

# Interfaces in the prosodic hierarchy: New structures and the phonological parser

**John Archibald**

*University of Calgary*

---

## *Acknowledgments*

I would like to thank Jen Mah for agreeing to have her acquisition of Japanese length contrast tracked so carefully, and for all her work in the lab. I would also like to thank the workshop organizers at the International Symposium on Bilingualism in Bristol, Conxita Lleó and Irene Vogel as well as all the participants for their stimulating discussion. Two anonymous reviewers also deserve credit for making suggestions which greatly improved the paper.

## **Abstract**

In this paper I investigate the assignation of syllabic structure to segments in second language learners invoking principles of parsing and learnability. Drawing on the model of Phillips (1996) in parsing, and the work of Fodor (1999) and Dresner (1999) in learnability, I discuss the implications for second language learning. I also look at the question of whether new phonological structure can be triggered if that structure is lacking in the first language. Drawing on evidence from the acquisition of (1) phonological features in Japanese, Chinese and French learners of English, as well as English learners of Czech, and (2) moraic structure by an English speaker learning Japanese consonant and vowel length contrasts, I argue that second language learners *can* trigger new prosodic structure. The process of acquisition is a combination of acquiring new structures, and mapping the interfaces of different levels of structure via a phonological parser.

## **Key words**

*L2 phonology*

*learnability*

*parsing*

## **1 Introduction**

Any attempt to investigate the acquisition of second language phonology must take seriously the question of what exactly is being acquired. What is the phonological model of the target grammar? In this paper, I am going to focus on the interfaces between certain levels of phonological structure. In particular, I will look at the interaction of segments, syllables, and moras. In order to address these questions, I will also look to a theory of parsing to see how this can inform our study of L2 phonology. Broadly speaking, I will discuss two questions:

1. How are existing phonological elements assigned to structures? In other words, what is the role of the parser, and what are its structural properties?

---

### **Address for correspondence**

Dept. of Linguistics/Language Research Center, Social Sciences 828, University of Calgary, Calgary, Alberta, CANADA, T2N 1N4; e-mail: <archibal@ucalgary.ca>.

---

2. Are new structures that are absent from the L1 triggered in the L2 representations?

I will first begin with a discussion of the question of triggering new structure, and then move on to illustrate the contribution of the parser.

## 2 The acquisition of new structures

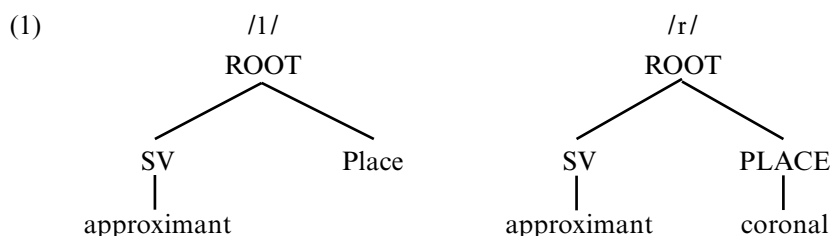
For quite some time, researchers have investigated the question of whether L2 learners can learn new elements that are not in their L1, and the stances adopted have been varied. The literature on the Critical Period Hypothesis (e.g., Singleton, 1989) suggested that postpubertal learners would retain L2 accents because they could not acquire the relevant features of the second language. Brown (2000) argues that if a feature is missing from the L1 that it will not be acquired in the L2. Flege (1992) argues that “new” sounds will be able to be acquired (because a new phonetic category could be set up) whereas “similar” sounds will not be acquired (because they will be assimilated to the L1 phonetic category). LaCharité and Prévost (1999) also suggest that some features that are absent from the L1 may be possible to acquire. Bongaerts, Mennen, & Van der Slik (2000) demonstrate that some postpubertal learners may achieve pronunciation skills that fall within the range of native speakers.

## 3 The acquisition of new features

In order to try to make sense of these varied positions, I would like to place the study of second language phonology within frameworks that are familiar to us from second language syntax. Historically in the field of second language acquisition, the question of whether L2 grammars were constrained by the same principles as other natural languages has oft been asked (e.g., White, 1989). Much of this research was conducted in the domain of syntax. Specifically, I would like to raise the question of Full Access/Full Transfer (Schwartz & Sprouse, 1996). Research such as Brown (2000) and Matthews (1997) would seem to suggest that learners do not have Full Access to UG in phonology insofar as they cannot trigger new features like [coronal] but *can* reset existing structures such as changing the settings of a branching onset, or the class of segments that can appear in a coda, or the type of foot from trochaic to iambic. On the other hand, we have research like Curtin, Goad and Pater (1998) which suggests that English speakers learning Thai, while they initially transfer their L1 lexical entries (which include [voice] but not [aspiration]), *can* change their lexical entries over time to include a feature [aspiration] that is not found in the L1 lexical entries. The Curtin et al. paper is useful in reminding us of the important distinction between a surface discrimination task (which may or may not be a true linguistic task insofar as it does not necessarily involve lexical access) and a deep lexical task. However, it does not illuminate the question of what types of cues are available to the learner in order to trigger new structure. Both Brown and Matthews’ work would seem to suggest that there is no surface cue that could *ever* trigger a feature like [coronal] in the Japanese learners.

## 4 Empirical results against triggering new features (no access)

Brown (2000) argued that Japanese and Chinese learners of English behaved differently when it came to acquiring an [l]/[r] contrast even though both L1's lacked the contrast. The Chinese learners were perceiving the distinction at a level indistinguishable from native speakers, while the Japanese learners were performing at a level close to chance. Brown argued that the differential performance had to do with the nature of their L1 phonological representations. The English liquid contrast is given in (1).

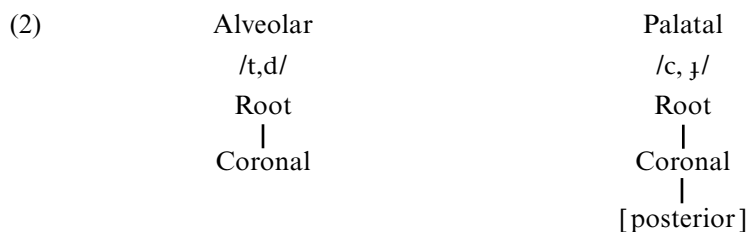


The goal of the learners, then, is to acquire a contrast based on the feature [coronal]. The Chinese speakers are argued to have the feature [coronal] in their phonological inventory in order to represent the contrast between [s] and [ʂ] (an alveolar vs. retroflex fricative). Brown suggests that the Chinese learners are able to build new phonological representations that make use of the existing [coronal] feature whereas the Japanese subjects are unable to trigger the same feature [coronal] as it is absent from their L1 inventory. This is what I will refer to as the redeploying of existing features.

## 5 Empirical evidence for redeploying existing features

Matthews (1997) argues that Japanese learners can improve in their discrimination of certain English sounds when the contrast depends on a feature found in their L1. His subjects showed improvement on the contrasts [b]/[v]; [s]/[θ]; and [θ]/[f] (because Japanese has contrasts which invoke [continuant] and [strident]) but no significant improvement on [l]/[r] (because Japanese lacks the [coronal] feature). Matthews' work, then suggests that L2 learners can improve on some contrasts which are not found in the L1.

Atkey (2000) also argued for the ability of L2 learners to acquire new distinctions that were based on featural content of the L1. She looked at the acquisition of Czech palatal stops by native speakers of English. The relevant Czech stops are shown in (2).



English has the relevant feature ([posterior]) to make the contrast in its fricative series as shown in (3).

(3) /s/	/ʃ/	/θ/
Alveolar	Alveo-palatal	Dental
Root   Coronal	Root   Coronal   [posterior]	Root   Coronal   [distributed]

Chart 1 shows that the subjects were able to accurately identify the alveolar versus palatal contrast (the numbers in parentheses after the subject's initials indicate the time that the subject has been in the Czech Republic).

### Chart 1

Percentage of Czech palatal stop tokens perceived correctly by English speakers

Position	Subjects					
	ML (0;3)	JD (0;5)	AD (0;11)	SW (0;11)	JA (1;0)	RK (10;0)
Initial	70	90	80	85	80	95
Medial	70	70	80	90	85	90
Final	20	30	50	70	70	80

## 6 Empirical evidence that some new features can be acquired (partial access)

LaCharité and Prévost (1999) argue that Brown's claim is too strong in that not *all* features which are inactive in the L1 are not learnable. They suggest that while articulator nodes (such as [coronal] or [dorsal]) might well not be learnable, terminal nodes (such as [distributed] or [back]) can be triggered if they are absent from the L1. Their evidence comes from the acquisition of English by Canadian francophones. They argue that the learners do better in acquiring [θ] (which depends on the terminal feature [distributed]) than they do in acquiring [h] (which depends on the articulatory feature [pharyngeal]).

## 7 Explanations

We must ask the question: why could Brown's Japanese subjects not acquire a contrast that required the feature [coronal]—not found in their L1 lexical entries, while Curtin et al.'s subjects could acquire a contrast that required the feature [aspiration]—also not found in their L1 lexical entries? Note that these are both cases where the learner would have access to positive evidence of the L2 properties. They would hear two liquids, and they would hear aspiration. As Young-Scholten's (1994) work has shown us, there are times when the types of evidence are crucial to our explanations, but this does not seem to be the case here.

It may well be the case then that the claim that “new structure cannot be acquired” is too strong. So, how can we account for these differences?

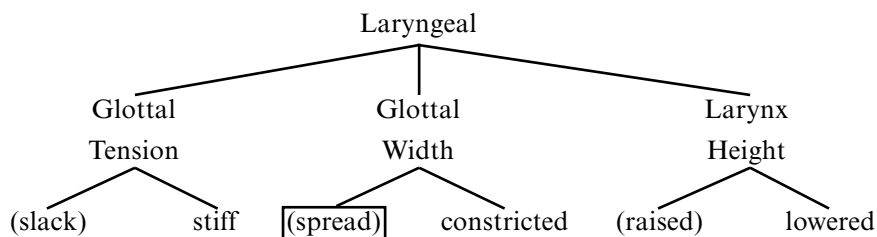
## 8 Phonological saliency

I would argue that we would have to reject the following account: L2 learners can make use of acoustic cues which are salient in the L2 input but they cannot make use of cues which are not salient. Saliency is derived from the mental representation and not just from the acoustic string. The distinction between [l] and [r] is salient to an English speaker because that is a contrastive feature for English speakers. It is not a salient acoustic feature for Japanese subjects because the sound pair is not contrastive in Japanese. Other sounds (like clicks) may well be notable in the acoustic string because they are so unlike any L1 sound, but that does not tell us that the learner is going to be able to set up a lexical entry which includes a nativelike *representation* of how to contrast a click series with other sounds.

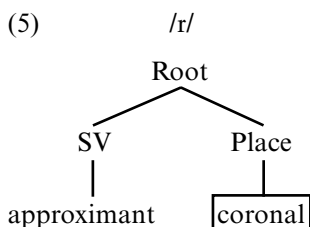
## 9 Feature structure

However, it may well be the fact that the features [aspiration] and [coronal] are found at different levels of embedding in the feature tree. Let us consider the types of structures involved, shown in (4). Avery and Idsardi (2000) present the following model for the structure of the Laryngeal node.

(4) Levels of Feature Structure



We can compare this to the feature structure required for the liquid contrast in English given in (5).



Once again, consistent with LaCharité and Prévost (1999), it seems that learners are better able to acquire contrasts more deeply embedded from the articulator nodes, as in (4).

## 10 Global determinacy

We must also remember the issue of global determinacy (Dresher & van der Hulst, 1995). Stated briefly, the issue of global determinacy acknowledges the fact that a learner has to acquire a mental representation for a sound based on information that may not be strictly local. Consider the difference between the representation of an /i/ in a three-vowel system as opposed to an /i/ in a seven-vowel system. The two vowels both sound like [i] when heard in isolation, but have very different structures. The learner would have to somehow note the difference in behavior of the two [i]s to set up the appropriate mental representation. Does this have any relevance to the aspiration case? We note that aspiration in English has a number of non-local effects, as shown in (6)

### (6) Aspiration in English

Aspiration is:

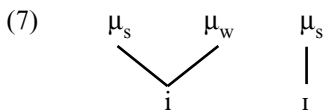
- not found after [s] in English.
- connected to stress placement.
- connected to foot structure.
- related to other VOT phenomena, for example, sonorant devoicing after voiceless stops.

It is possible that the learners in the Curtin et al. (1998) study were able to make use of these non-local cues toward aspiration in their L1 when setting up [aspiration] as part of the lexical entry in the L2.

I would argue that the basic difference may well be that the learner must rely on strictly local cues to acquire certain types of knowledge such as place of articulation (e.g., [uvular]). The perception of such things as formant transitions would be necessary to tell the learner about the articulation of these sounds. However, the learner could make use of global cues of the type shown in (6) to acquire other types of knowledge (e.g., [aspiration]).

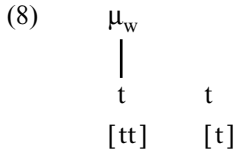
## 11 The acquisition of moraic structure

We now turn to a discussion of another interface: segments and moras. Mah and Archibald (2003) analyze some production data of an English speaker learning Japanese. It is a longitudinal case study that focuses on the acquisition of Japanese vowel and consonant length. English has a contrast between lax, monomoraic and tense, bimoraic vowels (which we could call short vs. long), but no contrast between long and short consonants (as English does not have phonologically contrasting consonantal length). These structures are given in (7).

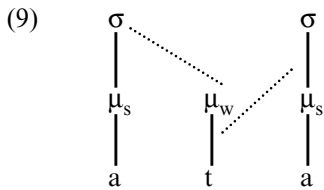


Following Zec (1995), I am assuming the difference between strong and weak moras. Japanese maintains contrasts in both consonant length and vowel length. The issue of representing consonantal length is not straightforward and requires elaboration.

The moraic structure of vowels must be stored as part of the lexical entry as minimal pairs depend on the contrast (e.g., ‘beat’ vs. ‘bit’). However, it has been argued (Hayes, 1989) that non-geminate consonants are not lexically associated with any mora while geminate consonants are associated with a single mora, as shown in (8).

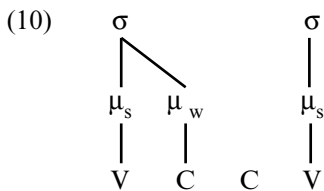


If moras are thought to bear a relation to phonetic timing then this correlates with the fact that long vowels are longer than long consonants. A geminate consonant would end up being associated with both the onset and the coda, as shown in (9).



We note that it is a universal property of vocalic moras to project a syllable node. To capture this, I have invoked Zec’s (1995) distinction of strong versus weak moras. The mora projected by a vowel is strong while the mora projected by a consonant is weak. We will return to this later when looking at the parsing procedures invoked by non-native speakers.

In quantity-sensitive languages, certain coda consonants will project an additional mora, as shown in (10).



This will capture the fact that bimoraic syllables can attract stress. When we compare the legal structures in both English and Japanese, we can see that the English speakers have an L1 grammar in which coda consonants are licensed by a weak mora for reasons of weight. In Japanese, geminate consonants are licensed by a weak mora. Therefore, we predict that English speakers will be able to acquire both the Japanese long vowels and the Japanese long consonants. These predictions are based not on the structure of the feature geometry of the two languages (as in the previous section) but on the licensing, or mapping, properties of the languages in question. We will return to this level of analysis later in the paper when we look at the parsing of consonant clusters.

## 12 Native Japanese ratios of long versus short consonants.

Han (1992) gives the durations for native Japanese geminates, presented in Chart 2.

### Chart 2

#### Japanese native speaker length ratios

Contrast	Ratio
[t] versus [tt]	3.00
[p] versus [pp]	2.71
[k] versus [kk]	2.80

Note then that native speakers' geminate consonants are about three times as long as their single consonants.

The subject of Mah & Archibald (2003) studied introductory Japanese in a university class. In the class there was little attention paid to pronunciation. Long vowels were referred to as "stretched" and geminates were described as "swallowing the first part." The data reported on here were collected after four months of study. Fifteen Japanese sentences were designed to elicit the targeted contrasts. The sentences were written on index cards in hiragana script. The goal here was to have the subject focus on decoding the hiragana rather than focusing on the pronunciation task. The subject read each sentence three times onto a Sony TCD-D100 DAT Recorder with a Sony ECM electret condenser microphone. The data were redigitized at a sampling rate of 22.2kHz using Soundscope 8. Wide-band spectrograms were made of the relevant sentences, and measurements were taken from these. Chart 3 provides the ratio of long to short consonants for the non-native speaker.

### Chart 3

#### Non-native speaker length ratios

Sound	Single	Geminate	Ratio
[t] vs. [tt]	.085	.332	3.95
[p] vs. [pp]	.098	.392	4.00
[k] vs. [kk]	.086	.333	3.87
[ʃ] vs. [ʃʃ]	.171	.304	1.77
[n] vs. [nn]	.102	.291	2.85
[m] vs. [mm]	.088	.296	3.36

Sound	Short	Long	Ratio
[a] vs. [aa]	.118	.295	2.5
[i] vs. [ii]	.106	.341	3.21
[u] vs. [uu]	.082	.219	2.67
[ɛ] vs. [ɛɛ]	.114	.351	3.07
[o] vs. [oo]	.148	.339	2.29

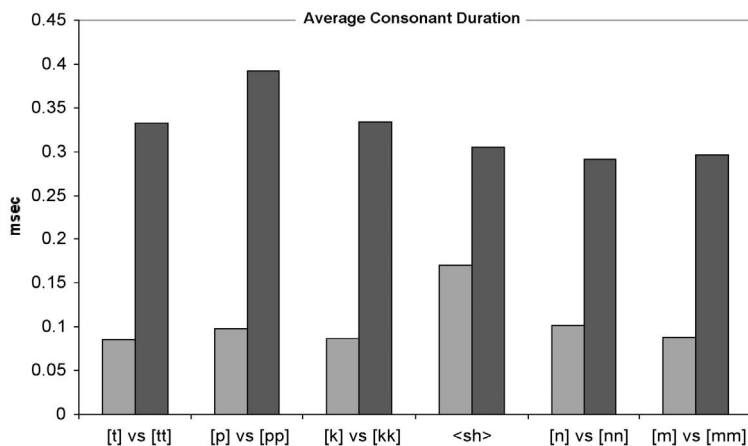
Note that the subject's production of the geminate consonants is quite close to the mean length of native speakers; quite close to three times as long as the singleton. The



exception to this pattern was found in the production of the fricative. While the ratio of short to long was fine for the stops, the fricatives were below 3:1. Graph 1 shows the comparisons.

**Graph 1**

Length ratios for different Japanese segments produced by a non-native speaker

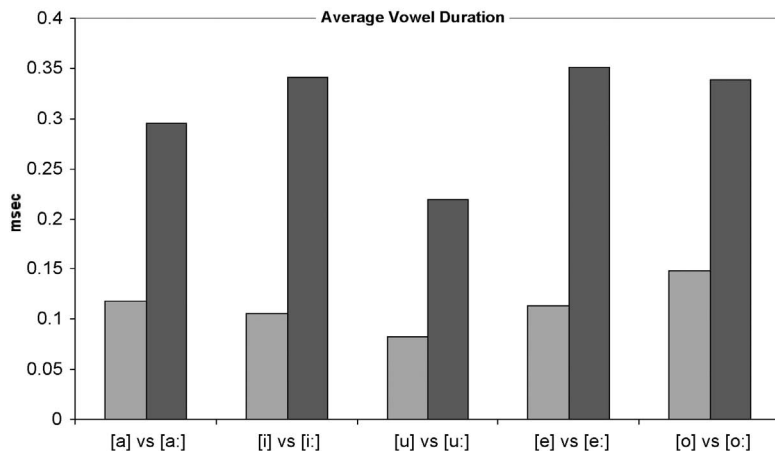


Two-tailed *t*-tests revealed that the geminate consonants were significantly longer than their corresponding singletons (all *p* values less than .001). These results suggest that the subject was able to acquire a new length contrast in spite of the fact that her L1 lacked a consonantal length contrast.

The production of long vowels was also considerably longer than the short vowels, as seen in Graph 2.

**Graph 2**

Non-native vowel length contrasts



Overall, the subject is producing a mean vowel duration contrast of 2.65: 1. Han (1992) reported the native speaker ratio as between 2:1 and 3:1. Again, for all contrasts, 2-tailed *t*-tests revealed that the long vowels were produced significantly longer than their corresponding short vowels (all *p* values < .002). While space does not permit a presentation of the spectral data, we would like to note that the subject is often making a vowel quality distinction where the native speakers would be making just a quantity distinction. In addition to length, she was also distinguishing long and short vowels by means of vowel quality.

After four months of study, our subject was able to make length contrast in both the consonantal and vocalic systems. This is consistent with our predictions based on the moraic structures of Japanese and English, given that weak moras are part of English prosodic structure.

## 13 A phonological parser

In this section, I am going to present some preliminary thoughts on how a theory of parsing may be incorporated into our field of L2 phonology, and how it may help us to explicate a theory of phonological learning. There are, of course, several learning algorithms which are already common currency in the field. See Dresher (1999) for a critique of several current learning algorithms.

I wish to pursue a theory of parsing that is analogous to a theory of syntactic parsing (e.g., Phillips, 1996) and see how far this can take us. I assume that learners are involved in parsing a phonological input string in a way analogous to the parsing of a sentential string where hierarchical structure is assigned (Fodor, 1999). Dekydsporter (2001) pursues this idea in the field of L2 syntax and semantics. Phillips (1996) like Fodor (1999) makes an assumption that he calls *PIG*: Parsing is Grammar. The tools that the learner has available for parsing are the tools the grammar provides (including *UG*). The grammar must change when the parse fails. However, what I wish to incorporate into our model is the idea of cue-based learning—that the action that is taken to change the grammar when a parse fails is not random but rather guided by principles of the grammar.

## 14 Phonological parsing

How do people acquire phonological representations? Let me begin by making a terminological distinction between surface cues and structural triggers. Surface cues are the product of general perceptual mechanisms and can be read off the input. Examples would be (1) that a stressed syllable is prominent and that may be perceived in the input (see Curtin, 1999 for a discussion of this issue in child language acquisition), or (2) the fact that aspiration occurs at the left edge of some prosodic domain. Structural triggers, on the other hand, cannot be read simply off the input (though they are derivable from the input, see Dresher, 1999). These triggers are more abstract and are provided by *UG*. I will return to concrete examples shortly.

It is worth noting that, even for what I will call deep triggers or structural triggers, Fodor (1999) has proposed that the triggers are provided by *UG* (reminiscent of the work

in phonology by Dresher) and that the learner is guided by these triggers in attempting to construct a parse of an input string. Therefore, it is plausible to assume that the kinds of structures assumed by UG, and the kinds of structures infants are sensitive to, can be used to parse the input string.

This is relevant to the kind of work done by Jusczyk and colleagues (e.g., Jusczyk, 1997) who have shown that, for example, infants (acquiring English) as young as nine months prefer to listen to trochaic patterns over iambic patterns. They also show a preference for pauses to occur at constituent boundaries.

Dresher (1999) provides the definition of a cue or structural trigger, given in (11).

(11) Cues

- UG associates every parameter with a cue.
- A cue is not an input sentence or form but is something that can be derived from input.
- Cues must be appropriate to their parameters in the sense that the cue must reflect a fundamental property of the parameter, rather than being fortuitously related to it.
- What the correct cue to any parameter is must be empirically determined (by the linguist—not the learner, to whom it is supplied by UG). There is thus not a parameter-independent general algorithm or parameter setting.

An example is shown in (12).

(12) Main Stress

Parameter: Project the {left/right}-most element of the line 1 constituent.

Cue: Scan a constituent sized window at the edge of a word. Main stress should consistently appear in either the left or the right window.

## 15 Parsing and learning

Fodor (1999) addresses some interesting questions in the field of learnability. She recasts the notion of what parameter setting is, and indeed what parameters and triggers are. For her, parameters and triggers are part of parse trees. In syntax at least, the presence of surface-true cues to abstract syntactic structure is problematic. As Gibson and Wexler (1994) have shown, many inputs can be parametrically ambiguous (in that more than one combination of parameter settings may produce them). This is what they designed their Triggering Learning Algorithm to deal with. When faced with an error (or the inability to parse a string) the grammar is arbitrarily changed and the parse is attempted again. Fodor proposes that a parameter value and its trigger are one and the same, and consist of a subpart of a tree. For example, the presence of [+finite] on a C node would be the structural trigger for [+V2]. To say that a language has parameter value P(v) is to say that its grammar employs the treelet that constitutes P(v) for generating sentences. To say that a particular sentence triggered the adoption of P(v) is to say that the parser found itself unable to parse that input without including P(v) in the structural analysis. To say that the learner set parameter P to value v is to say that the learner took P(v)

from the pool of treelets provided by UG (what she calls the Super-Grammar) and moved it into the current grammar. So, structural triggers can be thought of as a kind of “structural lexicon” provided by UG.

The question is, “how much is phonology like this?” A comparison between phonology and syntax is shown in Chart 4. We must ask “what can trigger structure?” And we must address the question of whether the process of assigning structure to a string of *sounds* is fundamentally different from that of assigning structure to a string of *words*, because perhaps while the input-to-syntax cues may be subtle and unreliable, it may be that the phonetic cues to phonological structure are more robust. I would argue that there is no fundamental difference. An inventory of phonological structure is still provided by UG, but the mapping problem may be more straightforward than, say, trying to determine the V2 status of a language based on its surface word order.

**Chart 4**

Syntactic and phonological parsing

<b>Syntax</b> <i>The parser assigns a syntactic structure to an input string (a sequence of morphemes)</i>	<b>Phonology</b> <i>The parser assigns a prosodic structure to an input string (a sequence of sounds)</i>
A structural trigger would be a treelet E.g., C [+finite]	A structural trigger would be a treelet E.g., Coda <div style="text-align: center;"> <math display="block">\begin{array}{c}   \\ H_w \\   \\ C_{SON} \end{array}</math> </div>

Let us look at some phonological structural triggers and see what they might look like. First we will look at Extrametricality because we can be fairly sure that there is no immediately apparent surface acoustic cue for extrametricality. An extrametrical segment is still pronounced, it just does not play a role in some other computation (like stress assignment or syllabification). Dresher (1999) proposes the cue given in (13).

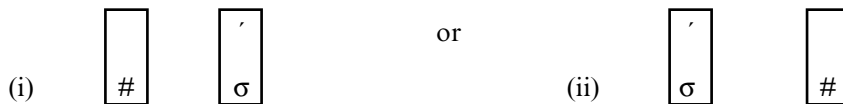
(13) Extrametricality

Parameter: A syllable on the {right/left} {is not/is} extrametrical.

Cue: Stress on a peripheral syllable rules out extrametricality on that side.

In Fodor’s terms then, we would have the parser scan the input and look for the kind of patterns shown in (14)

(14) Parsing for Extrametricality



The detection of 14 (i) would be the cue that there is no extrametricality on the left edge, while the detection of 14 (ii) would be the cue that there is no extrametricality on the right edge. UG would provide the knowledge that extrametricality is found on only edges of prosodic domains.

As Carroll (1999) reminds us, learning takes place when the parse fails. When the parser cannot handle the current input then something has to change.

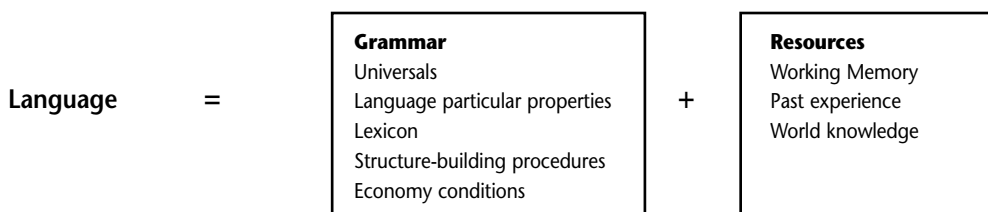
## 16 Syntactic parsing

According to Dekydtspotter (2001, p.97), “the parsing principles are the Universal Grammar principles at work in the parsing task.” As someone working on syntactic processing, he is interested in how the pragmatic module licenses or rejects a semantic representation determined by the syntactic parse in terms of the plausibility of the interpretation at a given stage. Rejection will force a revision in the syntactic module.

Phillips (1996) proposes too that the parser and the grammar are the same. His PIG model is shown in Figure 1.

**Figure 1**

The Parsing-Is-Grammar (PIG) model of Phillips (1996)



Phillips argues that structures are built from Left → Right, as dictated by the condition *Merge Right*, and that structure building is subject to the economy condition *Branch right*, shown in (15).

### (15) Parsing Conditions

**Merge Right:** new items must be introduced at the right edge of a structure.

**Branch Right:**

Metric: select the attachment that uses the shortest path(s) from the last item in the input to the current input item

Reference set: all attachments of a new item that are compatible with a given interpretation

Incorporation of new material is seen as ambiguity resolution; what choice does a comprehender make in an ambiguous situation, and how are things resolved when a mistake is realized? He defines the following notions, given in (16):

### (16) Grammaticality versus Parsability

**Grammatical:** a sound/meaning pair is grammatical if the grammar can generate a correspondence for that pair given potentially unbounded resources

**Parsable:** a sound-meaning pair is parsable with resources R if the grammar can generate a correspondence for that pair using only resources R.

When we bring this machinery back to phonology, we must ask: what is the

phonological analog of being interpretable? If parsing is grammar this should hold for phonology too. At the word level, a string would be interpretable if *lexical activation* takes place. At the phrase level, the lexical activation would be checked to see if the string is interpretable in the syntactic context.

## 17 Connecting segments and syllables

In this section, I demonstrate that phonological structure can be assigned on the basis of a strictly serial left-to-right parse (unlike approaches which begin by syllabifying the Nucleus). First let us consider the assignation of syllabic structure. If we deal with parsing at the word level then it can proceed top down following Phillips (1996). The learner is able to do the word segmentation tasks (i.e., we are assuming that the learner can pull words out of a continuous input stream) but also has to assign an internal structure to the word. Let us imagine a simple case where the learner must syllabify a word like ‘trip’. The process would be as shown in (17):

(17) Parsing ‘trip’

1. Link segmental content to lowest prosodic node.
2. Can the first element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

/

[t] < rip >

Where [x] stands for a parsed element, and < x > stands for an unparsed element.

3. Can the next element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

^

[t r] < ip >

4. Can the next element be assigned to the Onset? [Y/N]

[No] → Assign it to the Nucleus.

^ |

[t r] [i] < p >

5. Can the next element be assigned to the Nucleus? [Y/N]

[No] → Assign it to the Coda.

^ | \

[t r] [i] [p]

6. Lexical activation. The parse succeeds.

Under this mechanism, new structure is always assigned to the right. Immediately, this raises the question of how we deal with (1) left-edge effects (which have the potential for being problematic in a L → R parse), and (2) interlanguage effects?

### 17.1

#### Left-Edge Effects

Let us consider the parsing of a string like ‘sprig’, given in (18).

(18) Parsing ‘sprig’

1. Link segmental content to lowest prosodic node.

2. Can the first element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

/

[s] <prig>

3. Can the next element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

^

[s p] <rig>

4. Can the next element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

^

[spr] <ig>

5. This violates binary branching (following common assumptions of government phonology and minimalism), so delink the left-most element.

^

<s> [p r] <ig>

This build-then-repair procedure is not optimal, and I would propose an alternative; to incorporate a parser that has limited look ahead to a two-segment window. The algorithm would be:

a. Look one element to the right of the element being parsed. Is the next element consistent with the sonority sequencing generalization (SSG)? If [Yes], parse the first element. If [No], leave the first element unparsed.

b. Proceed to the next element.

This would also result in:

^

<s> [p r] <ig>

at this stage of the parse.

6. Can the next element be assigned to the Onset? [Y/N]

[No] → Assign it to the Nucleus.

^ |

<s> [p r] [i] <g>

7. Can the next element be assigned to the Nucleus? [Y/N]

[No] → Assign it to the Coda

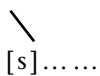


This still leaves the question of what happens to the [s]; does it ever get parsed at the word level? We know that s-clusters behave idiosyncratically in languages around the world. Kaye (1991/1992) provides a number of examples to argue that these clusters are always heterosyllabic. His arguments range from choice of definite article in Italian such as *il costollo scuro*—where words beginning with s-clusters behave like vowel-initial words—to y-glides in English *st[y]upid!* *\*fl[y]uid*—where we can see that the insertion of a y-glide to create a three-consonantal string is blocked in *fluid* but not in *stupid*. In both the Italian and the English examples, we see that s-clusters are behaving exceptionally. Here, the question is “at what stage do these consonants take on their special status”? Ultimately, I would agree with Kaye (1991/1992) that they behave as coda consonants to an empty-headed syllable<sup>1</sup>. In English, certain kinds of empty positions receive no phonetic interpretation. However, as we shall see when we look at the second language learners, these positions can be filled. For now I will remain agnostic as to whether the <s> is better described by (i) below where it is assigned to the coda in the first pass:

i). Looking at a two-element window, can the first element be assigned to an Onset?

[No] → Can it be assigned to a Nucleus?

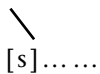
[No] → assign it to a coda



or by (ii) where the whole string is analyzed and then repaired:

(ii). <s> [pr] [i] [g]

This string still has an unparsed element; link it to a coda position.



The crucial aspect of this analysis is that the [s] ends up in a coda.

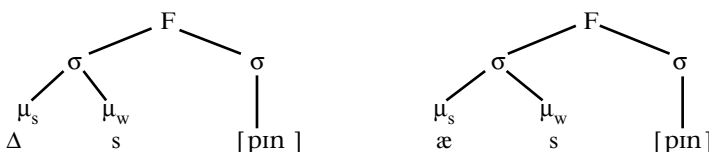
8. Lexical activation. The parse succeeds.

## 17.2

### Interlanguage effects

We now turn to the question of how this machinery accounts for second language data.

<sup>1</sup> If, as I suggest, the [s] is found in a coda position, the question emerges as to why the /t/ in a word like ‘stop’ is unaspirated. Noting the distribution of unaspirated stops in English is helpful: (1) after [s] (e.g., [spɪn]), and (2) syllable initially in unstressed syllables (e.g., [æspɪn] or [æspən]). These two environments seem to share a structural description: the stops are unaspirated when they are in the second syllable of a foot:





Let us look at some data (Broselow, 1988) from a learner whose L1 (Egyptian Arabic) does not allow onset consonant clusters acquiring an L2 (English) that does. Consider the word ‘trip’ in (19).

(19) Parsing ‘trip’ in a second language.

1. Link segmental content to lowest prosodic node.

2. Can the first element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

/

[t]<rip>

3. Can the next element be assigned to the Onset? [Y/N]

[No; L1 setting] → Assign it to the Nucleus. <FAIL>

Is it consistent with SSG? [Y/N]

[Yes] → leave the current element unparsed

/

[t]<r><ip>

4. Can the next element be assigned to the Nucleus? [Y/N]

[Yes] → Assign it to the Nucleus.

/ |

[t]<r>[i]<p>

5. Can the next element be assigned to the Nucleus? [Y/N]

[No] → Assign it to the Coda

/ ^

[t]<r>[i p]

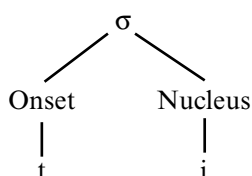
6. This structure is not allowed by the grammar, as the parse has failed. What is the repair strategy? The learner must fix the parse by projecting new prosodic structure (in this case a syllabic nucleus). The new nucleus goes to the left of the unparsed element.

^ ^

[t][i]<r>[i p]

This is consistent with Phillips’ (1996) principle of Merge Right. We have inserted a new item (the epenthetic [i]) and it will be attached at the right edge of the structure, as shown in (20):

(20)



Note that at this stage of proficiency, the learner is parsing the input (which includes a consonant cluster) and ends up inserting the epenthetic vowel. Later on, a different strategy of consulting the other structures available in UG (e.g., Branching Onsets) will occur and the nature of the input representation will change from [t]<r>[ip] to [tr][ip]. This acknowledges that initially the subject will have an incorrect input representation that has to change. That these early representations are inaccurate is shown by Dehaene-Lambertz, Dupoux and Gout (2000) who show via behavioral tasks and ERP analysis that Japanese speakers hear an epenthetic vowel between the two consonants in a string like ‘ebzo’ (which is an illicit Japanese string) even when no vowel is present in the input (see Matthews & Brown, this volume, too). The epenthetic vowel is not inserted late in a production routine, they actually hear it even when it’s not in the input. At a later stage of proficiency, the learner will realize that these vowels are not, in fact, in the input string, and the parse will fail. There are a few possible sources of this type of knowledge. One would be the existence of minimal pairs or near-minimal pairs related to the CVC sequence. Examples of such pairs are given in (21).

(21)

trade/tirade; train/terrain; plate/palate; plot/pilot; plow/pillow; present/percent;  
claps/collapse; clean/Colleen; dress/duress; drive/derive; sting/sitting

If the learner becomes aware of such pairs then there will be an impetus to realize that the pronunciation of *drive* with an epenthetic vowel between the [d] and the [r] cannot be right. It is also plausible to assume that literacy facts and the role of orthography in English play a role in making the learners aware that their L1 parsing strategies are failing in the parsing of onset clusters.

Knowledge of phonotactics resides in the knowledge of which classes of segments are licensed in which positions. In Japanese, codas must be sonorant. Therefore, when a Japanese speaker hears a sequence like ‘ebzo’, then as soon as the obstruent [b] is reached the learner would assume that the [b] was in an onset and would expect a following vowel. The learning that has to take place is to realize that obstruents may be licensed in a coda. This will cause the learner to consult the super-grammar to see what changes could be made. If the current grammar cannot assign a structure to anything except coda sonorants, then UG (the Super-Grammar) would provide another option. For example, the coda could be an obstruent. If the parse then succeeded, the treelet would be incorporated into the current grammar. Examples of phonological treelets are given in (22).

(22)



Let us now consider the parsing of an s-cluster sequence like ‘street’ by Egyptian learners, given in (23).

(23) Parsing ‘street’ in L2

1. Link segmental content to lowest prosodic node.

2. Can the first element be assigned to the Onset? [Y/N]

[Yes] → Assign it to the Onset.

/  
[s]<trit>

3. Can the next element be assigned to the Onset? [Y/N]

[No; L1 setting] → Assign it to the Nucleus. <FAIL>

Is it consistent with SSG? [Y/N]

[No] → delink first element

/  
<s>[t]<rit>

Note, that this process is necessary because a singleton [s] would be parsed legitimately into an onset.

4. Can the next element be assigned to the Onset? [Y/N]

[No; L1 setting] → Assign it to the Nucleus <FAIL>

Is it consistent with SSG? [Y/N]

[Yes] → leave the current element unparsed

/  
<s>[t]<r> <it>

5. Can the next element be assigned to the Nucleus? [Y/N]

[Yes] → Assign it to the Nucleus

/ |  
<s>[t]<r>[i]<t>

6. Can the next element be assigned to the Nucleus? [Y/N]

No → Assign it to the coda

/ ^  
<s>[t]<r>[i t]

7. There are still unparsed elements, but what is the repair strategy? The learner will fix the parse by projecting new prosodic structure, in this case a syllabic nucleus. The nucleus will be projected to the left of the unparsed element.

/ | ^  
<s>[t][i]<r>[i t]

Note that there are still some unparsed elements. As a result, the second pass occurs.

8. Can the <s> be assigned to the Onset?

[No] → Can it be assigned to the Nucleus?

[No] → Assign it to the Coda.

^ / | ^  
Δ [s][t][i]<r>[i t]

9. Can the <r> be assigned to the Onset?

[Yes]

$$\begin{array}{c} \wedge \quad / | \quad / \quad \wedge \\ \Delta [s][t][i][r] \quad [i \quad t] \end{array}$$

10. Fill the empty nucleus with an epenthetic [i].

$$\begin{array}{c} \wedge \quad / | \quad / \quad \wedge \\ [i][s][t][i][r] \quad [i \quad t] \end{array}$$

This treatment of the s-clusters appears to be consistent with the typological facts and with the parsing principles we have established.

11. Lexical activation. The parse succeeds.

Simple words with two or three consonants in a row would provide the second language learner with much of the information necessary to achieve the target grammar. Take medial examples like *apply*, *intrude*, *abstract* or *astronaut*. These examples would tell the learner *either* that obstruents must be licensed in codas in English *or* that onsets can branch. Either change in the grammar would be a move closer to the target. Compound words like *fieldtrip* or *eggplant* would also be good examples to force the reclassification of licensed coda consonants. Whether the L1 is Japanese or Arabic, there are data which would tell the learners that their current hypotheses were wrong.

## 18 Conclusion

The following are some of the main points I have argued for:

1. Our understanding of the learning theory necessary to account for the acquisition of L2 phonology can be enhanced with the notion of a phonological parser which proceeds left-to-right and is governed by grammatical principles (following Dresner, 1999; Fodor, 1999; Phillips, 1996).
2. Learning occurs when the parse fails.
3. When the parse fails, UG can be consulted as to what action to take.
4. L2 learners can indeed trigger new prosodic structure as well as alter the mapping principles that associate different levels of structure.

## References

- ATKEY, S. (2000). The acquisition of non-native segmental contrasts: A look at English speakers' acquisition of Czech palatal stops. In A. James & J. Leather (Eds.), *Proceedings of the New Sounds 2000 Conference* (pp. 22–31). University of Klagenfurt Press.
- AVERY, P., & IDSARDI, W. (2000). *On the laryngeal specification of tone*. Paper presented at the MOT Phonology Workshop, Toronto.
- BONGAERTS, T., MENNEN, S., & Van der SLIK, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, **54**, 298–308.
- BROSELOW, E. (1988). Prosodic phonology and the acquisition of a second language. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 295–308). Kluwer.

- BROWN, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Blackwell.
- CARROLL, S. (1999). Putting “input” in its proper place. *Second Language Research*, **15**, 337–388.
- CURTIN, S. (1999). Positional prominence in the acquisition of prosodic structure. In *Proceedings of the Annual Meeting of the Canadian Linguistics Association* (pp. 89–100). Cahiers Linguistiques d’Ottawa.
- CURTIN, S., GOAD, H., & PATER, J. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, **14**, 389–405.
- DEHAENE-LAMBERTZ, G., DUPOUX, E., & GOUT, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, **12**, 635–647.
- DEKYDTSPOTTER, L. (2001). The universal parser and interlanguage: Domain-specific mental organization in the comprehension of *combien* interrogatives in English-French interlanguage. *Second Language Research*, **17**, 91–143.
- DRESHER, E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry*, **30**, 27–67.
- DRESHER, E., & Van Der HULST, H. (1995). Global determinacy and learnability in phonology. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 1–22). Lawrence Erlbaum.
- FLEGE, J. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 565–604). York Press.
- FODOR, J. (1999). Learnability theory: Triggers for parsing with. In E. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 363–406). John Benjamins.
- GIBSON, E., & WEXLER, K. (1994). Triggers. *Linguistic Inquiry*, **25**, 407–454.
- HAN, M. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for non-native speakers. *Phonetica*, **49**, 102–127.
- HAYES, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, **20**, 253–306.
- JUSCZYK, P. (1997). *The discovery of spoken language*. MIT Press.
- KAYE, J. (1991/92). Do you believe in Magic? The story of S + C sequences. In *SOAS Working Papers in Linguistics and Phonetics*, University of London.
- LACHARITÉ, D., & PRÉVOST, P. (1999). The role of L1 and of teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of BUCLD 23* (pp. 373–385). Somerville: Cascadilla Press.
- MAH, J., & ARCHIBALD, J. (2003). The acquisition of L2 length contrasts. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *Generative Approaches to Second Language Acquisition (GASLA 2002)*, (pp. 208–212). Cascadilla Press.
- MATTHEWS, J. (1997). The influence of pronunciation training on the perception of second language contrasts. In J. Leather & A. James (Eds.), *New Sounds 97* (pp. 223–239). University of Klagenfurt Press.
- PHILLIPS, C. (1996). *Order and structure*. Ph.D. Dissertation, M.I.T.
- SCHWARTZ, B., & SPROUSE, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, **12**, 40–72.
- SINGLETON, D. (1989). *Language acquisition: The age factor*. Clevedon: Multilingual Matters 47.
- YOUNG-SCHOLTEN, M. (1994). On positive evidence and ultimate attainment in L2 phonology. *Second Language Research*, **10**, 193–214.

---

WHITE, L. (1989). *Universal grammar and second language acquisition*. Benjamins.  
ZEC, D. (1995). Sonority constraints on syllable structure. *Phonology*, **12**, 85–130.

---