The effect of dialectal tonal variation on lexical access and sentence comprehension

Yiya Chen, Leiden University

There has been a wealth of linguistic data explaining dialectal variation in languages across the globe. Growing evidence suggests that speech processing difficulties can arise for listeners due to variation in regional dialects. Most research on dialectal variation and speech processing, however, has focused on segmental variation. In this talk, I will introduce lexical tonal variations in the Mandarin dialects of Chinese and the tonal mapping differences between these dialects and Standard Chinese (or Standard Mandarin), the official language of China. Tonal acoustic patterns, perceptual judgement and reaction time data, as well as EEG-based brain responses from three sets of speech processing experiments will be presented, with which I will argue that despite the high-level of mutual intelligibility across the Mandarin varieties, there are effects of dialectal tonal variation not only on spoken word recognition but also on sentence comprehension.
This talk looks at the implementation and effects of boundary intonational tones in different contexts in a few Bantu languages. Constituents of varying sizes including subjects, topics and dislocated constituents are considered in different sentence types showing that there is a disparity between whether a boundary intonational tone targets a single syllable or a number of syllables. An important question that comes into play in these cases is the interaction of lexical tones and the intonational tones and whether lexical tone in some way influences the implementation of boundary tones. It will be argued that the level of the intonational phrase as maximal i-phrase or intermediate i-phrase plays a crucial role in the implementation. In addition, the discourse context is also significant so that the intended message is also cued by the boundary implementation type.

In Bemba, for example, subjects in declaratives show a consistent right edge boundary L%. The boundary tone does not appear to replace a lexical final high tone on the subject noun, but causes it to be realized at a lower register. We see this same pattern for a fronted clausal topic with the following main clause showing pitch contraction. By contrast, a fronted non-clausal object shows a continuation boundary H% which is superimposed onto the low tone of the final syllable of the fronted object. Contrastive topics, relative and complementizer clauses also show a boundary H%. At the rightmost edge of declaratives, for example, boundary tones tend not to be punctual but rather occur over a longer stretch as may be seen in declination. We will hypothesize that the explanation for the different marking/implementation of boundary tones for non-clausal fronted object topics vs. clausal topics and subjects follows from discourse information, which is directly correlated to the level of embedding within a maximal i-phrase.
Ramifications of accent and intonation in North Germanic
Tomas Riad

The North Germanic intonational system is taken to be constrained by the presence of a lexical contrast, usually referred to as accents 1 and 2 (Gussenhoven & van der Vliet 1999: 99). Also, the tonal system exhibits remarkable structural similarity across dialects. Other systems with a lexical tonal contrast, however, e.g. Central Franconian, exhibit broad tonal variation both within and across varieties. Thus, the presence of a contrast does not in itself account for the relative coherence of accent across North Germanic dialects, or for the limited variation within a given variety.

Instead, it must be properties of the particular tonal or intonational system that constrain the variation and that holds the potential for the apparent stability in certain respects, across dialects and perhaps also over time.

Another fact that likely ties in with this situation is the absence of an obvious function for the tonal contrast. One could have expected that the contrast would be lost more easily due to low functional load. But it seems that intense language contact is needed (Finnish, Low German) for the accent contrast to be lost. This further indicates that the tonal contrast as such does not really bear on the issue of stability of tonal structure, in North Germanic.

In this presentation I discuss the factors that appear stable across dialects, including tonal structure and association patterns, relationship between accents 1 and 2, distributional patterns in the lexicon, and geographic spread. The goal is to come closer to an understanding of the North Germanic tonal system that can account for both the internal stability and the coherence across dialects, in terms of the structural factors that constitute the phonological basis of the system.

The properties of the system will be illustrated with known facts across the typology (expression of focus, prosodic treatment of compounds, tonal variation), as well as by geographic/diachronic transitions in the areas of the Central Swedish and Dala varieties, where historical developments can be studied as dialect transitions today. The latter transition concerns the development from so-called two-peaked accent 2 to one-peaked accent 2. Corresponding such areas in the south of Sweden do not lend themselves to such direct study.

Alternative lexical representations in bilinguals: Evidence from Swedish tonal accent

Nadja Althaus\textsuperscript{1,2}, Allison Wetterlin\textsuperscript{2,3}, & Aditi Lahiri\textsuperscript{2}

\textsuperscript{1}University of East Anglia, Norwich, UK
\textsuperscript{2}University of Oxford, Oxford, UK
\textsuperscript{3}University of Agder, Kristiansand, Norway

How do bilinguals deal with suprasegmental features if these are only lexically contrastive in one of their languages? Swedish tonal accent presents a test case for this, as it carries low functional load (few minimal pairs exist, and not all regional varieties use tonal accent). We therefore investigated lexical access in monolingual and bilingual native Swedish speakers whose other native language is non-tonal.

In Swedish, Accent 2 (färja\textsubscript{2} -"ferry") is the most common pattern for disyllabic nouns and functions as a default (Lahiri et al., 2005). Accent 1 is less predictable for disyllables. Many disyllabic Accent 1 words are underlyingly monosyllabic (where all monosyllables have Accent 1) with an epenthetic vowel (e.g., /fingr/ > finger\textsubscript{1}) in the singular but become disyllabic and have Accent 2 in the plural (fingrar\textsubscript{2}). Truly disyllabic Accent 1 patterns (sesam\textsubscript{1} - "sesame") belong to specific semantic classes or are loanwords.

Our first set of experiments addressed lexical access via semantic fragment priming. Here, subjects heard the first syllable of a prime word, and then made a lexical decision regarding a visually presented target that was either semantically related to the prime or not. The prime fragment was either presented with the accurate or with the opposite tonal accent.

Experiment 1 used as prime words disyllables that have no accent competitor, such that the mispronounced prime was not the beginning of an existing lexical entry (e.g. ponny\textsubscript{1} -"pony" with fragments pon\textsubscript{1}, *pon\textsubscript{2}). As shown in Figure 1, monolinguals restricted priming to accurate accent pronunciations, i.e. pronunciations of related fragment primes with the opposite accent contour did not lead to faster reaction times than unrelated primes. Bilinguals' responses were similar to monolinguals for Accent 2 items. For Accent 1 items, by contrast, priming was not restricted to accurate pronunciation of the prime. In other words, mispronouncing an Accent 1 item with the Accent 2 contour still led to faster responses than unrelated primes.

![Figure 1. Experiment 1 results: Priming effects for monolinguals and bilinguals. For bilinguals, even mispronounced Accent 1 items exhibit priming, whereas for monolinguals mispronunciations do not prime.](image-url)

In Experiment 2 we tested whether this would change in the context of lexical competition. Here, we used primes that do have an accent competitor with segmentally
identical first syllable (e.g. paddel - "paddle" and padda - "toad", with fragments pad1 and pad2). As a consequence, mispronunciations are in fact first syllables of different words. Here, priming was restricted to accurate accent pronunciations even for bilinguals (Accent 1 and 2), as shown in Figure 2. This demonstrates that bilinguals do encode Accent 1 when the lexical neighbourhood requires it.

Experiments 1 and 2 did not tap into the differences between Accent 1 patterns that are underlying monosyllables (e.g., finger) and those that are truly disyllabic (e.g., tango). In Experiment 3 we therefore used eye tracking to investigate on-line processing of accent by monolinguals and bilinguals, in order to tap into more fine-grained differences between the categories. We used a fragment completion task (Felder et al., 2009) where subjects heard a first syllable fragment and then had to choose between two visually presented target words. The words were, as above, pairs of words whose first syllables contain the same segments but require opposing tonal accent. To analyse eye movements, we fitted a growth curve model to the time course patterns exhibited for the different groups of words and participants. Monolinguals' eye movements were consistent with the hypothesis that disyllabic Accent 1 is encoded in the lexicon, whereas underlying monosyllables (e.g. finger) and Accent 2 are unspecified. Bilinguals' eye movements indicated that Accent 2 is, as in monolinguals, not lexically specified. However, unlike monolinguals, bilinguals seemed to specify both types of Accent 1 in the lexicon, as indicated by the relative speed of processing.

Taken together, our results indicate that bilinguals struggle with the less predictable Accent 1 patterns. They encode Accent 1 in the lexicon - but only where lexical competition exists. They do not appear to discriminate between the different subcategories of Accent 1, indicating that they may not be sensitive to the morphophonological regularities the subcategories are associated with.

References


Phonetic realization of Swedish complex pitch accents as a function of accompanying head and eyebrow movements

Gilbert Ambrazaitis¹ & David House²

¹Department of Swedish, Linneaus University, Växjö, Sweden; ²Department of Speech, Music and Hearing, KTH (Royal Institute of Technology), Stockholm, Sweden

In this study we are examining prosodic prominence from a multimodal perspective. Our research question is whether the phonetic realization of tonal and intonational prominence varies as a function of accompanying head and eyebrow movements (beat gestures, cf. below). More specifically, we explore the realization of the tonal and intonational components of so-called focal or big accents in a Stockholm Swedish corpus of television news readings.

Swedish makes use of pitch contrasts at the lexical level, distinguishing between two so-called word accents (Accent 1, Accent 2). Orthogonal to this word accent contrast, many varieties of Swedish, including the Stockholm dialect studied here, distinguish between two pitch-related phonological prominence levels, where the higher level has been commonly referred to as a focal accent or more recently as a big accent [1], as opposed to the non-focal or small accent [1]. Crucially, the distinction between Accent 1 and Accent 2 is encoded at both levels. According to Bruce’s seminal analysis [2], the big accent can be conceived of as a complex pitch accent composed of the tonal configuration for the word accent (Accent 1 or 2) and a following high tone H (the sentence accent in his analysis [2]) which is realized as a rise in pitch from the accentual L (HL* in Accent 1, H*L in Accent 2). Although the details of tonal representation of Swedish word accents, as well as the question of the lexicality of tones involved, is much debated (e.g. [1][2][3]), there is a certain consensus on the compositional nature of big accents, as well as the assumption that the tonal components of a big accent relate to different prominence levels and different domains in the prosodic hierarchy of Swedish [2]. In this abstract, we simplify the debate by referring to the tones defining the word accents as tonal, and the subsequent rise (the H tone; i.e. the sentence accent [2]) that distinguishes small from big accents as intonational.

Previous research on co-speech gestures and audio-visual prosody strongly suggests that prosodic prominence is indeed an audio-visual, or multimodal, phenomenon, as pitch accents (verbal prominence cues) are frequently accompanied by movements of the hands, the head and certain facial areas (visual cues), also referred to as beat gestures, e.g. [4][5]. It has, moreover, been shown that visual and verbal prominence cues may co-occur in various constellations [6][7] and that beat gestures are more likely to occur with perceptually strong accents than with weak ones: Swerts and Krahmer [7] found in their study of Dutch news readings that the more accented a word was on an auditory scale, the more likely the word was to also be accompanied by a head movement, an eyebrow movement or both. Hence, we might predict a cumulative relation of verbal and visual prominence cues, i.e. a positive correlation between the acoustic strength of a pitch accent (e.g. in terms of segmental durations, F0 peak height or F0 ranges) and accompanying beat gestures. In this study, we test this prediction for the special case of complex (big, cf. above) pitch accents in Stockholm Swedish, thereby adding a prosodic-phonological dimension to our research question: Do we find a cumulative relation between the occurrence of head and eyebrow movements and (a) the tonal, (b) the intonational, or (c) both components of a big accent in Stockholm Swedish? Answering this question would add to our general understanding of gesture-speech integration, and more specifically of the interaction of visual and verbal prominence cues.

The present study is based on audio and video data of 60 brief news readings from Swedish Television (SVT Rapport), comprising 1936 words in total, or about 12 minutes of speech from five news anchors (two female, three male). The material was transcribed, segmented at the word level, and annotated for big accents (henceforth, BA), head beats (HB) and eyebrow
beats (EB) using a combination of ELAN and Praat. In a first step of annotation, the presence vs. absence of BA, HB and EB was judged upon on a word-basis. About half of the materials (30 files) were annotated by three labelers independently of each other. Inter-rater reliability was tested using Fleiss’ κ, and turned out fair to good (BA: κ = 0.77; HB: κ = 0.69; EB: κ = 0.72). For the purpose of this study, the analysis focuses on three conditions: (i) words produced with a BA only (i.e. without a beat gesture: 276 tokens in our material), (ii) words with BA co-occurring with a HB (BA+HB: 178 tokens) and (iii) words with BA co-occurring with both HB and EB (BA+HB+EB: 73 tokens). In a second step of annotation, tonal targets were labelled for all 527 tokens of interest: (H+)L* H- in case of Accent 1 and H*+L H- in case of Accent 2 (where H- is the sentence accent tone). In addition, as a baseline condition, tonal targets were labelled for a random selection of 102 non-focally accented words (small accents: 52 Accent 1, 50 Accent 2). Based on these tonal annotations, seven measures were derived to capture different aspects of the phonetic realization of the accentual fall (HL* or H*L respectively), and the sentence accent rise (H-): I/II – absolute peak height of HL*/H*L and H- (2 measures), III/IV – range of accentual fall and H- rise in semitones (2 measures), V – highest peak in word (= either HL*/H*L or H-), VI – largest range in word (= either accentual fall or H- rise), and VII – the difference between H- and accentual peak (HL*/H*L) in semitones.

The results reveal slight effects of the multimodal-prominence cluster (henceforth, MMP, i.e. conditions i-iii above plus the baseline), on measures I-VI, suggesting a tendency for a cumulative relation of verbal and visual prominence cues: i.e., the more visual cues accompanying, the higher the pitch peaks and the larger the rises and falls. These effects are strongest for the combined measures V and VI. All dependent variables were analyzed by means of linear mixed effects models assuming three fixed factors: MMP (cf. above), speaker sex, and word accent (Accent 1, Accent 2), as well as speaker as a random effect. According to a likelihood ratio test for each of the dependent variables, the effect of MMP was significant for measures V (p=0.015*) and VI (p=0.042*) suggesting an interaction of verbal and visual prominence cues that is reflected in both components of complex pitch accents in Stockholm Swedish – here referred to as the tonal fall and the intonational rise. The results are discussed both in the light of a proposed outline of a model of multimodal prominence production, and with reference to prosodic domains assumed in Swedish phonology.

References


Modelling variability in intonation: covariation, non-local effects, and cue trading

Amalia Arvaniti¹  Argyro Katsika²  Yang Li³

¹University of Kent  ²UC Santa Barbara  ³University of Cambridge & Amazon UK

We examined three focal pitch accents used in Greek declaratives [1]: H*, used for discourse-new information, L+H*, used for contrastive focus, and H*+L, which indicates speaker belief that the accented item should have been in the common ground. These accents are realized on one of the last three syllables of an utterance and are followed by L-L% edge tones. This makes them comparable but also prone to tonal crowding and thus variable in realization [2]. The challenge was to capture this variability in a systematic way that reflected its gradient nature.

Native speakers of standard Greek (9F, 2M) read four repetitions of dialogues designed to elicit the three accents on test words varying in stress placement; see (1). The test words were final in utterances that were either one or two content words long, for a total of 792 tokens [11 speakers × 3 accents × 3 test words × 2 lengths × 4 repetitions]. In the two-word utterances, the accent under investigation was preceded by a L*+H accent on the previous content word [1]; see (1a2). For each test word, the three-syllable interval ending at the offset of the stressed syllable was marked and its F0 extracted using STRAGHT [3]. The F0 curves of this interval, which is underlined in (1), underwent Functional Principal Component Analysis (henceforth PCA) following [4]. PCA returns the most dominant modes of variation in the functional form, called Functional Principal Components (PCs). Every input curve then receives a coefficient for identified PCs, representing the contribution of each PC to that curve’s realised shape.

PCA showed that 85.2% of the variance among the three accents can be captured by PC1 and PC2. As shown in Fig. 1, PC1 captures differences primarily in peak height and captures 62.1% of the variance; PC2 captures 23.1% of the variance and reflects a combination of contour shape and peak alignment (position of the peak in the three-syllable window). Thus, PCA allowed us to observe co-varying dependencies between peak scaling and alignment, and indicated that both are needed to distinguish the three accents, as each of the dimensions showed overlap between accents; see Fig. 2. Further, linear mixed-effects models showed that accent type, utterance length, and stress position significantly affected PC1 and PC2 in ways consistent with the description of the three accents [1], and with attested tonal crowding effects [2]. Finally, the data showed consistent differences that depended on distal contexts (such as the presence of a preceding pitch accent) leading to non-localized effects on the F0 trajectory, such as differences in scaling on the unaccented syllables in the analysis window captured in PC1 (Fig. 1). Additional PCA analysis of the first two PCs showed trade-off between the PCs and durational measures, especially the proportion of the three-syllable window taken up by the accented syllable, such that peak scaling got lower the greater this proportion was. PCA allowed us to capture these systematic effects that distinguish each accent from the others and to represent the attested variability along a small number of dimensions.

These results showcase the usefulness of data-driven parametrization of F0 curves through PCA, and have consequences for established practices in the study of intonation. They show that tonal alignment should not be prioritized over scaling, and that the two are not independent of each other. Accentual contrasts are shown to rely on a number of parameters (scaling, alignment, segmental duration, curve shape) that are in trading relations specific to each accent. Further, differences in realization may not be localized on accented syllables: here differences between accents are also evident on the unstressed syllables before the accented one. Overall, the results suggest that the established research focus on localized F0 targets and invariance as criteria for the phonological status of tonal events risks positing categories that are too fine-grained and capture phonetic variability rather than essential contrasts. Instead the results argue in favour of treating tonal events similarly to segments, i.e. as being expressed by a number of phonetic parameters that show variability and are in a trading relationship with each other.
Accent Sample dialogues

a. H*
   1 word utterance [laðoˈlemono] Lemon-and-oil-sauce.
   What will you have?

b. H*+L
   What should I do with all these lemons?
   [lemoˈnada] Lemonade. (What else can you make with them?)

c. L+H*
   Did you say their son has brown eyes?
   [yalaˈna] Blue! (Aren’t you paying attention?)

---


Contact-induced intonational variation is less well understood than segmental variation [1]. The few studies on the former [e.g. 2] document a range of synchronic outcomes, but do not address the role of contact in diachronic development. Our larger project investigates the implications of historical contact of Greek with Italian and Turkish for intonation change in Greek regional varieties. This study focuses on Greek-Turkish contact only.

Specifically, we compare the tunes of yes-no questions (YNQ) and continuation rises (CR) in Standard Modern Greek (SMG, as spoken in Athens) with those of Asia Minor Greek (AMG) and Turkish. The AMG material spans a century and 5 generations of speakers. As shown in Figure 1, the earliest available recordings date back to the turn of the 20th century, with the oldest speakers born in the Anatolian Peninsula in the 1890s and the youngest in the 1990s, who, unlike their great-grandparents, no longer have everyday contact with Turkish. The latter two varieties were chosen because their speakers cohabited and interacted in the Anatolian peninsula from the 11th century until 1923. Today second and third generation descendants of AMG speakers can still be found in mainland Greece.

Our methodology combines data-driven mathematical modelling of intonation $f_0$ shapes with the insights of (a) traditional diachronic investigations and (b) the Autosegmental-Metrical (AM) intonational model [3, 4]. For this talk we use archival and contemporary recordings of SMG, AMG and Turkish. Unlike the majority of studies in intonation, we engage with a natural speech corpus (1533 tokens; 312 AMG, 484 Turkish and 737 SMG) capturing emergent patterns and variability with Functional Data Analysis (FDA) [5, 6] of $f_0$ curves as well as the standard AM analysis.

Preliminary analysis indicates that the CR and YNQ tunes of AMG bear more resemblance to their Turkish than their SMG counterparts. Figure 2 (Left) exemplifies the type of comparison of the CR $f_0$ curves. While all three varieties end in a rise, in AMG (black disks) and Turkish (asterisks) there is a peak during the accentual syllable, which in SMG (open circles) this syllable is realized with a low plateau. Shape parameters (slope and curvature) of AMG pattern with Turkish (Figure 2 Right) and indicate a higher quadratic curvature than SMG. In Figure 3 (Left), the accented syllable [tɾe] of the YNQ [me to ˈtɾeno ˈirðate] ‘Did you come by train?’ is realized with a low plateau in SMG (blue), but with a rise in AMG (purple). Turkish (Figure 3 Right) has a rise similar to the AMG during the accented syllable [li] of [biˈlijoɾ] in the Turkish [inˈɡi̞lizde biˈlijoɾ muˈsunu] ‘Do you speak English’.

A detailed analysis of CRs and YNQs based on all 1533 tokens is presented in the poster.

The AM and FDA analyses converge on a consistent picture (cf. [6]), promising less labour intensive/costly and more reliable analyses of large datasets. The results presented here, taken alone, are not enough to shed light on questions about the role of long-term language contact in intonation change, the rate of this change, or the time span of contact influences; they form but a small part of a project in which a number of dialects will be compared along a chronological and a socio-cultural axis. On one hand, we aim to develop a model of diachronic change of intonation and establish patterns of attrition by comparing contemporary with mid- and early 20th century recordings. On the other, we intend to examine the extent of influence between Turkish and Greek (Athenian, AMG and Cypriot dialects), because the contact in each dialectal area differed in its temporal and socio-cultural characteristics. We intend to correlate these differences with the amount of intonational variability in each dialect.
Figure 2. Left: Model of the shape of CR in AMG (black disks) vs. Turkish (asterisks) and SMG (open circles). Dashed lines: contours modelled using cubic polynomial basis functions ($y = ax^3 + bx^2 + cx + d + \varepsilon$) fit very closely fit with the actual f0 curves; Right: CR shape parameters. The AMG examples cluster more with Turkish than with SMG.

Figure 3. YNQ comparison, accented syllable in rectangle. L: AMG (purple) and SMG (blue) [me to ‘treno ‘iɾðate]; R: Turkish [iŋ’gilizðe bi’lijoɾ mu’sumuz]

In this paper, I provide a novel analysis of floating tone and docking data from two Senufo (Niger-Congo, Gur) languages, Supyire and Syer. According to descriptive data from Carlson (1994), Supyire is analyzed as having four underlying level tones: high (H), low (L), strong mid (Ms), and weak mid (Mw). Additionally, certain noun root melodies found in the language involve a floating tone as shown in (1). Floating tones may either be root-initial (1a,b) or root-final, (1c,d).

(1) a. LMₘ (48(93))
   /L Mₛ.Mₘ/  'tail'
   nɛ.ŋɛ
   b. LH (55(115))
   /L H.Mₘₙ.Mₘ/  'twins'
   nɛ.mii
   c. ML (48(94))
   /Mₘₙₙₘₙₘₗ/  'goat'
   sika
   d. LMₘₙ (49(95))
   /L.L.L Mₘₙ/  'large hoe'
   ŋa.mii

The floating L in (1a,b) may dock onto any adjacent vowel, regardless of syntactic domain boundaries.

(2) ka u u sa [ [ nɛŋke num.bwɔ.he ] lwɔ ] na fi
L M H HL Mₘₙ.H M.M.L.M L M H

and she NARR go [vp[dp tail-DEF ADJ-big-DEF ] take ] PROG run.IMPFV

'She went and took the biggest tail and ran.' (223(87))

The final L in (1)c may dock rightwards (3)a, unless it is in the context of a definite suffix (3)b, in which case it docks leftwards onto the root-final vowel. Its docking is therefore morphologically conditioned.

(3) a. si.ka - pe.re
   Mₘₙₙₘₙₘₗ Lₘₙ
   goat - male
   'male goat' (58(123))
   b. si.ka -ŋi
   Mₛ.Mₘ₉ₙ Mₘ₉ₙ₉ₙ
   goat -DEF
   'the goat' (54(111))

Finally, the floating Mₘₙ tone in (1)d is also conditioned, but phonologically. When followed by a L, it also docks leftwards (4).

(4) tu.u.go  niŋ.kin
   L.L.Mₘ₉ₙ L.L Mₘ₉ₙ
   large.hoe one
   'one large hoe' (54(112))

These three distinct phonological patterns of floating tones vary with morphosyntactic construction. In her grammar of Syer, Dombrowski-Hahn (2015) notes comparable morphosyntactically conditioned alternations. For instance, the HL nominal melody involves an initial floating H tone as in (5).

(5) HL (98(48))
   bre 'game'
   /H L/  na 'man'
   /H L/  mро 'cloth'
   /H L/  gbәn 'pot'
That floating tone is not realized when the noun is in utterance-initial position (6a). When occurring after a definite marker (6b) or as a direct object (6c), floating H is realized on the noun itself. Finally, when an ML morpheme is suffixed to the noun (6d), the high tone docks onto the element that directly precedes it.

(6) a. brɛ nii m-brɛ
   L H.L LH.L
   /H L/
   game HAB IP-play
   'A game is usually played.' (96(37d))

b. ti brɛ
   M H.L
   /M H L/
   DEF7 game
   'the game' (96(41))

c. u maa gbän curu
   M L.L H.L L.M
   /H L/
   PR1 FUT.go pot wash
   'She is going to wash a pot.' (97(43))

d. pi mra -ɔb tyin
   H L.M L
   /L H L/ -ML/
   PR2.PRF cloth -PL weave
   'They wove cloths.' (98(46))

To model such facts, I use Cophonology Theory (Orgun 1996, Inkelas et al. 1997), which allows for multiple different phonological grammars, sensitive to construction, in the same language. I show that the docking of Senufo floating tones is conditioned differently at different levels of the syntactic structure. Namely, I argue in favor of a Def-cophonology, a DP-cophonology, a VP-cophonology, and a CP-cophonology, each with its own distinct constraint ranking. These cophonologies, along with a tone strength scale (Smolensky & Goldrick 2014, 2016) that is encoded in the grammar of the particular language, are responsible for the behaviors of floating tones observed above.

This work contributes to both the descriptive and theoretical literature. Empirically, it provides an account for complex tonal interaction data found in two related and both understudied Niger-Congo languages. By doing so, it provides a basis for a potential analysis of the phenomenon generalized across the Senufo subgroup of Gur languages.

Theoretically speaking, this paper participates in the ongoing debate of how to best encompass and represent different conditioning factors for phonological processes found in natural language. I argue in favor of Cophonology Theory over traditional Optimality Theory (Prince and Smolensky 1993/2004) and Indexed Constraint Theory (McCarthy and Prince 1995, Itô and Mester 1999, Pater 2007), since it is particularly well-equipped to account for phonological alternations that are conditioned in more than one way, like those observed in Supyire and Syer. Additionally, this work buttresses Inkelas (2015) by claiming that there is a confidence scale active in Senufo, where for each faithfulness constraint there exists a corresponding constraint that only targets representationally strong elements. This strength-based representational difference in Senufo tones adds to the growing body of work on gradient phonological representations (Smolensky et al 2014; Smolensky & Goldrick 2016).
Effects of Musical Aural Skills Training on L2 Learning of Mandarin Chinese Tones
Evan D. Bradley, Penn State University & Seth Wiener, Carnegie Mellon University

Music has powerful effects on the perceptual system, and musicians perceive sound, especially pitch, differently from non-musicians. A consequence of this is better perception of vocal pitch and enhanced ability by (non-tone language-speaking) musicians to learn lexical tone languages (Bradley, 2013). Music–language perception studies have typically compared musicians with many years of training to individuals with little to no formal music experience. It is not yet well established how much training or what skill level is required to see changes in linguistic pitch perception. Both training and individual differences in aptitude and personality play a role in the observed differences between musicians and non-musicians (Corrigall et al., 2013), but the neural changes which result from musical training suggest that such training could be harnessed for non-musical auditory perception (Kraus & Chandrasekaran, 2010). The goal of this project was to determine whether the observed differences between musicians and non-musicians can be exploited to improve linguistic abilities in the general population. Based on models of auditory perception (Patel, 2011), we adapted a computerized musical ear training program for novices, and evaluated its effect on tone perception by L2 Mandarin learners engaged in classroom language learning. Because the unique perceptual demands of music produce neural changes affecting language perception, we expect that as individuals improve their music pitch perception, their lexical tone perception ability will also increase.

Method: English-speaking students (n=11) in a 15-week introductory college Mandarin Chinese class participated in the experimental group. None had previously studied a tone language, and none were musicians (having studied any musical instrument for fewer than 5 years). In addition to typical class activities, students took tone and pitch perception tests 10 weeks apart, and participated in musical ear training during the intervening period. A control group of 11 students in another section of the same class took the pre- and post-tests, but did not participate in musical ear training.

Tone & Pitch Tests: Speeded tone perception tests were administered in E-Prime. In the 4-alternative-forced choice (4AFC) Tone Categorization Test, participants heard a Mandarin word (syllable ‘yu’) with one of four tones spoken by one male and one female native speaker, and matched it with a graphical arrow symbol depicting its pitch contour (e.g., →, ↗, ↘, ↝). In the Tone Discrimination (AX) Test, participants heard two words (both with syllable ‘yu’) spoken by the same speaker, and had to decide whether they had the same or different tones. Participants also took the Tonometric Adaptive Pitch Test (Mandell), which determines a difference threshold for pitch using synthesized pure tones. Two native speakers of Mandarin also took the three pre-tests to ensure naturalness of the stimuli and establish a baseline for native-like performance.

Musical Ear Training: Ear training is perceptual training to identify structural elements of music (like intervals and chords) through repeated listening and feedback. EarMaster, a commercial software program for musicians, was used to administer and track training. Participants were trained to recognize melodic piano intervals (two-note patterns defined by the distance between their pitches). Training tasks included interval discrimination (“which interval is larger?”) and interval naming (e.g., “is this a major second or a major third?”). EarMaster provides feedback to learners and tracks progress. Participants were assigned to practice for two 30-minute sessions each week for 8 consecutive weeks. Each week, training became progressively more difficult through the inclusion of more similar interval types (e.g., week 1: major second vs. octave; week 8: perfect fourth vs. perfect fifth).

Findings: The experimental group’s average pitch difference threshold (9Hz) did not change from pre- to post-test; the control group’s improved by 1.7Hz (from 10.7 to 9Hz). Both
groups improved in tone perception in both discrimination and categorization tasks: in tone discrimination (Fig 1), the groups reached a similar level of performance (~96% accuracy), but the control group started at a slightly lower level, meaning that they improved slightly more than the experimental group (5% vs. 3.5%); for tone categorization (Fig 2), the two groups again reached nearly identical performance (~75% accuracy), with the control group improving slightly more (12% vs. 9%), due to a lower score at pre-test.

Because students varied in their compliance with EarMaster training assignments, we tested the correlation between minutes of EarMaster training and post-test score and change in the 4AFC test (which was less subject to ceiling effects than AX). There was a small negative correlation between minutes spent on EarMaster Training and post-test 4AFC score ($r = -.31$) and change in 4AFC score ($r = -.30$), meaning those who trained more improved less in tone categorization and scored lower on the post test than those who trained more.

**Discussion:** The EarMaster intervention did not produce the expected boost in tone learning. Two primarily limitations of the intervention may have obscured any such effects:

1. The amount of musical training undertaken by learners—or the level of proficiency they were able to achieve—was not sufficient to produce change in pitch perception.
2. The time course of music-language crossover could not captured within the training period (8 weeks), either because crossover lags behind within-domain learning, or because simultaneous practice in the two modes was counterproductive or distracting (as suggested by the negative correlation between training time and tone learning).

So, while ample evidence still supports the idea that musical experience is relevant to tone language learning, capturing this effect within the confines of classroom L2 learning remains elusive, and should be undertaken with caution. Our future research will focus on factors affecting the relevance and time course of musical effects on lexical tone, and on making the training task more engaging for students.

**Selected References**

Mandell, J. Adaptive Pitch Test. Tonometric.com
Ezra Pound’s Cadences and the Tonal Grammar of Recitation
Patrik Bye, Nord University, Bodø (Norway)

In this talk I will argue for extending the autosegmental-metrical theory of intonation (AMP; e.g. Ladd 2008) to chant, a stylized form of intonation where tonal spans predominate, and which is found in certain types of poetry recitation. I will present results of an analysis, using Praat (Boersma & Weenink 2018), of a corpus of Ezra Pound’s readings of his own poetry, including the Cantos and other poems (Cantico del Sole, Hugh Selwyn Mauberley), all of which are available from the PennSound archive. I will outline a tonal grammar of Ezra Pound’s declamatory recitation style, describing also how Pound’s spoken melodies correlate with the arrangement into lines on the page. If there is time, I will draw brief comparisons to the recitation styles of other chanters, such as W. B. Yeats, Vachel Lindsay, and Allen Ginsberg.

In outline, Pound’s recitation is structured into intonation phrases (1) which in general, but not always, correspond to typographic lines. An example from ‘Cantico del Sole’ is shown in Figures 1 and 2, where there are two intonational phrases (1 and 2). In the most characteristic realization of the intonation phrase, there is a final low-high (or low-mid) contour, the L tone aligning with the head of the last prosodic word. The part of the t preceding the final rise is generally a sequence of level tone spans, rarely more than two. These generally decrease in pitch towards the final rise, suggesting a sequence of H tones. Each noninitial H-tone span is downstepped with respect to the immediately preceding one. The H-tone spans do not correspond in any obvious way with phonological phrases (φ), but downstep occurs on an accented syllable, e.g. in t in Fig. 1 on the first syllable of servant (second token) and peace. There seem to be no restrictions on the length of the initial H-tone span in t. Noninitial H-tone spans, on the other hand, are shorter, rarely spanning more than two prosodic word heads. The end of the last span in the sequence generally aligns with the penultimate prosodic word head in the t, and the interval between the penultimate H-bearing head and the final L-bearing head is interpolated (labeled <int> in Fig. 1). There are some variations on this basic pattern. For example, the final part of t may be another downstepped H-tone span (as in t). There may occasionally also be stretches of an t that do not form a H-tone span, and which instead evince pitch accents and interpolation between, as in nonchanted speech.

This study addresses two significant gaps. For one thing, little phonological work has been done on stylized intonation except the vocative chant (Liberman 1975), and so I aim to show that stylized intonation is worthy of investigation using the concepts and methods of phonological analysis. The second gap relates to the study of linguistic structure in poetics (Jakobson 1960; Kiparsky 1973). While metrical verse has received considerable attention from phonologists (e.g. Halle & Keyser 1966, et seqq. and inter alios), the same cannot be said for ‘free’ verse, where linguistic prosody does most of the work. This is partly due to the historical relegation of prosody to ‘mere performance’ in SPE (Chomsky & Halle 1968: 347), although this view is long since abandoned by phonologists working on prosodic structure and intonation. The significance of Pound’s recitation style for the study of linguistic poetics is that it represents an experiment in the use of a recurrent same other than metrical structure, and which cannot be inferred from the text—it resides in Pound’s performance alone. In addition to extending the coverage of AMP, this study also shows how linguistic poetics may bring new areas into its purview, while challenging the received stance that performance lies outside the domain of inquiry. In doing so, it feeds a growing interest from literary scholars in a poetry criticism informed by linguistic prosodic theory (e.g. Byers 1983; Gerber 2015).

Finally, chant also raises questions about the relationship between intonation and song, inviting consideration of the broader context of the emerging interest in language-music.

---

1 See Schuchard (2008) for the literary and cultural background of the use of chanting in poetry recitation.
2 See http://writing.upenn.edu/pennsound/.
connections (e.g. Pesetsky & Katz 2011), as well as the informational and affective meanings of chant as an evolved human signal (cf. Gussenhoven 2004; Mehr & Krasnow 2017).

**Examples**

Fig. 1: Cadences in ‘Cantico del Sole’

C1

H

H

H

int

u

C2

H

H

H

L

H%  

Fig. 2: Phonological representation of Fig. 1

**References**


French Sign Language (LSF) is built according to a prosodic structure just as spoken language (Blondel and Le Gac, 2007). But research on prosody in Sign Languages is fairly recent, especially regarding poetry. In this paper, I will present an experiment to study intonation groups in LSF poetry and identify the boundaries of these groups. Then I will submit an analysis as to how these boundaries are composed in LSF poetry and their function, and on the specificity of prosodic structure in a poetry context. Through this study, I intend to show how some aspects of LSF linguistics can be found in LSF poetry, especially in its intonation groups. (Catteau et al. 2016).

In works studying sign languages prosody, some authors consider that intonation is essentially marked by non-manual parameters (Sandler, 1999). For LSF, Blondel and Le Gac (2007) infer that duration, amplitude and speed of movement are intonation parameters, and that non-manual markers are not specific to signers (even though they are also intonation markers in Sign Languages), because they can be found in co-verbal gestures of hearing people. Furthermore, Boyes-Braem (1999) shows that labialization, pauses, hold, torso and hip rotation, as well as changes in facial expressions are involved in the structure of intonation boundaries in Swiss German Sign Language.

Based on this information about intonation construction in sign languages and on the model proposed by French specialists in vocal prosody (Vaissière, 2010; Delais-Roussarie, 2019), in my research I focus on defining an intonation group in LSF: a group of several signs whose opening and closing boundaries delimit sets that have contrastive properties. My hypotheses are that by studying LSF poetry one can find (i) prosodic structure parameters that are specific to LSF and (ii) parameters that are specific to sign languages poetry, like those described by Klima and Bellugi (1976) on between-sign movements.

To obtain movement information (duration, amplitude and speed), I decided to study motion captures of poetry data. In fact, motion capture can complement a simple analysis based on video and manual annotation, because it reveals precise information about the calculation of the degree of angular rotation of torso and hips, and the speed, acceleration or movement amplitude. This kind of methodology is thus relevant for studies on movement elongation (Tyrone et al. on American Sign Language, 2010), between-sign movement (Duarte, 2012, on LSF; Jantunen, 2013, on Finnish sign language) and emphasis (Tanaka and Van der Hulst, 2004).

For this study, I built a corpus with four deaf poets who interpreted their own poems. In addition to a basic camera, I used two different motion capture systems: (i) eleven poems were captured with the portative Noitom Perception Neuron 32 system (32 markers (12.5 mm x 13.1 mm x 4.3 mm and 1.2 g) each composed of a gyroscope, a magnetometer and an accelerometer); and (ii) five poems were captured with a Vicon system (composed of infrared lights, infrared camera and cutaneous hemispheric reflectors), thanks to a collaboration with the motion capture studio Mocaplab.

My first observations reveal that the degree of angular rotation of the torso and hips, and the speed, and amplitude of movements are the bases of intonation group boundaries in LSF poetry. When the torso and hips rotate, the signer can create several signing spaces by their
motions and these can be considered as different intonation groups. I consider that angular rotation can build an intonation group boundary. Moreover, amplitude and/or contrasts in speeds of movements between two successions of signs produce an emphasis effect. Using this, I infer that the changing of movement amplitude and speed forms an opening or a closing boundary of an intonation group.

I propose to analyze the peaks one can see on Figure 1 (generated by the motion capture data) as potential ways to detect boundaries. The same kind of curves can be generated with data regarding speed and amplitude of movements. My aim is to automatically detect boundaries of intonation groups with a combination of those three kinds of curves.

To conclude, according to the literature and the results, my hypothesis is that poems in LSF can be segmented in several successions of signs with opening and closing intonation boundaries that cause contrasts between two successive groups. I identify (i) those signs successions as intonation groups, and (ii) the angle rotation degree of torso and hips, speed and acceleration or movement amplitude as their boundaries’ components. Furthermore, I consider that those intonation groups segment the flow, organize it into groups and subgroups and, therefore, organize the information inside the poem.

Explaining the prosodic structure of LSF poetry helps building a better perception of the effects the poet wants to produce and the poem signification. This step is necessary to the next part of my research: understanding how to make a translation of an LSF poem into vocal French, especially at the prosodic level.

**Figure 1: Hips angle rotation of « Un Fruit », François Brajou’s poem.**

I propose to analyze the peaks one can see on Figure 1 (generated by the motion capture data) as potential ways to detect boundaries. The same kind of curves can be generated with data regarding speed and amplitude of movements. My aim is to automatically detect boundaries of intonation groups with a combination of those three kinds of curves.

To conclude, according to the literature and the results, my hypothesis is that poems in LSF can be segmented in several successions of signs with opening and closing intonation boundaries that cause contrasts between two successive groups. I identify (i) those signs successions as intonation groups, and (ii) the angle rotation degree of torso and hips, speed and acceleration or movement amplitude as their boundaries’ components. Furthermore, I consider that those intonation groups segment the flow, organize it into groups and subgroups and, therefore, organize the information inside the poem.

Explaining the prosodic structure of LSF poetry helps building a better perception of the effects the poet wants to produce and the poem signification. This step is necessary to the next part of my research: understanding how to make a translation of an LSF poem into vocal French, especially at the prosodic level.


How musical experience affects the role of vowel quality and stimulus duration played in tone perception by Mandarin listeners?

Si Chen¹, Yike Yang¹, Ratree Wayland², Yiqing Zhu²

¹ Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
² Department of Linguistics, University of Florida, FL, U. S.
sarahchen@polyu.edu.hk, yi-ke.yang@connect.polyu.hk, ratree@ufl.edu, yiqingzhu@ufl.edu

Pitch, the perceptual correlate of fundamental frequency (F0), plays a critical role in both music and speech. Musical experience may contribute to better identification of tones and sharper category boundaries in non-tonal (e.g. English) speakers (Alexander et al., 2005; Lee & Hung, 2008). However, it may not lead to differences in either overall or within-pair sensitivity to lexical tones for tonal (e.g. Mandarin) speakers (Zhao & Kuhl, 2015). Stimulus duration and vowel quality also plays a critical role in Categorical perception of pitch directions (Chen et al., 2017). Our study examines categorical perception of pitch stimuli modelled after Mandarin tones by 13 Mandarin musicians and 15 Mandarin non-musicians. We manipulated tones on both low and high vowels ([a] and [i]) produced by a male native Mandarin speaker to create 7-step, level-to-falling and level-to-rising pitch continua on both [a] and [i] vowels with 9 different durations (40 ms to 200 ms) using the pitch synchronous overlap add (PSOLA) method (Moulines & Laroche, 1995). In the identification task, all stimuli were presented in two blocks, one for falling pitch directions and one for rising pitch directions. In total, there were 1260 stimuli (5 repetitions * 7 steps * 9 duration * 2 vowels * 2 pitch directions). The order of presentation was randomized across participants.

For statistical analyses, we divided all stimuli into eight subgroups according to musical experience, pitch direction and vowel quality (FMMA, FMMI, RMMA, RMMI, FMNA, FMNI, RMNA, RMNI), where F and R stand for falling and rising pitch directions, MM and MN stand for Mandarin musicians and Mandarin non-musicians, and A and I stand for [a] and [i] vowels. Our results showed significant differences in the identification of pitch directions between Mandarin musicians and non-musicians for several stimulus duration values. Significant main effects of vowel quality and pitch directions as well as two-way and three-way interactions among vowel, musical experience and pitch directions were also found. The relationship between the sharpness of category boundary ($b_1$) and duration ($d$) is presented in Fig. 1 and Eq. (1) for Mandarin musicians to compare with the formula proposed by Chen et al. (2017) in Eq. (2) for Mandarin non-musicians.

Mandarin Musicians: $b_1 = 10.81 * d - 0.40$  (Eq. 1)
Mandarin Non-musicians: $b_1 = -50.65 * d^2 + 22.47 * d - 0.82$  (Eq. 2)

These two formulae suggest that sharpness of category boundary increased as stimulus duration increased for both Mandarin musicians and non-musicians. For Mandarin musicians, category boundary sharpness increased linearly, however, Mandarin non-musicians’ category boundary increased at a faster rate than musicians at first, but later slowed down as stimulus duration increases.
In addition, formulae (Eq.3-6) were also derived to capture the relationship between stimulus duration required to perceive 1 semitone (st) rise and fall by Mandarin musicians and compared them to those derived for Mandarin non-musicians by Chen et al. (2017). In all four formulae, \( t \) is the duration needed to perceive \( d \) st differences from level tones. To obtain these formulae, we fit linear mixed effects models to include a random effect of individuals.

Rising, Mandarin musicians: \( t = 62.85 + 10.09 \times d \) (Eq. 3)
Falling, Mandarin musicians: \( t = 42.81 + 10.39 \times d \) (Eq. 4)
Rising, Mandarin non-musicians: \( t = 66.76 + 8.39 \times d \) (Eq. 5)
Falling, Mandarin non-musicians: \( t = 99.45 + 5.29 \times d \) (Eq. 6)

These formulae revealed that in comparison to non-musicians, musicians require lesser time to perceive a pitch fall less than 12 st and a pitch rise less than 3st.

To explore the relationship between musicality, vowel quality and stimulus duration, we also derived formulae for the eight subgroups, and tested the intercept and slope values of Mandarin musicians and non-musicians. The results obtained indicate that the intercept and the slope values of the two groups did not significantly differ, suggesting that although Mandarin musicians may need shorter duration than Mandarin non-musicians to perceive a fall and a rise in many cases based on the formulae, the difference does not reach significance overall.

However, musicians did show differences in the time required to perceive a fall and a rise. Overall, they need more time to perceive a rise than a fall. However, unlike Mandarin non-musicians, Mandarin musicians’ sensitivity to both pitch directions varies as a function of vowel height. Specifically, they need a longer duration to perceive both the rising and the falling tones on a high vowel than on a low vowel when the rise is less than 10st and the fall is less than 13st. The results shed light on the effects from musical experience on categorical perception of tones, and its interaction between vowel quality and stimulus duration.

References


Towards a compositional analysis of intonational contours in questions

Elisabeth DELAIS-ROUSSARIE° & Claire BEYSSADE*
° UMR 6310-LLING, Université de Nantes  *UMR7023-SFL, Université Paris 8

Several approaches have been argued for to account for the relation between intonation and meaning, and more precisely to explain how meaning is conveyed by intonation. Phonological approaches, for instance, rely on the idea that a specific meaning is assigned to intonational units or tonal configurations (see, among others, Braun 2006 for a discussion on the meaning conveyed by specific pitch accents; and Pierrehumbert & Hirschberg 1990). Such approaches, however, cannot fully account for meaning differences observed. In French and German, for instance, rises at the end of the first prosodic phrase in the assertive sentences (1a) and (2a) are considered as similar at the phonological level as the rises at the end of comparable prosodic phrases in alternative questions ((1b) and (2b) respectively), but production experiments showed that rises are implemented differently in questions and in statements (see Michalski 2014 & 2017; Delais-Roussarie & Turco to appear), other phonological or phonetic differences being sometime also present in the prenuclear domain (see Petrone & Niebuhr 2014).

(1) a. Will Mone nachher zu Suse gehen, oder bei Narne bleiben (see Michalsky 2014)
   [ H*L  LH* H%/0%] [……]
   b. Will Mone nachher zu Suse gehen, kann ich nich bei Narne bleiben …)
   [ H*L  LH* H%/0%] [……]

(2) a. Il viendra demain ou la semaine prochaine. (uttered as a statement)
   [ (H*)  H* H-] [………]
   a. Il viendra demain ou la semaine prochaine. (uttered as an alternative questions)
   [ (H*)  H* H-] [………]

Following Gussenhoven 2004, Michalksy (2014 & 2017) analyses these differences as occurring at the phonetic level and considers that the meaning conveyed by intonation results from an interaction between abstract phonological forms to which a broad meaning is assigned (i.e. assertiveness, openness, etc.) and phonetic realization. To our mind, such an analysis does not account for the fact that the pragmatic value of an utterance is often gradient: there are, for instance, semantic and pragmatic differences between declarative polar question and interrogative polar question on the one hand, and confirmation request question and information-seeking question on the other, the syntactic form remaining identical. We will thus argue for a compositional approach of intonational meaning in which intonation contours result from the presence/absence of acoustic and phonetic cues, which can be considered as autosegment or element and as reflecting a precise semantico-pragmatic meaning.

Arguments in favour of such approach will be provided through an analysis of declarative and alternative questions. The rising contour occurring at the end of the first ip in (3a) and (4a) is implemented differently in terms of slope and pitch range than in (3b) and (4b), rises being more important in (3a) and (4a). The difference cannot result from a notion such as ‘interrogativity’ since all these sentences are questions.

(3) a. Grégoire va s’acheter un pantalon (H*H- + H) ou un gilet ?
   b. Est-ce que Grégoire va s’acheter un pantalon (H*H-) ou un gilet ?
(4) a. C’est un pantalon (H*H- + H) qu’il va acheter ?
b. Est-ce que c’est un pantalon (H* H-) qu’il va acheter ?

Following Beyssade & Marandin (2006), we thus consider that a speech act in context has a double impact: (i) speaker takes a public stance, and (ii) she/he expects an Addresssee reaction. Notions such as commitment and call-on-addresssee will be used to encode this double impact. Moreover, a division of labour, disentangling contributions of syntax (clause type), lexical semantics (particles), and prosody, come into play. By default, there is no difference between the content of speaker commitment and the content of the call on Addresssee. So no specific marking apart from the minor continuation (H*H- in F-ToBI) is expected at the end of the intermediate phrase in (3b) and (4b), the interrogative syntactic form indicating that the speaker is waiting for the addressee’s reaction. By contrast, in (3a) and (4a), the discrepancy between speaker commitment (with the use of declarative sentences) and the expectation toward the addressee forces to encode the call-on-addresssee by a rising element H. The latter has to be obligatory realized in addition to the ‘unmarked contour’ as shown in figure 1. It is this additional element that leads to the different phonetic implementation.

Figure 1: Pitch contours associated with (3a) and (3b) with the indication of the tonal events associated with the syllable [15]

This proposal allows making predictions on the intonational pattern expected in tag questions, and rhetorical questions. The latter are currently investigated and will also be presented.

Selected references
Narrow focus marking in monolingual and bilingual Spanish

Ingo Feldhausen\(^1\) and Maria Del Mar Vanrell\(^2\)

\(^1\)Goethe-Universität Frankfurt, \(^2\)Universitat de les Illes Balears

In this paper, we shed new light on the question of how narrow contrastive and information focus is realized in Central-Peninsular Spanish. We show that there is an important difference with respect to the use of both pitch accents and syntactic strategies for realizing focus between monolingual native speakers and German-Spanish bilingual speakers. Our five main findings:

(a) Bilingual speakers realize both types of focus almost always by stress shift, (1a), and the pitch accent is predominantly realized by L+H* (see section Bilinguals);

(b) Monolingual speakers, in turn, realize information focus by different strategies (cf. (1b), (2), and (3)), but stress shift is not a relevant option (see section No stress shift);

(c) Cleft constructions are used by monolinguals for both focus types even though there are certain preferences: (inverted) pseudo-clefts are favored for neutral focus, while simple clefts are preferred for contrastive focus (see section Cleft and focus type);

(d) Focus does not have to bear always sentential stress: in clefts, prosodic alignment can be a sufficient correlate of focus (see section Focus without sentential stress);

(e) Monolinguals typically realize the pitch accents by L+H* for non-final focused constituents and L* for final focused constituents.

We further argue that existing discrepancies between claims made in theoretical work on the one hand and in empirical work on the other can often be reduced to diatopic differences.

**Methodology:** We conducted a production test based on semi-spontaneous speech designed to elicit different focus readings (narrow informational and contrastive focus on the subject and (in)direct objects) by means of question-answer pairs from short picture stories. A total of 2508 contours were obtained (Monolinguals: 1848 = 24 short stories x 11 questions x 7 speakers; bilinguals: 660 = 12 x 11 x 5; all native speakers of Central-Peninsular Spanish).

**No stress shift:** There is an ongoing discussion on how focus is realized in Spanish. Theoretical work (such as Zubizarreta 1998, Gutiérrez-Bravo 2002) argues that neutrally focused elements must be located in sentence-final position (via p-movement, (1b)) in order to receive main stress by means of the Nuclear Stress Rule. Empirical studies, in turn, show that neutrally focused elements actually can be realized in situ (1a) and that this option reflects the predominant strategy for focus realization (e.g. Gabriel, Feldhausen & Pešková 2009, Gabriel 2010 for Argentinean Spanish; Muntendam 2013 for Andean Spanish; Hoot 2012 and Leal et al. 2018 for Mexican Spanish; Vanrell & Fernández-Soriano 2013 for Canarian Spanish). Our empirical results of the monolingual speakers (N=7) show that stress shift is not an option in Central-Peninsular Spanish and suggest that dialectal variation must be taken into account as a decisive factor involved in the variation of focus realization strategies.

**Cleft and focus type:** While the cleft constituent (such as Juan in (3)) is generally considered to be the contrastively focused element in Spanish (see, e.g., Zubizarreta 1998), Moreno Cabrera (1999: 4298f.) states that simple clefts on the one hand and (inverted) pseudo-clefts on the other hand have different information structural properties. Our study – as far as we know – represents the first empirical verification of this claim and confirms it; see Table 2.

**Focus without sentential stress:** It is generally accepted that focus in Spanish bears sentential stress (see, e.g., Ortiz-Lira 1994, Zubizarreta 1998 among many others). However, contrary to what has been claimed in the past, our results show the contrastively focused constituent in clefts such as (3a) does not always bear sentential stress – independently of the grammatical function of the clefted element. This experimentally corroborates the theoretical claim made by Moreno Cabrera (1999: 4298f.) who says that simple clefts in Spanish do not...
need a special intonational marking of the contrastive focus.

**Bilinguals:** We tested five early bilinguals who grew up and still live in Germany and who speak Central-Peninsular Spanish as a heritage language. The speakers show a clear preference for stress shift in both informational and contrastive focus, see Table 1 (in line with other studies on bilinguals, e.g. Leal et al. 2018). The realized focal pitch accent is almost always L+H*, but it is longer and more intense in contrastive contexts. Interestingly, the few instances of p-movement attested in the bilingual data occur with contrastive focus and not with information focus. Thus, the bilinguals clearly differ from the monolinguals. Future research will show whether the differences might be due to the influence of German (a language allowing for stress shift) or whether stress shift is a default strategy of bilinguals.

1. a. [f Los aLUMnos] se enfrentaron con la policia. ‘The students confronted the police’.
   b. Se enfrentaron con la policia [f los aLUMnos]. (Europ. Sp. /okLatinAmSp.)

2. [cf ManZAnas] compró Pedro (y no peras). Contrastive focus fronting
   ‘Pedro eats apples (and not pears).’

3. a. Es Juan el que viene. (Neutral focus)
   ‘It is Juan who comes.’
   b. El que viene es Juan. (Contrastive focus)
   c. Juan es el que viene. (Pseudo-clefts)

<table>
<thead>
<tr>
<th>Information focus</th>
<th>Bilinguals</th>
<th>Monolinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɾS]</td>
<td>Clefing 71.1%</td>
<td>Stress shift 77%</td>
</tr>
<tr>
<td></td>
<td>P-movement 14.5%</td>
<td>Clefing 18%</td>
</tr>
<tr>
<td>[ɾOoɔ]</td>
<td>P-movement 47.9%</td>
<td>Stress shift 83%</td>
</tr>
<tr>
<td></td>
<td>Clefing 23.3%</td>
<td>Clefing 15%</td>
</tr>
<tr>
<td>[ɾOoɔ]</td>
<td>Neutral WO 43.6%</td>
<td>Neutral WO 99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrastive focus</th>
<th>Bilinguals</th>
<th>Monolinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɾS]</td>
<td>Clefing 61.4%</td>
<td>Stress shift 72%</td>
</tr>
<tr>
<td></td>
<td>Focus fronting 15%</td>
<td>Clefing 23%</td>
</tr>
<tr>
<td>[ɾOoɔ]</td>
<td>Clefing 61.8%</td>
<td>Stress shift 63%</td>
</tr>
<tr>
<td></td>
<td>Focus fronting 23.6%</td>
<td>P-movement: 27%</td>
</tr>
<tr>
<td>[ɾOoɔ]</td>
<td>Clefing 41.2%</td>
<td>Neutral WO 87%</td>
</tr>
</tbody>
</table>

Table 1: Types and frequency of focus marking strategies in neutral focus (left panel) and contrastive focus (right panel) declaratives; types of clefts (see (3)) are not distinguished here.

<table>
<thead>
<tr>
<th></th>
<th>Neutral focus</th>
<th>Contrastive focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clefts 44.9%</td>
<td>Clefts 70.98%</td>
</tr>
<tr>
<td></td>
<td>Pseudo-clefts 13.4%</td>
<td>Pseudo-clefts 23.52%</td>
</tr>
<tr>
<td></td>
<td>Inverted pseudo-clefts 41.5%</td>
<td>Inverted pseudo-clefts 5.4%</td>
</tr>
</tbody>
</table>

Table 2: Types and frequency of cleft constructions attested in neutral and contrastive focus declaratives (monolingual speakers).

Pitch-to-segment Alignment in South Swedish and Mandarin Chinese: A Cross-language Comparison

Johan Frid\(^2\), Man Gao\(^1\), Malin Svensson Lundmark\(^2\) & Susanne Schötz\(^2\)
\(^2\)Lund University, \(^1\)Dalarna University

Studies of pitch-to-segment alignment provide robust evidence for the existence of stable alignment between prosodic unit and segment within a specific language. However, it has also been found that the alignment pattern is language-specific; for example, the segmental anchoring landmarks for f0 target of nuclear or pre-nuclear accent vary among languages or varieties of the same language \([1-4]\). Cross-linguistic comparisons of pitch-to-segment alignment have so far only been made among non-tonal (intonation) languages, but not between languages in which the prosodic events are used to signal different lexical meanings.

An attempt of comparing the alignment patterns in a word accent language (South Swedish) with those in a tone language (Mandarin Chinese) is made in this study, in order to further our understanding about the alignment between tonal event and segment string. Chinese uses five lexical tones to distinguish monosyllabic words, or to contrast meanings among otherwise homophonous syllables. Swedish uses pitch variations to differentiate certain pairs of words. In order to make the comparison possible and meaningful, disyllabic words in both languages are constructed to match in terms of segmental string, syllable structure (C1V1.C2V2C3) and tonal combinations. Phonologically, two possible tonal combinations of Chinese correspond to each of the two word accents in South Swedish. Take South Swedish Accent 1 (H*L) as an example, combination of H and L can be found in Chinese disyllabic words T1T3 (H-L) and T4T5 (HL-none). Since both tones in Swedish Accent 1 is realized within the first (stressed) syllable \([5]\), we hypothesize that the combination T4T5 (HL-none) may display similar alignment patterns as Accent 1 (H*L).

Simultaneous kinematic and acoustic data are collected from native speakers of South Swedish and Mandarin (two per each language) using ElectroMagnetic Articulograph (EMA) in the Lund University Humanities Lab. Target disyllabic words (ma.na(r)) were embedded in carrier sentences in both languages, and the informants were instructed to read each of the carrier sentence at normal speed eight to ten times. Temporal lags between the F0 peak/valley and the acoustic landmarks, and also between the F0 peak/valley and the algorithmically derived landmarks of consonant and vocalic gestures were computed and compared.

Preliminary analysis provides evidence in support of the hypothesis: the timing between f0 peak and C1 is similar between Swedish Accent 1 and the combination of T4T5 in Chinese. It is also observed that f0 peak maintains more consistent and synchronous alignment with articulatory landmarks on V1 than with acoustic landmarks in both languages. Findings of this study may possibly suggest that pitch-to-segment alignments are perhaps less varied between languages that use pitch to distinguish words, and further investigation seems warranted with full-scale study.


The intonation of vocative calls in Asturian
Eduardo García-Fernández
University of Massachusetts Amherst

Vocative marking in Asturian, a minority Ibero-Romance language spoken in Asturias (North Western Spain) is realized both intonationally and morphosyntactically. Intonationally they are realized with one of two intonation contours: L+H* (as in Figure 1) or H+L* (Figures 2 & 3). Morphosyntactically they can be marked by an optional phrase-initial particle a-, for example ¡A Manolo! (Manolo!). A previous production experiment confirmed that the implementation of the H+L* pitch accent is subject to phonological restrictions that affect the presence or absence of the particle. There is evidence that monosyllabic nouns/words with word-initial stress do not provide the necessary pre-tonic material that the leading H tone requires to align with (as in 2a below). In such cases, H+L* implementation is licensed by the addition of the particle a- to the vocative noun, which automatically provides the leading tone (H) with an anchoring location (2b). However, the particle a- can precede any vocative noun, regardless of stress placement (A Mari, A Marina) and is intrinsically associated with H+L* (therefore incompatible with L+H*).

When the noun has (at least) one pre-tonic syllable, H+L* may always be implemented without the aid of the particle a-, since the leading H tone is aligned with the pre-tonic syllable (example 2c). Even so, the presence of a- in these contexts is frequent (2d). In a previous experiment, the author found that, in an imitation task, participants failed to (re-)produce vocative calls represented in examples 1b, 1d, and 2a.

1. [L+H*]:
   a. Mari, …
   b. #A Mari, …
   c. Manolo, …
   d. #A Manolo, …

2. [H+L*]:
   a. #Mari, …
   b. A Mari, …
   c. Manolo, …
   d. A Manolo, …

Cross-linguistically, optionality in the presence of vocative particles allows for the encoding of pragmatic nuances about the degree of familiarity between speaker and addressee, as well as the formality of the utterance situation (Carvalho 2010; Hill 2007, 2014; Sonnenhauser & Noel Aziz Hanna 2013). The purpose of this study was to uncover the pragmatic restrictions that guide Asturian speakers in both tune choice and particle optionality. The fact that the occurrence of a- is inherently connected to H+L* raises the question as to whether this tune (in calls without a-) is used to convey pragmatic information typically encoded via vocative particles. Recent research shows that contour choice in isolated vocatives is also dependent on pragmatic meanings associated with differences in speaker-addressee relation (Borràs-Comes et. al, 2015; Hualde & Prieto 2015; Huttenlauch et. al. 2018). To explore Asturian speakers’ intuitions about the pragmatic use of vocative calls, 10 native speakers participated in an online survey that consisted of an open-response task. After listening to minimal pairs of utterance-initial vocative calls (based on the aforementioned phonological constraints), participants had to express their considerations about the difference in meaning between each pair. Results suggest that S(peakers) of Asturian exploit the variation in formal marking in vocative calls ([+/-particle] and [H+L*/L+H*]) to reflect different degrees of ‘callingness’ (seeking the attention of the A(ddressee)), represented in the following scale: a- > H+L* > L+H*. Participants’ intuitions indicate that there is a correlation between this scalar conception of the ‘calling’ meaning of the vocative and pragmatic nuances of S-A relationship. The presence of a- and the implementation of H+L* is perceived as informal in situations where S and A are on familiar terms. In turn, their use is not appropriate in formal contexts. On the other hand, speakers perceive the absence of a-
and the occurrence of L+H* as unspecified for formality. These intuitions derive from the pragmatic link between higher degrees of ‘callingness’ and an informal context of S-A familiarity. Thus, Asturian vocative calls not only serve the function of identifying A and getting her attention, but they also specify the intensity of the call by making use of the variability in formal marking. Furthermore, pragmatic nuances about S-A relationship arise when vocatives are used in context. This outcome is in keeping with recent accounts of a productive division of pragmatic labor under similar circumstances of variation in the marking of vocatives (e.g. Vanrell & Cabré 2011; Borràs-Comes et al, 2015; Hualde & Prieto 2015).

Figures 1, 2, and 3. Waveform, spectrogram, and F0 track of Asturian vocative calls ¡Manolo!, … (L+H* [left]; H+L* [center]) and ¡Á Monolo!, … (H+L* [right])

References
Two types of tonelessness: the case of Buli

Daniel Gleim, University of Leipzig [daniel.gleim@uni-leipzig.de]

Main claim: Opaque interactions of tone/stress and epenthesis have frequently been analysed by the ordering of operations, be it in a rule-based framework or in Harmonic Serialism. The doubly opaque pattern of epenthesis-tone spreading interaction in Buli (Gur, Akanlig-Pare & Kenstowicz 2002, Schwarz 2003,2005) poses a challenge for such a conception of opacity. I claim that while the inherent flexibility of ordering of operations in Harmonic Serialism can depose the central rule ordering paradox without making stipulative representational assumptions, it cannot account for different behaviour of epenthetic vowels on the one hand and underlying toneless vowels on the other. I show that the basic assumption of coloured containment – that epenthetic vowels are colourless, and that constraints may refer to the presence of colour – suffices to account for both types of opacity.

Data: Buli has three phonemic tones, L(ow) H(igh) and M(id); the tone bearing unit is the syllable. The most emblematic tonal process is Low Tone Spread (LTS), in which a low tone spreads one syllable to the right if it precedes a high tone. Mid tones are neither triggers nor targets in this process. If the syllable which L spreads to is followed by another high toned syllable (1-a), the targeted syllable surfaces as low. If it precedes something else – L or M tones, or nothing at all – it surfaces with a rising contour tone (1-b). All surface rises in Buli are derived by LTS.

(1) a. /ũ súî-kû/ → ũ sûkû ‘my catfish’ (Schwarz 2003, p. 06)
   b. /ŋôôŋkü têŋ/ → ŋôôŋkû têŋ ‘nahe der Eidechse’ (Schwarz 2005, p.106)

Epenthesis is employed to resyllabify coda consonants. The epenthetic vowel always shows up with a tone identical to the preceding tone. If it follows a high toned syllable affected by LTS, it has a low tone as well if it precedes another H (2-b), otherwise it gets a high tone (2-b). The tone on the epenthetic vowel is thus sensitive to the underlying preceding tone, the surface preceding tone and the following tone.

(2) a. /wâ nûr-mâ/ → wâ nurûma ‘his persons’ (Akanlig-Pare & Kenstowicz 2002, p.60)
   b. /wâ nûr/ → wâ nurú ‘his person’ (Akanlig-Pare & Kenstowicz 2002, p.60)

There are some underlying toneless affixes which also copy the tone from their left. Unlike the epenthetic vowels, they always surface with a high tone if preceded by an underlying high tone (3), no matter what tone follows them. The epenthetic vowel and the toneless affixes are therefore both tonally unspecified, yet their behaviour in LTS diverges.

(3) a. /wâ ģbâŋ-saʃ-à/ → wâ ģbâŋsanà ‘his book’ (Schwarz 2003, p. 07)
   b. /wâ vënta tiyè/ → wâ vëntûa tiyè ‘seine zweite Lüge’ (Schwarz 2005, p.124)

Assumptions: In order to derive these challenging patterns, I assume the concept of Morphological Colour (Revithiadou 2007, van Oostendorp 2006, Trommer 2011, Zimmermann 2017), which is a way of tracking morphological information through the phonology. Every morpheme has its unique colour (represented in the tableaux by optical colours). The phonology is not able to interfere with these colours. Elements inserted by phonology, like epenthetic vowels, are thus colourless (represented by grey). I assume that in a constraint-based framework constraints may be relativised to the presence of colour (cf. Zimmermann 2017). Analysis: The key to both the distribution of tone on the epenthetic vowels and the differences with the underlying toneless vowels lies in the colourlessness of the epenthetic vowel. In case of epenthesis, tones prefer to associate to a morphological, thus coloured, TBU (not violating $T \rightarrow \sigma$). But to be associated to an epenthetic vowel is still better than to be left a floating tone (violation of $T \rightarrow \sigma$). If the epenthetic vowel is followed by another high tone, the displaced high tone prefers to associate there, violating $T^r T$, a constraint against many-to-one association (4). If it is followed by something else, the high tone makes do with the epenthetic vowel, since a shift further to the right would violate high ranked constraints against falling tones or is altogether impossible (5).
The explanation for the different behaviour of underlying toneless affixes and epenthetic vowels follows accordingly: toneless affixes have a colour and are thus perfect targets for the displaced high tone. A violation of $^*T \rightarrow T$ is therefore unnecessary. The affix always shows up with the high tone, even if it could shift further to the right. **Alternative Analyses:** As Akanlig-Pare & Kenstowicz (2002) point out, the interaction of tone spreading and epenthesis is problematic for a standard rule-based framework, because it results in a rule-ordering paradox. They solve this problem with rather stipulative assumptions on the representation of Tone and TBUs, which is not necessary in my account. It has been argued extensively, that Harmonic Serialism offers a way to derive Stress-epenthesis opacities (Elfner 2016). It does indeed avoid the ordering paradox in Buli. However, HS cannot derive the different behaviour of underlying toneless suffixes: They are always predicted to pattern alike. **Summary:** Frameworks that base opaque tone-epenthesis interactions on ordering of operations have either problems with those interactions themselves (Rules) or with deriving the difference between epenthesis and underlying toneless affixes (HS). An account with morphological colour on the other hand can derive both patterns with simple assumptions: that epenthetic elements are colourless and that constraints may be relativised to the presence of colour.

The interplay of pragmatics and politeness in intonation

Stella Gryllia¹, Amalia Arvaniti², Mary Baltazani³

¹University of Leiden ²University of Kent ³University of Oxford

This paper investigates the role of pragmatics and politeness on intonation by examining the tunes used with Greek wh-questions [1, 2]. Briefly, [2] have shown that wh-questions are uttered either with a L*+H L-!H% tune, which is used to seek information, or with a L+H* L-L% tune, which in addition to the default questioning function can be used to turn a question into an implicit statement with negative connotations. Here this earlier analysis is tested with a large and variable corpus that also tests for effects of politeness, since [2] found that native Greek speakers consider L*+H L-!H% more polite than L+H* L-L%. The aim of the present study was to examine the effect that different pragmatic requirements have on the choice and phonetic realization of tune and explore how politeness considerations may affect both.

Nineteen speakers of Greek (10F) participated in a discourse completion task (DCT): they heard and read short scenarios ending in a wh-question. The DCTs were controlled for CONTEXT: scenarios led to the wh-questions being used either to request information or to make an implicit statement; see (1) and (2). The DCTs were also controlled for politeness, operationalized as POWER (inferior, equal, superior) and SOLIDARITY (solidary, non-solidary). (Degree of imposition was not included due to the nature of wh-questions.) There were 10 scenarios and questions per combination of factors, each produced 19 times (once per speaker), for a total of 2280 tokens. The questions were comparable in length and structure but not identical across conditions. Linear mixed effects models were run in R with CONTEXT, POWER, and SOLIDARITY as fixed effects, speaker and items as random factors, and a number of dependent variables (see below).

The results confirmed that CONTEXT (informational or non-informational) led to categorically different tunes in line with [2]: in informational contexts, (a) F0 started lower, (b) the accentual H was scaled lower and (c) aligned later, while (d) the scaling of the final boundary tone was higher; these results are consistent with a L*+H L-!H% tune being used in informational and L+H* L-L% in non-informational contexts. In addition, (e) the accented vowel of the wh-word was shorter and (f) the final vowel longer in informational than in non-informational contexts. A logistic regression with CONTEXT as the dependent variable and (a)-(f) as predictors showed that all variables except accentual H scaling (b) contributed to predicting the context in which a question was elicited (Fig. 1). In addition to categorical differences of tune type, there were significant gradient effects of politeness: differences in power and solidarity between speaker and addressee contributed to higher scaling of !H%, while non-solidary contexts also led to slower speaking rate and later alignment of the accentual H.

These results indicate that pragmatics (here, the context provided to speakers) affect tune choice. Phonetically, each tune was realized by a combination of parameters some of which cannot be predicted by the tune’s phonological representation or its associated metrical structure: this applied to duration differences in the accented and the final vowel. Politeness showed effects orthogonal to the tune (such as the speaking rate effect), but also affected the realization of the tune itself, namely the scaling and alignment of some tonal targets. Taken together, these results suggest that tunes and specific tonal events within tunes are realized by a constellation of phonetic parameters, some of which may be also affected by politeness. Overall, the results argue in favour of examining parameters beyond F0 in the study of intonation, and show that politeness must be taken into consideration as it may contribute to the phonetic variability of pitch contours.
(1) You return home from school hungry. You want to eat but only if dinner will be ready within half an hour, as you are going to the pool for swim practice and you don’t want to feel full. You see your mother boiling pasta and you ask:

[pote θa ˈine ˈetimi (i) makarɔˈnaða]
When will the pasta be ready?

(2) Whenever your dad does the grocery shopping, he gets very stingy and does not buy enough of anything. One day when you have a lot of friends over you run out of milk although you had warned your dad about running low. When a friend asks you for a double cappuccino, you tell your dad:

[meˈti na tu ˈftçakso ˈtora toŋ gapuˈtsino]
What will I make him a cappuccino with now?

![Figure 1. Plot of probability curves of all predictors in logistic regression; each curve indicates the probability that the event (non-informational context) occurs for each value of the predictor; boundary tone is treated as a nominal variable, where L%= 0 and !H% = 1.](image)


Perceptual evidence for the paralinguistic status of Mandarin interrogative intonation

Carlos Gussenhoven, Radboud University, Nijmegen, Netherlands

Prosodically encoded communication has widely been interpreted as based on distinct signal sources, one linguistic and one paralinguistic [8,3,6]. Support for this position from production and perception data has been elusive, but has been found in neuroimaging data [1,9]. I will present perception data from both identification and discrimination tasks for two Henan Mandarin lexical tone contrasts and an intonation contrast using monosyllabic stimuli. The results are consistent with a discrete interpretation of the tone contrasts and a gradient differentiation of the intonation contrast. The tone contrasts yielded sharper identification functions, while discrimination responses showed heightened sensitivity around identification switch points. Our interpretation was secured by two features in the experiments. First, stimuli within experimental sets either had rises or falls, so that direction of pitch movement cannot be held responsible for any of the differences found. Second, to ensure that phonetic forms signalling intonation differences and those signalling lexical differences did not intrinsically differ in identifiability or discriminability, we ran baseline experiments with 45 native Indonesian listeners, for whom all stimuli were largely equally discriminable and identifiable [4].

The Zhumadian variety of Henan Mandarin has four lexical tones which divide up into two rises (T2, T4) and two falls (T3, T1), each of which comes in a version with an early (T2, T3) and a late alignment (T4, T1) (where tone labels are based on etymological classes). Panel (a) in Fig. 1 shows the two lexical rises with interrogative and declarative intonation respectively, while panel (b) does the same for the lexical falls. This system was used to create two sets of four 7-step phonetic continua with the help of STRAIGHT [4], two of which represent a lexical contrast and two an intonational contrast, as shown in Fig. 2. Stimuli were evenly distributed over the 12 words in three tonal quadruplets so as to reduce the risk of boredom for listeners.

In the pitch discrimination experiment, stimuli were 72 pairs of step versions along the four continua which were 1, 2, 3 and 4 steps away for each of the pitch movement sets (falls or rises). Twenty-seven female and 13 male listeners (mean age 18.8, s.d 1.41) took part in this same or different task. Fig. 3 gives pooled results per step distance for word pairs (a) and intonation pairs (b). For the intonation pairs there is less discriminability towards the interrogative end of the continua. More importantly, for the word pairs there is an increase in sensitivity centrally in the continuum, in line with classic results for categorical perception [7]. The results of the word and intonation identification task, in which listeners were asked to choose one of two words and one of two intonations on each trial, are based on the scores for the lexical continua for which subjects chose the correct intonation as well as the scores for the intonation continua for which listeners chose the correct word (90.6% of the data). Identification scores are presented in panel (c) of Fig. 3 for the lexical and intonation continua separately. The Indonesian results for all continua closely match the Mandarin results for the intonation continua. The experiment shows that linguistic and paralinguistic f0 signals are cognitively distinct and that interrogative intonation can be paralinguistic.
Figure 1. Two lexical rises ((a); T3, T1) and two lexical falls ((b); T4, T2) with interrogative (?) and declarative intonation.

Figure 2. Two-by-two arrays of continua between the Early Rise and Late Rise (left) and the Early Fall and Late Fall (right) in declarative (!) and interrogative (?) intonations (after [2]).

Figure 3. Discrimination as a function of stimulus step (x-axis) for step sizes 1, 2, 3 and 4 separately for lexical continua (left) and intonation continua (centre). Identification as a function of stimulus step (x-axis) for lexical continua (blue line) and intonation continua (black line) (right). Zhumadian listeners.

The prosody of wh- questions in Fataluku

Tyler M. Heston
Payap University

This talk describes the intonational characteristics of wh- questions in Fataluku, an underdescribed Papuan language spoken in East Timor (ISO 639-3 ddg, IPA [fatáluku]). This work is based on first-hand fieldwork and situated within the autosegmental-metrical approach to intonational phonology (e.g., Ladd 2008). It is found that wh- questions differ prosodically from corresponding declaratives in terms of both their final boundary tone and their prosodic phrasing.

Fataluku’s intonational system shows little influence of stress or tone: rather, words are grouped into accentual phrases, each of which has a rising-falling f0 pattern peaking over the second mora of the phrase (Heston 2015). Past research has found that declaratives in isolation exhibit a short final syllable with a low f0, while questions have a much longer final syllable that bears a rising-falling f0 melody (Stoel 2007, Heston 2014); this work, however, did not address prosodic phrasing. Wh- question words may occur either in situ in Fataluku’s typical SOV structure (1), or with the wh- phrase fronted (2).

\[
\begin{align*}
\text{ā} & \quad \text{ina} & \text{fa’i?} & \quad \text{ina=’it} & \text{nami-moko} & \text{a} & \text{fa’i?} \\
(1) \quad \text{you} & \quad \text{what} & \quad \text{do} & \quad (2) \quad \text{what=FOC} & \text{male-child} & \text{SBJ} & \text{do} \\
\text{‘What did you do?’} & \quad \text{‘What did the boy do?’}
\end{align*}
\]

In this study, six native speakers of Fataluku (5M, 1F) were grouped into pairs, and asked to role play 19 pairs of wh- questions and answers. Sentences with various types of question words were collected, along with corresponding declaratives in broad focus. Four repetitions were collected, with speakers alternating roles. Utterances were transcribed by the author, who used the characteristic AP-initial rise identified by Heston (2015) to identify AP boundaries.

The intonational boundary tones identified here align in general with the results of Stoel (2007) and Heston (2014). As Stoel and Heston found, declaratives exhibit a short final syllable with a low f0, while wh- questions have a substantially longer final syllable. The final syllable typically exhibits a rising-falling melody; however, the final syllable of some wh- questions has a simple rise, or a rise with a small fall. I tentatively analyze these examples as truncated instances of the expected rising-falling boundary tone, though more research is needed to determine whether they realize an underlyingly distinct boundary tone.

The prosodic phrasing of wh- questions follows that of declaratives in broad outline, though wh- questions typically exhibit fewer prosodic boundaries than the corresponding declaratives. In the declaratives collected here, a subject generally forms an AP by itself, while a direct object forms an AP with the verb. The behavior of oblique-like elements varies; typically, each forms a separate AP, though they are sometimes included in the verb phrase AP. Wh- questions show two substantial deviations from the prosodic structure of the corresponding declaratives. In the first place, a fronted wh- phrase forms a separate AP. Wh- questions also frequently show dephrasing, whether the wh- word is in situ or fronted. The expected boundary after the subject is often absent, as well as the boundaries adjacent to the underlying position of the wh- word. Figures 1 and 2 illustrate both of these prosodic characteristics of wh- questions. Figure 1 shows a typical declarative utterance. In the corresponding wh- question, the only AP phrase boundary is the one immediately following the fronted wh- word. No AP boundary occurs after the subject ā ‘you’ or between the locational phrase ira na’en ‘at the water’ and the main verb phrase.

To sum up, wh- questions in Fataluku differ prosodically from corresponding declaratives by the presence of a lengthened final syllable bearing a rising-falling (or simply rising) f0 pattern, and by prosodic dephrasing. A fronted wh- phrase forms a distinct AP, but the expected phrase boundaries after the subject and adjacent to the underlying position of the wh- phrase are often absent. Future work is needed on best analysis of the apparently truncated final boundary tones, as well as on the relationship between the prosodic phrasing of wh- questions and focus intonation.

---

1 The clitic =it(u) ‘foc’ which appears on fronted wh- words is homophonous with the relativizer and may also be used to front non-wh- noun phrases.
Figure 1: The declarative *ana* *ira* *wêr* *i* *na*‘*en* *aficalafur* *aci* ‘I see the crocodile at the water’. Note the four H tones, which occur on the second mora of each of the four APs.

Figure 2: The question *ina‘it* *â* *ira* *na‘en* *aci* ‘What did you see at the water?’

References


Tonal neutralization of function words across Chinese dialects

Tzuchuan Huang
National Chengchi University

**Introduction:** This paper addresses from a typological perspective the types of Chinese function words that are (un)likely to have a neutral tone. In contrast to the previous studies — where homogeneity or language-specific variation among function words is assumed for tonal neutralization — this paper contends in cross-linguistic terms that function words are triply differentiated with respect to the likelihood of being neutral-toned, and that this ternary differentiation can be formally defined by syntactic categorization.

**Survey:** In the present study, seven Chinese dialects (comprising five branches of Sinitic languages) were surveyed regarding tonal neutralization of function words. A preliminary observation (1) is that function words can be categorized into three groups according to the proportion of being neutral-toned: (i) sentence-final particle (SFP) and topic marker (Top), termed Fa, which unexceptionally have a neutral tone, (ii) attributive marker (Atr), termed Fb, which is neutral-toned in most of the cases, and (iii) the remaining types (e.g. aspect marker/Asp, classifier/CL and pronoun/Pron), termed Fc, which are neutral-toned variably (represented by △) and restrictively. This classification clearly illustrates *implicational universals* — every dialect that has neutral-toned Fc has neutral-toned Fb, and every dialect that has neutral-toned Fb has neutral-toned Fa.

(1) Neutral-toned types of function words

<table>
<thead>
<tr>
<th>Dialects</th>
<th>Fnc types</th>
<th>Fa</th>
<th>Fb</th>
<th>Fc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SFT</td>
<td>Top</td>
<td>Atr</td>
<td>Asp</td>
</tr>
<tr>
<td>Shanghai</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ningbo</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>△</td>
</tr>
<tr>
<td>Raoping Hakka</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>△</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>✓</td>
<td>✓</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Sixian Hakka</td>
<td>✓</td>
<td>✓</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Cantonese</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Proposal:** Following Huang et al.’s (2009) categorial features for Mandarin Chinese (cf. Grimshaw 1991, 2000) — where noun (N), verb (V) and adjective (A) each have a set of related functional categories distinguished through incremental values of the feature [F] — the ternary classification of function words across the dialects is argued to be formally defined as follows (2): Fa is the highest V-related functional projection “clause typer” (CT), Fb the A-related functional projection, and Fc lower V-related functional projections plus all the N-related functional projections.
With those function words mapped hierarchically and categorically to the syntactic tree, the central proposal is attained: a function word that has higher projection is more prone to tonal neutralization. This idea is recast in an Optimality-Theoretic analysis, fleshed out through the interaction between Selkirk’s (2011) MATCH constraints and a currently proposed fixed constraint ranking *PARSE(Fa, PrW) >> *PARSE(Fb, PrW) >> *PARSE(Fc, PrW), which penalizes a function word that is higher in the syntactic hierarchy to form a prosodic word of its own (i.e. extrametricality of function words, see Selkirk 1996). The permutation of the given constraint set enables this paper to account for the typological variation of neutral tone across Chinese dialects.

References:
Match Theory and the Asymmetry Problem: Intonational phrase marking in Stockholm Swedish
Shinichiro Ishihara and Sara Myrberg (Lund University)

Introduction: This paper presents an analysis of the syntax-prosody mapping of intonational phrases (IPs) in Stockholm Swedish (SSw), which solves an apparent problem of Match Theory (MT, Selkirk 2011). MT claims that the syntax-prosody mapping be node (constituent) correspondence, via constraints of the type in (1). Thus, MT rejects the partial boundary correspondence assumed in Alignment Theory (AT, Selkirk 1986, McCarthy & Prince 1993, Selkirk 1996 inter alia).

(1) a. MATCH(CP,i) — MATCH-SP: A syntactic CP is mapped to prosody as an IP.
   b. MATCH(i,CP) — MATCH-PS: A prosodic IP is mapped to syntax as a CP.

The Asymmetry Problem: The current study concerns one of the crucial differences between the syntax-prosody mapping constraints in AT and those in MT, namely, the (un)availability of asymmetric ranking of L- and R-edge mapping constraints. In AT, mapping constraints (ALIGN-XP) are defined for L-edges and R-edges independently. L- and R-edge mapping constraints can therefore be ranked differently in relation to other constraints (asymmetric ranking, e.g. ALIGN-L >> X >> ALIGN-R). In MT, in contrast, L- and R-edges of prosodic constituents are obligatorily mapped to/from edges of syntactic constituents in a symmetric fashion. Under this theory, left and right edges are equally important. While MT is conceptually advantageous to AT due to its restrictiveness, MT faces empirical problems where data suggests asymmetric ranking of L- and R-edges.

IPs in Stockholm Swedish: SSw presents a potential problem for MT. SSw has tonal phenomena that indicate both left and right edges of IPs (Myrberg 2010, 2013, Myrberg & Riad 2015). A nuclear pitch accent plus a boundary tone indicates the right edge of an intonational phrase (IP), and the initiality accent (IA) marks the left edge of an IP (Roll et al. 2009, Myrberg 2013, Myrberg & Riad 2015). Data in Myrberg (2010, 2013) suggests that the syntactic structures in (1) exhibit different phrasing options, as shown in (2). A sentence with an sentence-initial subordinate CP (1a) can be prosodically realized with two coordinated IPs under one higher IP-node (phrasing pattern iii). Such a phrasing is not available for a sentence with a sentence-final subordinate CP (1b). This paper will present additional data, which confirms the phrasing difference between (2a) and (2b).

(2) a. Sentence-initial subordinate CP: [[ . . . ]CP . . . ]CP
   if zebras came closer then would Ida be able to touch at them.
   prosecutor the claimed that driver had not done anything wrong

Myrberg (2010) accounts for the phrasing difference by strictly ranking Align(i,CP)-L lower than other crucial constraints, whereas Align(CP,i)-R is freely ranked with respect to these constraints.

A Match-based Account: We argue that the difference in phrasing between (2a) and (2b) can be accounted for with MT, if the phrasing difference is described in terms of the location of heads of prosodic constituents (i.e. prosodic prominences), instead of mapping of edges. Prosodic constituents are either L- or R-headed. In our account, the distribution of the head is regulated by ALIGNHEAD-L/R given in (4) (cf. EDGEMOST, Prom-π, Edge-π...
Selkirk 2011:471). Note that ALIGNHead is not a syntax-prosody interface constraint, but a prosodic wellformedness constraint (PWC), as it requires a correspondence between two prosodic objects (prosodic heads and boundaries). In other words, we attribute the asymmetry problem to prosodic wellformedness, rather than to syntax–prosody mapping. In addition to ALIGNHead, *PHEAD (a *Struc constraint) is used to ban insertion of prosodic heads, and EQUALSISTERS bans sisterhood of unequal prosodic constituents.

(4) a. ALIGN(l, L/R, t-Head, L/R) — ALIGNHead-L/R [PWC]: Align the left/right boundary of every intonation phrase with its head. (Truckenbrodt 1995:119, Féry 2013:696, a.o.)
   b. *PHEAD(t) [PWC]: Avoid heads of the IP (t).
   c. EQUALSISTERS [PWC]: Sister nodes in prosodic structure are instantiations of the same prosodic category. (Myrberg 2013)

(5) a. ALIGNHead-R >> *PHEAD, MATCH-SP, MATCH-PS, EQUALSISTERS
   b. *PHEAD >> EQUALSISTERS

Free ranking of MATCH-SP, MATCH-PS, and EqSis results in phrasing optionality (cf. Myrberg 2010, 2013), and two crucial strict rankings are given in (6). (a) ensures that there are no headless IPs. (b) excludes the unwanted and unavailable phrasing for the sentence-final subordinate clause in (1b). The Tableaux in (6) show the most crucial ranking of the constraints in (1) and (4), which produces the phrasing pattern from (3iii), candidate c, for (2a), while (2b) yields a different prosodic output (*’s indicate the location of IP heads). Assuming the strict rankings in (5), no ranking will select candidate c for the final subordinate clause (2b). The basic generalization is that a prosodic constituent is less likely to be added when it triggers addition of a prosodic head. Thus, the addition of an IP is allowed only when it does not change the number of prosodic heads.

(6) MATCH-SP >> *PHEAD >> EqSis >> Match-PS

<table>
<thead>
<tr>
<th>Input: [...]</th>
<th>ALHD-R</th>
<th>MA-SP</th>
<th>*PHD</th>
<th>EqSis</th>
<th>MA-PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. {{ * } { }}</td>
<td></td>
<td></td>
<td>++</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. {{ * } { }}</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. {{ * } { * } { }}</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. {{ * } { * } { }}</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: [...]</th>
<th>ALHD-R</th>
<th>MA-SP</th>
<th>*PHD</th>
<th>EqSis</th>
<th>MA-PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. {{ * } { }}</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. {{ * } { }}</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. {{ * } { * } { }}</td>
<td></td>
<td></td>
<td>**!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. {{ * } { * } { }}</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conclusion:** The proposed analysis solves an apparent problem for Match Theory by reconsidering the effects of PWCs in relation to mapping constraints. It corroborates the basic idea of MT, namely, that any type of non-isomorphism between syntax and prosody should be derived by prosodic wellformedness constraints, or other interface constraints (e.g., information structure-related constraints), but never by syntax-prosody interference constraints. If our analysis is on the right track, it may even be hypothesized that Match constraints are the sole syntax–prosody interface constraints, and that no other constraints are needed which refer to syntactic categories (i.e., no constraints like Align-XP, Wrap-XP and Stress-XP).

Not all North Germanic tonal accent systems are privative: A closer look at Bergen (Norway) and Malung (Sweden)

Gjert Kristoffersen,
University of Bergen

At least since the 1990s, it has been claimed that the tonal accent distinction in Swedish and Norwegian can be characterized by privativity in the sense that Accent 2 consists of an initial tone that accent 1 lacks. Otherwise, the two melodies are identical. Based on the assumption that the accent contrast is specified tonally in the lexicon by the absence vs. presence of this initial tone, this initial tone is commonly referred to as a lexical tone (see e.g. Kristoffersen 2000: 253; Lorentz 1995; Riad 2014: 189).

The hypothesis itself has come to be known as the privativity hypothesis, and in its strongest form, it claims that all Swedish and Norwegian tonal dialects will belong to one of two major types, those where the lexical tone is a high tone and those where it is low (Riad 2014: 189). This entails that the phonological structure is the same across all dialects: A lexical H/L which associates with the stressed syllable in Accent 2 words, followed by a tone of the opposite value often referred to as the prominence tone. From this, a very simple and elegant basic typology can be derived: All tonal North Germanic dialects share the same basic system, which depending on the value of the initial, lexical Accent 2 tone can be divided into only two types.

According to Lorentz (1995), the dialect spoken in the city of Bergen in western Norway belongs to the type where the lexical tone is L. Thus the accent 2 melody is L*HL, while the accent 1 melody is H*L. Based on instrumental analysis of recordings made of three Bergen speakers, Kristoffersen (2006) argued that the presence of the initial L in accent 2 is not as clear as Lorentz seemed to assume in his 1995 paper. A proposal to the same effect is made by Hognestad (2012: 136 f.) regarding two dialects spoken not far south of Bergen, Tysnes and Stord.

Kristoffersen’s claim was based on a limited set of data, extracted from scripted recordings of high school students reading a varied set of target words in a limited set of carrier sentences. My talk will be a follow-up of Kristoffersen’s paper, where all the relevant data from nine
Bergen speakers are analyzed. I shall also analyze recorded speech data from Malung in Dalarna in Sweden. The outcomes for both dialects show that there seems to be scant evidence for an initial lexical L* in Accent 2. In utterance initial position, when followed by one or more unaccented, unstressed syllables, the Accent 2 contour in most cases shows a rise, not a dip or a level tone through the early part of the stressed syllable. In utterance medial position, when preceded by another accentual domain, the final L of the preceding accentual melody explains the presence of a trough before the rise towards the Accent 2 peak. Also here the rise in many cases starts before the onset of the stressed vowel. This again suggest interpolation, now from the final L of the preceding accent melody. Given these observations, the cases where the rise starts early in the stressed vowel can be more parsimoniously explained by optional rightwards spreading of the preceding L rather than by a presence of an additional and from an analytical point of view, superfluous L as part of the Accent 2 melody.

Consequently, contra e.g. Riad (2014: 189) the privativity hypothesis and the typology that can be derived from it, does not seem to hold when tested against the full range of tonal dialects. A richer typology, or an alternative one, is therefore called for.

Note that my claim is not that L*HL dialects do not exist. One plausible case appears to be the North-West Norwegian Sunnmøre variety analyzed in Abrahamsen (2003), another the South-West Norwegian Egersund variety analyzed in Hognestad (1997); (2012: 108–11).

For Bergen and Malung I argue that the accentual contrast in both dialects instead should be analyzed as a difference in timing of the same H*L contour, early in accent 1 and late in accent 2, that is, without an initial lexical L* being part of the representation, as assumed by the privativity hypothesis.

References


Cross-linguistic differences in the effects of distal prosody on speech segmentation

Carolyn Kroger\textsuperscript{a}, Ni-La Lê\textsuperscript{b}, and J. Devin McAuley\textsuperscript{a,c}

\textsuperscript{a}Department of Psychology, \textsuperscript{b}Department of Linguistics and Languages, \textsuperscript{c}Neuroscience Program (Michigan State University)

Research has found that, in addition to local prosodic cues occurring at or adjacent to word boundaries, prosodic temporal and pitch patterning in the distal (nonlocal) context also affects segmentation of lexical items. Here, we used an artificial language paradigm to examine the role of distal prosody in novel word segmentation for native English speakers and to consider how universal this effect is by testing non-native English speakers.

**Background:** Infants acquiring native language and adults learning a second language must map portions of continuous speech into meaningful units (e.g., words). Because speech does not contain reliable pauses between words, listeners use statistical cues (e.g., transitional probabilities) and prosodic patterns (e.g., intonational phrasing, stress) to segment words. While previous work focused primarily on local cues, there is evidence that distal prosodic patterns at the beginning of an utterance influence downstream word segmentation and lexical access (Dilley and McAuley, 2008; Morrill et al., 2015).

**Hypothesis:** The perceptual grouping hypothesis developed by Dilley and McAuley (2008) proposes that distal prosodic patterns create expectations about the grouping of later syllables into words. This hypothesis predicts disyllabic items congruent with the prosodic pattern at the beginning of an utterance are more likely to be perceived as words than incongruent disyllabic items. This congruency effect may not be universal, however, given previous findings on language specificity in speech perception (e.g., Tyler & Cutler, 2009; Yeung et al., 2012).

**Methods:** Adult participants initially heard a list of 12 disyllabic (target) words in a made-up language (Table 1). They then reported how well (on a six-point scale) they heard target words that were either present or absent in short utterances. Target words were either congruent or incongruent with the distal prosodic context (Figure 1). Testing non-native English speakers is ongoing.

**Results:** Participants distinguished between target word present and absent conditions. Signal detection analyses showed that native English speakers (N=40) better perceived target words in congruent distal prosodic contexts than in incongruent contexts. In contrast, non-native English speakers (N=24) show a smaller congruency effect (Figure 2). A subcategorization of non-native speakers revealed that speakers of tonal languages (Mandarin, Cantonese, Vietnamese, and Thai; N=12) effectively recognized the target words even in the incongruent condition while the non-tonal group (N=12) showed a pattern similar to native English speakers (Figure 3).

**Discussion:** Consistent with previous research, native English speakers use distal prosodic cues to group syllables into words. Effects of distal prosody were also present for speakers of other non-tonal languages, but not for native speakers of tonal languages. One possibility is that speakers of languages with lexical tones use different segmentation units from native speakers of non-tonal languages. In languages with lexical tones, (most) syllables carry tones and have lexical meanings, making syllables (not words) fundamental segmentation units, removing the need to perceptually group syllables. A related possibility is that speakers of tonal languages do group syllables, but emphasize local rather than distal cues for speech segmentation.
Table 1. Target words for language 1 and 2

<table>
<thead>
<tr>
<th>Language 1</th>
<th>dika</th>
<th>tepi</th>
<th>guba</th>
<th>duta</th>
<th>kobu</th>
<th>gope</th>
<th>kagu</th>
<th>tago</th>
<th>pedu</th>
<th>badi</th>
<th>bute</th>
<th>piko</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language 2</td>
<td>kadi</td>
<td>pite</td>
<td>bagu</td>
<td>tado</td>
<td>buko</td>
<td>pego</td>
<td>guka</td>
<td>gota</td>
<td>dupe</td>
<td>diba</td>
<td>tebu</td>
<td>kopi</td>
</tr>
</tbody>
</table>

Figure 1. Example stimuli illustrating distal prosody manipulation. Target words were either congruent (a) or incongruent (b) with the distal prosodic context (Morrill et al., 2015)

Figure 2. Empirical Receiver Operating Characteristic (ROC) curves for congruent (blue) and incongruent (black) target words for native English speakers (left) and non-native English speakers (right)

Figure 3. Empirical ROC curves for congruent (blue) and incongruent target words (black) for non-native English speakers: speakers of tonal languages (left) and speakers of non-tonal languages (right)

References


Origins of tone: aspiration and glottalization across Udihe

Natalia Kuznetsova & Elena Perekhvalskaya
(University of Turin and Institute for Linguistic Studies RAS & Institute for Linguistic Studies RAS)

The phonological status of vowel aspiration and glottalization in Udihe (a moribund Tungusic language in the Far East of Russia) has long sparked controversy among researchers. These phenomena have been considered as consonants, vocalic features, accents, or tones (Šneider 1936, Zinder 1948, Sunik 1968, Simonov 1988, Radčenko 1988, Kormušin 1998, Janhunen 1998, Nikolaeva and Tolskaya 2001). One of the reasons for this disagreement has been a dialectal heterogeneity of the language data used by researchers.

In 2007 and 2010, we undertook a comparative field research of Udihe dialectal phonetics and phonology with the very last speakers (about 45 hours of recordings collected). The questionnaire included about 1000 phrases for elicitation which contained data on syntagmatic and paradigmatic features of all Udihe phonemes and lexical suprasegmental features, as well as on known dialectal phonological isoglosses. The study showed that Udihe dialects form a continuum with no clear border with an adjacent Oroch language (Perekhvalskaya 2010). The evolution of aspiration and glottalization across the Udihe-Oroch dialectal continuum is attested from the north to the south of the area (Oroch language — Koppi dialect, a transitory Oroch/Udihe variety — Khor Udihe dialect — Bikin/Samarga Udihe dialect — Iman Udihe dialect) as follows:

Aspiration:

\[ V - s - V > V - h - V > V^hV > ^hV > ^V \]

Glottalization:

\[ V - q - V > V - ñ - V > V^ñV > ^ñV > ^ñV \]

Starting as consonants in the north, aspiration and glottalization turn first into vocalic and then syllabic prosodic features. Eventually, in the southernmost dialects at the border with Chinese, aspiration disappears and glottalization becomes a falling melody (cf. Janhunen 1998 on a possible Chinese influence on tonogenesis in Udihe). This process follows a cross-linguistically frequent evolutionary path of the transformation of consonants into a glottal stop, then into a laryngealized prosodic feature, and later into tone (for example, rising or falling, viz. a typological survey in Ivanov 1975).

The synchronic phonological status of these features presents an intriguing case especially at the last two aforementioned stages. The saliency of the theoretical notion of pitch-accent is in itself a topic of debates within functional phonology (Hyman 2009, van der Hulst 2011). However, pitch-accent is not the only type of uncanonical word-prosodic units found in the world languages, cf. e.g. a prosodic glottalization in Danish (stød) or a prosodic quantity in Estonian (Grønnun et al. 2013, Eek 1986). Such uncanonical cases challenge a traditional word-prosodic typology, which operates with only two main units, tone and stress (Kuznetsova, 2018).

We will address various phonetic and functional parameters of Udihe glottalization and pharyngealization and discuss to which extent they behave as (1) consonantal segments, (2) vocalic features and (3) word-prosodic units in each of the dialects studied. Their relation to the metrical stress in Udihe (and, in general, the relevance of the metrical stress notion to the Udihe phonology) will be also covered. We will show that at the last two stages of the aforementioned process pharyngealization and laryngealization are rather word-prosodic than segmental features, and at the same time they cannot be defined either as tone or as stress within any mainstream word-prosodic classification (e.g. Hyman 2006). Classical tonal system, such as in Chinese, would be the next developmental stage of this evolution.
References


Introduction

Questions in African languages show uncommon prosodic status such as lax question prosody in languages spoken in the Sudanic belt (Rialland 2007, 2009). Even so, the question prosody of African languages remains an understudied area. While some recent efforts, e.g. Downing & Rialland 2016 among others, improved our understanding in this area, we lack more detailed and nuanced view, in particular, concerning the prosodic behaviour of question elements (QEs), i.e. wh-words and question particles. The current study aims to address this gap by reporting on the prosody of questions elements in two Bantu languages: Shingazidja (G44a, Comoros) and Xitsonga (S53, South Africa).

Prosody of wh-words

In Shingazidja and Xitsonga, the syntactic behaviors of wh-words differ from one another. Even so, the prosody of these wh-words display similar characteristics.

In Shingazidja, the default position for a wh-word is the final position, as in other Bantu languages (e.g. Ekoti – Schadeberg & Mucanheia 2000; Nkore-Kiga – Taylor 1985: 139; Ha – Harjula 2004: 150). A wh-word in final position shows no association to prominence, whether it appears in the in-situ (1a) or ex-situ (1c) position. In the initial position, however the wh-word is associated with a sharp boost of the F0 in both cases (1b, 1d).

(1) a. ha-réme hindí 1.PFV-hit what ‘What did he hit?’
   b. hi↑ndí ha-réme what 1.PFV-hit ‘What is it that he hit?’
   c. ya-réma 'ndó 1.REL.PFV-hit who ‘Who hit?’
   d. ↑ndó ya-réma who 1.REL.PFV-hit ‘Who is it that hit?’

In Xitsonga, a wh-word appears in the in-situ position (2a), rather than the final position. In the initial ex-situ position, the wh-word displays an F0 boost (2b) as in Shingazidja.

(2) a. mi βón-é jíní élóndon you see-PFV what in.London ‘What did you see in London?’
   b. ī↑jíní mi βón-é élóndon TOP what you see-PFV in.London ‘What is it that you saw in London?’

Question particles

Previous studies on other Bantu languages such as Tswana (Zerbian 2016) report that the presence of a question element may not have any effect on the prosody of question. However, in Shingazidja and Xitsonga we observe some prosodic effects of question particles.
In Shingazidja, a question can be introduced by the question particle yé, with some inter-speaker variation on its obligatoriness. The QP yé is followed by a PPh boundary. In Shingazidja, a Yes-No question is obligatorily signaled by the presence of a superhigh tone that emerges on the penult, or the antepenult when the final syllable is high-toned (3) (author2 2017; see Figure 1 for pitch tracks).

(3) a. ŋɡo-fikir’ uká hā-li’mí
   2SG.IPFV-think.PRS that 1.PFV-cultivate
   ‘Do-you think he cultivated?’

In (3b), the tone of the particle shifts to the following syllable due to a postlexical constraint against the alignment of a tone to the left boundary of an IP. The shift of the tone is absent from some of the dialects in the language (the southern varieties), and it may lead to the creation of a tone plateau when a reduced number of syllables separate the tone of the particle from the superhigh tone. The QP can also be placed before a verb or the wh-word, as in the final pitch track of Figure 1, adding a focus to the following word, whose F0 is then raised.

In Xitsonga, a question particle xàná that appears both in yes-no questions as well as wh-questions. Speakers judge that xàná is optional, but the distribution of it is not yet fully known: a question can begin or end with xàná, also xàná appears at a constituent boundary in the middle of an interrogative sentence. While xàná shows prosodic patterns akin to in-situ wh-words in non-initial positions, the particle xàná has the prosody of an ex-situ wh-words (with an F0 boost) in the initial position.

[Figure 1. The left is (3a), the center is (3b), and the right has QP yé before the final verb in Shingazidja.]

**Conclusion** We have shown that QP yé in Shingazidja has prosodically complementary distribution with wh-words, while QP xàná in Xitsonga shares its prosodic distribution with wh-words. We propose that the initial position of a question in these two languages is prosodically prominent (F0 boost), but this requirement is violable as shown by QP yé in Shingazidja. We interpret this distribution as prosodic non-uniformity of QPs.
Tonal truncation and compression in Mandarin Chinese: effects of duration and phonological length

Yang Li, University of Cambridge, yangli.linguistics@gmail.com
Yixin Zhang, University of Cambridge, yz510@cam.ac.uk

Previous research has identified two main strategies of coping with durational constraint on F0 realisation: compression and truncation: the former an increase in F0 rates of change, allowing the range of F0 movement to be maintained in a shorter time; the latter typically an undershooting of F0 target without lowering F0 velocity (e.g. Grabe, Post, Nolan, & Farrar, 2000). Addressing the paucity of studies targeting tonal languages on this issue, this paper examined the impact of time constraints, in both the physical and the structural sense (phonological length), on the realisation of rising T2 and falling T4 in Mandarin Chinese.

Methodology: 10 speakers from Northern China were recorded repeating 18 familiar monosyllabic characters embedded in a carrier phrase or in isolation. These syllables vary in 3 levels of Length (CV, CVX and CVVN) and 2 Tones (T2, T4). Speaker were instructed to read the phrase at a normal and a fast speech rate, resulting in 3 levels of Rates (citation, slow, fast). Duration and full F0 curves were extracted from the target syllables for analysis.

Results: Rising T2 and falling T4 in Mandarin respond to durational pressure in distinct ways (See Figure 1). In contrast to previous literature (Xu, 1998), T4 seems to take a compressing strategy to adjust to shorter duration. The strategy of T2, on the other hand, is primarily a truncation of the initial falling section to allow the terminal high target to be maintained when duration is short.

The present data also reveal strong effects of phonological length on the realisation strategies. Figure 2 presents the F0 velocity on the rising section of T2 and falling section of T4, crossed for Length and Tone. In general, phonologically short (CV) syllables are exhibiting the strongest sign of compression – the shorter the duration, the higher the velocity – in both T2 and T4. This is corroborated by a significant effect of Length in a linear mixed effects model that contains an interaction term between Length and duration. This means, all else being equal (including duration), CV syllables compress F0 more than other syllable types.

Discussion: The implementation of Mandarin tones on durational constraints is both tone-dependent and structure-sensitive. We speculate that the structural dependency has a similar origin to the phonologized relation between duration and contour tone shape (Zhang, 2001): shorter syllables initially receive compressed F0 realisation, an effect which is reinforced in CV syllables due to its consistently short duration. The clear separation between the influence of the physical duration and the phonological length also adds to the evidence that non-contrastive, fine-grained phonetic details such as segmental duration and timing are supplied by language-specific phonetic component of grammar (Keating, 1990, 1994), which nevertheless must contain categorial labels of phonological units (Pierrehumbert, 2006).
Figure 1 Average F0 contour (semitone) on unnormalized time, by Tone and Length

Figure 2 Distribution of F0 velocity (semitone/ms) of each curve by Tone and Length

Reference
This paper investigates the tonal correspondence in language and music through text-setting strategy in Mandarin children songs. 10 Mandarin children songs are collected and analyzed how lyrics are set to music melody. The major findings are shown below.

First, the correspondence between Mandarin lyrics and music pitches is shown in the relative height of two Mandarin tones and the relative height of two music pitches. As shown in (1), the first Mandarin tone is a high-register tone and the second tone is a low-register tone, which together form a falling tone shape. As for the melody, the music pitches are shown in numbered musical notation. The shape of the two pitches is a falling shape. As a result, the relative height of the Mandarin tones corresponds with that of the music pitches.

Second, the correspondence between two syllables and their corresponding music pitches is also shown through obligatory contour principle (OCP), which triggers Mandarin tone sandhi. OCP demands that the first tone of two consecutive low tones changes into a high-register tone. According to Shih (1986), low tones undergo tone sandhi only when they branch to the same syntactic direction. However, as illustrated in (2), the simplified syntactic tree shows that the low tones li21 and you21 undergo tone sandhi even though they branch to the opposite direction. According to Chen’s (1984) foot formation rule, li21 and you21 form a foot (Ft) regardless that they do not form any syntactic constituent. A foot is a level in prosodic hierarchy where tone sandhi applies. The prosodic hierarchy is a distinct domain between syntax and phonology, which is framed by the Indirect Reference Hypothesis.

The sandhi form mentioned above also reflects on the music melody. As shown in (2), the sandhi form, li35, is higher than you21, which together forms a falling tone shape. Their corresponding music melody is also in falling shape. The correspondence between Mandarin tones and music pitches is in support of the Indirect Reference Hypothesis.

Third, the correspondence between Mandarin lyrics and music pitches is shown in monosyllabic tonal contour. As shown in (3), yao21 undergoes tone sandhi because it is followed by another low tone, wo21. The sandhi form of yao21 is yao35, which is a rising-contour tone. The corresponding music pitches of yao35 is also in a rising shape. This shows the output-to-output correspondence between mono-syllabic tonal contour and its corresponding music pitches.

(3) | Lyrics | ‘He bites me.’ |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ta55 yao21 wo21</td>
<td>Input</td>
</tr>
<tr>
<td>ta55 yao35 wo21</td>
<td>Output (sandhi form)</td>
</tr>
<tr>
<td>3SG CLF 1SG</td>
<td>Glossing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tone contour</th>
<th>Music pitches</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

To summarize, the output-to-output tonal correspondence in Mandarin tone and music pitch is shown in the relative height between two tones and their corresponding music pitches. Furthermore, the correspondence is shown in mono-syllabic tonal contour and its corresponding music pitches. The correspondence between Mandarin tone and music pitch also reflects the interface between syntax and phonology.

**References:**
Final stress is not represented in the French speakers’ mental lexicon

Amandine Michelas and Sophie Dufour
Aix-Marseille Université, CNRS, LPL, UMR 7309, 13100 Aix-en-Provence, France

In French, stress always affects the last syllable of a larger unit than the word (the Accentual Phrase, AP) and the same word can be stressed or not depending on its position within the AP (e.g. BANDEAU ‘headband’ is unstressed in \[\text{le ban'deau}\] _AP_ ‘the read headband’ while it is stressed in \[\text{le petit ban'deau}\] _AP_ ‘the little handband’). Contrary to speakers of languages such as Spanish, French speakers are never exposed to stress location contrast (e.g., 'bebe ‘he/she drinks’ vs. bebe ‘baby’ in Spanish), and they have difficulty distinguishing nonsense words that vary only in stress location (stress ‘‘deafness’’; Dupoux et al., 1997, 2001; Peperkamp et al., 2010). By contrast, French speakers are routinely exposed to words in their stressed and unstressed versions, and consequently they are fully capable of discriminating an unstressed word from its stressed counterpart, and this at a phonological level of processing (Michelas et al., 2016, 2018). In this study, we examined how French speakers represent stress information in their mental lexicon.

Considering models of spoken word recognition, an exemplar view (e.g., Goldinger, 1998) in which each word is associated with multiple tokens encoding fine-grained acoustic details, predicts that stress information would be integrated in the mental French lexicon. By contrast, an abstractionist view (e.g., McClelland & Elman, 1986) considers prosodic details as irrelevant for identification, and thus French words would be stored and processed independently of the presence of stress. Between these two extremist positions, hybrid models (e.g., Connine et al., 2008; Pierrehumbert, 2016) assume that several variants of the same word can be stored in memory, and thus regarding French mental lexicon both stressed and unstressed version of words could be stored.

In the present study, we conducted a long-term repetition priming experiment to examine whether French listeners encode words in a format that retains the presence of stress information. The repetition priming effect refers to the fact that words are recognized faster when they are encountered for a second time, and is generally interpreted as being due the repeated activation of the same form-based representation in memory (McLennan & Luce, 2005). Our hypotheses were as follow: If French words are stored regardless of stress information, no modulation in the magnitude of the repetition priming effect would be observed when the repeated words differed in the presence of stress. In contrast, if stress information is integrated into lexical representations, a greater repetition priming effect would be observed in case of a perfect acoustic match between the first and second presentation of the words, and thus when words are repeated with the same stress pattern.

Two blocks of 96 stimuli, 48 bisyllabic pseudowords and 48 bisyllabic words, were presented to 48 participants. The first block consisted of primes and the second block of targets. Within each block, half of the stimuli bore stress on their last syllable, and the other half was unstressed (Figure 1). In both the prime and the target blocks, participants were asked to make a lexical decision as quickly and accurately as possible with “word” responses using their dominant hand. Critically, the target block was composed of 16 words and 16 pseudowords consisting in the repetition of the primes in the same stress pattern (matched condition; _ban'deau/ban'deau_), 16 words and 16 pseudowords consisting in the repetition of the primes in a different stress pattern (mismatched condition; _ban'deau/ban'deau_), and 16 words and 16 pseudowords not used in the prime block and serving as control condition.

RTs on the target blocks were analyzed by means of mixed-effects modeling and are displayed in Figure 2. RTs were faster for target words preceded by match primes in comparison to control primes. RTs were also faster for target words preceded by mismatch
primes in comparison to control primes. However, no difference was observed between match and mismatch primes, thus indicating that a mismatch in stress between two repetitions of a same word has no impact on the magnitude of the repetition priming effect. Hence, French listeners appear to not store the presence of stress in the representation of words. In a more general way, our findings constitute a challenge for exemplar models of spoken word recognition.

Figure 1. Prosodic profile of the word bandeau ‘headband’ in its unstressed (a) and stressed (b) version.

Figure 2. Mean RTs and error bars for words and pseudowords in each condition.

References
Intonation of declaratives and questions in Luo
Zilpah Saul Ombijah (zo16605@essex.ac.uk)

This paper presents initial results of ongoing research on Luo intonation structure. Luo is a Western Nilotic language spoken in Kenya and Tanzania. This paper focuses on the Tanzanian variety where speakers are estimated to be around 170,000 (Rugemalira et al, 2009). Luo is a lexical tone language with pitch specification (i.e. High, Low, Falling and Rising) on every syllable. Tone in the language, apart from lexical distinction, has a crucial grammatical role of distinguishing perfective and imperfective aspects.

To investigate declaratives and questions, a robust data set was collected from four native speakers (2 males, 2 females) comprising a variety of declarative sentences and their associated polar questions as well as content questions, which were then compared. The declarative sentences comprised of 16 H toned, 16 L toned, 16 Falling and 7 Rising patterns. Polar questions comprised of 20 H toned, 20 L toned, 20 Falling and 6 Rising patterns, using the same target words as those of declaratives. By contrast, content questions comprised of 20 sentences of mixed tone patterns. Every speaker produced four tokens for each sentence which amounts to 564 tokens. In order to minimize tonal differences associated with specific words and structures, as suggested by Jun and Fletcher (2014), the same words and structures across different types of sentences were maintained as much as possible.

Initial observations show that declarative sentences with L tone sequences have slight declination where F0 gradually falls towards the end of an utterance, as indicated in Figure 1(a). Downstep is observed in both alternating and H tone sequences suggestive of automatic and non-automatic downstep, respectively. Figure 1(b) shows automatic downstep, which has been traditionally referred to as downdrift (Pierrehumbert, 1980; Gusenhoven, 2004). In this type of downstep an overt L tone lowers subsequent H(s). This also applies to sequences with falling contours which also have an L tone causing automatic downstep. Figures 2(a) & (b) illustrate a sequence of lexical Hs with non-automatic downstep where there is no overt L tone trigger. It is further shown that Luo has no final lowering (L9%) either phrase-finallly or utterance finally. Looking at Figure 1(a), L tone sequences have normal declination where F0 gradually falls towards the end of an utterance, without extra lowering or rapid fall at the end. Similarly, sentences ending in H tone as well as Rising tone also maintain their high pitch (above 100Hz).

With respect to questions using the question particle ‘be’ (for Yes-No questions) or ‘ŋa’= what (for interrogative questions), Luo appears to have two intonational structures for questions. One way is shown in Figure 1(b) where an interrogative question has the same structure as seen for declaratives, but is produced at a higher pitch register. In this case the cue for a question is register raising. The second alternative observed in some speakers is to leave the pitch relatively unchanged until the end of the utterance when a HL melody is added as illustrated with the Yes-No question in Figure 3(b). In this case the declarative sentence in 3(a) does not differ very much from its Yes-No question counterpart in 3(b), with both maintaining an F0 of just below 200Hz. These two question intonation patterns are not restricted to question type but rather show a variation in question marking intonation.

From the significant intonational patterns that emerge from this study, some of these suggest an interaction with syntactic structure; with non-automatic downstep in (2a) directly coinciding with XPs with downstep triggered at the VP and PP left edges.
1. (a) Female speaker SY Low tone sequence portraying declination; (b) mixed tone sequence portraying downstepped !Hs

Figure 1: (a) Female speaker SY Low tone sequence portraying declination; (b) mixed tone sequence portraying downstepped !Hs

2. (a) Female speaker SY High tone sequence; (b) Yes-No counterpart with pitch register expansion

Figure 2: (a) Female speaker SY High tone sequence; (b) Yes-No counterpart with pitch register expansion

3. (a) Male speaker SM High tone sequence; (b) Yes-No counterpart with final HL melody

Figure 3: (a) Male speaker SM High tone sequence; (b) Yes-No counterpart with final HL melody

References

Gussenhoven, Carlos (2004). The Phonology of Tone and Intonation. Cambridge: CUP.
In this paper, we examine prenuclear as well as nuclear F0 patterns in L2 Spanish yes-no questions, produced by L1 German and L1 Czech adult learners. We seek to determine to which extent L1 Czech and L1 German features are present in L2 Spanish intonation. The three languages (may) have non-inverted yes-no questions but cross-linguistic differences have been described in the realization of the first pitch accent (PA) and the nuclear contour (NC). In Spanish and German (head-prominence languages; [3]), questions are marked by a higher initial PA, when compared to statements, although there may be differences in the alignment of the high tone (see, e.g., [1], [2]). In Czech (head/edge-prominence language; our proposal in [3]’s typology), pitch accents do not show a very large pitch excursion, and initial accents show more reduced pitch range in questions than in statements. Our hypothesis is that German learners (GE) resemble controls more closely than Czech learners (CZ) in the realization of the initial PA. As concerns NC, although rising contours are present in the three languages, differences may be expected in the type of rise (see [4], [5], [6]); Table 1.

We tested 40 learners (20 per learner variety) and 5 controls by using a Discourse completion task to elicit semi-spontaneous speech. For the purposes of this study we are comparing five different yes-no questions as well as two statements across all speakers and languages. We labeled the initial (A1) and middle (A2) pitch accents and the nuclear configuration (nuclear accent A3, boundary tone BT), using the (Spanish) ToBI ([7]). We chose ToBI labels in order to provide simplified representations of tonal events, on the one hand, and to systematize and compare the patterns found in data, on the other hand. Finally, we measured the height of all pitch accents, and we calculated the pitch change for each tonal event.

Preliminary results obtained for one of the sentence types - an invitation yes-no question (¿Vamos a tomar una cerveza?, ‘Should we go have a beer?’) - show no clear between-group differences in the type of A1 (Figure 1a): the most frequent type for both groups was L*+H (a low tone on the accented syllable followed by a post-tonic rise). However, we found differences in the realizations of A2 (Figure 1b): whereas German learners’ most frequent realization is also L*+H, Czech learners produce sustained high tones (H*) or falling patterns (H*+L, H+L*). As for the nuclear configurations (Figure 2), both learner groups show a preference for a final high BT (()H%), and a greater variability of nuclear accents, which mainly depend on the previous PA’s realizations. As concerns pitch change, results confirm our hypothesis: significant differences (Figure 3) were found only in the realization of A1. German learners displayed a significantly larger pitch change in this position than Czech learners (β= 32.7, t= 2.3; p=0.02) but they did not differ significantly from controls. The reduction of initial pitch in L2 Spanish questions was typical for Czech learners.

There are several factors which may explain the variability found in L2, such as the input quality or input quantity, the length of the residence in a L2-speaking area, age of onset of acquisition, or personal motivation. Nevertheless, the observed differences between “Czech” and “German” Spanish learner varieties indicate that L1-based cross-linguistic influence (CLI) still plays a crucial role here. In comparison to the Czech learners, the German learners are slightly more consistent in their production of pitch accents and final contours. This could be related to the fact that German has a stronger, but Czech a weaker macro-rhythm (for the computation of the macro-rhythm see [3]). The higher variability in Czech learners can be attributed to the fact that Czech is a fixed stress language with initial prominence. Some participants (e.g., those exhibiting a H+L* accents) displaced the peak to the pretonic
syllable, as expected from CLI. These issues of phonetic implementation will be considered in future research closely together with the question as to whether and how L2 intonation patterns contribute to foreign accent.

<table>
<thead>
<tr>
<th>Language</th>
<th>Prenuclear (initial) position</th>
<th>Nuclear configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (Peninsular) Spanish</td>
<td>L*+H</td>
<td>L* H% (or L+!H* H%)</td>
</tr>
<tr>
<td>L1 Czech</td>
<td>L* Ha (accentual phrase)</td>
<td>L* H% (or L* H!H%)</td>
</tr>
<tr>
<td>L1 German</td>
<td>H* (late peak)</td>
<td>L*+H H% (or H*+L H%)</td>
</tr>
</tbody>
</table>

Table 1. Tonal patterns of yes-no questions in L1 Spanish, L1 Czech and L1 German.

![Figure 1a. Initial PA.](image1) ![1b. Second PA.](image2)

![Figure 2. Nuclear configurations.](image3)

![Figure 3. Pitch change by tonal event.](image4)

**References**

Czech yes-no questions from a typological perspective

Andrea Pešková, Laura Colantoni, Trudel Meisenburg

University of Osnabrück, University of Toronto

Cross-linguistically, interrogatives can be expressed by different strategies, which may range from initial to final pitch movements, duration, changes in word order, use of particles, or combinations of these cues (see, e.g., [1], [2], [3], [4]). In this paper we examine initial and final cues in Czech (unmarked) yes-no questions (also known as polar questions), describing not only F0 patterns, but also duration and pitch range of prenuclear as well as nuclear positions. To achieve a better understanding of the complexity of the interrogative structures, we compare polar questions with – syntactically identical – statements.

From a typological view, Czech is a Western-Slavic language, which has a fixed stress on the first syllable of the lexical word. If any (one-syllable) preposition precedes the noun, the pitch shifts to the left (e.g., na Moravu [‘namoravu]; ‘to Moravia’). Following [5] and [6], we consider Czech a phrase language or a head/edge-prominence language. Similarly to other languages with fixed stress, such as Bengali [7] or Hindi [8], Czech prosody is left-headed and its accentual phrase (AP) functions as a rhythmic unit exhibiting (mostly) a rising pitch, stretched from one stressed syllable to another (L*H; [9]). The rising pattern within the (prenuclear) AP is not the only possible phonetic realization (see, e.g., [10], [11] for a discussion), but more research is still needed. Although Czech has been described as a language with interrogative word order [1], statements and questions can be syntactically identical when the subject is dropped. In such cases, prosody is the only tool to mark sentence type.

To further our understanding regarding the prosodic cues that distinguish polar questions from statements, we conducted a production (reading) experiment with 21 Czech speakers of two different dialects: Moravian (eastern dialect) and Bohemian (western dialect). We selected four questions and statements with an unexpressed pronominal subject, so that the word order remains the same (e.g., Jedeme na Moravu? ‘Are we going to Moravia?’ vs. Jedeme na Moravu. ‘We are going to Moravia.’). All tokens ($N=168$) were transcribed and segmented in Praat ([12]), and the data were modelled within the Autosegmental Metrical (AM) framework ([13]). F0 values and duration were measured for the accented vowel and the AP-phrase accent in initial, middle and final positions. Linear mixed-effect regression models with pitch change and duration as the dependent variables, Sentence, Dialect and Tonal Event as fixed effects and Participants as random effect were used to compare the pitch and durational patterns across groups.

As expected ([9], [10], [11], [14]), the F0 patterns of Czech yes-no questions (rise, rise-fall) substantially differed from statements (fall) in the realization of nuclear contours (see Figure 1). Crucially, we found a significant difference in prenuclear position between statements and questions. Unlike head-prominence languages (e.g., Spanish, German or English), which use peak alignment and/or a high F0 peak to mark polar questions, Czech initial accentual phrases have a smaller pitch excursion ($p=.000$) and a shorter duration in questions than in statements ($p=.002$). Our findings support the claim that initial prosodic cues may also contribute to the identification of sentence types (see, e.g., [15], [16], [17]). However, the differences found in production need to be supported by perception data in order to determine whether Czech native speakers use these cues. As opposed to the variability in F0 final contours (see Table 1), which could be partially attributed to dialectal differences ([10], [14]), the shorter duration and/or pitch compression of APs in questions seem to be typical of all varieties of Czech. Although our results suggest that phrase
languages may favour pitch compression and local uses of duration for the marking of sentence type, this typological generalization needs to be corroborated by further research.

Figure 1: Spectrogram and F0 trace for the statement “Jedeme na Moravu” (left) and for the yes-no question “Jedeme na Moravu?” (right), (female speaker from Southeast Moravia).

<table>
<thead>
<tr>
<th>Type of sentence</th>
<th>Nuclear configurations</th>
<th>Bohemia (N=72)</th>
<th>Moravia (N=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no questions (N=84)</td>
<td>L*+H lH%</td>
<td>53%</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>H* LH%</td>
<td>22%</td>
<td>6.5%</td>
</tr>
<tr>
<td></td>
<td>L* H%</td>
<td>11%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>L*+H H%</td>
<td>11%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>L*+H LH%</td>
<td>3%</td>
<td>8.5%</td>
</tr>
<tr>
<td>Statements (N=84)</td>
<td>L* L%</td>
<td>40%</td>
<td>73%</td>
</tr>
<tr>
<td></td>
<td>H* L%</td>
<td>33%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>L*+H L%</td>
<td>17%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Table 1: Tonal inventory of Czech yes-no questions and statements

References
Lexical encoding of tone in a restricted tone language: Cross-linguistic evidence from Limburgian and Dutch toddlers and adults

Stefanie Ramachers, Susanne Brouwer, and Paula Fikkert

Centre for Language Studies, Radboud University, Nijmegen, the Netherlands

Psycholinguistic research on the acquisition and processing of lexical tones has typically investigated Asian tone languages such as Mandarin, Cantonese, or Thai. There is limited research dedicated to the acquisition of lexical tones in more restricted tone systems like Japanese, Swedish, and Limburgian.

In many Limburgian dialects, spoken in the south of the Netherlands, a word prosodic contrast can signal lexical and morphological differences. Relative to Mandarin, tone in Limburgian is not very important for lexical disambiguation: Fournier (2008) counted around 80 tonal minimal pairs. Moreover, there is only a two-way contrast, and lexical tones are restricted to minimally bimoraic syllables with primary stress (Gussenhoven, 2000).

Comparable to Swedish (Gussenhoven, 2004), Limburgian lexical tones co-occur with intonation tones, leading to different phonetic realizations as a function of information status, sentence type, and position in the intonational phrase (Gussenhoven, 2000). Studies on lexical tone acquisition in Swedish (Ota, 2003), Japanese (Ota, 2003; Ota et al., 2018; Yamamoto & Haryu, 2018), and Sesotho (Demuth, 1995) suggest that the reliability of the mapping between underlying tones and their surface realizations as well as the functional load has an impact on the speed of acquisition of tone. The Limburgian tone system might thus pose challenges to its learners. To investigate this, we compared 23 Limburgian (M age = 40.9 months) and 35 Dutch (M age = 36.8 months) toddlers as well as 14 Limburgian (M age = 53.6 yrs.) and 22 Dutch (M age = 23 yrs.) adults on their lexical encoding of tone in a word learning experiment. None of the Dutch participants had substantial exposure to Limburgian or any other tone language.

Following the procedure of Quam and Swingley (2010) and Singh, Hui, Chan, and Golinkoff (2014), participants learned two novel word-object mappings. After learning, word recognition was tested in correct pronunciation (CP) trials and mispronunciation (MP) trials featuring a pitch change. Word recognition is signaled by a naming effect, measured as an increase in the proportion of target looking upon naming of the target picture. Based on previous studies (e.g., Quam & Swingley, 2010; Singh et al., 2014; Singh, Goh, & Wewalaarachchi, 2015), we expected that Limburgian listeners would notice tonal MPs, but Dutch listeners would not.

A mixed ANOVA on the child data with Condition (CP vs. MP), Tone (Accent 1 vs. Accent 2), and Language (Limburgian vs. Dutch) yielded a significant main effect of Condition ($F(1,56) = 8.53, p = .005, \eta_p^2 = .13$), showing a significantly larger naming effect in CP trials ($M = .21, SD = .20$) than in MP trials ($M = .09, SD = .24$). No other effects or interactions were found (all $p$’s $> .1$). To investigate the strength of the MP, the naming effect in MP trials was compared to zero by
means of a one-sample $t$-test. The test revealed a significant positive naming effect ($M = .09$, $SD = .24$; $t(57) = 2.81$, $p < .01$, Cohen’s $d = .37$). The same analysis was run on the adult data, but did not yield any main effects or interactions (all $p$’s > .05). As in the CP trials, the naming effect in MP trials was significantly above zero ($M = .34$, $SD = .22$; $t(38) = 9.53$, $p < .001$, Cohen’s $d = 1.53$).

Our results partly deviate from studies on Mandarin (Singh et al., 2014, 2015), but are on a par with recent findings on lexical encoding of tone in Japanese (Ota et al., 2018; Yamamoto & Haryu, 2018), suggesting that functional load and phonetic variability may indeed influence the specificity of (early) lexical representations.

References


Word-initial CV coordination in a lexical pitch-accent language

Malin Svensson Lundmark¹, Johan Frid², Gilbert Ambrazaitis³ and Susanne Schötz⁴
¹Centre for Languages and Literature, Lund University, Lund, Sweden
²Lund University Humanities Lab, Lund University, Lund, Sweden
³Department of Swedish, Linnaeus University, Växjö, Sweden
⁴Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund University, Lund, Sweden

This study examines the coarticulatory anticipatory movements of the consonant and the vowel in the word-initial CV sequence /ma/. Coarticulation and coproduction of gestures is not merely an artefact, but a result of the anticipatory coordination of articulators reaching for phonemic targets [1]. Previous research has acknowledged the effect of prosody on intergestural coordination [2–3], but the integration of tones with the articulation of the CV sequence is still understudied. The few existing relevant studies have been based on small-scale data sets and have used a variety of measures [4–8].

The present study is an investigation on CV coordination in a lexical pitch-accent language (Swedish) with the binary tonal word accent distinction (HL tone and LH tone, respectively). Kinematic recordings have been made in an EMA (Carstens AG501) at The Lund University Humanities Laboratory. To this end, spatiotemporal measurements of bilabial and tongue body data, as well as acoustic measurements on f0, from 19 speakers (1200 tokens) are used to examine articulatory gestures in word-initial CV sequences (/ma/). Articulatory measurements include: bilabial intervals (bilabial closure and release), time lag between onsets of the bilabial closure and the lowering and retraction of the tongue (CV time lag), the interval of the lowering and retraction of the tongue (palatal wide interval), as well as the height of a stable tongue body landmark (tongue body height) (Figure 1).

Results from mixed effects regression models reveal a significant effect of word accent on the tongue body height and on the palatal wide interval, suggesting a link between tongue body movement and f0. Results also reveal a significant difference between word accents on the bilabial release, but not on the bilabial closure or on the CV time lag measurement. However, CV time lag do display a significant difference between Accent 1 and Accent 2 when the retraction of the tongue as well as the acceleration of the lips are included. Furthermore, previous studies have indicated a CV coordination difference between tone and intonation languages. However, given the contradictory CV time lag results of our study we suggest that previous cross-linguistic findings can partly be explained by the choice of measurement landmarks, and, because of speaker-dependent patterns on intergestural coordination, by the choice of speakers.

Subtle speech production variation between languages results in systematic differences in phonemic targets [9]. Our study shows that the different tongue body movements found between Accent 1 and Accent 2 affect the articulatory coordination of the phonologically similar /ma/-sequences. Presumably, assorted coarticulatory patterns of similar word onsets result in systematically varied phonemic targets, and in variations in acoustic space, which may be used by the listener as anticipatory cues to the word accents.
Figure 1. Trajectories of articulatory movements (EMA-data): An illustration of the spatiotemporal measurements on the bilabial (Lip aperture) and tongue body movements of the target CV sequence /ma/.


The effect of focus on syllable duration of trisyllabic sequences in Mandarin

Maolin Wang
Jinan University, Guangzhou, China
wmljd@126.com

In this study, the focal effect on the temporal pattern of trisyllabic sequences in Mandarin is investigated. The trisyllabic sequences are words and phrases where neutral tone syllables are included. It has been proved that in Mandarin, neutral tone syllables are metrically weak (W), and normal tone syllables are metrically strong (Lin and Yan, 1980), so in this study, the focal effect on the duration of strong and weak syllables will be examined.

Four types of trisyllabic sequences are investigated: those with no neutral tone syllables (SSS), those with one neutral tone at the third syllable (SSW), those with one neutral tone at the second syllable (SWS), and those with two neutral tones at the second and third syllables (SWW). They are embedded in sentence medial position. Focus was induced by leading questions, which were recorded beforehand and played to the subjects at the time of recording. Eight native speakers participated in the recording, who read the sentences after hearing the questions. Annotation was done by an experienced annotator using the software of Praat (www.praat.org), and syllable durations under the focal and unfocused conditions were compared.

It is shown that, for the four sequence types, all the syllables lengthen under focused condition, whether strong or weak syllables. For the SSS sequences, the final syllable lengthens the most, which is consistent to the results of Chen (2006). For the SWS sequences, similar to the SSS sequences, the lengthening of the final syllable is more than the other two syllables. For the SSW sequences, the second syllable lengthens the most, and for the SWW sequence, the first syllable lengthens most. Strong syllables lengthen more than weak syllables, and final syllable lengthens more than the other syllables. Mandarin is not a stress language, but its focal lengthening pattern is similar to that of stress language.

References
Exploring perception of pre-final prosodic cues to turn ends in German

Margaret Zellers
Institut für Skandinavistik, Frisistik, und Allgemeine Sprachwissenschaft
University of Kiel, Germany

It has been clear since at least Sacks et al.’s (1974) well-known work on the structure of turn-taking in interaction that people in conversations mostly conform to a set of turn-taking rules which prioritize having one person talk at a time, and generally keep overlapping speech as well as long silent pauses, or “gaps”, to a minimum. Work by Stivers et al. (2009) and Heldner & Edlund (2010) has shown that the most frequent duration of silent between-turn gaps in a variety of languages is about 200 milliseconds (ms); overlapping speech represents only 5-10% of talking time (Shriberg et al., 2001; Levinson & Torreira, 2015). Would-be speakers must carefully time the onset of their speech if they wish to maintain smooth interaction, and must therefore have ways of predicting when and how another speaker’s turn is coming to a close. A variety of different prosodic signals for the prediction of an upcoming transition space, with scope ranging from intonation to speech rate to voice quality, have been investigated across languages in the region preceding the offset of speech (Koiso et al., 1998; Gravano & Hirschberg, 2009, 2011; Hjalmarsson, 2011; Zellers, 2016; Brusco et al., 2017, inter alia).

A criticism of much of the previous work that has been carried out on the investigation of prosody at turn ends is that these cues arise too late to be used. The finding that the most frequent between-turn gap is around 200 ms belies the prediction that would-be speakers wait for silence to arise before beginning to speak. This short amount of time is insufficient for the motor planning necessary for speech, which requires a minimum of 600 ms even for the production of single words (Indefrey & Levelt, 2004; Schnurr et al., 2006; Indefrey, 2011). Turn-end cues, potentially including prosodic variations, must thus be available substantially earlier than the offset of speech. However, there is not yet any fixed span of time or unit that can be definitively identified as the critical region for the initiation of pre-turn-end marking. Levinson & Torreira (2015) thus propose the final 500 ms of speech as the crucial span for prosodic signalling. Zellers (2016) finds that turn-end-related phonetic variation can be initiated a full second before a potential turn boundary. Qualitatively-oriented researchers have sometimes identified turn-final prosodic variation in terms of phonological structures, e.g. the final foot (Local et al., 1986), or the final pitch accent (Wells & Macfarlane, 1998).

The current study investigates the role that prosodic features arising earlier than the final 500 ms of speech have on German listeners’ interpretation of speakers’ intention to continue speaking or to release the floor. Conversational turns taken from the Kiel Corpus of Spontaneous Speech (IPDS 1995-1997) were resynthesized to have different speech rate and intonational characteristics in the portion of the turn preceding the final 500 ms or final pitch accent, with this final stretch held constant for each base stimulus. Speech rate could be fast, normal (i.e. the original rate), or slow; additionally, the whole pitch range of the early portion of the turn could be raised or lowered by 2 semitones (st) compared to the original pitch, and the peak of the penultimate pitch accent could be raised or lowered by 3 st compared to the original pitch. Participants heard pairs of versions of the same turn, and were asked to respond either which version sounded more like the speaker wanted to keep speaking, or which version sounded more like the speaker intended to stop speaking and let someone else talk. Pilot results suggest that increased pitch peak height or raised range in the pre-final portion of the turn led listeners to anticipate that the speaker would want to continue speaking, while slower speech and a lower pitch peak in this portion led listeners to expect that the speaker would want to stop speaking. It is yet to be determined whether the prosodic signalling in this region is itself crucial, or whether the contrast with the final portion of the turn is the most important aspect of such interactional prosodic signalling.
References
Anticipatory Dissimilation in Non-clitic Neutral Tones in Mandarin

Yixin Zhang, University of Cambridge, yz510@cam.ac.uk

This paper investigates anticipatory dissimilation before non-clitic neutral tones in Mandarin. Anticipatory dissimilation in tone languages refers to the phenomenon whereby a (relatively) low pitch target will trigger a pre-target high point in the preceding tone (cf. Xu, 1997). This phenomenon has been systematically reported for several tone languages including Mandarin (Xu, 1997), but only in the context of the four citation tones (T1, high-level; T2, low-rising; T3, low-dipping; T4, high-falling). In Mandarin, there are also morphemes that do not carry any citation tone (CT) but have a neutral tone (NT). The neutral tones carried by the non-clitic morphemes (non-clitic NTs) are usually regarded as the results of NT sandhi, namely, when a CT-bearing morpheme loses its lexical importance in a particular context, it will lose its CT and carry a NT instead (cf. Shen, 1992).

The underlying CTs (i.e. the CTs lost in non-clitic NTs) have been found to have no obvious $f_0$-based influence on the NT pitch (e.g. Lee and Zee, 2014). However, we have found previously that (relatively) low targets in CTs underlying NTs seem to promote anticipatory dissimilation, particularly when the low target is adjacent. T1 preceding NTs with underlying T2 (low-rising tone) and T3 (low-dipping tone) is higher than T1 preceding the underlying T4 (high-falling tone). T1 is lowest preceding NT with underlying T1 (high-level tone).

To further examine this phenomenon, I have recorded 5 native Northern Mandarin speakers with disyllabic words exemplifying T1 + NT1 (NT with underlying T1), T1 + NT3 (NT with underlying T3), T4 + NT1 (NT with underlying T1) and T4 + NT3 (NT with underlying T3). The words were embedded in short stories. Phrasal position and focus were controlled. The preceding T1 was chosen to represent the level tones ending with a high target and T4 to represent the contour tones ending with a low target; the underlying T1 was used to stand for the underlying tones without low targets and the underlying T3 for those with low targets.

A preliminary analysis of the results with Mixed-design Anova shows that the anticipatory dissimilation was promoted by the underlying T3 but not the underlying T1 in both the preceding T1 and the preceding T4. In accordance with previous literature, the anticipatory dissimilation was mainly manifested through pitch height in the preceding T1 (Xu, 1997) but through pitch range in the preceding T4 (Kratochvil, 1984), namely, T1 preceding NT1 was significantly lower than T1 preceding NT3 while the falling contours of the T4 before NT1 were truncated compared to the T4 before NT3 (Figure 1). However, the dissimilatory effect caused by the underlying CTs is not as strong as that caused by the surface CTs since the average pitch height of the T4 before NT3 was not significantly higher than the T4 preceding NT1.
Figure 1 Pitch Contours of the disyllabic NT words: The first 3 time points of the 2nd syllable are still regarded as CT-carrying points rather than NT-carrying points because of the carry-over effect, i.e. the full realization of a 1st-syllable CT (especially the contour tones) tends to include at least the first quarter of the following syllable (cf. Xu, 1997).

A further analysis of the results shows clear individual difference in the realization of the anticipatory dissimilation. All the participants produced higher T1 before NT1 but the truncation of the preceding T4, i.e. the smaller pitch range of the T4 before NT1, was only clearly seen in 3 participants. This asymmetry is not unexpected as the previous research has reported that low tonal targets are more resistant than high tonal targets to contextual influences (e.g. Xu, 1997).

This preliminary study on the anticipatory dissimilation before the non-clitic NTs shows that the anticipatory effect may be caused by the low target in either the surface or the underlying form. However, the manifestation (through pitch height or contour) and the extent of dissimilation vary across different types of the preceding tones and different individuals.

Reference: