of the final IP) and thus analyzed as forming IPs (Truckenbrodt & Féry 2015: 27–29). One important difference to Ladd’s experiment is that because in German, the nuclear accent in an

IP is upstepped (Féry & Kügler 2008), the scaling of this upstep[52] is also hypothesized to be affected by the proposed hierarchical prosodic structure (Truckenbrodt & Féry 2015: 24), in addition to the scaling on the IP-initial accents. Their results confirm Ladd's in that the difference in boundary strength is reflected in the scaling of the initial accent between the two conditions (see A in Figure 12), and no such effect is found in the control condition, which shows downtrend with partial reset across the three IPs, as in Swedish ( Truckenbrodt & Féry 2015: 30). In order to model this effect, they use the concept of phrasal reference lines, first proposed in van den Berg et al. (1992) for Dutch. There it was observed that peak values in downstep contexts seem to orient themselves towards an abstract reference line extending across a phrasal constituent. Truckenbrodt & Féry (2015: 31–32) analyze their results as showing that a phonological process of downstep is responsible for the observed effects. Downstep is implemented as a lowering of the respective phrasal reference lines. It only applies between sister nodes in the prosodic structure (as also proposed by Ladd 1988: 541–542), lowering the second sister, and its effects manifest as less strong between higher constituents than between lower ones.[53] This is schematically represented[54] in B in Figure 12. In the AX condition, lowering takes place between the IPs A and X (=B + C) and between B and C. In the XC condition, it takes place within X, i.e. between A and B, and between X and C, but not between B and C, as they are not sisters. The actual pooled values reflect this: the initial peaks in B and C in the AX condition are both lowered with respect to the preceding initial peak, whereas in the XC condition, they are at the same level, and B is lower here than in the AX condition.

The results on the upstepped peaks also confirm the general predictions of this model but show some additional variation. In the AX condition, upstepped IP-final nuclear peaks are roughly at the same height as the preceding initial peaks (no significant differences in paired t-tests), but in the XC condition, the height of the nuclear accent in the second clause (B), labeled H4 in Figure 13, varies. For two speakers, it is significantly higher than the preceding IP-initial peak H3 and not significantly different from H1, the initial peak in the first clause, while for two others, it is the other way around, with H4 being at broadly the same height as H3

[52] The nuclear accent in the experiment sentences always fell on a lexical item signifying a car name in each IP, which was contrasted with a different item of the same type in the other clauses.

[53] Correlating boundary strength with amount of lowering is the proposal by Ladd (1988), reformulated via depth of embedding in Féry & Truckenbrodt (2005). Truckenbrodt & Féry (2015: 32–33) relate the difference between the initial accents in B in the two conditions to an additional stipulation of final lowering (Liberman & Pierrehumbert 1984) applying to the phrasal reference line of the last sister under a node. The different proposals do not make any differing predictions with regards to the data discussed in Truckenbrodt & Féry (2015).

[54] In reality, the reference lines should presumably not be thought of as horizontal, but as always including some slowly increasing amount of decay, the effect of phonetic declination over any utterance.

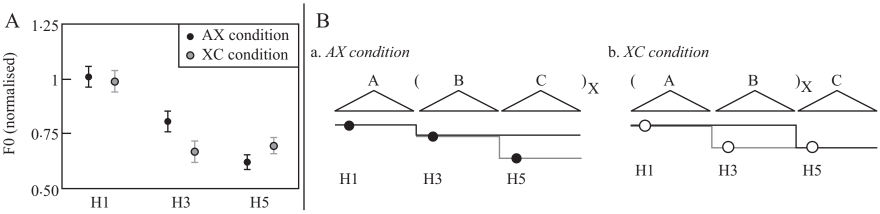


**Figure 12: A:** Mean values and 95% confidence intervals for the IP-initial peaks, normalized and pooled values of speakers S1–S5 for the two conditions AX (n= 67) and XC (n= 66). *H1*, *H3*, and *H5* are the respective initial peaks of the clauses A, B and C, from Truckenbrodt & Féry (2015: 31).
**B:** Schematized phrasal reference lines and IP-initial peaks for the two experimental conditions, from Truckenbrodt & Féry (2015: 32).

and lower than H1 (Truckenbrodt & Féry 2015: 35). This is reflected in the average value for H4 in the XC condition seen in B in Figure 13. The interpretation is that the height of H4 is variably orientated either towards the higher reference line of X, or the downstepped one of B in the XC condition, as shown in A of Figure 13. The fifth speaker always upstepped H4 almost to the height of H1 in both conditions, pointing once again to the need for further studies with more speakers to illuminate the role of inter-speaker variation in these phenomena.

### 3.6.2.4 European Portuguese

European Portuguese is a Romance language for which comparable empirical findings about prosodic boundary strength have been connected to a hypothesis about prosodic recursion (Frota 2012, 2014). This argument is based on gradual differences corresponding to boundary strengths not only with respect to pitch scaling, but also phrase-final lengthening and even the likelihood of the application of segmental prosodic processes. The IP in European Portuguese, according to Frota (2014: 11–13), is the domain of a number of processes, including the following: a) it is the domain at which at least one pitch accent is obligatory, on the strongest word in the final PhP in the IP; b) it has an obligatory boundary tone at its right edge and an optional one at its left edge; c) word-final fricatives are voiced when followed by a word-initial vowel within the IP, but voiceless at its edges; d) it is the domain of final lengthening; and e) clitic elements have a tendency to be realized in their stronger forms when positioned at the left edge of the IP. The distribution of pitch accents and boundary tones (a+b) is shown to agree with the phrasing cued by fricative voicing (c) and final lengthening (d) in (22)a and b and the corresponding pitch tracks given in Figure 14. The tonic and posttonic syllables of the words preceding an IP-boundary as marked in (22)a and b are lengthened (underlined in the examples), a pitch accent is realized on the final stressed syllable followed by a boundary

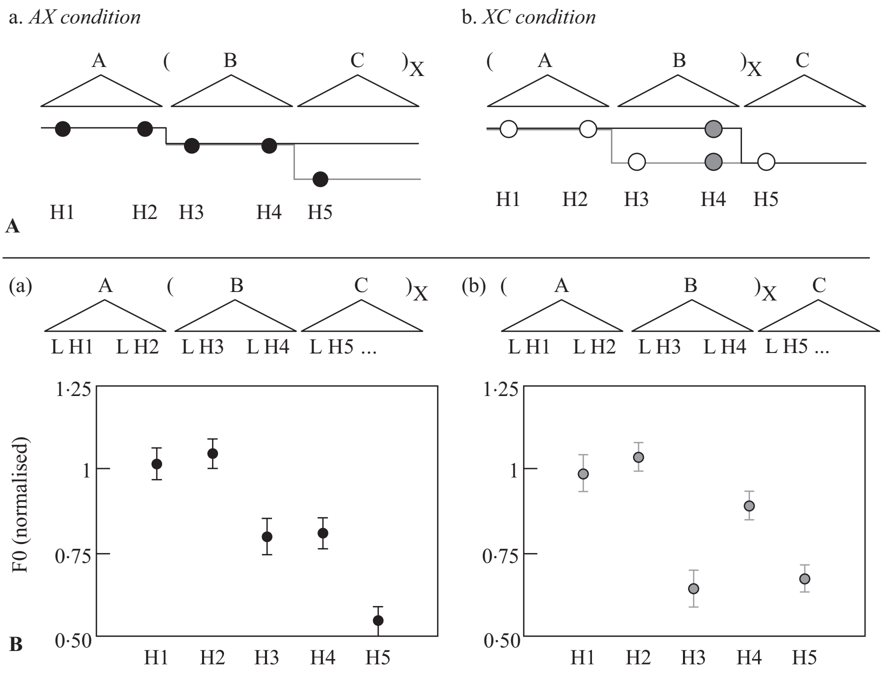


**Figure 13: A:** Schematized phrasal reference lines and upstepped nuclear peak values for the two experimental conditions, from Truckenbrodt & Féry (2015: 34). **B:** Mean values and *95%* confidence intervals for upstepped nuclear peaks *H2* (clause A) and *H4* (clause B) in relation to the three IP-initial peak values *H1, H3*, and *H5* in the normalized pooled data of speakers S1–S4. (a) AX condition (n=51); (b) XC condition (n=50), from Truckenbrodt & Féry (2015: 35).

tone, as evidenced by the pitch movements in Figure 14, and IP-final fricatives are realized as voiceless [ʃ], while IP-internally, they are realized as voiced [z].

The difference between (22)a and (22)b is proposed to be that while in (22)a, all IPs are of equal status, in (22)b the “short IP” (Frota (2014: 11)) *as alunas* “the students” is embedded within the larger IP extending until *até onde sabemos*, forming a recursive “compound prosodic domain” (Ladd 2008: 297–309; Frota 2012, 2014). The short IP is argued not to be an entirely different category, e.g. an intermediate phrase, because the differences to a longer IP are mostly only gradual: while fricative voicing extends throughout the larger IP domain in (22)b (visible also in the clearly different spectral quality of the segment corresponding to the fricative in B compared to A in the original pitch tracks[55] in Frota 2014: 13), final lengthening

[55] Voice onset time (VOT), the phonetic exponent of voicing, is of course also a continuous parameter. It might turn out that both VOT and the place difference between [z] and [ʃ] as realized in terms

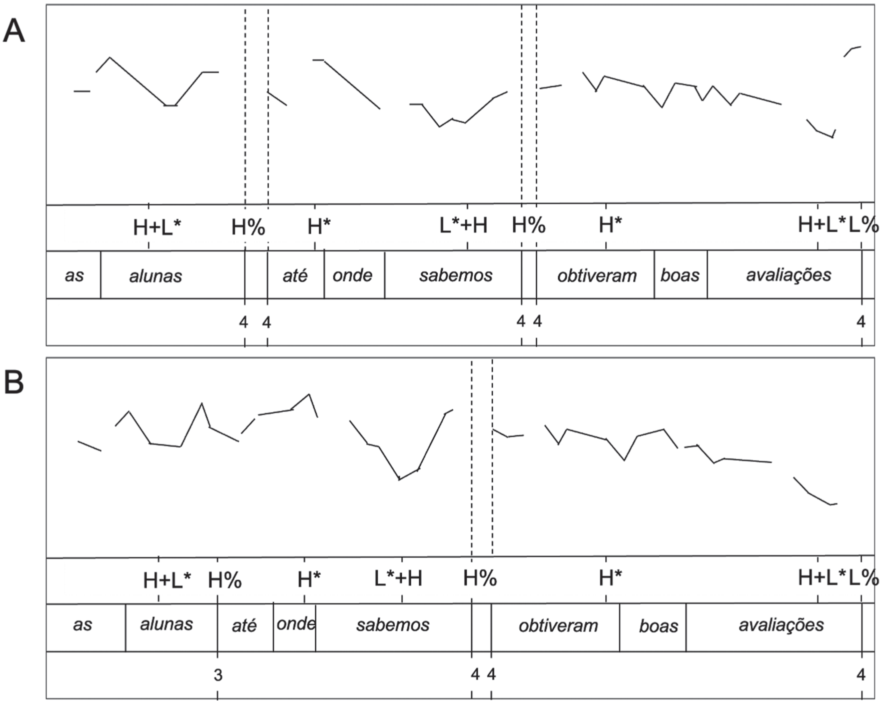


**Figure 14:** Two versions of an elicited European Portuguese utterance, corresponding to the proposed phrasing structures in (22)a (A) and (22)b (B). Adapted from Frota (2014: 13).

and pitch scaling are decreased (but still present) at the right edge of the short IP in comparison to the right edges of larger IPs, as exemplified in Figure 14.

(22) European Portuguese example sentences with different proposed phrasings (Frota 2014: 14)

a. { a[z] aluna[ʃ] }$_{\text{IP}}$ { até onde sabemo[ʃ] }$_{\text{IP}}$ { obtiveram boa[z] avaliaçõe[ʃ] }$_{\text{IP}}$

b. {{ a[z] aluna[z] }$_{\text{IP}}$ { até onde sabemo[ʃ] }$_{\text{IP}}$ }$_{\text{IP}}$ {obtiveram boa[z] avaliaçõe[ʃ] }$_{\text{IP}}$

“The students, as far as we know, have got good grades”

For the fifth boundary cue given above, the realization of clitics in their strong or reduced form (e), the data presented in Frota (2014) does not directly bear on the question of whether the IP is recursive or not in European Portuguese. However,

of spectral quality are actually also subject to gradual effects of boundary strength in European Portuguese utterances like these.

the percentages given with the examples in (23) indicate that this is also not a discrete but gradual difference, which only comes out when quantified: in (23)a, the clitic element in question *a os* "to the" is placed IP-medially, and realized in the strong, i.e. diphthongized, form [awʃ] in 20% of the elicitations, and in the reduced monophtongized one [ɔʃ] in 80% of them. When the clause *a os jornalistas* "to the journalists" receives its own phrase, either fronted as in (23)b or in the same position ((23)c), which in both cases effects a topical interpretation according to Frota (2014: 14), so that the clitic elements are IP-initial, this number goes up to 88% and 92%, respectively. Although cases testing this are not reported on in Frota (2014), the difference between 20% and around 90% is large enough to hypothetically allow for intermediate prevalences (say, around 55%) that might characterize the behaviour of cases that are in-between: at the left edge of the second element in a compound IP (comparable to the beginning of *até onde sabemos* in (22)b), for example. In such a position, a clause like *a os jornalistas* would be adjacent to an initial boundary, unlike in (23)a, but this boundary would be weaker than the boundary the clause is adjacent to in (23)b and (23)c following Frota's analysis. Should this turn out to be the case, it would indicate that boundary strength of recursive domains is also expressed by the quantitative prevalence of an otherwise discrete phenomenon.[56]

(23) Three European Portuguese test sentences with reported prevalence in realization as its strong form [awʃ] instead of the reduced form [ɔʃ] of the clitics *a os* "to the", from Frota (2014: 14)

a. [ a[z] angolana[z] ofereceram especiaria[z] [ɔʃ] jornalista[ʃ] ]$_{IP}$
'The Angolan women offered spices to the journalists' (20% occurrences diphthongized)

b. [ [awʃ] jornalista[ʃ] ]$_{IP}$ [ a[z] angolana[z] ofereceram especiaria[ʃ] ]$_{IP}$
'To the journalists, the Angolan women offered spices' (88% occurrences diphthongized)

c. [ a[z] angolana[z] ofereceram especiaria[ʃ] ]$_{IP}$ [ [awʃ] jornalista[ʃ] ]$_{IP}$
'The Angolan women offeres spices, to the journalists' (92% occurrences diphthongized)

[56] Although this is presumably a simplification. Just as VOT and fricative spectral energy are located on a continuum, so are the differences in formant frequency change between [ɔ] and [aw]. If it were the case that non-maximal IP boundaries really resulted in an intermediate rate of monophtongization, it would be interesting to see whether measurements of these continuous parameters also revealed intermediate realizations. This would then presumably be reflected quantitatively as differing rates of realization of a discrete variable because of our categorical perception.

#### 3.6.2.5 (Tokyo) Japanese

For (Tokyo) Japanese, Kubozono (1989) reports on findings that again also involve tonal scaling. It is usually assumed that the intermediate phrase in Japanese is the domain of downstep (Beckman & Pierrehumbert 1986; Venditti 2005), such that the pitch accents and initial rise of a second accentual phrase within the same intermediate phrase are realized at a lower pitch than those of the preceding one. However, when they are grouped into two separate intermediate phrases, the second AP's tones are realized at the same height as that of the preceding one, and reset takes place. In addition, the perceived disjuncture between words belonging to two separate intermediate phrases is larger than between two APs, e.g. pauses are more likely to occur or to be longer[57] (Venditti 2005: 176, 185–186). Kubozono's (1989) findings show that phrases composed of right-branching complex noun phrases containing only accented words (i.e. made up of as many APs, since it is usually taken to be definitional of APs to contain a single lexical pitch accent, cf. Ito & Mester 2012: 284) such as (24)a, realize a "pitch boost", an effect similar or equivalent to upstep, on the second accent, i.e. the one following the stronger boundary, while those composed of left-branching noun phrases, such as (24) b, do not. This upstep also takes place on the third accent in phrases realizing balanced structures, such as (24)c, again after the stronger boundary.

(24) Japanese test phrases with prosodic bracketing, adapted from Kubozono (1989)

a. right-branching complex noun phrase[58]

| | | |
|---|---|---|
| ${}_{iP}[{}_{AP}[\text{kowa'i}]_{AP}]_{iP}$ | ${}_{iP}[{}_{AP}[\text{me'-no}]_{AP}$ | ${}_{AP}[\text{ya'mai}]_{AP}]_{iP}$ |
| terrible | eye-GEN | disease |

"terrible eye disease" (Kubozono 1989: 42)

[57] Note however that Venditti (2005: 186) calls pauses neither a necessary nor a sufficient condition for the perception of a disjuncture of the kind that occurs between intermediate phrases. She also points out that there are cases where the perceived disjuncture does not match with the tonal cues assumed for a given category.

[58] Kubozono (1989) locates the difference between these three types of sentences in the syntax, and assumes that this is then what is reflected in differing prosodic structures. The argument for syntactic difference however is only built on the lexical semantics and world knowledge: in *kowai me-no yamai* "terrible eye disease", *kowai* "terrible" is only understood to modify *me-no yamai* because that is the most likely interpretation. A marginal interpretation of "the disease of terrible eyes" is presumably possible. If *kowai* were replaced by *akai* "red", the noun phrase would most likely be understood as "the disease of red eyes" and correspondingly parsed and phrased as [akai me-no][yamai], because redness is something more likely to be said of eyes than of an eye disease. Similarly, in the other two examples, replacement of one of the lexical items by another of the same class, without changing anything in the observable syntactic structure, would result in an interpretation with different branching. It would be interesting to see if different kinds of prosodic phrasing were available in ambiguous cases, such as a variant on (24)b, *kanbashii ayu-no nioi*, where *kanbashii* "aromatic" could modify either *ayu* or *ayu-no nioi*.

b. left-branching complex noun phrase

| | | |
|---|---|---|
| $_{iP}$[$_{AP}$[naˈmano]$_{AP}$ | $_{AP}$[aˈyu-no]$_{AP}$]$_{iP}$ | $_{iP}$[$_{AP}$[nioˈi]$_{AP}$]$_{iP}$ |
| raw | type.of.fish-GEN | smell |

“smell of raw *Ayu*” (Kubozono 1989: 44)

c. balanced complex noun phrase

| | | | |
|---|---|---|---|
| $_{iP}$[$_{AP}$[naˈoko-no]$_{AP}$ | $_{AP}$[aˈni-no]$_{AP}$]$_{iP}$ | $_{iP}$[$_{AP}$[aoˈi]$_{AP}$ | $_{AP}$[eriˈmaki]$_{AP}$]$_{iP}$ |
| Naoko-GEN | older.brother-GEN | blue | muffler |

“Naoko’s older brother’s blue muffler” (Kubozono 1989: 51)

Parallel effects of prosodic boundary strength on the scaling of the AP-initial rise were found by Selkirk et al. (2003) in a reading task with differing syntactic structures. They found that scaling depended on whether the AP was initial or non-initial in an iP. Based on these findings by Kubozono (1989) and Selkirk et al. (2003), Ito & Mester (2012) propose to reduce the two levels of the AP and iP in Japanese into one, as recursive instantiations of ϕ, the phonological phrase. In their interpretation, downstep applies at that level. Because it is definitional for the minimal ϕ (=AP) to contain only one lexical pitch accent (and one phrase-initial accent) and downstep only applies between two pitch accents, they argue that the effect of downstep is only visible at a recursive instance of a non-minimal ϕ, i.e. an ϕ that contains at least another ϕ, and the scaling of the AP-initial rise is increased with increasing levels of recursion of the ϕ (Ito & Mester 2012: 286). The same boundary strengthening effect is what they invoke to explain the upstep in Kubozono’s data (Ito & Mester 2012: 294).

Independent of prosodic recursion, all of these studies provide evidence that initial prosodic strengthening (Fougeron & Keating 1997, cf. also 3.2.2) also affects pitch scaling and is sensitive to differing boundary strengths, including differences that are not predicted by the traditional prosodic hierarchy. They also provide evidence that pitch scaling, especially in the case of regular downtrend, cannot be modeled purely locally, in contrast to what Pierrehumbert (1980); Liberman & Pierrehumbert (1984) had suggested, and that it instead seems to make reference to a hierarchical organization of larger prosodic units. From there, assuming prosodic recursion proceeds by two main arguments: for one, the fact that the prosodic unit of the smaller subparts and that of the whole utterance do not differ in their discrete cues (boundary tones or constraints on accent placement within them) but only gradually in their boundary strength as cued by pitch scaling of the boundary-adjacent movements, duration and degrees of pitch reset, is seen as most compatible with an account taking them all as instantiations of the same recursive phrasal category (Ladd 1986: 327–328; Frota 2012: 260–261; Ito & Mester 2012: 294; Myrberg 2013: 108–110; Truckenbrodt & Féry 2015: 37). Secondly, if the component units were taken to be categorically different from the overarching unit based on these gradual differences, then sufficiently long utterances with sufficiently complex structure would require postulating a large number of additional categories ad hoc, with no principled way of accounting for their similarities (the fact that they all share the same boundary phenomena with only gradual differences, cf. Ladd 2008: 294–295).

### 3.6.3 Separating arguments for prosodic recursion from assumptions of universality and a close prosody-syntax mapping

I would like to restrict arguments for recursive prosodic phrasing to those based on the kind of evidence described in the preceding sections, because they require the fewest theoretical presuppositions. There are other works invoking recursive prosodic structure at a level above the prosodic word for the analysis of their observations, but they are based on additional theoretical assumptions. For example, Elfner (2015) proposes recursive phonological phrases to account for the distribution of phrase-initial LH and phrase-final HL pitch accents in Connemara Irish: the HL accent appears on the stressed syllable of the final word in the phonological phrase, but the LH accent only appears on the stressed syllable of the leftmost word in a ϕ that is not minimal, i.e. one that dominates another ϕ (Elfner 2015: 1180, 1182), similar to the argument in Ito & Mester (2012). While her analysis seems well-suited to the data she presents, this sort of evidence is of a somewhat different nature: instead of gradual differences between instantiations of what is the same prosodic category by all other cues, as we have seen in the other cases discussed in this section, in the Irish case under Elfner's proposal, the different recursive levels of ϕ actually differ categorically, in their tonal makeup (LH initial in all ϕs, LH initial and HL final in all non-minimal ϕs). Bickel et al. (2009); Schiering et al. (2010) argue for up to 4 prosodic units close to the prosodic word in Limbu (Sino-Tibetan), based on the observation that they all serve as the application domain of quite distinct processes. They discuss and explicitly discard the option to solve this puzzle via a recursive prosodic word domain, precisely because the domains in question do not have the same phonological properties (Bickel et al. 2009: 50). However, the necessity for a recursive solution here only even arises when there is some incentive to apply the same set of prosodic categories universally for all languages. That is clearly the case for Ito & Mester (2012: 287–289), who posit recursive versions of the prosodic word (ω), phonological phrase (ϕ) and intonational phrase (ι) as universal interface categories with the syntax and who explicitly reject the postulation of prosodic categories based only on observable phonetic effects. Elfner (2015: 1203) equally voices some optimism that such a universal reduction would facilitate typological comparison because it is “uncontroversial to assume that all languages distinguish an intermediate category” between ω and ι.[59]

While this assertion is perhaps premature in the face of reliable empirical studies on the prosody of probably less than 5% of all living languages, it shows that some of the driving motivation for the proposal of prosodic recursion lies in this potential for universalization as well as the close correspondence with syntax. If these are not presupposed axiomatically, it is not problematic to assume 4 distinct categories resembling the prosodic word in Limbu, for example, if they can indeed all be shown to be the domains of distinct processes. The evidence on prosodic boundary strengthening we have discussed is of a different kind, because if recursion is not assumed in those cases then the number of categories of the prosodic hierarchy will potentially rise indefinitely, assuming that gradual boundary differences can still be found in sufficiently long and complex utterances (which hasn’t been tested so far).[60] This is conceptually problematic even if no assumptions about the universality of prosodic categories or a tight mapping between syntax and prosody are made. It is clear in any case that prosodic recursion must have its limits, as argued for by both van der Hulst (2010) and Ladd (2008: 297–299): no one has ever proposed an intonational phrase inside a syllable, for instance, and the relative “flatness” of prosodic structure as compared to the syntactic structure seems to be still agreed upon by everyone. In addition to clear evidence of abundant mismatch between syntax and prosody, there is also evidence that the same kind of syntactic embedding does not correlate with similarly recursive prosodic structures in different languages: Féry & Schubö (2010) found evidence from downstep indicating recursive prosodic phrases in German utterances with syntactic center-embedding, but not for similar Hindi utterances.

In the interest of adequate descriptions for each language under investigation, universal claims should be made very cautiously, while the general idea of different prosodic levels that stand in a hierarchical relation to each other and serve as

[59] This is somewhat ironic because for the analysis of her Irish examples, all full utterances, she makes no recourse to a second phrasal category (namely ι) at all, having no need for anything above a maximal ϕ in terms of unexplained cues or structural analysis.

[60] For Korean, Jun (2007) takes this path: she proposes the intermediate phrase based on findings about cues that resemble those for the AP, but in a stronger version (increase in tonal scaling, but depending on the AP-initial tones, which are in turn dependent on the AP-initial segments, cf. Jun 2007: 160–161). Above the iP there is also an IP, which is clearly separate because it has different cues (additional boundary tones, cf. Jun 2007: 155). Whether it is more parsimonious to take the Korean intermediate level also as a recursive version of the AP, as Ito & Mester (2012) propose for Japanese, or as its own category, as Jun (2007) does, is so long a question of theoretical choice as it is unclear whether the cue strength for it differentiates only between two versions (AP and iP) or indefinitely and correlating with the number of proposed levels of recursive AP-embedding.

the domain of phonological rule application should always be kept in mind. Ladd (2008: 298–299) proposes a weakened version of the SLH, in which the different levels are still ranked, but where compounding of levels is allowed in order to reduce the number of categories necessary for the adequate description of each language, explicitly with the motivation that this would “allow us to identify any given boundary as being of one category or another on purely phonetic and phonological grounds, without as it were looking over our shoulder at the theoretical consequences for prosodic structure or for syntax-prosody mapping”.

### 3.6.4 What is (not) known about prosodic recursion in Spanish and Quechua

As far as I have been able to determine, similar empirical studies have not been done on Spanish (or on any variety of Quechua). I will review some results from related studies on Spanish, but the issue of prosodic recursion based on empirical evidence itself so fas has not been tackled. Garrido et al. (1995) attempt to correlate degree of final lengthening, pitch reset and pre-boundary pitch contours in read Spanish utterances with differing syntactic boundaries[61] but only find non-significant tendencies for final lengthening. Because they treat pitch reset and pre-boundary contours only as categorical variables (presence or not of reset and a pre-boundary peak), their findings in this respect are not interpretable for the issue at hand. Rao (2008) investigates phrasing patterns in Barcelona Spanish, via SVO sentences read by 18 speakers that systematically differ in terms of the length (in prosodic words) and complexity (depth of syntactic branching) of both subject and object. He assumes that under a tight mapping between syntax and prosody and following a proposal originally from Ladd (1986), postnominal appositives should form a recursive phrase together with their head NP (Rao 2008: 100–101). He also provides the pitch track of an individual example where tonal scaling might reflect a difference in boundary strength:

In Figure 15, pitch is scaled clearly higher at the end of the postnominal adjectival modifier *inteligente y gordo de Barcelona* than after *el Javier*, with which it forms an NP together. However, Rao (2008: 101) does not claim that this is recursive prosodic phrasing, unlike in the case of an appositive. He also does not make any statements about whether this scaling difference is an individual occurrence or a robust finding in this or similar conditions. Instead, he simply assumes that whenever a phrasing such as [head noun]$_{\text{PhP}}$ [appositive]$_{\text{PhP}}$ occurs, this constitutes recur-

[61] They do not refer to categories of the prosodic hierarchy, but under a close syntax-prosody mapping their different syntactic boundaries would likely result in recursive phrasing.

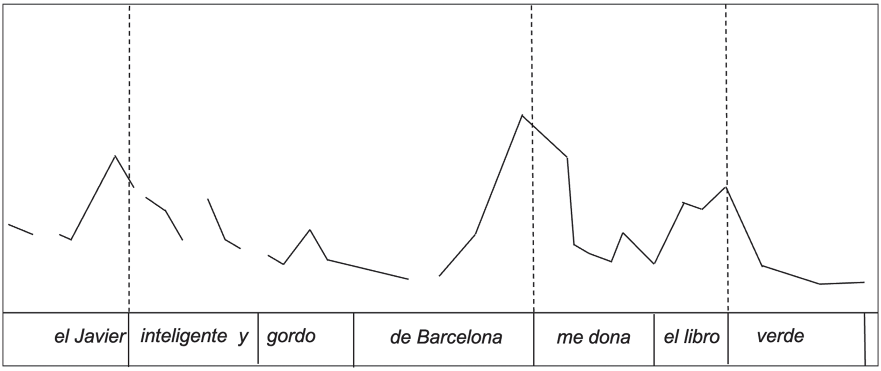


**Figure 15:** Pitch track of a read Barcelona Spanish utterance from Rao (2008: 97), *el Javier inteligente y gordo de Barcelóna me dona el libro verde* “The intelligent and fat Javier from Barcelona donates the green book to me”. The phrasing according to Rao 2008 is (El Javier)ϕ(inteligente y gordo de Barcelona)ϕ(me dona el libro)ϕ(verde)ϕ.

sive prosodic PhP phrasing. That cannot be counted as empirical evidence for it. His criteria for identifying a PhP boundary (equivalent to an iP boundary) include the presence of continuation rises, decreases in pitch range, phrase-final lengthening and pauses (Rao 2008: 93), but when presenting the results, only the presence or absence of a boundary is marked, and no information on the cues given. Interestingly, his results show that even in the case of appositives, the “recursive” phrasing is never the only, and sometimes not even the most frequent, pattern (cf. e.g. Rao 2008: 107). At a broader level, they show that syntactically motivated phrasing constraints aiming to map relevant syntactic boundaries to PhP boundaries can frequently be overridden by prosodic constraints aiming to create balanced structures or those conforming to minimality or maximality constraints on the PhP, and that “as syntactic branching of utterances increases, prosodic concepts seem to play a more crucial role than syntactic conditions in determining the parsing of phrases” (Rao 2008: 120). In this regard, these results reported for Barcelona Spanish are similar to those made for Castilian Spanish in Prieto (2006) and for Lima Spanish in Rao (2007) based on similar methodology[62] (read speech, 3 speakers from Madrid and one from Burgos in Prieto 2006, 3 Lima Spanish speakers in Rao 2007). In both of those, phrasing patterns are explored via OT constraints, and

[62] Prieto (2006: 42) uses slightly different cues for the boundaries of phonological phrases than Rao (2007, 2008): her criteria are “the perception of a prominent stress (together with a level 2 phrase break in the ToBI framework)” as well as optionally a boundary rise. In general, the status of the phonological phrase / intermediate phrase in Spanish is not settled, and no phrasing studies based on spontaneous speech have been undertaken (Hualde & Prieto 2015: 359–360).

their results also suggest that the purely prosodic ones often outrank those aiming for a close mapping between prosody and syntax. It has also to be noted that only in the study on Lima Spanish do any of the experimental conditions reach something approaching categorical preference for a given phrasing pattern, and then always the one most clearly following prosodic constraints instead of ones promoting a close mapping (Rao 2007: 96–97, 100–102). Overall, this is less strongly the case for the Castilian Spanish study, where considerable variation in pattern preference between the speakers is reported (Prieto 2006: 46).

In sum, the question of recursive prosody in Spanish (and Quechua) is still open to be answered via further research. These studies however suggest that the assumption of a tight mapping between syntax and prosody might not be the best point of departure to go looking for prosodic recursion, since they show prosodic phrasing in Spanish quite clearly to be subject to a number of constraints caring little about such a mapping.[63] It must also be noted that all the studies discussed in this section (not only on Spanish) are based on read speech by less than ten speakers each (except Rao 2008), and yet they also report on some considerable variability in their results. No comparable studies addressing prosodic recursion have been done with more naturalistic data to my knowledge. In section 5.2, I will argue for prosodic recursion at the level of the IP based on data from Huari Spanish uttterances that are both semi-spontaneous and quite long and complex, also at the level of information structure.

## 3.7 Information structure and prosody

In the preceding sections, we discussed relevant aspects of prosodic typology mostly without referring to the dimension of meaning. In this section, I will first give a brief overview over the types of meaning intonation can convey and then concentrate on information structure. I will give an introduction to the model of context and information structure that I follow in this work (section 3.7.1), lay out how it can be applied to the study of understudied languages without having to make prior assumptions about a relation between information structural categories and formal means of expression (3.7.2), and then conclude with (what is known about) the relation between information structure and prosodic cues in Spanish and Quechua (section 3.7.3).

[63] The equal status of prosodic and syntactic constraints for prosodic structure is also emphasized in Féry (2015).

Prosody and intonation aid in the signaling of a variety of propositional and non-propositional aspects of meaning. Prosodic phrasing helps to disambiguate semantic and syntactic structures as well as word recognition both in spoken language (cf. e.g. Nibert 1999, 2000; Serrano 2010 on Spanish, cf. also Brown et al. 2015) and probably even in silent reading (the “implicit prosody hypothesis”, cf. Fodor 2002, the works contained in Frazier & Gibson 2015). Intonation, especially contour choice, has been shown not only to play a decisive role in signaling utterance type in many languages, but also to convey a number of propositional attitudes and other non-at-issue content in several languages (among others, cf. Ward & Hirschberg 1985; Hirschberg & Ward 1992 on pragmatic meanings classified as epistemic uncertainty and incredulity in English; Vanrell et al. 2013; Vanrell et al. 2017 on uncertainty and evidential meanings in Catalan polar questions; Grice & Savino 1997; Grice & Savino 2004; Savino 2012 on uncertainty and the difference between information-seeking and confirmation-seeking polar questions in Bari Italian; Venditti et al. 1998 on meanings including incredulity and insistence in Japanese; Wollermann et al. 2010; Wollermann et al. 2014 on uncertainty in German declaratives; Escandell-Vidal 2017 on evidential meanings in polar questions and Fliessbach 2023 on mirativity and obviousness in Spanish). Contour choice can also have an effect on the salience of conversational implicatures (Kurumada et al. 2014, Prieto 2015, Marneffe & Tonhauser 2019). Not least, prosodic cues are essential for the smooth working of the turn-taking and repair systems, two strong candidates for perhaps truly universal systematic features of human language (Stivers et al. 2009; Bögels & Torreira 2015; Dingemanse et al. 2015; Levinson 2016).

### 3.7.1 Context and the question under discussion

Here we will only treat one particular aspect of how prosody and intonation help to relate individual utterances to their linguistic and non-linguistic context, the signaling of information structure. This is because it will be the most relevant aspect in the analysis of the Huari Spanish and Quechua data. Other aspects of intonational meaning will also feature occasionally, but not be treated systematically. The present work is mainly on prosody and not on formal pragmatics, but because these aspects of linguistic context are crucial factors for understanding prosody in spontaneous, ongoing discourses, some background to them will be provided here. Information structure relates to how the information contained in utterances is packaged with regards to the momentary knowledge states of participants, the discourse progression up to the point of utterance, and the intentions of participants for the discourse progression in the future (Chafe 1976; Krifka 2007). I assume a context model similar to the ones of Farkas & Bruce

(2010) and Roberts (2012b). One of the main components of these models is the *common ground* (CG, Stalnaker 1974, 2002). The common ground can be defined as the proposition specifying the set of propositions on which the participants have reached an agreement with regards to its truth value. More accurately, it is the proposition which speaker A believes that speaker B believes that speaker A believes that speaker B believes . . . specifies the set of propositions which the speakers have agreed upon. This formulation preserves the fact that the common ground is always a doxastic projection by each individual participant about what they believe the common ground to be. Crucially, it is also necessary to include a performative aspect here in that the common ground is what speakers act like it were the common ground: this can be out of pretense or forgetfulness or for another reason (Stalnaker 2002: 704–705). However, it is also assumed that speakers are basically rational in the pursuit of their intentions and communicating cooperatively in the sense of Grice (1989); this assumption effectively allows speakers to ascribe reasons (intentions) to their interlocutor's behaviour and to keep calculating it rationally even when it deviates from what they themselves would expect their interlocutor to behave like given the state of CG, instead of having to assume that they behave randomly.[64]

What is not in CG can be seen as the set of propositions for which no truth value has yet been agreed upon. That is to say, these propositions define the set of all possible worlds that are compatible with CG (the context set in Roberts 2012b: 4). The assertion by one speaker of a proposition, publicly committing to a truth value for it, and the acceptance of this assertion by the interlocutor(s), moves the proposition to CG, reducing the context set. This movement of propositions from context set to CG is the basic drive of the discourse progression.[65] In order to model its directedness and coherence in a hierarchical structure, Roberts (2012b) employs

[64] Recasting this point somewhat, one could say that the social act of committing to an action is more relevant here than whatever mental states are responsible for it: acting like a certain proposition is part or not part of CG is a commitment by a speaker to another to be consistent with respect to this position, and breaking such commitments has the social consequence of not being treated as a reliable communication partner anymore (cf. Geurts 2019).

[65] This is necessarily an extremely simplified picture. Not only are there speech acts other than assertions and questions, such as directives, that have the aim of getting another speaker to change the way things in the world are instead of increasing shared knowledge about how they are, but there are also assertions with additional presupposed or non-at-issue modal meanings such as surprise and obviousness that force interlocutors to revise some of the assumptions already shared by them and contained in CG (surprise / mirativity) or that make statements about the fact that the content of an assertion is already contained in or entailed by CG and therefore not newsworthy (obviousness). See Reich (2018); Fliessbach (2023) for a discussion of these meanings.

the device of the *question under discussion* (QUD, also used in Farkas & Bruce 2010; Ginzburg 2012, and several others). In her QUD model, discourse progresses via the posing and answering of (mostly implicit) questions. By accepting a QUD as the current one, discourse participants agree to keep making attempts at answering it until they either agree that it has been fully answered or that they must leave it unanswerable for the moment (Roberts 2012b: 6–7). Questions are only felicitous when their answers are not entailed by CG[66] and when they can be construed to be *relevant*, that is when answers to them contribute to providing (partial) answers to the current QUD in the discourse[67] (Simons et al. 2010: 316–317; Roberts 2012b: 21), with a partial answer defined as providing an evaluation of at least one of the propositions comprising the *alternative set of the question*, i.e. all possible answers to it (Roberts 2012b: 11–12, based on the semantics of questions first proposed in Hamblin 1958, 1973). Assertions are also only felicitous when they are not entailed by CG and when they are relevant, providing a (partial) answer to the current QUD. The entailment and relevance conditions define permissible *question strategies*, i.e. those that aim to completely answer sub(-ordinate) questions in order to partially answer super(-ordinate) ones; once a sub-QUD has been answered, the super-QUD it is entailed by becomes the current QUD again, giving discourse overall a hierarchical QUD structure. The question alternative set is pragmatically restricted by the discourse context, as in (25), where the context reduces the question alternative set to three members from the potentially infinite number of propositions about Alice winning some prize.

(25) *Context*: Alice, Claire, and David each won a prize at a sheep shearing competition. They bring home a pair of golden scissors, a silver tuft of sheep's wool, and a bar of lanolin soap.
*Question*: What did Alice win?
*Pragmatically restricted question alternative set*: {Alice won the golden scissors, Alice won the silver tuft of wool, Alice won the lanolin soap}

[66] This is again a simplification. In practice, questions whose answers are entailed by CG do occur and are likely to provoke an assertion that is marked for obviousness (cf. Fliessbach 2023). Effectively they and their responses represent highly marked discourse moves.

[67] Based on a discussion of utterances containing epistemic modals and evidentials in discourse, Roberts (2015: 52–53, 2017: 30–31) introduces a revised notion of relevance which allows discourse moves (questions and assertions) to be indirectly relevant, i.e. felicitous, when they provide an assessment in terms of the *likelihood* that a relevant proposition is true or false (see also Simons et al. 2010: 316–317, note 3).

Propositions forming the question alternative set of a wh-question all share a similar structure in that they only differ with regards to which element replaces the wh-element in the question, with the appropriate type of entities specified by the choice of the wh-element (cf. Roberts 2012b: 10). Polar questions have a question alternative set comprising only two members that only differ with respect to their absolute polarity (cf. especially Farkas & Bruce 2010 for a treatment of polar questions in discourse). Alternative questions have effectively the same kind of question alternative set as wh-questions, with a specified constituent in the question to be replaced by one in the answer, but the answers treated as likely by the speaker are usually fully listed in the question, and the restriction on appropriate types of entities is provided only by context, not the choice of wh-element (e.g. *Do you prefer Canada, going home, or Chadwick Boseman?*; cf. on alternative questions Sadock & Zwicky 1985: 178–179; Riester & Shiohara 2018: 292). A felicitous answer to a question must be the assertion of a proposition that is a member of the question alternative set; this *question-answer coherence* is pragmatically ensured via the Gricean maxim of relevance/relation, according to Roberts (2012b: 21) (cf. also Grice 1989; Krifka 2007: 22–23).

While this conception is probably not fully capable of providing a realistic model of discourse (cf. Riester et al. 2018: 405 for this view as well as e.g. Roberts 2012a: 8–14 for a number of outstanding issues), it couples two major mechanisms: discourse coherence (via question-answer coherence and the hierarchical QUD structure) and the internal division of utterances via their information structural status. The latter has two aspects: the division between what is and what isn't *at-issue* and the one between *focus* and *background*. The former of these describes a meaning distinction at the level of propositions: in an utterance, only those propositions are *at-issue* (AI) that are relevant with regards to the current QUD (Simons et al. 2010: 317; Beaver et al. 2017: 280; Roberts 2017: 9), those that aren't are *non at-issue* (NAI). The AI/NAI-division is a broad label encompassing aspects of a range of phenomena such as pragmatic presuppositions, appositives, non-restrictive relative clauses, evidential meanings, conventional implicatures, and the projection behaviour of factives (cf. e.g. Tonhauser et al. 2013; Faller 2014; Bianchi et al. 2016; Tonhauser et al. 2018; cf. Potts 2015 for an overview). One diagnostic for at-issueness is whether content can be directly targeted by polarity particles in responses, according to Roberts (2015: 44–46, 2017: 9). According to this diagnostic, some of the meanings conveyed by intonation are also non-at-issue (cf. for this view also Potts 2005: 26, 36–37; Prieto 2015):

(26) *Context*: Alice, Bob, and Doreen are living together. They are going to have friends over for dinner.
*Alice*: Who's coming for dinner?
a) *Bob*: Jonathan (L*H L-H%)
*Doreen*: a. No [he's not coming]
b. # No [you're not uncertain]
c. Why so unsure about it? You told me he was coming five minutes ago!
b) *Bob*: I'm not sure about Jonathan
*Doreen*: Yes [you are – you told me he was coming five minutes ago]

In (26)a), Bob answers Alice's question with the uncertainty contour on *Jonathan*. Doreen can deny the truth of the at-issue proposition that John is coming (answering the current explicit QUD, (26)a)a), but she cannot deny the non-at-issue content that Bob is uncertain conveyed by the intonational contour directly by using only the response particle *no* ((26)a)b). In order to do that, she has to perform a marked[68] move requiring much more explicit elaboration, like (26)a)c. That it is not the fact that she cannot target Bob's uncertainty for reason of it constituting a subjective mental state accessible only to him is shown in (26)b), where the uncertainty meaning is lexically conveyed in the main clause and at-issue (in the sense that Bob's public commitment about being unsure has a bearing on the probability that the answer to the relevant question *Is Jonathan coming?* can be taken to be true, cf. Roberts 2015, 2017). For Roberts (2017: 7), the focus-background division is part of the mechanism that can separate AI from NAI content, as in (27):

(27) (from Roberts 2017: 7)
Context: A, B, and C are discussing the race that B and C just watched, in which their mutual friend Mary was supposed to be a participant.
A: How did Mary run?
B: She ran [QUICKLY]$_F$—one of her best times ever!

[68] Farkas & Bruce 2010 provide reasoning about what kind of conversational moves are more or less marked depending on context. They model the relatively privileged status of some *responses* to certain *provocations* using the device of the *projected set*, which contains the immediate future state(s) a context will assume upon an unmarked response to a given provocation. For example, they argue that the unmarked response to an assertion is to accept it (and thus for the asserted proposition to enter CG), and that to a neutral polar question is to commit to either of the propositional alternatives provided by the polarity, but that biased polar question types differ in precisely this property, i.e. that they only contain one of the polar alternative propositions in the projected set (Farkas & Bruce 2010: 93, 96–97).

C: a. No she didn’t!
b. Hey, wait a minute! Mary didn’t run at all! / But Mary didn’t run at all!

In (27), the explicit QUD posed by A already presupposes that Mary ran, it asks only for how she ran, which is answered in B’s response by the focused adverb *quickly*. It is only this assertion about Mary’s quickness which C can target in their contradiction by saying *no she didn’t*; if they wanted to challenge the backgrounded content that she ran, they would again have to resort to a more marked response like the ones in (27)b. In this conception, the focus-background division is a special case of the AI-NAI division as applied to a particular type of content, the *proffered* content: “the compositionally calculated truth conditional content of the expression; what it contributes to what is asserted, asked or directed by an utterance in which it occurs” (Roberts 2017: 6). Only (parts of) the proffered content can then be at-issue at all. Other types of content, the presupposed and the auxiliary content, which include conventionalised implicatures and the type of meaning conveyed by intonation in (26), for example, are always non-at-issue. Under this view, focus and background can essentially be subsumed via cross-classifying types of contents and at-issueness: focus would be [+at-issue, +proffered] and background [-at-issue, +proffered]. Others, although using the same terminology, draw the distinction elsewhere: Riester et al. (2018); Riester & Shiohara (2018); Riester (2019) develop a model for information structural annotation of actual conversations based on QUD structure, but they reserve the label of “non-at-issue” for those types of content Roberts would class as presupposed and auxiliary, and explicitly exclude backgrounded but proffered material from it (cf. Riester et al. 2018: 428). Although Roberts’ proposal is theoretically neat because it achieves a comprehensive unification, I will here adopt the convention followed by Riester and colleagues, simply because their goals align more immediately with my own in this work: to operationalize the QUD model for the categorization of segments of actual spoken conversation according to information structure, independently of linguistic form, so that prosodic exponents can then be related to them without entering a circular argumentation (cf. Riester et al. 2018: 405–406). In the following, I will use their definitions to discuss the information structural notions of *focus*, *background,* and *topic*.

### 3.7.2 QUD-based annotation of information structure independent of form

Focus is very heterogeneously defined in the linguistic literature (cf. Krifka & Musan 2012: 6–7, 17–18; Riester & Shiohara 2018: 290–291 for overviews). The definition adopted by Riester et al. (2018: 417); Riester & Shiohara (2018: 291–292); Riester (2019: 166) is to say that the *focus* in an utterance is exactly that part which answers the current QUD, while its counterpart, the *background*, is material that is proffered but already given in the question itself. I will also adopt this definition of focus here. It has the crucial advantage of being entirely based on context (as described in QUD terms).

(28) QUD-focus relation in wh-questions
  a. Q: What happened?
     A: [Sharon sheared her sheep on the first Saturday in September]$_F$
  b. Q: What did Sharon do?
     A: Sharon [sheared her sheep on the first Saturday in September]$_F$
  c. Q: When did Sharon shear her sheep?
     A: Sharon sheared her sheep [on the first Saturday in September]$_F$
  d. Q: Who sheared their sheep on the first Saturday in September?
     A: [Sharon]$_F$ sheared her sheep on the first Saturday in September
  e. Q: When in September did Sharon shear her sheep?
     A: Sharon sheared her sheep [on the first Saturday]$_F$ in September

(29) QUD-focus relation in polar and alternative questions
  a. Q: Did Sharon shear her sheep on the first Saturday in September?
     A: [Yes]$_F$ / [Yes, she did]$_F$ / [Yes, Sharon sheared her sheep on the first Saturday in September]$_F$
  b. Q: Did Sharon shear her sheep or her goats on the first Saturday in September?
  c. A: Sharon sheared [her sheep]$_F$ on the first Saturday in September

In (28), the same utterance, *Sharon sheared her sheep on the first Saturday in September* has different focal domains depending on the preceding question context. The focused element(s) correspond exactly to the open variable in the wh-question and are marked by square brackets with a subscripted F. All material that is already given in the question is not part of the focus in the answer and thus backgrounded. As mentioned above, in the conception by Roberts (2017), it would also be non-at-issue. In actual conversation, the full answer utterances would most likely not be produced. Instead, most or all of the material in the background would probably be left out, but note that the focal material could not be further reduced given the QUD. This is true also in (29)a, where if the two first variants aren't chosen, then the full utterance has to be produced to be felicitous. The bracketed notation for the focal material is intended merely as a notational aid: the property of being in focus is bestowed entirely by the context in this conception, and no specific focus marking is necessary, although various expressive means, a number of them prosodic, are employed to various degrees in different languages to aid in the interpretation of which material is focal and which is backgrounded (see below).

In that, this focus definition differs from another very influential one, originally proposed in Rooth (1985, 1992) and defended in Krifka (2007); Krifka & Musan (2012), namely that focus indicates the presence of alternatives relevant for the interpretation. Because of the way the context is modeled, with felicitous answers effectively coming from the question alternative set of the QUD (Roberts 2012b), the two definitions cover a lot of common ground. However, while in the definition adopted here, focus is essentially just a label for that part of an utterance which directly corresponds to the QUD – and this means that alternatives will be relevant for its interpretation by definition –, in the other it is something (a feature) which actively acts to indicate that alternatives should be taken into consideration for interpretation. This is made explicit in the revised definition given in Krifka (2007: 19): "A property F of an expression α is a Focus property iff F signals (a) that alternatives of (parts of) the expression α or (b) alternatives of the denotation of (parts of) α are relevant for the interpretation of α." Here, focus is a *property of an expression* which *signals* a meaning, just like e.g. a morphological tempus marker. After some discussion that alternatives are also relevant in the interpretation of other expressions lacking focus, Krifka & Musan (2012: 7) refine the definition given there by saying that it "means to say that focus especially stresses and points out the existence of particular alternatives", which again implies the presence of some kind of focus signifier. From this perspective, it also makes sense to state, as they do, that it seems that some African languages do not make "use of focus to mark question-answer coherence" (Krifka & Musan 2012: 10, note 4) given findings that these languages do not employ any expressional means to mark the difference between focus and background in question-answer pairs. From the perspective of the definition adopted here, the argument is subtly different: the answer to the QUD, as long as it can be clearly established, is the focus by definition – but it is clearly an open question typologically whether and how individual languages will employ any means that aid to reconstruct such a context in interpretation.

It has repeatedly been observed that languages differ with respect to the expressional dimension of focus, and plausible hypotheses have been made about what makes focus marking more likely, e.g. based on an asserted proposition's status with respect to expectations built up in the discourse (Zimmermann 2008), or on a notion of scale of strength applied to different proposed types of foci (Féry 2013: 689–690). However, if focus is fundamentally conceived of as some kind of property of expressions, then making any kind of comparison between what can count as a focus marker becomes a circular enterprise because context cannot serve as a true third of comparison, but only as an insufficient indicator. This circularity and its gloomy consequences for the investigation of information structure across languages is the explicit reason why Riester et al. (2018: 406); Riester & Shiohara (2018: 290) define focus and all other information structural notions strictly based on context instead of on any linguistic form. Doing so means to assume that the underlying mechanism of information transmission and the essential pragmatics concerned with its packaging are effectively universal or at least sufficiently similar across languages, to the extent that a model can accurately describe them (Riester et al. 2018: 405–406). Effectively, the assumption is that this mechanism is shaped directly by universal aspects of human communication, such as the Gricean cooperative principle and the maxims resulting from it.[69] In this regard, the position adopted here from Riester and colleagues, locating the universality in the pragmatics, is different from both the Rooth/Krifka position, which locates it in some part of the expressional machinery, presumably in the syntax, and that taken by Matić & Wedgwood (2013), who deny any universality to focus as an information structural category. Whichever of these positions might eventually turn out to be true, the approach by Riester and colleagues has the clear methodological advantage that focus and other IS-notions can be identified across languages and from context alone, and thus allows to correlate definable context conditions with whatever means of expression a language might or might not employ. In order to do that in the analysis of naturally occurring conversation, criteria for the identification of implicit QUDs must be set down. This is highly important because in many types of conversations, explicit questions are quite rare, and instead, sequences of assertions occur, that nonetheless are taken to follow a hierarchical QUD structure (Riester & Shiohara 2018: 292). For them, the implicit QUDs have to be reconstructed, as it were, in order to be able to make statements about the information structural division of the assertion at hand.

The first and broadest of these criteria is question-answer congruence, i.e. that the observed assertion must be a felicitous answer to the implicit QUD (Riester et al. 2018: 411–412). Question-answer congruence allows any of (30)a-c (and some others) to be identified as implicit QUDs for the assertion in (30), but prohibits questions such as (30)d-e because they are not answered by the assertion.

[69] Presumably, this can also be brought into relation with Chafe (1994: 109–110)'s „one new idea", and also with how information transmission is described by Shannon (1948), namely that it eliminates possible states a system could assume. If no indication in an utterance were available from context-based expectation which the relevant location for the elimination of such possibilites is, all transmitted material would be equally informative and no inference about the intended direction of the discourse were possible, which would arguably divest humans of a large part of their intention-reading skills, effectively grounding communication to a halt.

(30) Question-answer congruence

*Assertion:*

Alice went down the rabbithole

*Question:*

a. What happened?
b. Where did Alice go?
c. Who went down the rabbithole?
d. # What is a rabbithole?
e. # When did Alice go down the rabbithole?

Possible implicit QUDs are further constrained by two principles that capture a crucial difference between implicit QUDs and explicit ones: implicit QUDs, not being actually performed discourse moves, cannot possibly introduce any new material, they can consist only of material that is already given[70] in the context (and combine it with operators, such as a wh-element). This principle is called *Q-Givenness* in Riester (2019: 174). The second principle is called *Maximize-Q-Anaphoricity* (Riester et al. 2018: 412; Riester 2019: 175) and it states that implicit QUDs should not only consist only of given material, but that they "should contain as much given (or salient) material as possible". "As possible" is intended to be constrained by the preceding context and the observed assertion: this means that all material that can be determined as given in the answer (from the preceding discourse context) will be contained in the reconstructed implicit QUD. Effectively, *Maximize-Q-Anaphoricity* thus works to narrow the focus as much as possible both in the implicit QUD and in turn, in the assertion (cf. Riester 2019: 175).[71] We can see how the three principles act together in identifying implicit questions in the analysis of a short example excerpt from the Quechua corpus:

[70] With a definition of givenness adopted from Baumann & Riester (2012), where it is broadly defined as an expression or a referent having been made available in the preceding discourse context, with a separation between referential and lexical givenness and a temporal decay of cognitive activation included in the model. See also Riester (2019: 174, note 8); cf. section 6.2.3.3 for a more detailed discussion.

[71] Thus, while implicit QUDs are considerably constrained by the assertions they are reconstructed from, explicit questions, with only general constraints of coherence placed on them, can truly change the direction of discourse (cf. Riester 2019: 174–175).

(31) TP03_KP04_MT_Q_2640–2704[72]

| *time* | *KP04 (with the path on the map)* | *TP03 (without the path on the map)* |
|---|---|---|
| *264.0* | alli-m<br>good-ASS<br>*alright* | |
| *265.0* | naa tsawra tillaku-pita-qa<br>PSSP then lightning-ABL-TOP<br>*well then from the lightning* | |
| *266.8* | | ya<br>yes<br>*yes* |
| *267.5* | pasa-rku-y manka-yaq<br>pass-DIR-INF pot-TERM<br>*go up to the pot* | |
| *268.7* | manka-pa hana-n-pa-m pasa-n<br>pot-GEN above-3-GEN-ASS pass-3<br>*it goes above the pot* | |

(31) is a short excerpt from a *maptask* experiment forming part of the corpora analyzed in this work. As Riester et al. (2018: 408) emphasize, it is very important that a corpus analyzed in this way should be well understood by the analyst. This also includes knowledge about the type of conversation (in this case, the kind of experimental communication game), its overall context, and the speakers involved. General information on these matters is provided in sections 2.3 and 2.4. Here, it is relevant to know that the speakers are in the process of tracing the path for a second time already, after the first attempt led to confusion (because two landmarks are intentionally different on the two speakers' maps, cf. Figure 188 and Figure 189 in Appendix B). Their discourse proceeds from landmark to landmark, having begun with one where they were sure to be still in agreement. At 264.0, KP04 signals that the discussion about the preceding landmark is complete and that they can move on to the next issue; in terms of Farkas & Bruce (2010), the Table of current issues is empty (but in terms of QUD structure following Roberts 2012b, very superordinate QUDs such as WHAT IS THE OVERALL TRAJECTORY OF THE PATH? are still active). We will proceed backwards through the example, for reasons that become clear soon. At 268.7, the only new element in the assertion is *hana-n-pa-m* "above"; all the other elements are given already in the preceding turn. Thus by the principle of *Q-Given-*

[72] https://osf.io/d7tzc

*ness*, the meanings encoded by *manka-pa* "of the pot" and *pasa-n* "s/he/it passes" are included in the implicit QUD, a good candidate for which would be something like WHERE IN RELATION TO THE POT DOES THE PATH GO?, so that only *hana-n-pa-m* is in focus in the assertion. Note that also *Maximize-Q-Anaphoricity* prevents WHERE DOES THE PATH GO NOW?, with a correspondingly broader focus [manka-pa hana-n-pa-m]$_F$. Moving one turn backwards to 267.5, although this is a directive, we can apply basically the same kind of reasoning: *manka-yaq* "to the pot" is not given in the context, so the implicit QUD must at least ask for it and it is in focus. Regarding *pasa-rku-y* "move in a direction", strictly speaking it is not given in the preceding context and thus the most apt implicit QUD could be something like WHAT NEXT?, effecting the broadest possible focus. On the other hand, it could be argued that the action of passing from one place to the other in this conversational game is so commonplace that it is always salient and thus given. Then, by *Q-Givenness* and *Maximize-Q-Anaphoricity*, the implicit QUD would be something like WHERE TO GO?, with a narrower focus in the assertion just on *manka-yaq*. Deciding this issue can be facilitated by considering the entire corpus, but in the end it is probably more important that the analytical decision of how to treat such verbs and the actions denoted by them should be consistent. Moving backwards by yet another step to the peninitial turn (in the excerpt) by the same speaker, a less simple decision awaits us. At a first glance, we could reconstruct an implicit QUD like FROM WHERE THEN? for *naa tsawra tillakupitaqa* "well then from the lightning" at 265.0, with *Maximize-Q-Anaphoricity* working to include *tsawra* "then", and perhaps even the meaning encoded by the ablative suffix *–pita*, into it. However, this turn is not an assertion by itself. The only argument for that could be that it is directly followed by backchanneling by the other speaker, but if the turns by KP04 at 265.0 and 267.5 were produced in a single utterance, it would be clear that *tillakupitaqa* does not constitute an independent speech act. Instead, it seems to fulfill a topical function here:[73] the directive at 267.5 is referentially disambiguated only by its presence. I want to argue further that it is a contrastive topic, albeit a type called *implicit contrastive topics* by Riester & Shiohara (2018: 300–301).

[73] The suffix *–qa* or its variants is usually identified as a „topic marker" in Quechua grammars (cf. amongst others Parker 1976; Weber 1989; Cusihuamán 2001; Adelaar & Muysken 2004), and the gloss used here also indicates this. However, topic definitions are almost as numerous as focus ones (see Roberts 2011), and Weber (1989) describes uses of it for Huallaga Quechua that do not square well with all of them. In addition, in keeping with the approach laid out by Riester and colleagues, I try to refrain as much as possible from building arguments for information structural analysis from evidence based on formal linguistic features, instead of context-content relations. Section 6.4 will be particularly concerned with disentangling the cues for IS from different domains from context-based interpretation in Huari Quechua.

Contrastive topics, according to Büring (2003); Roberts (2012b), are the result of a complex strategy of inquiry that involves parallel subquestions. That is to say, when a complex QUD containing two wh-variables, such as WHO SHEARED WHICH SHEEP? is to be answered, then this is often done by asking the subquestions that result from filling one of the variables with the individual members of its answer set, such as SEAN SHEARED WHICH SHEEP?, SIAN SHEARED WHICH SHEEP?, SHANE SHEARED WHICH SHEEP?. This implies providing (partial) answers to the first variable. Thus the superquestion WHO SHEARED WHICH SHEEP? is answered by first answering WHO SHEARED SHEEP?, and then substituting the answers into the superquestion, yielding subquestions that are parallel. This is a viable strategy because each subquestion is entailed by the superquestion, each answer to a subquestion is a partial answer to the superquestion and thus relevant. In the answers to the subquestions, yielding parallel structures such as *Sean sheared Bessy, Berta, and Balu; Sian sheared Billy, Bartholomew, and Boris*; *Shane sheared Bob, Bonita, and Brenda*, the sets of sheep {*Bessy, Berta, Balu*}, {*Barbara, Bartholomew, Boris*}, and {*Bob, Bonita, Brenda*} are at-issue with regards to the current QUD, and the shearers *Sean*, *Sian*, and *Shane*, as contrastive topics, are at-issue with regards to the preceding QUD. According to Riester & Shiohara (2018: 295); Riester et al. (2018: 422–423), the parallel structure of such assertions indicates the complex relationship they are in, namely that they aren't just relevant for the current QUD, but answer such a superquestion in concert.[74] However, it is precisely this overt parallelism which at first sight seems to be missing in our example (31). Riester & Shiohara (2018: 300) remark that in their Sumbawa corpus, such overtly parallel structures do not occur, and it seems likely that they are in general quite rare in naturally occurring discourse (they do come up on occasion in our own corpora). They nevertheless argue that turns similar to our example should be analyzed as containing *implicit* contrastive topics because, while only realizing one of the parallel answers overtly, their interpretation in context implies the existence of such relevant (with regards to a super-QUD in the discourse) parallel predications made about other entities that would also be contrastive topics. They cite considerations that continuing topics, which are not contrasted, usually do not need to be overtly expressed at all (of course depending on the possibility for the language in question to not realize given arguments), and that when such elements are then overtly realized, they often indicate a topical change, i.e. a contrast, and such an implicit parallel structure if this is corroborated by the broader context (Riester & Shiohara 2018: 301). In our example, the

[74] Cf. also the definition for so-called *delimitators*, comprising contrastive topics and frame-setters together, in Krifka & Musan (2012: 32–34), which also emphasizes that such structures cannot be interpreted purely locally.

case can be argued quite well that *tillakupitaqa* is an implicit contrastive topic. In the preceding discourse, speaker KP04 (the instructor in the *maptask*) had already tried to move the conversation to where the path leads from the lightning (as part of the repeated instruction going stepwise from landmark to landmark), this was then followed by the other speaker (TP03) backtracking because he could not follow the instructions KP04 was giving subsequent to moving from the lightning on his map. This is the issue that KP04 signals as having been settled by uttering *allim* at 264.0, and with the turn at 265.0, he thus indicates that the following path instructions are to be understood with reference to the lightning (again), as opposed to the subsequent landmark that left TP03 stranded. Because of this, and the stepwise progression from landmark to landmark, it seems reasonable to assume that a superquestion something like FROM WHICH LANDMARK, WHERE DOES THE PATH GO? plays an important role in the structure of their discourse, and that the assertion at 265.0 and 267.5 is to be understood as a partial answer to that, with *tillakupitaqa* an implicit contrastive topic in the sense laid out above. In summary, we can see that the approach pioneered by Riester and colleagues can be used to make quite a bit of headway[75] into an information-structural analysis of spontaneous speech corpora without making reference to formal means of encoding them, and it will be used in this way in parts of this work to separate these meaning-based categories from potential prosodic cues.

### 3.7.3 Prosodic cues and information structure

A strictly context-based definition of information structural categories still leaves the question open whether formal means of expression signal information structure directly. However, a view often adopted in the literature on prosody and intonation is that this signaling is only indirect (Ladd 2008; Calhoun 2010b; Kügler & Calhoun 2020). According to this view, a constituent e.g. being in focus is not automatically linked to some kind of prosodic exponent; instead, information structural division of utterances is signaled indirectly via metrical and prosodic structure, which is what is cued by phonetic exponents in turn. In the following two sections, I will review evidence for this hypothesis and refine the relation between information structure and prosodic cues first for Spanish, and then for what is known in this regard about (Cuzco) Quechua.

[75] Although the analysis especially of implicit contrastive topics is not yet fully satisfying, as Riester & Shiohara (2018: 301) admit. My own analysis will often be based on reasoning drawing on a consideration of larger parts of the nonlocal discourse context, as exemplified above.

#### 3.7.3.1 Spanish

In fact, much of the literature on prosodic focus marking in Spanish can be read as evidence for precisely this indirect relation. Cases in point are Gabriel (2007) and Vanrell & Fernández Soriano (2018). Both investigate prosodic and syntactic strategies of signaling information structure in assertions via a question-based elicitation paradigm involving speakers from several varieties of Spanish.[76] They find that the prosodic cues involved in signaling the information structural division between background and focus in declaratives include a high iP-boundary tone (H-) occurring at the right edge of the prefocal background material,[77] an LH* pitch accent on the final prosodic word of the focal material, a low iP-boundary tone (L-) at its right edge, and pitch accent compression or deaccentuation on the postfocal background material. While all of these cues were attested in productions from speakers of all varieties in both studies, they were also all optional to some extent. H- at the right edge of the prefocal background occurred in 81.3% of cases (113 of 139 possible utterances), with individual occurrence rates ranging from about 55% (3 speakers) to 100% (4 speakers) and the other speakers distributed evenly among intermediate rates in Gabriel (2007: 276–277). Across speakers, occurrence was more frequent when the focus domain was narrower (i.e. the focal material consisting of fewer elements in correspondence to the QUD) than when it was broader. Postfocal deaccentuation/pitch accent compression (cf. section 3.4.6) has slightly lower occurrence rates, ranging from 40% to 83% across the relevant elicitation contexts and also showing strong individual differences between speakers. Again, the width of the focus domain seemed to play a role, with a group of utterances where the focal material consisted of more than a single prosodic word never occurring with deaccentuation on the available postfocal backgrounded material (Gabriel 2007: 282). Vanrell & Fernández Soriano (2018) report on similar tendencies with regards to these two cues, but do not offer precise numbers. With regards to the realization of

[76] In the case of Gabriel (2007): 18 speakers, of which 14 from various Spanish regions, all judged to be speakers of “standard Castilian Spanish”, and one each from El Salvador, Colombia, Mexico, Argentina, of which only the last two are judged to not be speaking “standard” Spanish by Gabriel (2007: 269). In contrast, Vanrell & Fernández Soriano (2018: 38) present data from 9 speakers, 2 from the Canarian Islands, 2 from the Basque country with Basque as L1, another two from the Basque country with Spanish as L1, three from Madrid; they take them to speak at least four different varieties of European Spanish.

[77] Gabriel (2007: 280–281) also finds evidence that the segmental prosodic process of *sinalepha*, the assimilation of adjacent vowels across word boundaries, is blocked between the right edge of the final phrase containing backgrounded material and the left edge of the phrase containing focal material. For a study on variant phonetic realizations of the H- boundary tone, see Gabriel et al. (2011).

the focal material itself, both studies find that one alternative to the LH* L- contour is L* L%, which occurs only when the focal material is final in the utterance (but then not always, cf. also Hualde & Prieto 2015: 364). In the data reported on by Vanrell & Fernández Soriano (2018: 54), it is also not restricted to cases of broad focus. Vanrell & Fernández Soriano (2018: 43) attest to a third option, L<H* H-, which they describe as a "rise throughout the accented syllable which continues to the end of the intermediate phrase". This contour they only describe in cases where the focal constituent is a subject and occupies an initial position due to it-clefting (e.g. *Fue [Juanita]*$_{\mathrm{F}}$ *la que vio siete armadillos*);[78] it is not followed by postfocal deaccentuation or compression (Vanrell & Fernández Soriano 2018: 43–54). The latter fact suggests a possible connection between the presence of the L- and subsequent deaccentuation or compression.

Both studies interpret their results in such a way that the right edge of the focal material seeks to align with the right edge of an iP, and with the most prominent position in that iP. In view of the definition of nuclearity adopted here from Ladd (2008) (cf. section 3.4.4), where the nuclear accent is the only obligatory and final one in an iP, we can simply say that focus is preferentially aligned with a nuclear accent. What the results of these studies (and many others) further show is that it is misleading to speak of prosodic cues for focus. The cues discussed above are not only used in the contexts described: H- and L- are simply cues for iP-phrasing, which is certainly not wholly determined by information structure. Deaccentuation and locally reduced pitch span also occur in contexts that are not postfocal (cf. Ortega-Llebaria & Prieto 2007; Torreira et al. 2014). Summing up, the division between focus and background is clearly one important factor influencing iP-level phrasing and nuclear accent placement in Spanish, but its signaling does not make use of a unique phrasing strategy, or one that is reserved exclusively for this function.

[78] Note that by and large, syntactic strategies are found to be at least as variably associated with information-structural configurations as prosodic ones in both of the studies: preferences for different types of clefts, fronting, *in-situ* word order (with prosodic marking as discussed in the text) and the postposing of the focal material (*p-movement* in Zubizarreta 1998) vary somewhat across focus type and variety, but mostly across speakers, and are thus only ever tendencies, with the gist being that clefts are preferentially used in contrasting or correcting contexts, and that *in-situ* word order is overall far more preferred than p-movement, which is rather marginal (81/182 for in-situ vs 26/182 for p-movement in "information focus" contexts, 114/300 for in-situ vs 34/300 for p-movement in "contrastive focus" contexts in Vanrell & Fernández Soriano 2018: 48,54; cf. Gabriel 2007: 283, 289–290). However, some type of syntactic strategy resulting in a deviation from unmarked word order seems to be employed in roughly half of the cases.

While results on phrasing in Spanish indicate that focus is associated with the highest prominence at the iP-level, variably expressed by a range of phonetic cues, results on the paradigmatic choice between nuclear configurations also indicate that focus is not the determining factor for the choice of pitch accents and boundary tones. Nuclear configurations occur in focus position, but the choice of configuration can encode additional pragmatic meaning such as illocutionary force and modal evaluative meaning.

Returning to the three focal nuclear configurations just discussed, both the paradigmatic choice between LH* L-, L* L%, and L<H* H-, and indeed of these three variants against other nuclear configurations, does not differentiate between what is focal and what isn't.[79] Between those three, on the one hand, the choice is largely determined by position, with L* L% occurring IP-finally, LH* L- IP-medially, and L<H* H- seemingly optionally when important material is phrased separately IP-initially.[80] With regards to the contour conventionally transcribed as L* L%, its most

[79] Or between „informational“ focus and „contrastive“ focus (a focal choice between a finite set of salient candidates), as is sometimes claimed: both Gabriel (2007) and Vanrell & Fernández Soriano (2018) explicitly test for this difference and come to the conclusion that it does not affect prosodic realization in any of the dimensions discussed here (“contrastive” focus is also sometimes related to a notion of “emphasis” that is supposedly expressed in increased pitch scaling).

[80] Gabriel (2007: 285–287) also finds this contour in part of his data elicited through a reading task, where the context together with the word order of the elicitation sentence forces a reading in which the focal constituent is fronted. He dismisses instances of L<H* H- on such fronted constituents that denote the answer to the (explicitly given) QUD as an inappropriate reinterpretation of them as (left-dislocated) topics by the speakers. Given that fronting as a focusing strategy is never attested in his more spontaneous data and that in clefts, the occurring tonal movement is analyzed as LH* L-H% (Gabriel 2007: 283), this characterization is certainly not unreasonable. However, fronting (without claims about its intonational specification) has been suggested to convey a mirative or counterexpectational import in Spanish (Leonetti & Escandell-Vidal 2009; Reich 2018; Cruschina 2019). The lack of such a pragmatic specification in the elicitation contexts might go some way in explaining its scarcity in the data reported on by Gabriel (2007), and in the light of the findings by Vanrell & Fernández Soriano (2018), it seems plausible that L<H* H- is not a misinterpretation (under the purely contextual definition of focus adopted here, that is strictly speaking impossible if the experimental task was not itself misunderstood), but simply an option available for phrasing important IP- or utterance-initial material in its own iP. That could occur both in it-clefts and constructions with initial topics, so that “important material” would have to remain as placeholder until further research provides more precise insights. Under the view adopted here, prosodic configuration, syntax, and discourse context together would then act together to yield an interpretation disambiguating the information structural role of the fronted constituent.

An additional point could be made regarding the interpretation of L<H* H- based on its phonetic realization: Gabriel et al. (2011) show that the phonetic realization of H- has a range of variants, in most of which the pitch is affected not only in the final syllable of the phrase, but already in the preceding ones including the tonic and posttonic. That makes it difficult to tell whether the

conspicuous feature is the absence of any pitch excursion on the nuclear accented syllable, as noted by Hualde & Prieto (2015: 364), who also call this an "accent without tonal correlates". That such a featureless contour is regularly interpreted to cue rightmost prominence in the IP can best be explained by the strong expectational bias for this metrical configuration, an observation already partly encapsulated in the formulation of the "nuclear stress rule" (Chomsky & Halle 1991 [1968]: 17–25). Prosody corresponding to this unmarked rightmost default prominence does not seem to need to be cued overtly most of the time, only marked deviations from it do (cf. Ladd 2008: 223, 257–259). This explains why focus in prefinal positions is associated with the acoustically more prominent nuclear configurations LH*L- and L<H* H-.

The paradigmatic choice between these three and other nuclear configurations, on the other hand, seems to encode pragmatic meaning at the illocutionary level and that of additional non-at-issue meanings: Fliessbach (2023) shows conclusively that non-at-issue meanings conventionally described as mirativity and obviousness, as well as whether an assertion is an agreeing or disagreeing response or a neutral provocation, have a clear effect on the nuclear configuration in declaratives even when focus is kept totally constant. A similar intuition also informs the data in Table 3 from Hualde & Prieto (2015) and much of the literature contributing to it. This means that LH* L-/L<H* H-/L*L% should be seen to signal something like neutral or unmarked declarative, to which focus position is orthogonal.

Based on this discussion, a more appropriate conceptualization for the relation between information structure and prosody is the following: the suprasegmental realization across the entire utterance cues a prosodic structure and a prominence profile (i.e., a metrical structure) that can be brought into relation with information structurally relevant divisions of the utterance. This cueing is asymmetrical because it is based on default expectations: as seen above, the expected = unmarked case of rightmost prominence at the highest level of metrical structure is often not given acoustically prominent pitch cues at all (cf. Calhoun 2010b: 11–13). The argument can further be made that even the nuclear accent is truly a metrical-prosodic category whose association to focus is only preferential, but not categorical (cf. Ladd 2008: 263–273). Evidence that this is true for Spanish comes from Calhoun et al. (2018), with similar arguments made for English in Calhoun (2010b). Calhoun et al. (2018) investigate the prosodic and syntactic realization in Venezuelan Spanish utterances (n=651 from 9 speakers from Valera) consisting of intransitive sentences under different information structural conditions and accounting

pitch accent on the iP-final word is LH* or L<H* based on peak alignment. While Vanrell & Fernández Soriano (2018) opt for the delayed version (L<H*), Gabriel et al. (2011) analyse it as LH*.

for the difference between unaccusative and unergative verbs. They performed a picture-based elicitation task that allowed speakers to produce utterances freely in terms of syntax and prosodic realization. See (32) for examples.

(32) Unergative and unaccusative Spanish example sentences, after Calhoun et al. (2018)
  a. La chica estornudó  *unergative*[81]
   "The girl sneezed"
  b. La chica apareció / Apareció la chica  *unaccusative*
   "The girl left"

Utterances were coded by the three authors for the position of nuclear stress (i.e. highest prominence in the utterance, on either the initial or final of the two lexical words) independently of the elicitation context. Acoustic analysis shows that this coding corresponds to a clear mean F0 difference (final stressed syllable 23 Hz lower) in favour of the initial accent in the case of initial stress and roughly equal pitch height (final stressed syllable 1 Hz lower) in the case of final stress, but only a relatively smaller difference in syllable length (final stressed syllable 50 ms longer) for initial stress, as opposed to final stress (final stressed syllable 67 ms longer, Calhoun et al. 2018: 16). That the coding thus corresponds to such a complex acoustic correlate is in itself evidence that strength relations as modeled via metrical structure are above all based on complex, context-dependent expectations (see also section 3.3.2).

In the analysis of their data, Calhoun et al. (2018: 18–20) find that in the overall majority (56%), their intransitive utterances have the word order subject-verb ((32) a and the first alternative of (32)b) with rightmost nuclear stress. All in all, 16% of utterances have the word order verb-subject (the second alternative in (32)b), mostly also with final nuclear stress. The remaining 28% exhibit subject-verb word order with initial nuclear stress. These ratios vary considerably across conditions, with contexts eliciting a correction (corrective focus) on the subject, those in which the subject is focused as answer to the QUD but not in correction (information focus), and those asking the question "what happened?" (broad focus) having significantly different rates of nuclear stress on the subject (53.3%, 32.5%, and 12.8%, respectively). Note that even in the corrective focus condition, 20.7% of utterances

[81] The classification here is taken from Calhoun et al. (2018: 7–10) and the works their discussion is based on. In generative syntax, unergative verbs are those whose single argument is VP-external, while the single argument of unaccusative verbs is VP-internal. This has been related to semantic verb types, such as ones denoting an uncontrolled process (unergative), or changes of location (unaccusative).

still have subject-verb order with final nuclear stress, i.e. the focal position is not marked. For information focus, this number increases to 49%, so that nearly half of all occurrences do not mark the focal position (the remaining 26% and 18.5%, respectively, have verb-subject order). In contrast, in the broad focus contexts, less than 30% (28.8%) of occurrences do not have subject-verb order with rightmost stress. Only here, the verb type also accounts for a large and significant difference: with unergative verbs in the broad focus condition, 88% are subject-verb with rightmost stress, whereas with unaccusative verbs, only 54% are, while 19% have the same word order with initial stress, and 28% have verb-subject word order. In the condition of information focus, the difference due to verb types is much smaller and just about reaches significance (p= 0.04), while it entirely disappears in the corrective focus contexts. This means that even though both information structure as induced through discourse context and verb type (to a lesser degree) have demonstrable effects on nuclear stress placement, there is a clear default to place it rightmost, even when this means providing no overt cues to focus position. The relationship between prosodic/metrical structure, semantics and syntax in the signaling of information structure is clearly not categorical, but can only be conceived of as probabilistic and heavily shaped by contextual expectations (cf. Calhoun 2010b; the assumption of a distributional relation between information structure, prosody, and syntax is also evident in the use of stochastic OT for their modeling by Gabriel 2007). The role of expectations is also implicitly invoked by Calhoun et al. (2018: 22) when they say that the corrective focus condition is more often overtly cued because it is more informative. Such degrees of informativity can be modeled in a discourse model such as the one by Farkas & Bruce (2010), where the assertion p of a correction against a proposition q which is already in the commitment set of the interlocutor would entail an empty projected set and would therefore be a marked move. Since the acceptance of such a move necessitates the removal of q from the commitment set of the interlocutor in addition to entering p into CG, it is more informative than an assertion that is not a correction. A plausible hypothesis is therefore that the more unexpected (i.e. informative) and therefore pragmatically marked a move is, the more likely it is to also receive a prosodic form that is also marked in the sense that it cues a metrical structure that is not the default.

Interesting further implications arise from the conception of the nuclear accent as belonging minimally at the level of the iP, and the distributive relation shaped by expectations between metrical/prosodic structure and information structure: it makes the prediction that several nuclear accents can stand in a metrical relationship with each other, and that utterances produced with several of them can still be assigned e.g. an unmarked, i.e. rightmost relationship overall, and thus receive essentially the same interpretation in terms of information-structural division as smaller utterances consisting of only a single iP/IP (see also Ladd 2008: 271).

Calhoun (2010b: 6) claims that this is the case for English and uses the constructed example given in Figure 16 to demonstrate that what she calls an "emphatic rendition"[82] of *Arun bought a Porsche*, with each content word produced as a separate iP/IP (she assumes only a single phrasal category), can still be an answer to "What happened?", in that *bought* and *a Porsche*, each a phrase with a nuclear accent, are assigned the relation w-s at a higher node, which then receives the highest prominence at the highest node, assigning *Arun* w only at this level. While the idea in itself is compelling, it somewhat suffers from being not much more than a thought experiment in this fashion. From an intonational perspective, investigating such utterances consisting of several iP/IPs is also compelling in terms of questions about how prosodic structure might deal with this and what it might mean for the discussion about recursive prosodic structure (cf. section 3.6). As a first approximation to how such multi-iP/IP utterances might actually play out, we can consider the two utterances in Figure 17 and Figure 18, which are basically attempts at recreating something like Calhoun's invented example for (Lima) Spanish.[83]

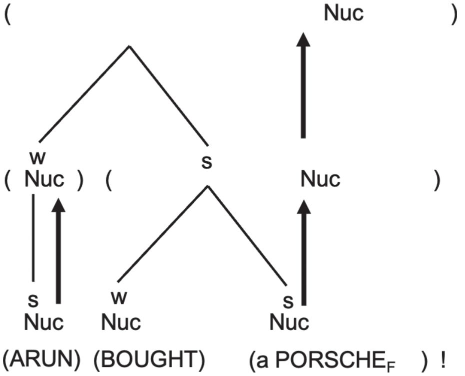


**Figure 16:** Metrical structure of an „emphatic rendition" of the sentence Arun bought a Porsche, adapted from Calhoun (2010b: 6).

[82] The context Calhoun (2010b: 5) evokes for the constructed example is that „the speakers are so surprised they produce every word in a separate phrase". This is perhaps not fully convincing, especially without further intonational specification, but intuitively, I would agree that such utterances where each or nearly each word is phrased separately, do exist.

[83] The speaker is the same speaker from Lima who also produced the utterances shown in Figure 7, Raúl Bendezú Araujo, who was kind and patient enough to record himself speaking them according to my instructions, for which I owe him thanks. They are of course totally artificial, but only serve to illustrate the problem here.

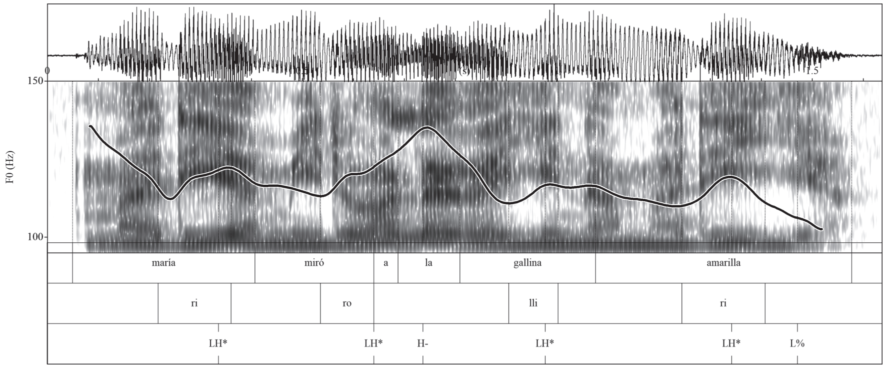


**Figure 17:** *María miró a la gallina amarilla* ‘Maria looked at the yellow chicken’, intended as an answer to “What happened?”, spoken by a Spanish speaker from Lima, normal rendition.

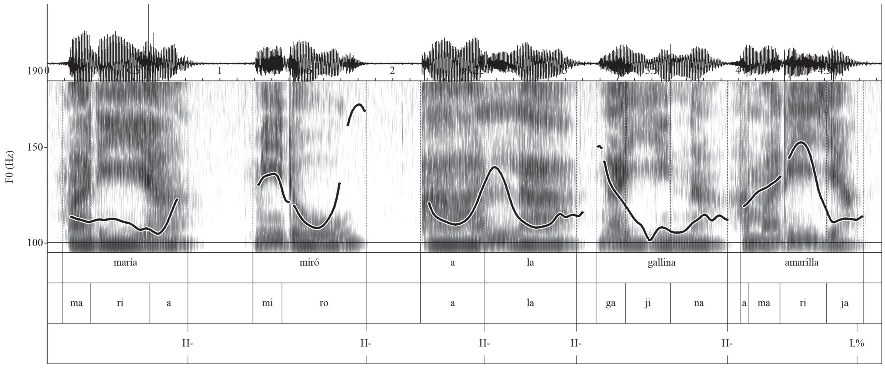


**Figure 18:** *María miró a la gallina amarilla* ‘Maria looked at the yellow chicken’, intended as an answer to “What happened?”, spoken by a Spanish speaker from Lima, “insistent” rendition.

Both Figure 17 and Figure 18 are productions of the utterance *María miró la gallina amarilla* “Maria looked at the yellow chicken” as answer to the question “What happened?”. While in the case of Figure 17, the speaker was told to produce the utterance as naturally as possible in this context, Figure 18 is the result of being asked to produce the utterance to the same QUD, but explicitly with added insistence, speaking slowly and hyperenunciating, as if talking to someone hard of hearing or slow on the uptake. I have refrained from transcribing pitch accents in Figure 18, both because Lima Spanish has not really been given an intonational account in the AM framework to my knowledge, apart from the inclusion of equally unanalyzed individual examples in the *Atlas interactivo de la entonación del español* (Prieto &

Roseano 2009–2013), and because the example is so obviously artificial.[84] It is still instructive to make a comparison with Figure 17. There, each content word realizes an LH* pitch accent, and a high rising pitch movement suggests the presence of a H- on or after *a la*, indicating that an iP-level boundary occurs there, while finally, a L% can be safely assumed. In comparison, Figure 18 seems to consist of a far greater number of phrases, visible not only from the pauses but also from the accompanying pitch movements indicating boundary tones, which I have tentatively transcribed. What is interesting is on the one hand that seemingly each single prosodic word (including the clitics *a la*, which seem to have been promoted to prosodic word status) is indeed realized as an individual iP and thus with its own nuclear accent, and that on the other hand, prominence relations between them can still be made out. The boundary rise delimiting *miró* is scaled far higher than any of the others, suggesting perhaps a stronger boundary there (roughly corresponding to the position of the H- in the "natural" Figure 17). The pitch accent on the final word, *amarilla*, is also quite certainly different from the preceding ones, further suggesting a structure above the level of the individual iPs, presumably an IP, with respect to which this pitch accent is final. Assuming that Figure 18 is indeed an appropriate answer to the question "what happened?", then this incidental comparison does seem to bear out the prediction by Calhoun (2010b) regarding the assignment of prominence relations at successive levels above that of the one at which the nuclear accent is assigned. In section 5.2 I will introduce a set of less artificial elicited Huari Spanish utterances that share the property of apparently assigning prominence relations at multiple levels of "nuclear accents". I will provide an analysis there for them that connects the indirect relation between phrasing, prominence, and information structural categories with the question of prosodic recursion in Spanish.

Broadening the focus somewhat, we can finally turn to Kügler & Calhoun (2020) to see how the prosodic cues to information structure found for Spanish relate to a wider typological context. They describe three main strategies in which information-structural categories (mainly focus) can be signaled in this indirect fashion via prosodic structure crosslinguistically. In our discussion of Spanish we have seen all three of them employed, namely prominence (focus seeks to align with the highest stress in the phrase which is often realized with the most acoustically prominent pitch movement), phrasing (focus seeks to align with prosodic edges) and pitch register (the focus-background division of the utterance corresponds at least partially

[84] We might note that the majority of identifiable pitch accents seem to be falling, perhaps instances of HL*, which occurs as part of HL* L% in Table 3 with the label "insistent explanation/insistent request".

to an asymmetrical difference in pitch register and span, in that postfocally, pitch movements are reduced and occurring at a lower register). Languages seem to differ in the degree to which they prefer either of these strategies and implement them.

#### 3.7.3.2 (Cuzco) Quechua

In the case of Quechua, very little research has been done regarding prosodic cues for information structure. Cole (1982: 210–211) describes for Imbabura Quechua (Quechua IIB, Ecuador) that there is a single main intonational peak in an utterance and that it is normally located on the final word, followed by a final fall independent of whether the utterance is a declarative, polar interrogative, or wh-interrogative. The main peak can move to a non-final word if that word is "emphasized or contrasted". If an utterance consists of several breath groups (i.e. smaller phrases), they can each bear a main peak. For Cuzco Quechua (Quechua IIC, southern Peru), Cusihuamán (2001: 79–81) does not mention a similar shift in the position of a main pitch peak according to the position of what is felt to be the "contrasted" word; he only describes pitch movement on the last two syllables of the utterance (using a four-level system), indicating that while declaratives, interrogatives, and imperatives have a final fall (with different pitch spans), exclamative utterances have elevated pitch only on the final syllable and end high. These affirmations are based on impressionistic data; the only works that I am aware of that deal with prosodic cues for information structure on an instrumental basis are again concerned only with Cuzco Quechua. O'Rourke (2005: 61–68) tentatively proposes that there are no prosodic cues to information focus[85] in declaratives, neither in terms of peak

[85] Her criterion for whether a constituent is in focus is that it is non-initial and marked by the evidential suffix *–m/mi*, after a proposal by Muysken (1995) that the evidentials *–mi/-chi/-shi* mark focus on the constituent they attach to (in initial position he states that they can also have scope over the entire clause). It seems however, that the relationship between the position of focus and the presence of the evidentials can at most be characterized in such a way that if there is an evidential in a sentence, the constituent it attaches to is focal, but not the other way round: this much is clear from the frequent attestation of sentences without any evidentials. The characterization in Weber (1989: 427–429) for Huallaga Huanuco Quechua (Quechua I, central Peru) is more cautious: he asserts that broadly, what parts of a sentence are thematic (i.e. topical), and which are rhematic (i.e. focal), can be determined from the interplay between the distribution of the "topic marker" *–qa*, the evidentials, and the position of the verb. Optional *–qa*-marked initial constituents are thematic, followed by a rhematic part which may contain constituents marked with the evidentials, and then the verb, followed optionally by further *–qa*-marked thematic constituents. He explicitly warns against simply identifying the evidential-marked constituent as the last or first rhematic one and also attests sentences with more than one evidential, which is ungrammatical in Muysken (1995: 381–382). This suggests a relationship (perhaps dependent on the language variety) between focus position and evidential marking that is quite similar to the distributional relationship be-

alignment nor of postfocal downstep or deaccentuation, but stresses that further research is needed because her analysis is based on only a small number of individual examples. O'Rourke (2009) somewhat qualifies this assessment. She proposes a regular LH* pitch accent[86] on stressed syllables (taken by her to be the penult in Cuzco Quechua) in declaratives, combined with iP-initial L- and optional iP-final H- in non-final, and iP-final L-, in utterance-final iPs (O'Rourke 2009: 308–309). Overall, peaks are aligned within the tonic syllable and on average show downstep across the utterance, but based again on the observation of individual utterances, this downstep pattern is said to be overruled by the focal constituent (again identified via the presence of *–n/mi* marking) having the highest peak in the utterance, suggesting highest prominence (O'Rourke 2009: 302–304, 307–308). Again, no evidence for postfocal deaccentuation or compression is found.

I want to conclude the section on prosodic cues and information structure by discussing in some more depth Muntendam & Torreira (2016), a study that to date is unique not only in that it experimentally investigates Quechua prosody under different information structural conditions, but also in providing comparative data from two Spanish varieties. Because their findings are so relevant for the present work, I will also address some important methodological shortcomings, but it should be clear that their pioneering contribution is extremely valuable to the aims of my own study because in many aspects it represents the only comparable work on at least related varieties. Muntendam & Torreira (2016) investigate the effect of information structure on prosody in Cuzco Quechua and Cuzco Spanish by the same bilingual speakers (16 speakers), and "Peninsular" Spanish (7 from Castile and Leon, 1 from Murcia), using a question-based task to elicit short utterances mainly of a noun phrase made up of an adjective and a noun in contexts of broad focus, contrastive focus on the noun, and contrastive focus on the adjective. Speakers participated in pairs and were given a stack of cards containing preformulated questions and coloured objects from which answers were to be built. As can be seen from (33)b) and c), "contrastive focus" on the adjective and the noun were elicited by asking a polar question in which the other element in the noun phrase was given

tween focus and prosodic cues argued for here. Inarguably, the evidentials and other morphosyntactic devices in Quechua interacting with focus all convey an additional (paradigmatic) meaning that is orthogonal to their use as focus markers, see e.g. Faller (2002, 2003, 2014); Behrens (2012); Bendezú Araujo (2021) for accounts.

[86] Actually she proposes an L*H for pitch accents in prefinal, and an LH* for those in final position in the utterance, based on peak alignment data. However, she also admits to the possibility that since peaks are nearly always realized within the stressed syllable in all positions, the pitch accent could also be taken to be LH* in general, and the alignment difference as phonetic (O'Rourke 2009: 309, note 16). The classification is clearly tentative and awaits further research.

and correct, while the adjective or the noun, respectively, did not correspond to the object shown on the card for the answering speaker and were thus intended to be corrected[87] in the elicited assertions. Broad focus was elicited by asking a wh-question about what object the answering speaker had on their card ((33)a)).

(33) Example elicitation dialogues in Spanish,[88] from Muntendam & Torreira (2016: 75)
a. Q: ¿Qué tienes?
"What have you got?"
A: Tengo una luna morada
"I have a purple moon"
b. Q: ¿Tienes una flor morada?
"Have you got a purple flower?"
A: No, tengo una luna morada
"No, I have a purple moon"
c. Q: ¿Tienes una luna negra?
"Have you got a black moon?"
A: No, tengo una luna morada
"No, I have a purple moon"

For each language variety, Muntendam & Torreira (2016) identify several attested intonational contours in the responses and their frequency of occurrence[89] across

[87] Note that according to Farkas & Bruce (2010: 96), polar questions such as those asked in (33) b) and c) are unbiased with regards to their response (whether the proposition asked for is true or not). The answers elicited here thus constitute reversals, which are different from denials in the sense that no commitment has been made in the provocation. It could thus be argued that the difference between the "corrected" and the "uncorrected" element in the responses is simply one of relative givenness, which is very variably cued prosodically across languages (Cruttenden 2006; Calhoun 2010b). However, it could also be considered that these responses form a particular subclass of reversals that might be called partial reversals, in analogy to the partial denials that are described as possible responses to assertions (Farkas & Bruce 2010: 99–100).

[88] Example elicitation dialogues for Cuzco Quechua are not provided in Muntendam & Torreira (2016). They would have been interesting with regards to possible differences in the placement of the polar question marker *–chu* (*-ku* in Conchucos Quechua) which can attach to different constituents and possibly be used to mark the question focus (specify the QUD with respect to which constituent it asks about, cf. O'Rourke 2005: 183). As far as I can tell, the intonational form of the elicitation question, produced by the participants themselves, was not controlled for. This applies also to the Spanish part of the experiment, for which it has been shown that the intonational form of the question can have a significant influence on the form of the response (Fliessbach 2023).

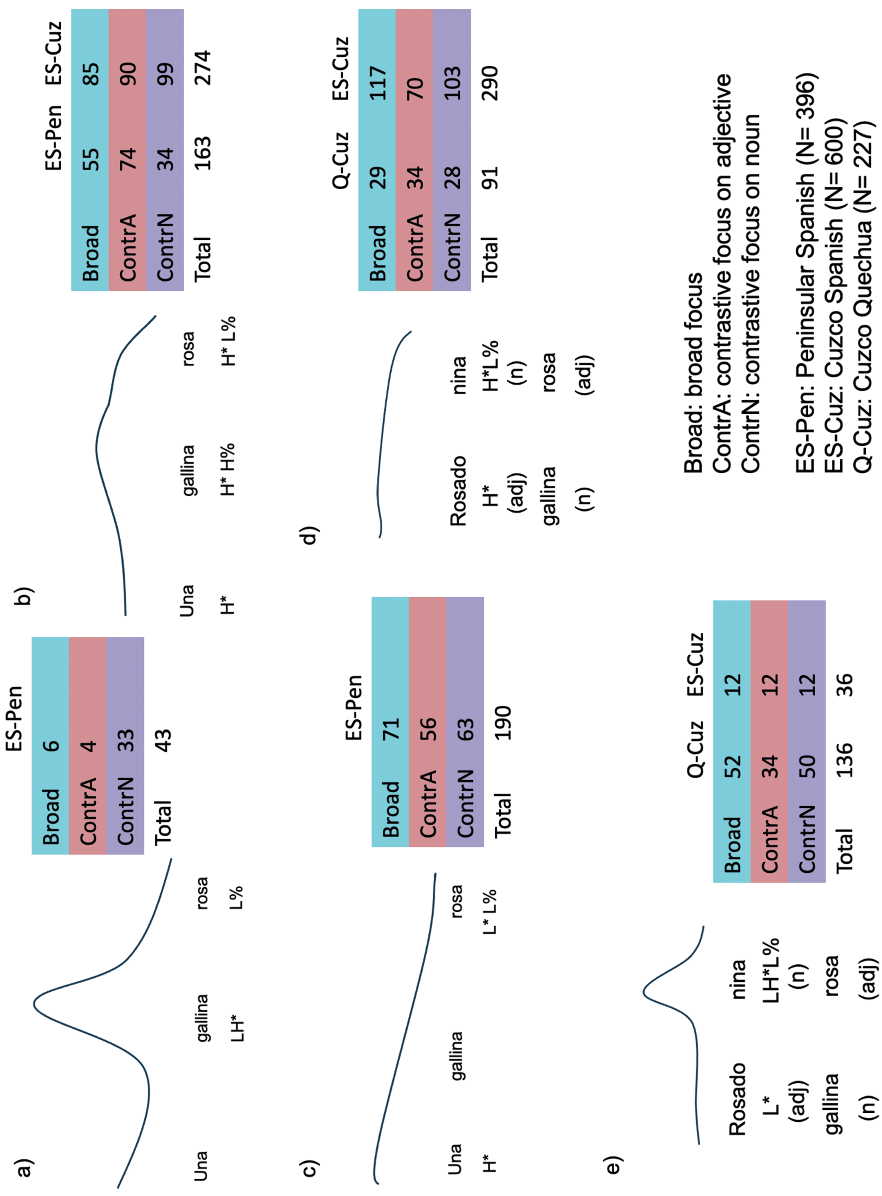
[89] Originally, speakers produced 20 target utterances per condition, yielding 960 utterances in bilingual Cuzco Quechua and Spanish each, and 480 utterances in Peninsular Spanish (Muntend-

the experimental conditions, reproduced together in Figure 19. For Peninsular Spanish, the results fall well in line with previous findings as well as with the theory of an indirect relationship between information structure and prosodic cues.

As can be seen from Figure 19, all three attested contours[90] for Peninsular Spanish occur in all three experimental conditions. Only tendencies can be made out in both directions of association: neither is an experimental condition exclusively linked to a single contour, nor an observed contour exclusively occurring in only one condition (with the (c) contour in the ContrN condition however coming closest). The same observation can be made correspondingly for the other two language varieties, making a direct encoding of information structure via prosody unlikely (Muntendam & Torreira 2016: 78, 84–85). The three contours identified

am & Torreira 2016: 75). For all languages, utterances including hesitations or longer pauses were excluded. Only for the Quechua data, all utterances containing case markers and all utterances in which the target NP containing the adjective and the noun were not utterance-final were also excluded, because "this is the natural position for the corresponding NPs in Spanish" (Muntendam & Torreira 2016: 76–77). For Peninsular and Cuzco Spanish, the elimination resulted in 396 and 600 remaining utterances for analysis respectively, but in the case of Quechua, only 227 utterances were included in the final analysis, a reduction by more than 75%. Assuming that a similar amount of elimination due to hesitation took place in the Quechua data and the Cuzco Spanish data, leaving some 600 utterances, then still more than 62% of the remaining data must have been eliminated only due to the presence of case markers or non-NP-final word order. It is difficult to assess how likely the utterances were to include case markers, because no Quechua example sentences are given in the paper and several constructions are possible in Quechua to convey the content of the examples in (33), some of which would not regularly have to contain any case markers at all. On the other hand, it can be assumed that the target utterances in Quechua were as simple as the Spanish ones, effectively consisting of only a verb and the target NP. That means that whether the target NP is final comes down to a binary option, and potentially the excluded occurrences represent the unmarked majority option, in which the verb is final. This issue is not addressed at all in the paper, but eliminating the word order option that represents the majority of cases should be a cause for concern. As it stands, there are already a number of studies about word order in Quechua, including the Cuzco variety, that all point to there existing a relation between word order and information structure, and a trend for verb-finality (Wölck 1972; Weber 1989; Muntendam 2010; Sánchez 2010). In the light of these assessments, it seems likely that the order V-NP is itself a marked word order and/or a relevant cue to information structure. The decision to eliminate it from the analysis perhaps thus means that data from a marked word order in Quechua is compared to that from an unmarked word order in Spanish. If this is the case, it could cast doubts on both the comparability of the findings on Quechua to those on the Spanish varieties (internal validity) and the possibility for generalization of the Quechua analysis beyond this sample of data (external validity).

[90] Muntendam & Torreira (2016: 76) do not provide any details on how the contours were identified, and how ambiguous cases and disagreements in annotation were handled. They do assert, however, that the ToBI-style pitch accent and boundary tone labels "only serve [. . .] the practical purpose of distinguishing the contours" (Muntendam & Torreira 2016: 77, note 5) in their data, presumably as opposed to representing a fragment of grammatical analysis.



a)

| | ES-Pen |
|---|---|
| Broad | 6 |
| ContrA | 4 |
| ContrN | 33 |
| Total | 43 |

b)

| | ES-Pen | ES-Cuz |
|---|---|---|
| Broad | 55 | 85 |
| ContrA | 74 | 90 |
| ContrN | 34 | 99 |
| Total | 163 | 274 |

c)

| | ES-Pen |
|---|---|
| Broad | 71 |
| ContrA | 56 |
| ContrN | 63 |
| Total | 190 |

d)

| | Q-Cuz | ES-Cuz |
|---|---|---|
| Broad | 29 | 117 |
| ContrA | 34 | 70 |
| ContrN | 28 | 103 |
| Total | 91 | 290 |

e)

| | Q-Cuz | ES-Cuz |
|---|---|---|
| Broad | 52 | 12 |
| ContrA | 34 | 12 |
| ContrN | 50 | 12 |
| Total | 136 | 36 |

**Figure 19:** Schematized versions of attested intonational contours for noun phrases consisting of a noun and an adjective, *una gallina rosa* “a pink chicken”/ *rosado nina* “pink fire”, from Peninsular Spanish (a, b, c), Cuzco bilingual Quechua (d, e), and Cuzco bilingual Spanish (b, d, e). Note the reversed word order between the two languages. Tables give number of occurrences of the identified intonational contours, per language and experimental condition, all adapted from Muntendam & Torreira (2016: 77–78, 81, 83).

for Peninsular Spanish are familiar and in broad agreement with the literature already discussed above. The (c) contour in Figure 19 is easily identified as the typical Spanish declarative nuclear contour in which the last word does not form an observable pitch peak at all, leading to an interpretation of final prominence by virtue of the expectational bias in that direction. Its status as an unmarked default is here confirmed by the fact that it is not only the most frequently used contour in the broad focus condition, but also in the condition with contrastive focus on the (prefinal) noun, and the most frequent overall (48% of all cases). Both other contours are arguably prosodically more complex because, if we equate the (a)-contour with the LH* L- contour from Gabriel (2007); Vanrell & Fernández Soriano (2018) and others, they involve an additional division of the IP into two iPs. The (b)-contour with the high prefocal iP-boundary tone is significantly more frequent in the condition with contrastive focus on the final adjective than the (c)-contour (Muntendam & Torreira 2016: 78), and the (a)-contour is clearly used most frequently (but not exclusively) in the condition with contrastive focus on the prefinal noun. Crucially, the relatively fewer occurrences of the (a)-contour in the condition of contrastive focus on the prefinal noun (33/130 occurrences, or 25%, in the ContrN condition) compared to the (b)-contour in the condition of contrastive focus on the final adjective (74/134 occurrences, or 55%, in the ContrA condition) cannot be explained via pragmatic or information structural markedness or degrees of informativity. In both conditions, focus is “contrastive” in the same way, but the (a)-contour is additionally marked in that the nuclear accent is not rightmost in the IP, i.e. the differential in the rate of occurrence can be explained when we make reference to the prosodic and metrical structure independently of the experimental conditions standing in for information structure. The expectation-based default of rightmost highest prominence seems strong enough to prevent a contour from being realized that would align focus position with a prefinal highest prominence in the majority of cases. The flipside is that this naturally makes an occurrence of such a contour that much more markedly informative, as already suggested in the discussion of Calhoun et al. (2018). These results thus provide further evidence for the independent relevance of prosodic and metrical structure, and the proposal about the nature of its relationship to both phonetic cues and information structure, which should be characterized as indirect: probabilistic, which is to say that from the perspective of an individual event, it is only possible to say that a certain information structural configuration will result in one of several realizations with a certain likelihood; and distributional, which is to say that only when observing many events can an association between an information structural configuration and a certain realization be made out, and intervening factors identified, in the form of trends in the distribution (cf. Calhoun 2010b).

Broadly, the same generalizations can be drawn from the Cuzco Quechua results in Figure 19: both contours are used in all three conditions, but there is a preference for the contour with clearer rightmost prominence (e) in the conditions in which focus is broad or rightmost (on the noun). Effectively, there does seem to be a tendency to differentiate between prefinal and final prominence, but no additional differentiation corresponding to the information structural division of broad vs. narrow final focus, as achieved in Spanish through the preferential use of contour (c), without iP-phrasing within the IP, vs (b), where the given material is phrased off with a H-. An open question is whether this is due to Cuzco Quechua not differentiating between broad and final narrow focus, or because this phrasing option is not available, or not used to separate given from new material. The Quechua results are also interesting in further ways. Both contours, (d), and (e), are at odds with the analysis of Cuzco Quechua declarative intonation in O'Rourke (2009). There, only a bitonal rising pitch accent LH* (see note 85) is proposed, but here the contours include the two monotonal pitch accents L* and H*. In addition, the high beginning in the (d) contour is also incompatible with the analysis in O'Rourke (2009: 304–305, 308–309), where a low boundary tone L- is proposed to be initial in every phrase. This suggests either that one of the analyses is incorrect, or that the contours in Muntendam & Torreira (2016) are not full contours in that they only characterize partial phrases. The discrepancy cannot be entirely explained away by saying that O'Rourke (2005, 2009) does not cover cases of contrastive focus; the L- LH* LH* L% contours we would expect from her analysis also do not occur here in the broad focus condition. It should be noted that since low pitch accents are notoriously difficult to identify and not much is known about how obligatory pitch accentuation is in prefinal position in Cuzco Quechua, the analysis of the L* in the (e)-contour is especially worthy of future investigation.

Cuzco Spanish, finally, makes use of contours (b), (d), and (e), also in general supporting the hypothesis about an indirect, distributional relationship between information structure and prosody because all contours occur in all conditions, with the difference in preference for (d) in the broad focus condition vs the preference for (b) in the condition with contrastive focus on the final adjective found to be statistically significant, but not the difference between the two contrastive conditions (Muntendam & Torreira 2016: 82–83). That is to say, Cuzco Spanish seems to mainly make a difference between a simple IP and one in which given material is phrased off in a separate iP. Based on the presence of contours (d) and (e) in both Cuzco Quechua and Spanish, Muntendam & Torreira (2016: 86), claim that this is evidence for cross-linguistic influence from Quechua to Spanish and even that this influence is "unidirectional: Spanish adopts prosodic features from Quechua but not the other way around". I would argue that their own evidence does not support this assertion, and that more research on the intonational phonology of Cuzco Quechua, but also on neighbouring varieties of Quechua or Spanish, is needed before any conclusions about directionality of influence can be drawn. Firstly, given the lack of a full intonational analysis of the contours in Quechua, with the proclaimed use of the ToBI-like labels solely for "the practical purpose of distinguishing the contours" (Muntendam & Torreira 2016: 77, note 5) but no information provided on the criteria for contour identification, comparing Quechua contours with Spanish ones cannot go beyond establishing a superficial phonetic similarity. Secondly, even if that issue were settled, there are no grounds for claiming that the (d)- and (e)-contours are in any sense "original" in Cuzco Quechua, and only "adopted" in Cuzco Spanish. The only statement supported by the facts is that the Cuzco speakers seem to use two of the attested contours in both of the languages spoken by them, and one in only one of them. The two "shared" contours might just as well "originate" from their use of Spanish and have made its way into Quechua, or be a shared innovation of the speaker community transgressing language boundaries, in the way of "diasystematic constructions" (Höder 2014a, 2018) in prosody, since virtually nothing else is known about the prosody of neighbouring varieties, of Quechua or Spanish.

In sum, even though considerable issues remain, Muntendam & Torreira (2016) make important headway into the study of prosody and IS in Cuzco Quechua. It seems that some of the same conclusions on the relation between information structure and prosody can be drawn as for Spanish: it appears to be distributional, insofar as there is not a categorical mapping between a contour type and a focus type.[91] Whether it is also indirectly mediated via metrical structure remains yet to be seen. For Conchucos Quechua, no comparable study exists, and not much more than basic facts of its prosody are known. In section 6.1, I will develop an account of Huari Quechua prosodic and intonational structure based on quantitative and qualitative data that takes some elements from O'Rourke's (2009) analysis, especially her use of initial and final iP-boundary tones, but greatly reduces the role of word stress and pitch accents. I will furthermore lay out how this proposal relates to information structure in sections 6.2 and 6.4, and develop an OT-model of intonation that describes a prosodic variation space between the attested forms of Huari Spanish and Quechua in sections 5.3 and 6.3. The insights gained there might even help shed some light on the outstanding issues in Cuzco Quechua and Spanish.

[91] Roessig (2021: 81–87) also comes to the conclusion that no one-to-one mapping between focus types and pitch accent types seems to exist in West Germanic languages. Evidence from other studies considered there indicates that the distribution of continuous parametres like relative peak alignment and tonal onglide interact with the distribution of categorical pitch accent types across focus types to ensure that they can still be recognized correctly in perception.