Intonation as a quantifier-free logical interpretation of metrical and prosodic structure

Hyunjung Joo and Adam Jardine
Department of Linguistics
Rutgers University
{hyunjung.joo,adam.jardine}@rutgers.edu

Abstract

This study views *intonation* as a quantifier-free (QF) logical interpretation of a metrical and prosodic structure. Under logical transductions, tones in intonational melodies can be interpreted as literal copies of prosodic elements, with their association to TBUs being a local process. The head-prominence intonational pattern in American English can be defined by copying both accented syllables (heads) and phrasal boundaries, whereas the edge-prominence pattern in Seoul Korean was defined by copying only phrasal boundaries (edges). For Tokyo Japanese, lexical pitch accents are defined by copying accented moras, and post-lexical tones by copying phrasal boundaries. This QF interpretation of intonation structure enabled restrictive predictions about computational complexity and typology of intonation.

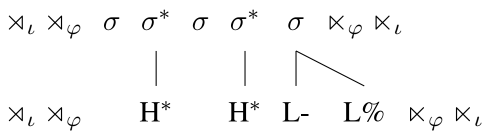
1 Introduction

How can we define what it means to be a possible intonational pattern in a precise way? Here we view *intonation* as a quantifier-free (QF) logical interpretation of a metrical and prosodic structure (Chandlee and Lindell, to appear; Strother-Garcia, 2019). Importantly, in this framework, tones in intonational melodies are viewed as literal copies of elements in the metrical and prosodic structure, such as accented syllables or phrasal boundaries, and they are always linked *locally* to their tone-bearing units (TBUs). Importantly, because QF is a very weak logic, a theory of intonation built around QF interpretations makes strong predictions about what is a possible intonational pattern. We show support for these predictions by showing that major intonational patterns are QF interpretations.

In the Autosegmental-metrical (AM) theory of intonation (e.g., Pierrehumbert, 1980), intonation can be defined as a sequence of Highs (Hs) and Lows (Ls). The tones in intonation are associated with their TBUs within the nested prosodic domains. Languages may vary depending on which prosodic elements, such as prominent syllables and/or phrasal boundaries, are used for intonation.

For example, in American English, intonational tones are associated with *metrically strong positions* and phrasal boundaries in an utterance. (1) shows an utterance "an orange ball gown" produced with intonation. Within an intermediate phrase (ip; $\rtimes_\varphi/\ltimes_\varphi$), pitch accents (H*) are associated with accented syllables (σ^*) and a phrasal tone (L-) is associated with the final syllable of the ip. Then, within an Intonational Phrase (IP; $\rtimes_\iota/\ltimes_\iota$), a boundary tone (L%) is associated with the final syllable of the IP.

(1) [[ən ɔɹɪndʒ bɔl gaʊn]$_\varphi$]$_\iota$



Jardine (2017) showed that autosegmental representation of lexical tones and their TBUs is an *interpretation* of the toned syllables in the input structure, using *logical transductions* (Courcelle, 1994; Engelfriet and Hoogeboom, 2001; Filiot and Reynier, 2016). Also, the tone–TBU association patterns in tonal languages have been studied in terms of their local nature and computational complexity (Chandlee and Jardine, 2019a; Chandlee and Jardine, 2021; Koser et al., 2019). Then, how can we define the autosegmental representation of intonation using logical interpretation and what does this say about the computational nature of intonation?

We extend Jardine (2017) and Strother-Garcia (2019) by viewing AM representations as additional structure imposed on an input string. In doing so, we find that intonational tones and their associations with TBUs are always local to accents and boundaries if we make reference to a metrical grid and a prosodic structure. That is, the accented syllables and boundaries in the input structure can be interpreted as intonational tones in the output structure, which are always linked to their TBUs that are near the accents and boundaries.

Also, there exists another evidence supporting the QF logical interpretation of intonation. Not every logically possible intonational pattern is attested. For example, there are no patterns like Midpoint Pathology (Hyde, 2008; Eisner, 1997), in which tones are associated to a center-most TBU, for the intonational patterns. Computing such a tonal sequence demands memory proportional to the sequence length, exceeding the regular complexity bound of phonology (Heinz and Idsardi, 2011; Johnson, 1972; Kaplan and Kay, 1994) and thus far exceeding the power of QF.

Therefore, we can start with a hypothesis that intonation can be a QF logical interpretation of a metrical and prosodic structure, by examining three different intonation patterns: a *head-prominence* language, American English; an *edge-prominence* language, Seoul Korean; a *lexical pitch accent* language, Tokyo Japanese.

Based on this local nature of intonation, we can posit a theory that makes restrictive predictions about the intonational typology and measure the complexity of intonational structures, as the connections between logical interpretations and computational complexity are well-studied (Filiot and Reynier, 2016). This has been fruitfully applied to the study of phonological representations (Strother-Garcia, 2019; Jardine, 2017; Jardine et al., 2021).

2 Preliminaries

2.1 String models and logic

We define a finite alphabet of symbols as Σ and the set of all strings over σ^*. We use two boundary symbols $\rtimes, \ltimes$ to indicate the beginning and the end of strings. For example, for $\Sigma = \{C, V\}$, $\rtimes CCV \ltimes$ is a string over Σ delineated with boundaries.

We can describe strings and other structures with *models* in the following way (Enderton, 2001; Libkin, 2004). A *signature* is a set $\{R_1, ..., R_m, f_1, ..., f_n\}$ of named relation and function symbols. (We do not use signatures with constant symbols.) A *model* is thus an instantiation $\langle D; R_1, ..., R_m, f_1, ..., f_n\rangle$ of this set of relations and functions with a domain D of elements.

For example, in strings over an alphabet Σ, we can describe them with a signature $\{P_\sigma \in \Sigma, p, s\}$. where each $P_{\sigma\in\Sigma}$ is a unary relation that refers to a set of positions over the domain D for each σ in the alphabet. The predecessor and successor functions are p and s that return the immediately preceding and immediately following element in the string, respectively. For example, in $\{P_C, P_V, p, s\}$, P_C and P_V refer to the sets of positions over the domain D for C and V, respectively. With this signature the string $\rtimes CCV \ltimes$ can be defined with the following string model:

$$\begin{aligned}
&\langle D = \{0, 1, 2, 3, 4\};\\
&P_C = \{1, 2\}, P_V = \{3\}, P_\rtimes = \{0\}, P_\ltimes = \{4\};\\
&\quad p = \{(0, 1), (1, 2), (2, 3), (3, 4)\};\\
&\quad s = \{(1, 0), (2, 1), (3, 2), (4, 3)\}\rangle
\end{aligned}$$

From a signature we immediately get a *first order* (FO) predicate logic in the usual way. Briefly, x, y, ... denoting *variables* that range over positions in a string; $\sigma(x)$ for each $\sigma \in \Sigma$ denoting *atomic predicates* which are true when x is interpreted as positions in the unary relation P_σ of a model; and FO formulae are are built recursively out of the logical connectives $\neg, \vee, \wedge, \rightarrow$ and quantifiers $\exists, \forall$. A *free variable* is a variable not bound by a quantifier. QF is the fragment of FO in which no quantifiers appear.

2.2 Logical transductions

Based on the input string that we've just defined, we can build a larger model using *logical transductions* (Courcelle, 1994; Engelfriet and Hoogeboom, 2001; Filiot and Reynier, 2016). We *interpret* the input structure into a finite number of *copies* in the output structure, using FO formulas. Via a logical transduction τ, the domain of the input structure (Σ) in the signature ($\mathcal{S}_i$) is extended in the output structure (Γ) in the signature ($\mathcal{S}_o$), which is represented with copies (Cs) of the input domain. Following Strother-Garcia (2019), we use syllable structure as an example, as shown in Figure 1.

The output structure Γ is defined with relations R' satisfied for any transduction τ if $\langle D'; R'_1, ..., R'_n\rangle$ is based on the input signature $\mathcal{S}_i$. For instance, $C_o(x) \overset{\text{def}}{=} C_i(x)$ means a consonant x appears in the output if and only if it exists in the input.

The domain D' of the output structure Γ is expanded by copying input elements n times, creating *n* copies of each input element. Unary relations R' are represented as R'^n for $n \in C$ (e.g., $C^0_o(x)$ denotes a consonant in the 0th copy). Binary relations R' are represented as $R'^{m,n}$ for $m, n \in C$

(e.g., $\mathcal{A}_o^{0,1}(x,y)$ denotes an association between x in the 0th copy and y in the 1st copy). The order p_o of the output structure Γ over the domain D' is defined separately for the 0th copy and the 1st to nth copies ($n > 0$), while preserving the order of the input structure Σ over D for both copies, following Chandlee and Jardine (2019b).

For the 0th copy, $p_o(x^0) \stackrel{\text{def}}{=} \mathrm{p}_i(x)$, such that the output order p of the elements in the 0th copy of Γ works the same as that in the input structure Σ, just like an identity function. For all the set of nth copies except for the 0th copy, $p_o(d^n) \stackrel{\text{def}}{=} \mathrm{e}^m (n, m > 0)$ if and only if $(p(d) = e) \vee (d \approx e \wedge p(n) = m)$. That is, for any elements $d, e \in D$ and for the copies $m, n \in C$, the element e^m precedes the element d^n in the output, with two conditions. The first condition is that if there are two distinct domain elements, we follow the order of the elements, such that if the element e precedes the element d in the input, the element e^m always precedes the element d^n in the output. However, the second condition is that if there are two identical domain elements in different copies, we follow the order of the copies such that if the mth copy precedes nth copy, the element e in the mth copy always precedes the element d in the nth copy.

From the $\rtimes CCV \ltimes$ string, we can build a syllable structure in the output, using logical transductions. The input strings are copied twice in the output (C0, C1) and each node with a free FO variable x is defined accordingly. The order of copies in the input, as determined by the predecessor and successor functions p_i and s_i, is preserved in the output using p_o and s_o.

$$C_o^0(x) = C_i(x) \quad V_o^0(x) = V_i(x)$$
$$\rtimes_o^0(x) = \rtimes_i(x) \quad \ltimes_o^0(x) = \ltimes_i(x)$$
$$\sigma_o^1(x) = V_i(x)$$
$$\begin{aligned}\mathcal{A}_o^{0,1}(x,y) = {} & C_i(x) \wedge V_i(y) \wedge y \approx s(s(x))) \vee \\ & (C_i(x) \wedge V_i(y) \wedge y \approx s(x)) \vee \\ & (V_i(x) \wedge V_i(y) \wedge y \approx x)\end{aligned}$$

In the first copy (C0), every C and every V in the input has one copy with the same label in the output. Also, boundaries in the output, $\rtimes_o^0(x)$ and $\ltimes_o^0(x)$ are the same as in the input. Importantly, for the second copy (C1), syllables in the output, $\sigma_o^1(x)$, is defined from a vowel in the input, $V_i(x)$, showing that every syllable is a *reflection* of nucleus.

Then, we can establish some relations between the output copies to build phonological structures.

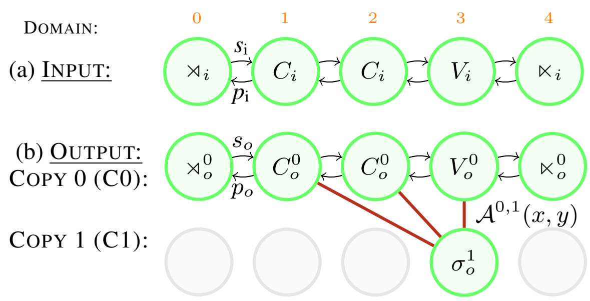


Figure 1: The illustration of a logical transduction from the input string $\rtimes CCV \ltimes$ to the output syllable structure.

$\mathcal{A}_o^{a,b}(x,y)$ defines an association relationship between the output copies over two free variables x and y, where a and b indicate the copies in the output. $\mathcal{A}_o^{0,1}(x,y)$ associates the two Cs and V in C0 with the syllable in C1, respectively. In this way, phonological structure building can be seen as an interpretation of a more basic structure.

Defining phonological processes with logical transductions allows us to measure computational complexity within the regular upper bound of phonology. Chandlee (2014) and Chandlee and Heinz (2018) showed that local phonological processes can be defined using input strictly local (ISL) functions, which are a proper subset of regular functions and are characterized by quantifier-free (QF) first-order logic. Chandlee and Jardine (2019b) showed that the subsequential functions for both local and long-distance phonological processes can be better characterized using QF first-order logic with a least fixed-point operator (QFLFP), further restricting them to a subset of the subsequential functions. As most phonological mappings are ISL (Chandlee, 2014; Chandlee and Heinz, 2018) and thus QF-definable (Chandlee and Lindell, to appear), a strong initial hypothesis for tone-TBU mappings in intonation is that they should be QF-definable. We investigate this hypothesis below.

3 Intonation as quantifier-free interpretation

Now turning to the intonational structures, we define a logical interpretation for intonation. Importantly, tones in intonational melodies are viewed as *copies* of elements in the metrical and prosodic structure, such as accented TBUs or boundaries. The source of intonational melodies is computationally defined as prosodic elements, but they are associated with their *local* TBUs in order to be realized as the actual tones.

Intonation involves two key stages of transduction: first, creating tonal slots with unspecified tones (Ts) via a *melodic transduction*, and second, filling these slots with specified tonal sequence with Hs and Ls via a *declarative meaning transduction*. While this section primarily focuses on outlining the properties and relations for melodic transduction, the details of the meaning transduction will be specified for each intonational pattern following the melodic stage.

For the melodic transduction, the input signature ($\mathcal{S}_i$) is $\{\sigma, \sigma^*, \ltimes_\varphi, \ltimes_\iota, \ltimes_\varphi, \ltimes_\iota, p, s, p^*, s^*\}$ and the output signature ($\mathcal{S}_o$) is $\{\sigma, \sigma^*, T, T^*, \ltimes_\varphi, \ltimes_\iota, \rtimes_\varphi, \rtimes_\iota, \mathcal{A}, p, s, p^*, s^*\}$, where each property and relation symbol in the signature is as follows: σ and σ^* for TBUs; $\rtimes_\varphi$ and $\ltimes_\varphi$ for ip boundary; $\rtimes_\iota$ and $\ltimes_\iota$ for IP boundary; T for tones other than pitch accent tones (nonstarred tones); T^* for pitch accent tones (starred tones). $\mathcal{A}$ is a binary association relation for tone and TBU.

For the unary relations, we can find the set of positions for each symbol with a variable x in the input structure. For example, $\sigma(x)$ is true when x is a syllable; $T(x)$ is true when x is a tone, etc.

As for the binary relations, in addition to p and s, we also define special *predecessor* and *successor* functions, p^* and s^*, to define the relations in the tier that is projected from the set of the selected elements such as metrically strong TBUs and phrasal boundaries. We use two tiers to represent a metrical grid: one for all the strings and the other for the starred elements and phrasal boundaries, as shown in Table 1. While the *nonstarred* function $p(x)$ works locally on the first tier, the *starred* function $p^*(x)$ works locally in the second tier. Similarly, $s(x)$ and $s^*(x)$ work the same way but in different directions.

$p^*(x)$	$\rtimes_\iota$	$\rtimes_\varphi$		σ_*		σ_*		$\ltimes_\varphi$	$\ltimes_\iota$
$p(x)$	$\rtimes_\iota$	$\rtimes_\varphi$	σ	σ	σ	σ	σ	$\ltimes_\varphi$	$\ltimes_\iota$
	$\rtimes_\iota$	$\rtimes_\varphi$	ən	ɔ*	ɹɪnʤ	bɔl*	gaʊn	$\ltimes_\varphi$	$\ltimes_\iota$

Table 1: A metrical grid using a tier-based representation.

Now, we will now look at three case studies, each focusing on a different intonational pattern.

3.1 American English

3.1.1 Basic intonational pattern

American English is a *head-prominence* intonational language (Beckman and Pierrehumbert, 1986), where *metrically strong positions* receive pitch accents in a phrase. For example, as shown in (2), the accented syllables (σ^*) are associated with pitch accents (H*) within an ip. A phrase tone (L-) is also associated at the right edge of the ip. Within an IP, the largest prosodic domain, a boundary tone (L%) is also associated with the right edge of the IP. The actual f0 contour of an English declarative for (2) is provided in Figure 2.

(2) $\rtimes_\iota$ $\rtimes_\varphi$ σ σ^* σ σ^* σ $\ltimes_\varphi$ $\ltimes_\iota$

H* H* L- L%

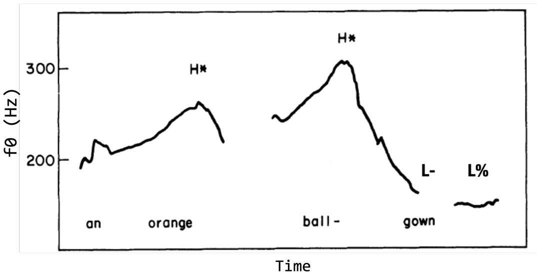


Figure 2: An actual f0 contour of a declarative intonational pattern in American English, extracted from Beckman and Pierrehumbert (1986).

3.1.2 Melodic transduction

Step 1: Copying The input is a string that consists of syllables (σ, σ^*) and boundaries ($\rtimes_\iota/\ltimes_\iota$, $\rtimes_\varphi/\ltimes_\varphi$). As defined in the formulas below, the outputs are four *copies* of the input, which are also illustrated in Figure 3. For the first copy (C0), everything in the input is copied such that syllables and ip and IP boundaries in the output are interpreted the same as those in the input.

$$\sigma_o^0(x) \overset{\text{def}}{=} \sigma_i(x) \qquad \sigma_o^{*0}(x) \overset{\text{def}}{=} \sigma_i^*(x)$$
$$\rtimes_{\varphi_o}^0(x) \overset{\text{def}}{=} \rtimes_{\varphi_i}^0(x) \qquad \ltimes_{\varphi_o}^0(x) \overset{\text{def}}{=} \ltimes_{\varphi_i}^0(x)$$
$$\rtimes_{\iota_o}^0(x) \overset{\text{def}}{=} \rtimes_{\iota_i}^0(x) \qquad \ltimes_{\iota_o}^0(x) \overset{\text{def}}{=} \ltimes_{\iota_o}^0(x)$$

In the formulas for the remaining copies (C1-C3) below, only starred syllables and boundaries are copied and interpreted as tones, reflecting the head-prominence characteristics of American English intonational patterns. In C1, starred syllables in the input, $\sigma_i^*(x)$, are realized as pitch accents in the output, $T_o^{*1}(x)$. In C2, ip boundary at the right edge, $\ltimes_{\varphi_i}(x)$, is realized as a phrasal tone, $T_o^2(x)$. In C3, IP boundaries at the left or right edge, $\ltimes_{\iota_i}(x) \vee \rtimes_{\iota_i}(x)$, are realized as boundary

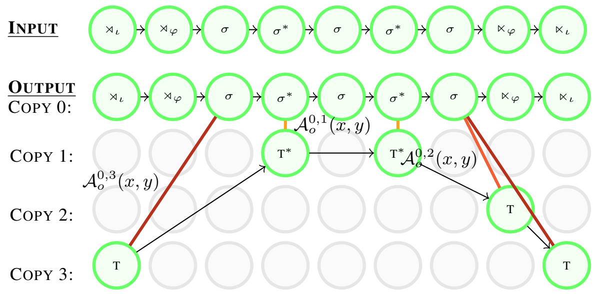


Figure 3: Melodic transduction of American English intonation.

tones, $T_o^3(x)$.

$$T_o^{*1}(x) \stackrel{\text{def}}{=} \sigma_i^*(x)$$
$$T_o^2(x) \stackrel{\text{def}}{=} \ltimes_{\varphi_i}(x)$$
$$T_o^3(x) \stackrel{\text{def}}{=} \ltimes_{\iota i}(x) \vee \rtimes_{\iota_i}(x)$$

Thus, tones in American English are direct copies of starred syllables and phrasal boundaries.

Step 2: Tone-TBU association Importantly, tones in the melodic tiers (C1-C3) are associated with syllables in the segmental tier (C0), as defined below. $\mathcal{A}_o^{0,1}(x,y)$ specifies the association between pitch accents in C1 and their TBUs in C0 if they are at the same position in the input. For phrasal and boundary tones, tones are linked to syllables near boundaries. Specifically, $\mathcal{A}_o^{0,2}(x,y)$ defines the association between phrasal tones at the right edge and the phrase-final syllables just before that edge. Similarly, $\mathcal{A}_o^{0,3}(x,y)$ links boundary tones to their TBUs: tones from the left edge are linked to the first syllable, while those from the right edge are linked to the last syllable in an utterance. Thus, tone-TBU association is computed using only predecessor or successor functions, showing a local logical characterization without quantifiers.

$$\mathcal{A}_o^{0,1}(x,y) \stackrel{\text{def}}{=} x \approx y$$
$$\mathcal{A}_o^{0,2}(x,y) \stackrel{\text{def}}{=} \sigma_i(x) \wedge \ltimes_{\varphi_i}(y) \wedge y \approx s(x)$$
$$\mathcal{A}_o^{0,3}(x,y) \stackrel{\text{def}}{=} (\sigma_i(x) \wedge \rtimes_{\iota_i}(y) \wedge y \approx p(p(x))) \vee (\sigma_i(x) \wedge \ltimes_{\iota_i}(y) \wedge y \approx s(s(x)))$$

3.1.3 Declarative meaning transduction

In the melodic transduction, we have made the slots for the tones that are associated with their TBUs. The remaining step is to compute the *meaning* of a declarative sentence in English, which is specified as H* H* L- L% tonal sequence in Figure 2. As shown Figure 4, we use another simple transduction that changes the unspecified tones (T/T^*) into actual tones (H^*/L), using these simple formulas: $H_o^*(x) = T_i^*(x)$ and $L_o(x) = T_i(x)$.

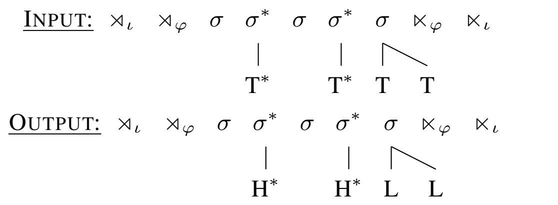


Figure 4: Declarative meaning transduction of American English intonation.

With these melodic and declarative transductions, we can logically define the intonational tones associated with their TBUs in the output based on the strings in the input.

3.1.4 Summary

Results showed that American English intonation can be defined as a QF logical interpretation of a metrical and prosodic structure. The melodies in the output were copies of starred syllables and boundaries in the input. Crucially, copying the starred syllables was able to capture the head-prominence characteristic in American English intonation, showing that the pitch accents in the melodies were the direct reflections of the *heads* of the prosodic unit – starred syllables. Also, the tone-TBU associations were defined *locally* from the input structure without using any quantifiers.

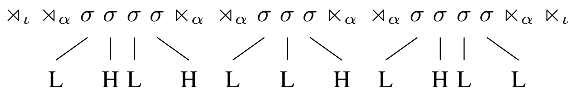
3.2 Seoul Korean

3.2.1 Basic intonational pattern

Seoul Korean is an *edge-prominence* intonational language (Jun, 2006), where phrasal boundaries are marked with prominence without any pitch accents. Basically, a typical tonal pattern is LH...LH in an Accentual Phrase (AP). But when the initial segment of an AP is an aspirated or a tense consonant, the tonal pattern is HH...LH. An Intonational Phrase (IP) consists of more than one AP.

In (3), LH tones are associated with the first two and last two syllables. However, in the final AP, the L% boundary tone overrides the phrase-final H tone at the end of an utterance. If a phrase has fewer than four syllables, one of the tones may not be realized. Edge tones—LH at the left edge and LH at the right edge—plays a crucial role in the intonational pattern of Seoul Korean. An actual f0 contour of a Korean declarative for (3) is provided in Figure 5.

(3)



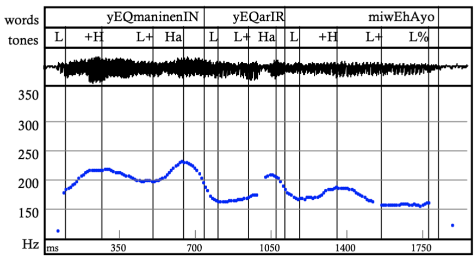


Figure 5: An actual f0 contour of one of the declarative intonational patterns in Seoul Korean, extracted from Jun (2006).

3.2.2 Melodic transduction

Step 1: Copying The input is a string that consists of syllables and boundaries. The outputs are four *copies* of the input, defined in the formula below. As shown in Figure 6, for C0, everything in the input is copied such that syllables and boundaries in the output is interpreted the same as those in the input. In Seoul Korean, the stiffness feature for aspirated or tense consonants ([+stiff]) is specified in the syllable, allowing retrieval during tonal contour computation (e.g., HH...LH).

$$\sigma^0_o(x) = \sigma_i(x) \qquad \sigma^0_{F_o}(x) = \sigma_{F_i}(x) \ (F = \text{[+stiff]})$$
$$\ltimes^0_{\alpha_o}(x) = \ltimes^0_{\alpha_i}(x) \qquad \ltimes^0_{\iota_o}(x) = \ltimes^0_{\iota_i}(x)$$
$$\rtimes^0_{\alpha_o}(x) = \rtimes^0_{\alpha_i}(x) \qquad \rtimes^0_{\iota_o}(x) = \rtimes^0_{\iota_i}(x)$$

As for C1-C3, only boundaries are copied and interpreted as tones, showing a crucial characteristic for the edge-prominence intonational property. Both C1 and C2 shows that AP boundaries at the left or right edge in the input, $\rtimes_{\alpha_i}(x) \vee \ltimes_{\alpha_i}(x)$, are realized as tones in the output, $T^1_o(x)$ and $T^2_o(x)$. A boundary at the end of an utterance, $\ltimes_{\iota_i}(x)$, is realized as a boundary tone, $T^3_o(x)$.

$$T^1_o(x) = \rtimes_{\alpha_i}(x) \vee \ltimes_{\alpha_i}(x)$$
$$T^2_o(x) = \rtimes_{\alpha_i}(x) \vee \ltimes_{\alpha_i}(x)$$
$$T^3_o(x) = \ltimes_{\iota_i}(x)$$

Thus, tones in Seoul Korean are simply direct copies of elements in the prosodic structure, which are only phrasal boundaries.

Step 2: Tone-TBU association The tones in the melodic tiers (C1-C3) are linked to the syllables in the segmental tier (C0). First, $\mathcal{A}^{0,1}_o(x,y)$associates a phrasal tone in C1 with either the first syllable of an AP or the second-to-last syllable of an AP in C0. $\mathcal{A}^{0,2}_o(x,y)$ links a phrasal tone in C2 to the second syllable of an AP, if it is preceded by a left edge of an AP or followed by the last syllable of an AP in C0. Finally, $\mathcal{A}^{0,3}_o(x,y)$ links a boundary tone in C3 to the last syllable before the boundary. The boundary tone in C3 overrides the AP-final phrasal tone in C2, reflecting the hierarchy of boundary tones over phrasal tones.

$$\begin{aligned}
\mathcal{A}^{0,1}_o(x,y) &= \sigma(x) \wedge (\rtimes_\alpha(y) \wedge\ y \approx p(x)) \\
&\quad \vee (\ltimes_\alpha(y) \wedge\ y \approx s(s(x))) \\
\mathcal{A}^{0,2}_o(x,y) &= \sigma(x) \wedge \neg(\ltimes_\iota(y) \wedge y \approx s(s(x))) \\
&\quad \wedge (\rtimes_\alpha(y) \wedge y \approx p(p(x))) \\
&\quad \vee (\rtimes_\alpha(y) \wedge y \approx s(s(x))) \\
\mathcal{A}^{0,3}_o(x,y) &= \sigma(x) \wedge \ltimes_\iota(y) \wedge y \approx s(s(x))
\end{aligned}$$

3.2.3 Declarative meaning transduction

After the melodic transduction, the unspecified tones (Ts) are filled with Hs and Ls for the declarative in Seoul Korean, as shown in Figure 7. The input signatures are $\{\sigma, \ltimes_\varphi, \ltimes_\iota, \rtimes_\varphi, \rtimes_\iota, T\}$ and the output signatures are $\{\sigma, \ltimes_\varphi, \ltimes_\iota, \rtimes_\varphi, \rtimes_\iota, H, L\}$. The formulas are as follows: $L_o(x) = T_i(x) \wedge (\rtimes_\alpha(p(x)) \vee \ltimes_\alpha(s(s(x))))$ and $H_o(x) = T_i(x) \wedge (\rtimes_\alpha(p(p(x))) \wedge \neg H(s(x))) \vee \ltimes_\alpha(s(x)))$.

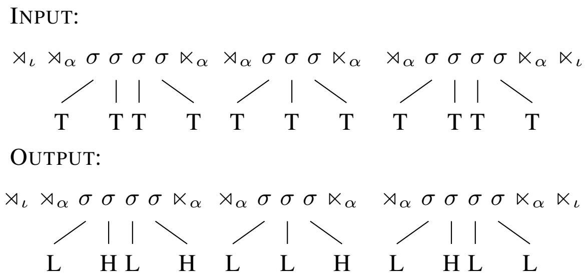


Figure 7: Declarative meaning transduction of Seoul Korean intonation.

3.2.4 Summary

Seoul Korean intonational pattern can be defined using logical interpretation of a prosodic structure. The melodies in the output were copies of *only* boundaries from the input, capturing the edge-prominence characteristic of Seoul Korean intonation. This reflects the edge tones as direct representations of phrasal edges. Similar to American English, the tone-TBU associations were defined *locally* from the input without quantifiers.

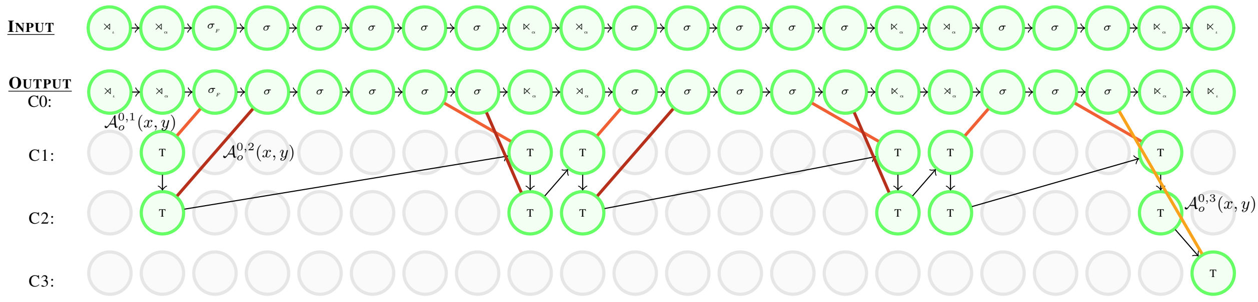


Figure 6: Melodic transduction of intonation in Seoul Korean

3.3 Tokyo Japanese

3.3.1 Basic intonational pattern

Tokyo Japanese is a lexical pitch accent language (Beckman and Pierrehumbert, 1986), where tones are lexically specified for particular moras, while other tones are defined in the phrase-level. The typical intonational pattern in Tokyo Japanese is a rising pitch pattern at the beginning of an Accentual Phrase (AP), which depends on where the lexical pitch accent H*L is realized. The actual f0 contour of a Japanese declarative for (4) is shown in Figure 9.

(4)

$\rtimes_\iota\ \rtimes_\alpha\ \mu\ \ \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \ltimes_\alpha \rtimes_\iota\ \rtimes_\alpha\ \mu\ \ \mu\ \mu\ \mu\ \mu\ \mu\ \ltimes_\alpha\ \ltimes_\iota$

L H* L   L   H* L   L

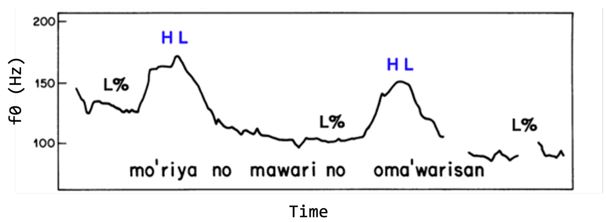


Figure 9: An f0 contour for a declarative intonation in Tokyo Japanese, extracted from Beckman and Pierrehumbert (1986).

When the first syllable of the first lexical item in an AP is *accented*, H*L is associated to the first mora of the accented syllable, with H* realized on the first mora and L on the second. This realization prevents an L% boundary tone and a phrasal H tone from associating with the first and second moras of the AP. Instead, the L% boundary tone of the preceding AP is linked to its final mora rather than the first mora of the current AP.

When the first syllable of the first lexical item in an AP *unaccented* (e.g., omáwarisan), a phrasal H tone is usually linked to the second sonorant mora and L% boundary tone of the preceding AP is associated to the first mora of the following AP.

Lastly, L% boundary tone is inserted at the beginning of the utterance as a whole. A postlexical rule deletes all accents after the first accent in an AP, which is known as deaccentuation.

3.3.2 Melodic transduction

Step 1: Copying The input is a string that consists of moras (μ, μ^*) and boundaries ($\rtimes_\iota$/$\ltimes_\iota$, $\rtimes_\alpha$/$\ltimes_\alpha$), as defined below. The outputs are five *copies* of the input, as shown in Figure 5. For C0, everything in the input is copied such that moras and boundaries in the output are interpreted the same as those in the input.

$$\mu_o^0(x) = \mu_i(x) \qquad \mu_o^{*0}(x) = \mu_i^*(x)$$
$$\ltimes_{\alpha_o}^0(x) = \ltimes_{\iota_i}^0(x) \qquad \ltimes_{\iota_o}^0(x) = \ltimes_{\iota_i}^0(x)$$
$$\rtimes_{\alpha_o}^0(x) = \rtimes_{\alpha_i}^0(x) \qquad \rtimes_{\iota_o}^0(x) = \rtimes_{\iota_i}^0(x)$$
$$H_o^{*1}(x) = \mu_i^*(x) \qquad L_o^2(x) = \mu_i^*(x)$$
$$T_o^3(x) = \rtimes_{\alpha_i}(x) \qquad T_o^4(x) = \rtimes_{\iota_i}(x) \wedge \ltimes_{\alpha_i}(x)$$

In C1 and C2, the HL lexical pitch accents ($H_o^{*1}(x)$ and $L_o^2(x)$) in the output are derived directly from the starred moras ($\mu_i^*(x)$) in the input, as they are lexically specified. This allows the actual HL tones to be computed in the output without creating unspecified tone slots like T. In C3 and C4, phrasal tones ($T_o^3(x)$) are derived from the left edge of an AP boundary ($\rtimes_{\alpha_i}(x)$), while boundary tones ($T_o^4(x)$) are derived from the left edge of an IP boundary ($\rtimes_{\iota_i}(x)$) or the right edge of an AP boundary ($\ltimes_{\alpha_i}(x)$). This direct mapping of input moras to lexical pitch accents and unspecified tones to post-lexical tones reflects Tokyo Japanese's pitch accent patterns.

Step 2: Tone-TBU association The tones in the melodic tiers (C1-C4) are associated with moras in the segmental tier (C0). For lexical pitch accents in the last AP, only the first pitch accent sequence (H^* in C1 and L in C2) is realized, while others are deaccented. This association is defined by

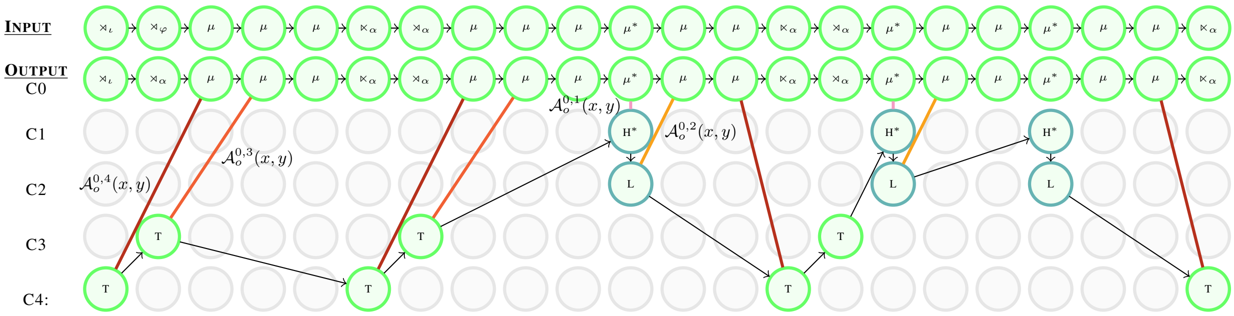


Figure 8: Melodic transduction of Tokyo Japanese intonation.

$\mathcal{A}_o^{0,1}(x,y)$, linking the first starred mora after the left edge of an AP boundary with the H^* using the p^* function. Similarly, $\mathcal{A}_o^{0,2}(x,y)$ links L to the next mora. Subsequent pitch accent sequences in the last AP are not associated with their TBUs. $\mathcal{A}_o^{0,3}(x,y)$ associates the phrasal tones with the second mora in an AP only when not followed by a lexical pitch accent. Therefore, if the following elements are the lexical pitch accents, the phrasal tones cannot be realized. As for the boundary tones, $\mathcal{A}_{\text{o}}{}^{0,4}(x,y)$ associates the boundary tones with the first mora in an AP or with the last mora of the preceding AP or the final AP.

$$\mathcal{A}_o^{0,1}(x,y) = \mu_i^*(x) \wedge \rtimes_{\alpha_i}(y) \wedge y \approx p^*(x)$$
$$\mathcal{A}_o^{0,2}(x,y) = \mu_i(x) \wedge \rtimes_{\alpha_i}(y) \wedge y \approx p^*(x)$$
$$\mathcal{A}_o^{0,3}(x,y) = \mu_i(x) \wedge (\rtimes_{\alpha_i}(y) \wedge y \approx s(s(x))) \wedge \neg(\mu^*(y) \wedge y \approx s(x))$$
$$\mathcal{A}_o^{0,4}(x,y) = \mu_i(x) \wedge (\rtimes_{\iota_i}(y) \vee \rtimes_{\alpha_i}(y) \wedge y \approx p(p(x))) \vee (\ltimes_{\alpha_i}(y) \wedge y \approx s(x))$$

3.3.3 Declarative meaning transduction

After the melodic transduction, the unspecified post-lexical tones (Ts) are filled with Hs and Ls for the declarative in Tokyo Japanese in Figure 10. Note that the lexical pitch accents are already filled with H^* and L. The input signatures are $\{\mu, \mu^*, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, T, H^*, L\}$ and the output signatures are $\{\mu, \mu^*, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, H^*, H, L\}$. The formula is as follows: $L_o(x) = T_i(x)$.

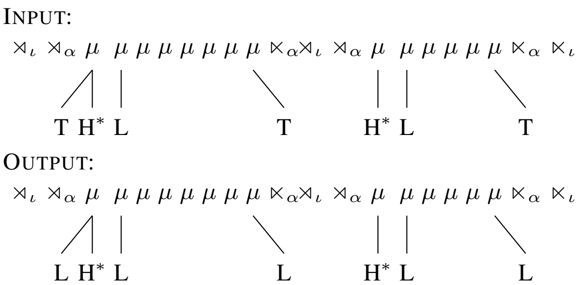


Figure 10: Declarative meaning transduction of Tokyo Japanese intonation.

3.3.4 Summary

Results showed that the intonational pattern in Tokyo Japanese can be defined using a QF logical interpretation of a prosodic structure. Unlike the post-lexical (head-prominence and edge-prominence) intonational patterns in American English and Seoul Korean, copying starred moras *directly* to specified tones—H* and L—was able to capture the lexically specified pitch accent in Tokyo Japanese. Also, copying boundaries was able to capture the realization of post-lexical (phrasal) tones. This process reflects the typical initial rising pitch in an AP in Tokyo Japanese. Even with deaccentuation, where only the first lexical pitch accent in an AP is realized, tone-TBU associations were defined *locally* without quantifiers, by making reference to tier-based representation.

4 Discussion

By defining the intonational structure as a QF logical interpretation of a metrical and prosodic structure that are ISL, we were able to create an *intonational theory* that is restrictive enough to characterize different intonational patterns.

From the typological view of intonation, the head-prominence intonational pattern in American English was defined with the copies of both starred syllables (i.e., heads) and boundaries, whereas the edge-prominence pattern in Seoul Korean was defined with the copies of only boundaries (i.e., edges). The lexical pitch accent pattern in Tokyo Japanese was defined with both copies of starred moras for the lexical pitch accent and copies of phrasal boundaries for the post-lexical tones.

This suggests that the prosodic elements in the input strings are not realized the same way, but the way they are logically interpreted leads to the characterization of different metrical and prosodic realizations in intonation.

Crucially, the computational nature of intonational tone-TBU association patterns found to be characterized as QF logical interpretations. As for the Melodic Transduction, the tone-TBU associations in both American English and Seoul Korean were analyzed in a strictly local manner, without the need of quantifiers. Even in the case of Tokyo Japanese, where prosodic elements like starred TBUs and boundaries may appear non-local, the QF logical interpretations are achieved by preserving the input order in the output (Chandlee and Jardine, 2019b) and using tier-based predecessor and successor functions (p^*, s^*). Furthermore, the use of these starred ordering functions captures the hierarchical structure of TBUs, reflecting their relative prominence, in line with the AM theory's view. Even within the class of QF logical interpretations, typological distinctions can be observed (Danis, 2025). The intonational patterns of Tokyo Japanese are found to be more complex, requiring the use of p^* and s^*, whereas those of American English and Seoul Korean can be captured without using such functions.

As for the Declarative Transduction, at least for American English, Seoul Korean, and Tokyo Japanese, H and L sequences were defined using FO logic without quantifiers. Notably, no case required even-numbered starred syllables to be H tones. This result can be extended to Question Transduction with similar tonal sequence except for an H boundary tone at the end of an IP. This QF logical characterization confirmed that intonational patterns are also ISL functions like most of other phonological mappings within the regular upper bound of phonology (Chandlee, 2014; Chandlee and Jardine, 2019b; Chandlee and Lindell, to appear).

Based on these results, we may able to ask several questions to predict the intonational patterns: 1) what kind of prosodic elements are being copied in the output? Is it a head of a constituent? Is it a phrasal boundary? Or are they both?; 2) when are the tones specified during the derivation from the input to the output? Is it directly specified from the input to the output in a melodic transduction? Or is it specified during the meaning transduction? These questions can provide valuable predictions of possible intonational patterns in the typology.

Further research is needed to generalize the locality of intonational patterns by examining more languages within the same intonational categories. For instance, Spanish is another head-prominence intonational language (Beckman et al., 2002), where the stressed syllable receives pitch accents (e.g., H*, L*+H) within an ip, and boundary tones (L%, H%) are realized at the end of an IP. The intonational pattern in Spanish may possibly seem to function similarly to that in American English, as the heads of constituents serve as main prosodic elements. In contrast, French is known for marking prominence at the edges of an AP (/LHiLH*/ (Jun and Fougeron, 2000), where the phrase-final H* on the last full vowel signals the edge of an AP, while the initial accent Hi is optionally realized. Boundary tones (H%, L%) are realized on the final syllable of an IP. The phrase-final edge-prominence properties in French can be compared to those in other edge-prominence languages like Seoul Korean.

As for lexical pitch accent patterns, Lekeitio Basque may exhibit similar patterns as in Tokyo Japanese. That is, in Lekeitio Basque, a H*+L lexical pitch accent is realized in an AP and a %L boundary tone is realized on the first syllable of an AP (Elordieta, 1998). An IP begins and ends with boundary tones (L%, H%). Due to the absence of a deaccentuation pattern, tonal computation in Lekeitio Basque may be less complex than in Tokyo Japanese. Likewise, we need further analyses on the intonational pattern of other languages to generalize our results that intonation is a QF logical interpretation of a metrical and prosodic structure that are defined locally. But in this way, we can provide a theory of intonation that makes restrictive predictions about the typology of intonation and measure the complexity of intonational structures.

5 Conclusion

The present study explored how the tone-TBU association patterns in intonation can be defined using a QF logical interpretation of a metrical and prosodic structure. Tones were construed as literal copies of prosodic elements, such as starred syllables or boundaries, and their associations with TBUs were defined locally without quantifiers. Head-and edge-prominence intonational patterns were QF metrical grids, whereas lexical pitch accent patterns were more complex. By defining intonation as a logical interpretation, we were able to understand the computational nature of intonation and predict the typology of intonation, contributing the theory of intonational and computational phonology.

References

Mary E. Beckman, Manuel Díaz-Campos, Joanne T. McGory, and Terrell A. Morgan. 2002. Intonation across spanish, in the tones and break indices framework. In Carlos Gussenhoven and Natasha Warner, editors, *Laboratory Phonology 7*, pages 545–576. Mouton de Gruyter, Berlin.

Mary E. Beckman and Janet B. Pierrehumbert. 1986. Intonational structure in japanese and english. *Phonology*, 3:255–309.

Jane Chandlee. 2014. *Strictly Local Phonological Processes*. Ph.D. thesis, University of Delaware.

Jane Chandlee and Jeffrey Heinz. 2018. Strict locality and phonological maps. *Linguistic Inquiry*, 49(1):23–60.

Jane Chandlee and Adam Jardine. 2019a. Autosegmental input strictly local functions. *Transactions of the Association for Computational Linguistics*, 7:157–168.

Jane Chandlee and Adam Jardine. 2019b. Quantifier-free least-fixed point functions for phonology. In *Proceedings of MOL 16*, pages 50–62.

Jane Chandlee and Adam Jardine. 2021. Input and output locality and representation. *Glossa: A Journal of General Linguistics*, 6(1):90.

Jane Chandlee and Steven Lindell. to appear. Logical perspectives on strictly local transformations. In Jeffrey Heinz, editor, *Doing Computational Phonology*. Oxford University Press.

Bruno Courcelle. 1994. Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science*, 126:53–75.

Nick Danis. 2025. Logical transductions are not sufficient for notational equivalence. In *Proceedings of the Annual Meetings on Phonology*, volume 1. University of Massachusetts Amherst Libraries.

Jason Eisner. 1997. What constraints should ot allow? Paper presented at the Annual Meeting of the Linguistic Society of America, January 1997, in Chicago.

Gorka Elordieta. 1998. Intonation in a pitch accent variety of basque. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, 32(2):511–569.

Herbert. B. Enderton. 2001. *A Mathematical Introduction to Logic*. Elsevier.

Joost Engelfriet and Hendrik Jan Hoogeboom. 2001. Mso definable string transductions and two-way finite-state transducers. *ACM Transactions on Computational Logic (TOCL)*, 2(2):216–254.

Emmanuel Filiot and Pierre-Alain Reynier. 2016. Transducers, logic and algebra for functions of finite words. *ACM SIGLOG News*, 3(3):4–19.

Jeffrey Heinz and William Idsardi. 2011. Sentence and word complexity. *Science*, 333(6040):295–297.

Brett Hyde. 2008. Alignment continued: distance-sensitivity, order-sensitivity, and the midpoint pathology. Manuscript, Washington University.

Adam Jardine. 2017. On the logical complexity of autosegmental representations. In *Proceedings of MOL 15*, pages 22–35.

Adam Jardine, Natalie Danis, and Luca Iacoponi. 2021. A formal investigation of q-theory in comparison to autosegmental representations. *Linguistic Inquiry*, 52(2):333–358.

C. Douglas Johnson. 1972. *Formal aspects of phonological description*. Mouton, The Hague.

Sun-Ah Jun. 2006. Intonational phonology of seoul korean revisited. *Japanese-Korean Linguistics*, 14:15–26.

Sun-Ah Jun and Cécile Fougeron. 2000. A phonological model of french intonation. In Merle Horne, editor, *Intonation: Analysis, modelling and technology*, pages 209–242. Springer Netherlands, Dordrecht.

Ronald M. Kaplan and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, 20(3):331–378.

Nathan Koser, Christopher Oakden, and Adam Jardine. 2019. Tone association and output locality in non-linear structures. In *Supplemental Proceedings of the 2019 Annual Meeting on Phonology*. Linguistic Society of America.

Leonid Libkin. 2004. *Elements of Finite Model Theory*, volume 41. Springer, Heidelberg.

Janet B. Pierrehumbert. 1980. *The phonology and phonetics of English intonation*. Ph.D. thesis, Massachusetts Institute of Technology.

Kristine Strother-Garcia. 2019. *Using model theory in phonology: A novel characterization of syllable structure and syllabification*. Ph.D. thesis, University of Delaware.