

Elisabeth Selkirk
Department of Linguistics
University of Massachusetts
Amherst, MA 01003 USA

The Syntax-Phonology Interface

1. Introduction

An individual word, pronounced in isolation, will have a characteristic pronunciation. Yet the pronunciation of a sentence, made up of a sequence of words, is not merely a stringing together of these individual pronunciations. The pronunciation of a sentence, i.e. its phonetic realization, is a function of the surface phonological representation of the sentence. Standard models of the organization of a grammar hold that this phonological representation interfaces with the surface syntactic representation of the sentence, the latter being made up of the sequence of the underlying phonological representations of the words and morphemes that are actually pronounced and their organization into a syntactic word and phrase structure. This organization of the grammar permits in principle that the syntactic representation of a sentence may influence its phonological representation, and therefore allows an explanation for why sentence phonology is not simply the phonology of the individual words of the sentence strung together. Consider, for example, the sentence pairs *No people will go* vs. *No, people will go* and *People will go happily* vs. *People will go, happily*. In both cases, commas indicate a difference in syntactic structure within the pairs. That syntactic structure difference is reflected in a difference in pronunciation, specifically here in a difference in intonation. Effects like these of syntax on phonology have been most commonly studied, in part because earlier models of generative grammar sanctioned only effects in this direction. It is also possible that effects might go in the opposite direction, with phonological principles constraining the range of acceptable syntactic representations at the interface; some recent research explores this question.

Two major lines of thinking on the influence of syntax on phonology can be distinguished. The *prosodic structure hypothesis* holds that the phonological representation of a sentence is organized into a prosodic constituent structure which is independent of, but related to, the surface syntactic structure of a sentence (Selkirk 1986, 1995, Nespor and Vogel 1986, Truckenbrodt 1999). It claims moreover that the syntax of a sentence can impinge directly only on this prosodic structure, namely on the organization into phonological words or phrases and on the distribution of prosodic heads (or stresses). Other apparent effects of syntax on phonology, e.g. the choice of segmental variants or the placement of certain phrasal tones, are claimed to be indirect, mediated by the prosodic structure organization. And since the phonetic interpretation of a sentence is based on surface phonological representation, it follows that there are no direct effects of syntax on phonetic form. If the prosodic structure hypothesis embodies the correct theory of the domain structure for sentence phonology, then the theory of grammar must distinguish between the surface phonological representation of a sentence (PR), which would include this prosodic structure, and its surface syntactic representation (PF). The *direct access hypothesis* (Kaisse 1985, Odden 1995), by contrast, holds that phenomena

of sentence phonology and phonetics may be directly defined off of the surface syntactic word and phrase structure. It does not require a theory of grammar that crucially distinguishes the surface syntactic representation of a sentence from its surface phonological representation, at least in its hierarchical structure. Evidence appears to favor the prosodic structure hypothesis.

2. The Prosodic Structure Hypothesis

A prosodic structure is a well-formed labelled bracketing or tree. The constituents of prosodic structure belong to distinct prosodic categories, arranged in a *prosodic hierarchy*:

(1) The Prosodic Hierarchy:

- Utterance (Utt)
- Intonational Phrase (IP)
- Major Phonological Phrase (MaP)
- Minor Phonological Phrase (MiP)
- Prosodic Word (PWd)
- Foot
- Syllable

In the unmarked case, it is claimed, prosodic structure is *strictly layered*, in the sense that a constituent of a higher level in the hierarchy immediately dominates only constituents of the next level down in the hierarchy. In addition, within a prosodic constituent, in the unmarked case, one of the daughter constituents constitutes the *prosodic head*, the locus of prominence or stress. Compelling support for the claim that the domain structure of sentence phonology and phonetics has the formal properties of such a structure comes from evidence of *domain convergence* and *domain layering* within individual languages. Yet a direct access approach could potentially model these effects. Crucial support for the prosodic structure hypothesis comes from the fact that the constraints which define the hierarchical phonological domain structure are heterogeneous in type, including prosodic structure markedness constraints which are properly phonological in character, appealing only to properties of surface phonological representation, and syntax-prosodic structure interface constraints, which call for features of the syntactic representation to be reflected in phonological representation. Effects of prosodic markedness constraints show that the domain structure of surface phonology is not strictly determined by the syntax, only influenced by it.

2.1 Domain Convergence.

In English declarative sentences, the right edge of a major phrase (MaP) in prosodic structure is marked by the presence of a low tone. In (2ab) a L- marks the low target tone found on the final syllable of a MaP.

- (2) a. (No^{L-}) (animals are allowed^{L-}) = No, animals are allowed.
 [÷ænpmlz]
- b. (No animals are allowed^{L-}) = No animals are allowed.
 [ænpmlz]

We also see a segmental reflex of that same MaP organization: a glottal stop appears as the onset of the first syllable of *animals*, [\div ænɪmlz], in (1a), where it is phrase-initial, but not in (1b). This glottalization effect is arguably phonetic in character (Dilly and Shattuck-Hufnagel 1996). In (3ab) we see an additional MaP effect: a contrast in the pronunciation of the function word *to*.

- (3) a. (They're allowed to graze there^L) (by law^L)
 [t̪]
 b. (They're allowed to^L) (by law^L)
 [tuw]

To appears in its stressless, vowel-reduced weak form [t̪] in (3a), where it is medial within the MaP, but in its stressed, full vowel strong form [tuw] when it appears at the right edge of MaP in (3b) due to the ellipsis of the following verb phrase. In English, monosyllabic function words like *to* generally appear in strong form only at the right edge of MaP (or when they bear contrastive stress) (Selkirk 1995). These three different phenomena which converge on major phrase in English together show that both edges of these constituents are simultaneously relevant to defining phonetic and phonological phenomena.

2.1.2 In Bengali sentence phonology (Hayes and Lahiri 1991), MaP is marked at its right edge by a H- tone. The left edge of MaP is the locus of further phenomena: the word that is leftmost in a MaP is the head of that phrase, and a Low pitch accent, L*, is located on the initial syllable of this stressed word:

- (4) (ʃæ^{L*} moli^H) (ra^{L*} m-er bari^H) (i̇^Hu^{L*} ketβ^Hil^{H-L%})
 Shamoli Ram's house entered
 Shamoli entered Ram's house.

In addition, there are segmental assimilation phenomena which apply optionally both within and between words. These assimilations operate across the span of the MaP, but are blocked if the sequence of segments belongs to different phrases:

- (5) (ø^{L*} mor^H) (tβa^{L*} dor^H) (ta^{L*} ra-ke^H) (dʲe^{L*} tβ^He^{H-L%})
 [r] [tβ] [r] [t]
 Amor scarf Tara -to gave Amor gave the scarf to Tara.

 (6) (ø^{L*} mor tβador tara-ke^H) (dʲe^{L*} tβ^He^{H-L%})
 [tβtβ] [tt] Amor gave the scarf to Tara.

The faster pronunciation in (6) is organized into fewer MaP than in the more deliberate pronunciation of the same sentence in (5). This difference in domain organization is indicated both by the patterns of tone distribution and by the assimilation of /r/. This Bengali phrasal domain, then, shows right and left domain-edge phenomena, as well as phenomena defined on the prominent head of the domain and across the span of the domain.

2.2 Domain Layering

The edge of a clause in the syntax typically coincides with the edge of what prosodic structure theory calls the intonational phrase (IP), the prosodic constituent immediately superordinate to the MaP in the prosodic hierarchy. In English, a final rising tonal contour on the last syllable of a word (notated $L-H\%$), sometime referred to as the continuation rise, is taken to indicate the presence of the right edge of IP (Beckman and Pierrehumbert 1986), indicated with curly braces.

- (7) $\{($ (Since herds of grazing cows $^{L-}$) (have been allowed in her meadow $^{L-}$) $^{H\%}$),
 $\}$
 $\{$ (her flowers have all disappeared $^{L-}$) $\}$
 $\}$
- [h|dz] [v] []
[h] [v]

The right edge of an IP is typically accompanied by a short pause or a significant lengthening of the final syllable. In English (and other languages) the left edge of IP is moreover the site of phonetic strengthening effects that are potentially different in kind or in degree from those seen with the major phrase (Keating et al 1998). For example, an underlying /h/ in English will typically fail to be pronounced if it is in an unstressed syllable (compare stressed *herds* to stressless *her* in the first clause), but at the beginning of an IP /h/ must necessarily be pronounced, even in a stressless syllable (compare stressless *her* in the second clause to stressless *her* in the first). By contrast, there is the possibility of absence of /h/ in the stressless *have* which is MaP-initial in the first clause. This difference in appearance of [h] provides further testimony of the difference in the two levels of domain structure.

In Japanese, a rise (LH) in pitch which is referred to as Initial Lowering is found at the left edge of a minor phrase (MiP) (Poser 1984, Kubozono 1993), indicated with angle brackets in (8b).

- (8) a. Yamaai-no yamagoya-no uraniwa-no umagoya-ni kabi-ga hae $^{H^*}$ -ta
[[[mountain village-GEN] hut-GEN] backyard-GEN] barn-LOC] [mold-NOM] [grow-PAST]
‘The barn in the backyard of a hut in a mountain village grew moldy.’
- b. ($^L <^H$ Yamaai-no yamagoya-no $>^L <^H$ uraniwa-no umagoya-ni $>^L <^H$ kabi-ga hae $^{H^*}$ \downarrow -ta $>^L$) L

(8b) shows an organization into binary MiP. In (9b), by contrast, where all the words have a lexical accent, we see that there are as many instances of MiP, and hence of Initial Lowering, as there are of accents:

- (9) a. Yama $^{H^*}$ -gata-no yama $^{H^*}$ -dera-no one: $^{H^*}$ -san-ga mayone: $^{H^*}$ -zu-o ho:ba $^{H^*}$ -tteiru-wa
[[[Yamagata-GEN] mountain temple-GEN] young-lady-NOM] [[mayonnaise-ACC] filling-her-mouth]
The young lady from mountain temple in Yamagata was filling her mouth with mayonnaise.
- b. ($^L <^H$ yama $^{H^*}$ \downarrow -gata $>^L <^H$ yama $^{H^*}$ \downarrow -cdera $>^L <^H$ one: $^{H^*}$ \downarrow -san-ga $>^L <^H$ mayone: $^{H^*}$ \downarrow -zu-o $>^L$) L
 $^L <^H$ ho:ba $^{H^*}$ \downarrow -tteiruwa $>^L$) L

The difference in patterns of minor phrasing seen in (8) and (9) are the result of prosodic markedness constraints to be discussed below.

At the level of MaP, the level above MiP in the prosodic hierarchy, two tonal phenomena, are defined: catathesis (↓), and upward pitch reset (↑) (Poser 1984, Pierrehumbert and Beckman 1988). Catathesis is an accent-induced downstepping of the pitch range, while upward pitch reset returns the pitch range to a higher level. Only material within the same MaP undergoes catathesis. Upwards reset, seen in both (13) and (14), appears at the left edge of MaP. These examples, then, illustrate the basic strict layering of unmarked hierarchical prosodic structure: here a MaP is parsed into a sequence of MiP.

2.3 Nonsyntactic determinants of phonological and phonetic domain structure

A direct access theory could model patterns of edge-sensitive phonetic and phonological phenomena as long as the edges were definable solely in syntactic terms. But demonstrably nonsyntactic factors on phonological domain organization like prosodic markedness constraints and speech rate show the influence of syntax is indirect.

2.3.1 The interaction of prosodic markedness constraints and interface constraints

In optimality theory (Prince and Smolensky 1993), phonological markedness constraints call for some ideal phonological target output shape. Of particular importance to our topic, markedness constraints compete with a class of constraints called faithfulness constraints, which call for properties of the surface phonological representation to reflect, or be identical to, properties of a related grammatical representation. Syntax-phonology interface constraints can be understood as a variety of faithfulness constraint, one of the input-output variety, if surface syntactic representation (PF) and surface phonological representation (PR) are in an input-output relation, or one of an output-output variety, if the PF and PR relation are evaluated in parallel, not serially.

In the examples of Japanese minor phrasing cited in (8) and (9) prosodic markedness constraints complement the effects of syntax-phonology interface constraints, producing phrasing where it would not be called for by the syntax. A prosodic constraint calling for a MiP to be binary in its PWD composition, call it Binary MiP, results in the binary phrasing seen in (8). Another prosodic constraint, call it MiP Accent, calls for a MiP to contain at most one accent; it is responsible for the appearance of the nonbinary MiP in (14). In Japanese MiP Accent takes precedence over Binary MiP, as in (9), a relation that can be expressed in the grammar of Japanese with an optimality-theoretic constraint ranking: MiP Accent >> Binary MiP.

The ideal binary shape for MiP, which appears in (8), fails to appear in sentences with somewhat different syntactic structures, showing that certain syntax-phonology interface constraints also outrank the Binary MiP markedness constraint in the grammar of Japanese. The simple three word subject-object verb sentence (10) divides into two major phrases (MaP) at the break between subject and object, with the result that the subject consists of a nonbinary minor phrase (MiP), in violation of Binary MiP:

(10) a. Inayama-ga yuujin-o yonda

[[Inayama-NOM] [[friend-ACC] call-PST]]
Mr. Inayama called his friend.

b. (^L<^HInayama-ga>)^L↑(^L<^Hyuujin-o yonda>)^L

The MaP break here is imposed by a syntax-phonology interface constraint expressible as the alignment constraint (Selkirk and Tateishi 1988, 1991):

(11) Align-L (XP, MaP)

The left edge of a maximal projection in syntactic representation (PF) corresponds to the left edge of a major phrase in surface phonological representation (PR).

The noun phrase (12), which consists of an accentless Adjective-Adjective-Noun sequence, consists of a single MaP (as shown by lack of internal upward pitch range rest), but nonetheless shows a MiP break between the two adjectives, leaving the first MiP nonbinary.

(12) a. Amai akai ame-ga
[sweet [red [candies]]]

b. (^L<^Hamai>^L<^Hakai ame-ga>)^L NOT: * (^L<^Hamai akai ame>)^L

At play here is an interface constraint sensitive to syntactic branching (Kubozono 1993):

(13) Align- L (X^{br}, MiP)

The left edge of a branching constituent in syntactic structure corresponds to the left edge of a minor phrase in prosodic structure.

These two syntax-phonology alignment constraints outrank Binary MiP in the grammar of Japanese, causing the violations of Binary MiP seen in (10) and (12):

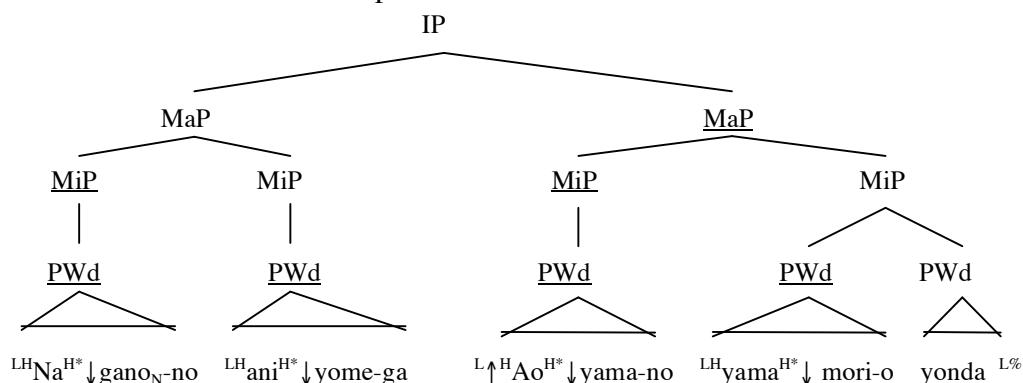
(14) Align-L (XP, MaP), Align- L (X^{br}, MiP) >> Binary MiP

It is because neither of these interface constraints is applicable in the left branching structure of the subject noun phrase in (8) that we see the effects of the lower ranked Binary MiP emerge there. In other words, where the lower ranked Binary MiP is not forced to be violated by the higher ranked interface constraints, it produces effects that are *complementary* to the effects of the interface constraints.

We also find cases where an expected effect of a syntax-phonology interface constraint *fails to appear* because of pressure from a competing prosodic markedness constraint. This can be modelled by ranking the prosodic constraint higher than the relevant interface constraint. (15) shows the expected MaP break at the left edge of VP in a neutral Focus sentence in Japanese.

(15) A neutral Focus sentence in Japanese

PR:

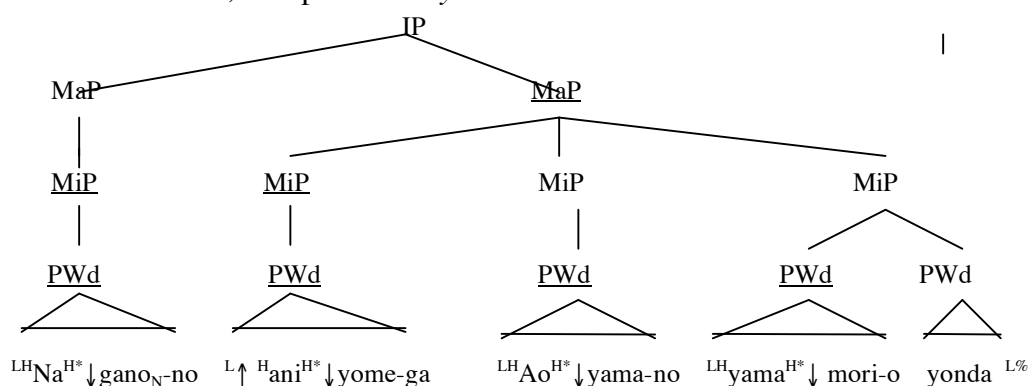


PF: [[[Na^{H*}gano_N-no] ani^{H*}yome-ga]_{NP} [[[Ao^{H*}yama-no] yama^{H*}mori-o] yonda]_{VP}]_S
 [[Nagano-GEN sister-in-law-NOM] [[Aoyama-GEN] mountain guard-ACC] called]]
 A sister-in-law from Nagano called a mountain guard who is in Aoyama.

(16) contains a sentence which differs only in having a contrastive Focus on the last word of the subject NP, *ani'yome*. The phonological domain structure of this sentence shows an absence of the syntactically expected MaP break at the left edge of the VP, as well as the unexpected presence of a MaP break at the left edge of the Focus constituent (Poser 1984, Pierrehumbert and Beckman 1988, Nagahara 1994).

(16) The same sentence, except with *aniyome* as FOCUS

PR:



PF: [[[Na^{H*}gano_N-no] [ani^{H*}yome-ga]_{FOCUS}]_{NP} [[[Ao^{H*}yama-no] yama^{H*}mori-o] yonda]_{VP}]

Truckenbrodt 1995 argues that the source of these phrasing effects in Focus sentences lies in a syntax-phonology interface constraint calling for a Focus to be more prosodically prominent than any other element within the Focus domain.

(17) Focus Prominence

The string in PR which corresponds to a string in PF dominated by Focus must contain a prosodic prominence that is greater than any other prominence contained in the Focus domain.

In the case where the domain of the Focus is the entire sentence, as in (16), respect of this constraint means that the Focus element and all the prosodic phrases that dominate it will be the head(s) of the intonational phrase corresponding to that sentence in PR (underlined). A prominent constituent is subject to a prosodic markedness constraint that calls for it to be aligned with an edge of the superordinate constituent (McCarthy and Prince 1993). The violation in (17) of the interface constraint (11) calling for a MaP edge at the left edge of VP is claimed to be driven by the higher-ranked prosodic constraint Align-R (MaP_{head}, IP). This constraint calls for the prominent MaP in the IP that dominates the Focus in (17) to align with the right edge of the intonational phrase, and so prevents the existence of any other MaP following the Focus. This analysis is captured in the ranking in (18).

(18) Focus Prominence, Align-R (MaP_{head}, IP) >> Align-L (XP, MaP)

So, here, a syntactic effect is overridden by a prosodic effect.

2.4 *Speech rate*

One additional sort of nonsyntactic factor on phonological domain structure that has been recognized for some time in the literature is speech rate. In Bengali, a faster speech rate gives rise to larger phonological phrases (see above). Speech rate plays a role as well in MiP organization in Japanese; in slower speech, single words may easily constitute a MiP (Selkirk and Tateishi 1988). Since syntactic structure does not vary with speech rate, these rate-based differences in phrasing give additional evidence for an autonomous, nonsyntactic, prosodic structural representation of phonological domains.

3. **The Syntax-Prosodic Structure Interface: Serialist or Parallelist?**

The classic generative model of the relation between syntax and phonology is an input-output model, in which surface syntactic representation, the output of the syntactic component, is input to the phonological component, whose output is the surface phonological representation. Most approaches to modelling the syntax-prosodic structure interface have assumed this serialist, uni-directional, input-output model, with the prediction that syntax may influence phonology, but not vice-versa (see, e.g. Miller et al 1997 on phonology-free syntax). They have posited constraints, algorithms or rules relating the two representations which have the general form: If the surface syntactic representation of the sentence has property S, then the surface phonological representation will have property P. In the Align/Wrap theory of the interface (Selkirk 1986, 1995, Truckenbrodt 1999), the demarcational Align class of constraints calls for the edge of any syntactic constituent of type C_s in the surface syntax to align with the edge of a prosodic constituent of type C_p in the surface phonology, while the cohesional Wrap class of constraints calls for a syntactic constituent of type C_s in the surface syntax to be contained within a prosodic constituent of type C_p in the surface phonology. There have been other, relation-based, proposals concerning the nature of syntax-prosodic structure interface constraints as well, including Nespor and Vogel 1986, McHugh 1990, Kim 1997, all understood to be consistent with the prosodic structure hypothesis, and with the input-output architecture.

An input-output model of the syntax-phonology interface makes the essential claim that constraints on phonological output representation do not interact with constraints on the surface syntactic representation. Two sorts of apparent challenges for the input-output model have recurred in recent literature. One sort concerns clitics whose distribution is governed by both syntactic and prosodic constraints, such as the second position clitics of Serbo-Croatian (Inkelas and Zec 1990). Yet because the syntax itself affords options in the positioning of second position clitics, it is not necessary to construe the prosodic subcategorization of a clitic as outranking a constraint on syntactic word order. Another sort of challenge concerns the alleged dependence of word order on the positioning of focus-related prosodic prominences in the sentence, as suggested by Vallduví 1991 for Catalan or Zubizarreta 1998 for Spanish. However, since the latter sort of phenomena can potentially receive a purely syntactic treatment in which the Focus properties of syntactic representation are crucial in determining word order (see e.g. Grimshaw and Samek-Lodovici 1998), this sort of case is not yet compelling. It remains to be seen whether the syntax-phonology interface is an input-output relation or whether it is instead a two-way street.

References

- Beckman, M. and J. Pierrehumbert (1986). Intonational structure in English and Japanese. *Phonology* 3: 255-309.
- Dilley, L. and S. Shattuck-Hufnagel (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24: 423-444.
- Grimshaw, J. and V. Samek-Lodovici (1998). Optimal Subjects and Subject Universals. In *Is the Best Good Enough? Optimality and Competition in Syntax*. P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds.), Cambridge MA, MIT Press: 193-220.
- Hayes, B. and A. Lahiri (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9: 47-96.
- Inkelas, S. and D. Zec (1990). Prosodically Constrained Syntax In *The Phonology-Syntax Connection*. S. Inkelas and D. Zec (eds.). Chicago, CSLI: 365-378.
- Kaisse, E. M. (1985). *Connected Speech: The Interaction of Syntax and Phonology*. San Diego, Academic Press.
- Keating, P., C. Fougeron, et al. (1998). Domain-Initial Articulatory Strengthening. *Papers in Laboratory Phonology* 6. Cambridge, Cambridge University Press.
- Kim, N.-J. (1997). *Tones, Segments, and Their Interaction in North Kyungsang Korean*. Ph. D. dissertation, Department of Linguistics, The Ohio State University.
- Kubozono, H. (1993). *The Organization of Japanese Prosody*. Tokyo, Kurosio Publishers

- McCarthy, J. J. and A. Prince (1993). Generalized Alignment. *Yearbook of Morphology*. G. Booij and J. v. Marle (eds.). Dordrecht, Kluwer: 79-153.
- McHugh, B. (1990). The phrasal cycle in Kirunjo Chaga tonology. In *The Phonology-Syntax Connection*. S. Inkelas and D. Zec (eds.). Chicago, U of Chicago Press: 217-242.
- Nagahara, H. (1994). *Phonological Phrasing in Japanese*, Ph.D dissertation, Department of Linguistics, UCLA.
- Nespor, M. and I. Vogel (1986). *Prosodic Phonology*. Dordrecht, Foris.
- Odden, D. (1995). Phonology at the Phrasal Level in Bantu. In *Bantu Phonology and Morphology*. F. Katamba (ed.). München-Newcastle, LINCOM EUROPA: 40-68.
- Pierrehumbert, J. and M. Beckman (1988). *Japanese Tone Structure*. Cambridge, MA, MIT Press.
- Poser, W. J. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese*, Ph.D dissertation, Department of Linguistics, MIT.
- Prince, A. and P. Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Manuscript, New Brunswick, NJ, Rutgers U. Cntr for Cog. Science.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology* 3: 371-405.
- Selkirk, E. (1995). The prosodic structure of function words. In *Papers in Optimality Theory*. J. Beckman, L. Walsh Dickey and S. Urbanczyk (eds.). Amherst, MA, GLSA Publications: 439-470.
- Selkirk, E. and K. Tateishi (1988). Constraints on minor phrase formation in Japanese. In *Proc. of the 24th Annual Meeting of the Chicago Linguistics Society*, Chicago, CLS.
- Selkirk, E. and K. Tateishi (1991). Syntax and downstep in Japanese. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*. C. Georgopoulos and R. Ishihara (eds.). Dordrecht, Kluwer: 519-544.
- Truckenbrodt, H. (1995). *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*, Ph.D. dissertation, Department of Linguistics, MIT.
- Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219-256.
- Vallduví, E. (1991). The role of plasticity in the association of focus and prominence. In *Proceedings of the Eastern States Conference on Linguistics (ESCOL)* 7, 295-306.
- Zubizarreta, M. L. (1998). *Prosody, Focus and Word Order*. Cambridge, The MIT Press.