Acquisition of L2 Japanese WH questions: Evidence of Phonological Contiguity and Non-Shallow Structures

In this paper we look at some of the structural properties of L2 Japanese WH questions. In Japanese the WH words are licensed to remain in situ by the prosodic contiguity (Richards, 2016) properties of the phrases which have no prosodic boundaries between the WH word and the question particle. In a rehearsed-reading, sentence production task, we look to see whether non-native speakers of Japanese who are learning the L2 in university classes in North America are able to acquire grammars which are constrained by such universal properties as Match Theory (Elfner, 2015; Selkirk, 2011) and Contiguity Theory (Richards, 2010, 2016). While linear mixed effects analyses of the pitch contours reveal that the L2ers have not acquired nativelike the phonetic implementation distinction of the documented pitch boost on WH words compared to non-WH DPs, our data show that the subjects have acquired the pitch compression patterns indicative of having no prosodic phrases intervening between the WH word and the question particle. This property of Japanese WH questions is not taught in their classes, and, thus we argue, that the data are supportive of the position that interlanguage grammars are constrained by universal grammatical properties such as the prosodic contiguity of WH-phrase licensing. We also present these results as being counter to the Shallow Structure Hypothesis (Clahsen & Felser, 2017).
Acquisition of L2 Japanese WH questions: Evidence of Phonological Contiguity and Non-Shallow Structures

John Archibald & Nicole Croteau
University of Victoria

Abstract
In this paper we look at some of the structural properties of L2 Japanese WH questions. In Japanese the WH words are licensed to remain in situ by the prosodic contiguity (Richards, 2016) properties of the phrases which have no prosodic boundaries between the WH word and the question particle. In a rehearsed-reading, sentence production task, we look to see whether non-native speakers of Japanese who are learning the L2 in university classes in North America are able to acquire grammars which are constrained by such universal properties as Match Theory (Elfner, 2015; Selkirk, 2011) and Contiguity Theory (Richards, 2010, 2016). While linear mixed effects analyses of the pitch contours reveal that the L2ers have not acquired the phonetic implementation distinction of the documented pitch boost on WH words compared to non-WH DPs, our data show that the subjects have acquired the pitch compression patterns indicative of having no prosodic phrases intervening between the WH word and the question particle. This property of Japanese WH questions is not taught in their classes, and, thus we argue, that the data are supportive of the position that interlanguage grammars are constrained by universal grammatical properties such as the prosodic contiguity of WH-phrase licensing. We also present these results as being counter to the Shallow Structure Hypothesis (Clahsen & Felser, 2017).

The field of second language acquisition (SLA) looks to probe the properties of the grammars of second language (L2) learners. While there is much research that investigates a single module such as phonology (Archibald 1998, Cabrelli Amaro 2017) or syntax (White 2003), there has also been much written on the grammatical interfaces (White 2011; Montrul 2011; Sorace 2011; Goad and White 2004). Considerable attention has been given to the morphology/syntax interface (Franceschina 2001), and also to the syntax/pragmatics interface (Sorace and Filiaci 2006), but there has been less attention given to the phonology/syntax interface (cf. Fodor 2002) in SLA, though, of course, it has been

1 We would like to acknowledge feedback we have received from three anonymous reviewers and the Associate Editor, all of whom made suggestions which both strengthened the analysis, and made it accessible to a wider readership. We also received feedback from audiences at the Canadian Linguistic Association in Calgary, and GASLA in Southampton on preliminary versions of this paper. Thanks to Heather Goad and Joey Windsor for their comments at those conferences. Tess Nolan was invaluable in providing phonetic support. We would also like to thank Mary Lesperance for statistical advice.
addressed in the theoretical literature (Elfner 2015). This paper looks at one aspect of the phonology/syntax interface related to the formation of WH-questions.

Richards (2010, 2016) articulates a theory which accounts for the variation between (a) languages which move WH elements to the left periphery, and (b) languages which allow WH elements to stay in place. English generates WH questions via WH-Movement as shown in (1).

(1) *Whom* should Bob call?

Japanese, on the other hand, allows the WH word to remain *in situ*, as shown in (2):

(2) Mito-ga *nani-o* katta no?
    Mito-NOM what-ACC bought +Q
    'What did Mito buy?'

In this paper, we investigate the second language acquisition of Japanese WH questions in order to explore whether the interlanguage grammars of the learners are subject to the grammatical principles laid out in the syntactic and phonological literature.

1.0 The Syntax/Phonology Interface

1.1 Contiguity Theory

Richards (2010, 2016) argues that these are two strategies to achieve the *contiguity* of the WH word and the interrogative feature [+Q]. In English, surface linear adjacency achieves contiguity as a result of moving the WH word to the Spec of CP position at the left edge (adjacent to where the phonologically null [+Q] feature resides), as shown in (3).
In Japanese, there are two noteworthy differences from the English structure, as we see in (4), “What did Mito buy?”.
First, we note that the [+Q] feature is spelled out by a question particle (which in (4) is no). Secondly, the WH word does not move, to the periphery (either left or right) but contiguity is achieved in another fashion: *prosodic* contiguity. Richards demonstrates that (a) there are no prosodic boundaries between the WH word (in this case *nani-o*) and the [+Q] feature, and (b) there is what he calls a phonetic boost on the WH phrases (compared to non-WH DPs).

Richards draws on work on the prosody of Japanese WH questions (Deguchi and...)

---

2 The particle *-ka* is used when there is a politeness marker in the sentence (see Miyagawa, 2012).
Kitagawa, 2002; Ishihara, 2003; Sughara, 2003; Hirotani, 2005). He provides pitch tracks of the Tokyo dialect which show a pitch boost on the WH element (like nanio) compared to DP direct objects (like nanika). Our pitch tracks are almost identical to those given in Richards.

The structure created by Merge$^3$ would be (Japanese being right-headed) as shown in (5):

(5) “What did Naoya buy?”

More schematically, the constituent structure would be that in (6):

(6) [DP] [[DP WH] [V] VP] C

The strategy to achieve prosodic contiguity that Japanese invokes is to eliminate the prosodic boundaries between the WH phrase and the Complementizer, and generate the prosodic structure given in (7):

(7) (ϕ DP) (ϕ WH C)

Let us look at how this plays out when there are more elements between the WH phrase and the Complementizer. Richards (2010) looks at Minor Phrases (which dominate

---

$^3$ Merge is the operation which combines categories into hierarchically-structured units. For example, taking [the] and [sandwich] and building the new constituent [the sandwich]. Then taking [ate] and [the sandwich] to build [ate the sandwich].
prosodic words). He follows Kubozono (2006) in the analysis that Japanese Minor Phrases can be recursive; multiple Minor Phrases are dominated by an overarching Minor Phrase, as shown in (8).

(8) “drank what at the bar?”

He refers to the highest Minor Phrase as the WH-domain. Given a WH-phrase $\alpha$ and a complementizer $C$ where $\alpha$ takes scope, $\alpha$ and $C$ must be separated by as few Minor Phrase boundaries as possible, for a given level of Minor Phrasing$^4$.

This phonological phrasing would be generated in a fashion consistent with Match Theory (Selkirk, 2011; Elfner, 2015). Match Theory probes the syntax-phonology interface. In attempting to describe the domains of certain phonological phenomena, sometimes syntactic structures appear to be relevant. For example, English

---

$^4$ A reviewer points out that scrambling of the WH-phrase to the position adjacent to the [+Q] feature should also improve prosodic contiguity. We acknowledge that there are many technical aspects of Richards’ model which we are not addressing in this paper. Space does not permit us to explore the issues which scrambling raises.
phrasal stress falls on the rightmost element in a syntactic phrase (Chomsky & Halle, 1968) as shown in (9):

(9) [A syntactician] and [a phonologist] [walk into a bár]

Yet, other phonological phenomena are not predictable by syntactic structures, but rather by prosodic domains. Some patterns apply at the highest level of phonological utterance, such as RP-English r-insertion (Nespor and Vogel, 1986) which applies when there are adjacent vowels which can be adjacent even across sentences, as shown in (10).

10. neve[r] again
the giant Panda[r], as you know, is an endangered species
That’s a nice ca[r]. Is it yours?

Other changes, such as Tuscan spirantization (where stops become fricatives) are somewhat more restricted insofar as parenthetic material blocks the change. The domain of application, then, is the intonational phrase, as shown in (11). The orthographic ‘c’ words which are underscored are produced as the fricative [h] while the orthographic ‘c’ words which are not underscored are produced as the stop [k].

11. i canarini congolesi costano molto cari
‘Congolese canaries are very expensive.’

i canarini congolesi, come sai, costano molto cari
‘Congolese canaries, as you know, are very expensive.’

English stress retraction occurs within a phonological phrase but not across the boundary of two phonological phrases. The sentence in (12a) shows stress retraction while (12b) does not.

12a. [pérésséres gladly] (cf persevéres)
12b. persevéres [gládly and diligently]

Recent research has demonstrated that there is a systematic correspondence between syntactic and prosodic structure. These correspondences are articulated in Match Theory.

1.2 Match Theory
Richards’ work is couched within the larger literature of the syntax-prosody interface (e.g., Selkirk, 2011; Tyler, 2019). Match Theory, broadly speaking, begins with the assumption of syntax-prosody isomorphism. Phonological structure is modelled via the Prosodic Hierarchy (Nespor & Vogel, 1986) with modifications to earlier proposals which recognize that (a) not all languages may encode all levels (see Özcelik, 2017) and (b) that recursive structures are permissible. The theory accounts for the correspondences between phonological and syntactic structures. The preferential mapping is between (a) syntactic phrases (XPs) and phonological phrases (ϕ), and (b) syntactic heads (X0s) and prosodic words (ω)⁵. Furthermore, it is assumed that CPs are mapped on to intonational phrases (τ).

Within such a model (e.g., Tyler (2019), an NP such as shallow structure would have the structural mapping shown in (13) at spell out⁶:

(13)

```
NP
 /\    
 Adj0  N0
 shallow structure
```

This transparent mapping between the syntax and the phonology which would generate the following structures for the sentences we are investigating (where the arrows indicate the mapping between the syntactic structure and the phonological structure given in (14):

---

⁵ These preferences are, in fact, modelled as violable constraints, the details of which are not crucial to our argument. In some languages, for example, single-word DPs behave as full phonological phrases.

⁶ We use the sparser Bare Phrase Structure trees to highlight the isomorphic mapping of heads and phrases.
Thus, the prosody preserves the hierarchical syntactic structure. For the structure to be well-formed the in situ WH phrase must be followed by a minimal number of intervening \( \phi \)-phrases.

Let us look at a specific example from Richards (2016). We begin with the declarative sentence, “Naoya drank something in the bar” shown in (15):
(15)

Naoya-ga da
Naoya-NOM drank

something-ACC

drunk

bar-LOC
This syntactic structure would be mapped on the prosodic tree shown in (16):

(16)

Richards argues that such prosodic trees\(^7\) are generated by the narrow syntax as the derivation proceeds, not, as is standardly assumed after the narrow-syntactic derivation is finished. Thus, abstract phonology is not merely a late spell-out routine.

Now let us consider the corresponding interrogative sentence “What did Naoya drink at the bar?”, shown in (17).

---

\(^7\) A reviewer notes that we have drawn prosodic trees with phonological recursion. Following van der Hulst (2010), Féry (2010), we also assume that phonological structure is recursive.
(17)

```
(17) Naoya-ga da
    Naoya-NOM
    DP
    VP
      V' nomiya-de
      PP nani-o nom
      V
    T
    TP
      C [+Q] no
    CP
  ```
The generation of the prosodic tree proceeds through several intermediate steps which need not concern us here to arrive at the structure given in (18):

(18)

It is worth noting here some of Richards' (2010, 2016) assumptions about the interaction of the syntax with the phonology. Note that he assumes that WH-movement takes place only if the prosody requires it. While he maintains that WH-movement is a syntactic, not a late phonetic, operation, he is agnostic as to the mechanism which allows the syntax to ‘look ahead’ to the phonology.

1.1 What Is To Be Acquired?

The research question we seek to investigate in this paper is: are advanced L2 speakers
able to set up a phonological grammar with no prosodic boundaries between the WH word and the Question complementizer ([+Q]) in order to properly license WH in situ as would be the case if Richards’ (2016) Contiguity Theory holds for interlanguage grammars?

In order to understand the operationalization of this question, let us compare two sentences given in (19) and (20), where the direct objects are in bold, and the minor phonological phrases are in italics.

(19) Naoya-wa nani-o nomiya-de nonda no?
Naoya-TOP what-ACC bar-LOC drank +Q
ナオヤは、何を飲み屋で飲んだの?
'What did Naoya drink at the bar?'

(20) Naoya-ga nanika-o nomiya-de nonda.
Naoya-NOM something-ACC bar-LOC drank
ナオヤが、何かを飲み屋で飲んだ。
'Naoya drank something at the bar.'

For native speakers of Japanese, we expect higher pitch on the objects which are WH words (such as nanio in (19)) compared with objects which are DPs (such as nanika in (20)). Readers should note that we use the label DP (recognizing that WH words are, in fact, found within DPs) to refer to DPs which lack a WH word thus allowing us to compare the behaviour of WH phrases with DPs. The spectrograms in Figures 1 and 2 are taken from one of the native speaker controls in this study (a female voice is shown). Statistical analyses will be reported in section 3 but for now a visual comparison will suffice.

---

8 A reviewer notes that it would be interesting to see data from Yes-No questions which have a nanika DP for comparison. We certainly agree but will have to leave this for a follow-up study. Richards (2010: 147) shows the pitch track of one Yes-No question in Fukuoa Japanese (from Smith, 2005) which shows a pitch rise between the DP and the end of the sentence.
Figure 1. Pitch track of the WH question shown in (14).

Note the pitch spike on the WH phrase (nanio). Compare this with the declarative sentence of (15) produced by the same speaker shown in Figure 2.

Figure 2. Pitch track of declarative sentence of (15) with DP object.

Note the lower pitch (209Hz) on the DP object (nanika) – at the intersection of the red lines. These pitch tracks are virtually identical to those given in Richards (2010: 144). Given this convergence, we assume that our production task is revealing the same
properties of Japanese pitch as Ishihara (2003), and we, thus, assume that native speakers of Japanese implement this pitch boost on WH phrases.

Furthermore, we find no pitch peaks on the prosodic words (e.g., *nomiya de nonda*) which come between the WH word and the [+Q] particle (as in (14))\(^9\). The pitch tracks in Figures 3 and 4 are taken from one of our native speaker controls\(^10\).

![Pitch Compression](image)

**Figure 3.** A native speaker recorded in our study demonstrating pitch compression.

This is in contrast to the pitch peak we see in the declarative sentence:

---

\(^9\) A reviewer suggests that this pitch compression could be the result of pragmatic effects. We acknowledge that information structure can affect intonation, however, for these data, our subjects were told to assume a neutral discourse context.

\(^10\) A reviewer wonders if the complementizer might better be analysed as being in a higher prosodic phrase. We are following Richards (2010: 145) where he defines the WH-domain as [C \[WH\]], “in which C and WH are not separated by prosodic boundaries.”
Figure 4. Native speaker pitch rise in declarative sentence.

2. Methodology

In this study, we looked to see whether non-native speakers (NNS) of Japanese are able to acquire L2 grammars which observe this property of phonological contiguity. Sixteen instructed learners of Japanese (9 self-assessed advanced proficiency; 7 self-assessed intermediate proficiency) were audio recorded. In addition, there were 6 native speakers of Japanese to act as controls in a subsequent analysis of prosodic contiguity.

The subjects\(^{11}\) were recruited as volunteers from announcements in senior-level Japanese classes at a Canadian university\(^{12}\). The advanced group consisted of 4 males and 5 females while the intermediate group consisted of 4 females and 3 males. Twelve were native speakers of English while 4 were native speakers of other languages (Cantonese, Mandarin, or Korean). Furthermore, this heterogeneity of L1, in our view, is not a major impediment insofar as we are not looking to explore any aspects of L1 transfer but rather

---

\(^{11}\) Appendix B contains a summary of the subject profiles.

\(^{12}\) There would have been considerable variation in the number of classroom hours of instruction the subjects received. Not all of them were Japanese majors. Roughly, they would have taken 4-6 courses in Japanese language with each course having approximately 39 hours of class time.
whether the L2 production is nativelike with respect to the phonological and phonetic properties of the WH domain. The subject pool is also mixed in that there are both males and females. We have grouped them together (after transforming the data) for analysis because nothing in the analysis depends on the absolute pitch of the production. Rather, it is the pitch change that is of interest. All analyses were performed in R version 3.2.3 (R Core Team, 2015) with the package ‘lme4’ version 1.1-12 (Bates, 2015).

2.1 The Task

The subjects were given in advance 19 Japanese sentences (see Appendix A\textsuperscript{13}) which they were asked to rehearse at home in preparation for reading aloud assuming a neutral discourse context. If, during recording in the lab, they made a slip, they were allowed to record again. The vast majority of sentences were read only once. Sometimes subjects requested a second attempt. To us, this is an indication that the task was feasible, and not too demanding on the learners. After they finished a sentence, it was played back for them, and were asked if they wished to record it again. If they did not, we proceeded to the next sentence. The final version was the one used for acoustic analysis. In this way, we tried to ensure that the production we analyzed was the best possible production that the subject could produce. Two examples are given in (21) and (22)\textsuperscript{14}.

\begin{equation}
\text{(21) Miki-wa kinō nani-o kai-mashi-ta ka?}
\begin{align*}
\text{Miki-TOP yesterday what-ACC buy-POL-PAST +Q} \\
\text{ミキは、昨日何を買いましたか?}
\end{align*}
\end{equation}

\begin{equation}
\text{\textquoteleft What did Miki buy yesterday?\textquoteright }
\end{equation}

\begin{equation}
\text{(22) Miki-wa kinō hon-o kai-mashi-ta.}
\begin{align*}
\text{Miki-TOP yesterday book-ACC buy-POL-PAST} \\
\text{ミキは、昨日本を買いました。}
\end{align*}
\end{equation}

\begin{equation}
\text{\textquoteleft Miki bought a book yesterday.\textquoteright }
\end{equation}

\textsuperscript{13} These sentences were vetted (and the orthography confirmed) by a linguistics graduate student who is a native speaker of Japanese. A reviewer also points out that since the stimuli have both Roman and Japanese orthography, we cannot tease apart the different reading strategies. We agree, but our goal was to get the best, most natural production that each subject was capable of, and for that reason decide to include both orthographies.

\textsuperscript{14} The morpheme-by-morpheme gloss was not part of their reading text. Both roman and Japanese characters were included, however.
Sentences were recorded on Audacity at 44.1kHz, and pitch tracking was done via Praat 6.0.09 (Boersma & Weenink, 2019) after exporting the .wav files from Audacity. An example of the pitch tracking is shown in Figure 5, where we note the pitch rise on `nanio` and `no`.

![Figure 5. Visualization of the pitch patterns.](image)

The pitch peak was measured and entered into a spreadsheet for calculation. Given that we were comparing different lexical items, we controlled for pitch accent in the stimuli. All of the target words (both WH and DP) contained both high and low pitch accents, so any difference which could potentially emerge in the average across subjects and across lexical items cannot be attributed to some words having either all high or all low pitch accents\(^\text{15}\).

### 2.2 The Structure of the Argument

Native speakers of Japanese boost the pitch on WH words (compared to non-WH DPs) and compress the pitch between the WH word and the Complementizer. Richards (2016) builds a theory in which (a) the WH in situ is licensed by prosodic contiguity, and (b) the phonology/syntax interface is mediated by an architecture following Match Theory (Selkirk, 2011; Elfner, 2015), which predicts that the phonological and syntactic constituents should be largely isomorphic. Our study is designed to investigate these properties in second language learners. If the L2 subjects are producing prosodic contours which are nativelike in that they reflect L2 Japanese syntax, this is something that they

\(^{15}\) Our native speaker consultant gives the following pitch accent patterns on the key words: `nanio` (HLL), `nanikao` (HLLL), `pizao` (HLL), `kino`: (LHH).
could not have learned easily from classroom input. We have given our subjects a reading task (not a grammaticality or acceptability judgement task), and make the assumption (Levelt, 1989) that the L2ers are generating a syntactic representation in production. So, given the syntactic structure of WH in situ, we are looking to see if the produced utterances reflect nativelike prosody. There is much literature which demonstrates that syntactic constituent structure is invoked in silent reading tasks (e.g., Juffs & Harrington, 1998; Pliatsikas & Marinis, 2013; Kim, Montrul and Yoon, 2015), and we maintain this is true in production tasks as well.

There are four main reasons why this is a question of interest:

1. It will add to the literature exploring whether interlanguage grammars are constrained by the same principles as primary languages
2. It will add to the literature exploring whether non-native speakers can attain L2 grammatical properties which are (a) not taught in class, and (b) not easily read off the input
3. It will provide further empirical data relevant to the Shallow Structure Hypothesis (Clahsen & Felser, 2017) which argues that second language learners often have problems processing abstract hierarchical grammatical constituents in real time.
4. It will provide data relevant to the question of whether L2 phonetic ability and L2 phonological knowledge proceed along the same developmental path.

Let us examine each of these contributions in a bit more depth.

2.2.1 Interlanguages as Natural Languages

This is a thread which has been addressed often in the literature (Bley-Vroman, 2009) under a number of guises but it really comes down to the issue of domain specificity. Many researchers have probed the question of whether interlanguage grammars are constrained by universal linguistic principles (White, 2003). Of particular interest are the questions of whether these universal properties are evident when there is no overt evidence for them in the L1. Will, for example, an L1-speaker who has no overt WH movement show
knowledge of constraints on WH movement in the interlanguage? These poverty of the stimulus cases are summarized nicely in Özcelik & Sprouse (2016); Özcelik (2017). The broader question is whether the knowledge of the L2 is the product of domain-general learning procedures (Wulff & Ellis, 2018) or whether we must invoke domain-specific algorithms. Our position here is that a domain-general hypothesis-testing model would have difficulty arriving at the hypothesis the WH word is licensed to remain in situ because the pitch is compressed in the syntactic WH domain; it must be domain-specific.

2.2.2. The Role of Input

While we take a modular, rationalist epistemological stance, we do not deny that the environmental input plays a critical role in L2 learning. See Yang et al., (2017), Yang (2017), and Lidz & Gagliardi (2015) for discussion of these issues. What is potentially interesting about the current study is that the subjects are classroom learners of Japanese as a Foreign Language. These learners have not received very much Japanese input (compared to naturalistic learners), and the structures in question (i.e., prosodic contiguity) have not been taught explicitly in class at all. Thus, if they acquire targetlike phonology and syntax while studying a foreign language at university, to us it seems highly unlikely that this learning is the result of a simple mapping16 of the classroom input. To us, such a result would argue for domain-specific constraints on the nature of the interlanguage grammar.

2.2.3. Shallow Structure

Clahsen & Felser (2006) presented a hypothesis (the shallow structure hypothesis) which attempted to account for some of the non-nativelike properties of interlanguage processing and interlanguage grammars by positing shallower, less-embedded syntactic representations in the non-native speakers. Clahsen & Felser (2017) revisit the model, and note that the ‘core claim of the SSH was that unlike native speakers, even highly proficient

---

16 A reviewer raises the question of what a ‘simple’ mapping is, or what is ‘easily’ read off the input. See Carroll (2001) for an in depth discussion of these issues. In brief, we would assume that phonetic properties such as nasality or stridency are robust phonetic properties easily made use of by second language learners. In contrast, the patterns of pitch variation in a sentence has many underlying correlates such as pitch accent, sentence type (question or statement), or information structure. For a learner to notice that there were no pitch rises between two particular syntactic categories in certain sentence types is, in our view, something that would require much more complex analysis.
L2 speakers tend to have problems building or manipulating abstract syntactic representations in real time and are guided more strongly than native speakers by semantic, pragmatic, probabilistic, or surface-level information.” Relevant to our project, they make explicit that the claims of the SSH are expected to hold in production tasks as well as comprehension tasks.

They also argue that since the original research in 2006 (which focused primarily on syntactic processing) there has been empirical evidence provided (Clahsen et al., 2010) which suggests that morphological processing is also affected. Thus, it would not be a stretch to infer that they would suspect that non-native speakers would also have difficulty “building or manipulating” abstract phonological representations. And while they acknowledge that non-native speakers can have hierarchical grammatical representations, they emphasize that the L2ers underuse abstract hierarchical representations in real-time processing. We will argue that our research subjects are constructing and manipulating abstract syntactic and phonological categories in their production of L2 Japanese WH questions. To the best of our knowledge, this is the first study to empirically test the SSH in the domain of phonology.

2.2.4 Phonetics & Phonology

We are investigating (1) the phonetic property of the implementation of a pitch-level boost on WH words, and (2) the grammatical property of phonological contiguity. The logical possibilities are that our subjects will demonstrate targetlike performance in both, neither, or one. It is our expectation that the production of these classroom learners will be accented phonetically, but that they have more of a chance of acquiring the target phonology.

\[^{17}\] A reviewer cautions us against over-interpreting our data with respect to the Shallow Structure Hypothesis. We feel the data are relevant for two reasons. (1) While it is logically possible that a researcher could assume shallow morphological and syntactic structure but complex phonological structure, we know of no one who takes this stance. (2) Given the nature of the syntax-phonology interface, our data provide an argument that the complex phonology is a diagnostic of complex syntax, and this is also a counter-argument to the Shallow Structure Hypothesis.
2.3 Hypotheses

Before presenting the results, let us formalize our hypotheses. For all of these, we are agnostic as to the role proficiency will play but will analyze the data to see whether the self-assessed proficiency level plays a role in their production.

1. We predict that our subjects will be more nativelike on the categorical phonological phenomena than the gradient phonetic elements. Specifically, we predict that they will have no prosodic phrases marked by pitch peaks between the WH word and the complementizer.

2. We predict that our subjects will be non-nativelike on the gradient phonetic phenomenon of the pitch boost on the WH word compared to DP objects. This is not because we feel the property is unlearnable, but rather that our subjects are classroom learners in a foreign language environment. We do not expect them to have had the necessary amount of input\(^{18}\) to have acquired this phonetic property.

3. We predict that, contra the Shallow Structure Hypothesis, our subjects will show evidence of abstract, recursive phonological constituents in their interlanguage which explain the properties of their Japanese L2 production.

3. Results

The initial measurements were made and recorded in Hertz. Then, in order to normalize the differences between male and female voices, we converted the Hertz scale to semitones using the formula \((12\log_2(\text{Hz}/100))\). Furthermore, as noted earlier, the difference between L1-English subjects and L1-other subjects was not significant \((p=0.315)\). This was based on a likelihood ratio test (LRT) of two nested linear mixed models: a full model with a random subject effect and fixed effects for object, first language, and the interaction of object and first language, and a reduced model without the interaction term. When it comes to the pitch differences on WH-objects versus DP-objects we will report on all subjects together.

\(^{18}\) A reviewer wonders about the quantification of an input threshold. While we feel it is beyond the scope of this paper to explore how much input might lead to the acquisition of these phonetic properties, we would note that the heritage learner literature (e.g., Kupisch & Rothman, 2016) is exploring these issues fruitfully.
3.1 Pitch Boost

Let us now turn to a discussion of the first component of Richards' notion of contiguity by looking at the instantiation of pitch boost by comparing sentences (in particular the bolded items) of the type shown in (23)-(25).

(23) a. Noboru-wa **piza-o**\(^{19}\) mottekitandesu ka?  
Noboru-TOP pizza-ACC bring - POL [+Q]  
Did Noburo bring pizza?

(23) b. Tarō-wa **nani-o** mottekitandesu ka?  
Taro-TOP what-ACC bring - POL [+Q]  
What did Taro bring?

(24) a. Miki-wa **kino nani-o** kai-mash-ita ka?  
Miki-TOP **yesterday what-ACC** buy-POL-PAST [+Q]  
What did Miki buy yesterday?

(24) b. Ponyo-wa **kino hon-o** kai-mash-ita ka?  
Ponyo-TOP **yesterday book-ACC** buy-POL-PAST [+Q]  
Did Ponyo buy a book yesterday?

(25) a. Naoya-ga **nanika-o** nomiya-de nonda.  
Naoya-NOM. Something-ACC bar-LOC drank  
Naoya drank something at the bar.

(25) b. Naoya-wa **nani-o** nomiya-de nonda no?  
Naoya-TOP **what-ACC. Bar-LOC drank** [+Q]  
What did Naoya drink at the bar?

We compared (by paired t-test, since we have paired observations for the data in sentence pairs 1-3) the average pitch on DP direct objects versus WH direct objects in these sentences across all sixteen subjects, shown in Table 1\(^{20}\). None of the three pairs of sentences had a significant difference in pitch.

---

\(^{19}\) A reviewer has noted that *piza* is a loanword and that this may raise issues of loanword phonology. This is an important point, and we are aware of the rich literature on the nature of the bilingual lexicon. However, this was the only loanword in our task, and we will have no reason to believe it lead to exceptional pitch results.

\(^{20}\) Our research design was not created to see if our non-native subjects were behaving in a nativelike fashion. The group sizes were not equal. Furthermore, it turned out that we had dialect variation in our native speaker group; variation which possibly affects the pitch boost phenomenon. For this reason, we are not reporting on the native/non-native comparison data. We accept the claim in the literature that native speakers utilise the
Table 1. WH vs DP pitch levels (in semitones) by sentence pair.

<table>
<thead>
<tr>
<th>Sentence Pair</th>
<th>Object</th>
<th>Average pitch (sd)</th>
<th>Average difference in pitch (95% CI)</th>
<th>Significance from paired t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DP</td>
<td>12.98 (6.79)</td>
<td>-0.80 (-2.57, 0.97)</td>
<td>p=0.349</td>
</tr>
<tr>
<td></td>
<td>WH</td>
<td>12.18 (7.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DP</td>
<td>13.59 (6.55)</td>
<td>-0.04 (-1.17, 1.08)</td>
<td>p=0.935</td>
</tr>
<tr>
<td></td>
<td>WH</td>
<td>13.54 (7.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DP</td>
<td>11.16 (6.65)</td>
<td>-0.90 (-2.36, 0.56)</td>
<td>p=0.207</td>
</tr>
<tr>
<td></td>
<td>WH</td>
<td>10.26 (6.44)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The preceding comparisons are illustrative but let us also compare across all of the sentences. Remember that Richards would predict that the pitch of WH phrases would be boosted compared to comparable non-WH phrases (which we call DPs for sake of contrast). This comparison is shown in Table 2.

Table 2. WH vs DP pitch levels all sentences.

| All L2 Subjects' DP Direct Objects Average | 12.90 |
| All L2 Subjects' WH Objects Average       | 12.78 |

Across all non-native Japanese speaking subjects there is no difference in the mean pitch of WH and DP words (p=0.801 on a paired t-test). We used a linear mixed model to account for the repeated measures on the same subjects, and this also revealed no significant difference between WH and DP (p=0.745, LRT of full model with random subject effect and fixed object effect and reduced model without the object effect). Furthermore, after controlling for proficiency level (as a fixed effect with the categories being either Intermediate or Advanced), we do not see differences between sentence types (DP vs WH) (p=0.197, LRT of full model with random subject effect and fixed effects for sentence type, proficiency, and the interaction of sentence type and proficiency, and a reduced model.

pitch boost on the WH-phrase. Our concern is whether our L2ers treat the WH-phrases differently from the DPs.
without the interaction term), indicating that the intermediate subjects were not behaving differently from the advanced subjects.

The lack of significant difference between the two sentence types, has, we think, two interesting implications. First, that the non-native speakers were not phonetically implementing the WH pitch boost found in native speakers. Secondly, this reveals that we are not seeing English echo questions being produced here ("Ponyo bought WHAT?"), but rather L2 Japanese WH questions. The L2 learners are not transferring English echo-question intonation here.

For in addition to WH-moved questions ("What should Ponyo eat?"), English allows WH in situ questions with a particular pragmatic force. A sentence such as *Ponyo ate WHAT?* could be used to indicate shock or surprise at what Ponyo ate. It could also be used to indicate uncertainty on the part of the listener as to what was actually said. In either case, the *what* receives a boost in pitch to indicate the marked interpretation. If the L1 English speakers were transferring the intonation from these structures then we might have expected a pitch boost on the Japanese in situ utterances. However, we see no significant boost. We have no reason to doubt that our subjects are producing Japanese WH questions in our production task.

### 3.2 Prosodic Structure

Let us turn to the second component of contiguity theory: prosodic contiguity.21

Table 3 shows the data from two of the advanced speakers for the sentence given in (20b). The boxed cells clearly indicate the level pitch between the WH Phrase (*nani-o*) and the question particle (*no*).

<table>
<thead>
<tr>
<th>Subject #</th>
<th>nani-o</th>
<th>nomiya-de</th>
<th>nonda</th>
<th>no [+Q]</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>5.94</td>
<td>0.511</td>
<td>1.33</td>
<td>5.82</td>
</tr>
<tr>
<td>S15</td>
<td>20.5</td>
<td>15.3</td>
<td>15.3</td>
<td>17.83</td>
</tr>
</tbody>
</table>

21 For this analysis, we had an additional subject, thus n=17.
This almost-completely level pitch contour between the WH word and [+Q] is clearly consistent with Richards' hypothesis in that there is no prosodic boundary (as would be indicated by a pitch rise) in what he calls the *WH domain*. For these speakers, we posit the structures (from Richards 2010) given in (26):

\[
(26) \quad [\text{DP WH DP [VP]}]_C \\
(\text{MinP}) (\text{MinP}) \\
(\text{MinP}) \quad \text{WH domain}
\]

The advanced subjects shown in Table 3 clearly show a nativelike prosodic pattern insofar as they maintain a level pitch contour between the WH word and the question particle, with higher pitch on the WH word and on the question particle. This particular sentence pair is used for illustration but we see the same pattern in other sentences.

In order to probe the question more thoroughly, we needed to operationalize the diagnostic of a prosodic boundary. As before, the pitch values key points of the target sentences were read off Praat. We looked at the 4 sentences given in (27):

\[
(27) \\
a. \quad \text{Naoya-wa nani-o nomiya-de nonda no?} \\
   \text{Naoya-TOP what-ACC bar-LOC drank Q} \\
   \text{What did Naoya drink at the bar?} \\
b. \quad \text{Taro-wa nani-o mottekita-ndesu ka?} \\
   \text{Taro-TOP what-ACC bought-POL Q} \\
   \text{What did Taro bring?} \\
c. \quad \text{Dare-ga nani-o kaimasita ka?} \\
   \text{who-NOM what-ACC bought-POL Q} \\
   \text{Who bought what?} \\
d. \quad \text{Miki-wa kino nani-o kaimasita ka?} \\
   \text{Miki-TOP yesterday what-ACC bought-POL Q} \\
   \text{What did Miki buy yesterday?}
\]

In each of these sentences, we looked at the change in pitch from Word1 to Word4 as indicated in Table 4:
Table 4. Sentential pitch patterns.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Word1</th>
<th>Word2</th>
<th>Word3</th>
<th>Word4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nani-o</td>
<td>nomiya-de</td>
<td>nonda</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>nani-o</td>
<td>mottekita</td>
<td>ndesu</td>
<td>ka</td>
</tr>
<tr>
<td>3</td>
<td>dare-ga</td>
<td>nani-o</td>
<td>kaimasita</td>
<td>ka</td>
</tr>
<tr>
<td>4</td>
<td>kino</td>
<td>nani-o</td>
<td>kaimasita</td>
<td>ka</td>
</tr>
</tbody>
</table>

In order to analyze our data set as a whole, we had to merge the male and female voices which, of course, have different pitch ranges. First, we needed to determine that the variation in scores was the same in the male and female groups. To do this, we conducted an F-test of equality of variances for the two groups. Two-sample F-tests for the equality of variances were performed at the alpha = .05 level, and the hypothesis (that the variances are equal) was not rejected. Given that the variances were equal, we were justified in adjusting the location of the mean of one of the groups rather than scaling the data. After converting Hertz scores to semitones, the males’ pitch was on average 11 semitones lower for each sentence word compared to the females’ pitch. Therefore, in order to make comparison across all subjects, we added 11 to each male measurement such that the ratio of means of the females and males would be approximately 1 indicating a similar mean pitch across the two genders. As a result of this shift, any differences which are observed between the pitch of words, are less likely to be the result of differences between the pitch of male and female voices. The working hypothesis for us to test with respect to phonological contiguity, then, would be that there would not be a pitch increase from Word1 to Word2 or from Word2 to Word3. In other words, we predict that in terms of pitch Word3<Word2, Word3< Word1, and Word2<Word1.

---

22 All of the variances fell within the range of 0.2958693 and 4.5960944 for the critical value limits of the F distribution.
Linear Mixed Effects (LME) Analyses

For each word pair comparison, we fit two linear mixed models to the data: a full model where speakers were random effects and the fixed effects were word, proficiency, and the word-proficiency interaction, followed by a reduced model with the proficiency terms removed. The fit of the full model is examined to see whether the intermediate subjects were behaving differently from the advanced subjects while an LRT was performed to probe the proficiency effect.

Word2/Word1

The full model showed that the intermediate and advanced groups did not behave significantly differently (p=0.823) and the LRT revealed that proficiency was not a significant factor (p=0.975) in explaining the pitch relationship of Word2 to Word1. We are thus justified in combining the intermediate and advanced subjects. Both the reduced and full models showed that the average pitch of Word2 is significantly less than the average pitch of Word1 (p<0.001, p=0.016 respectively). Results are given in Table 5.

Table 5. Results of linear mixed models for the comparison of Word1 and Word2.

<table>
<thead>
<tr>
<th></th>
<th>Full Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td>17.37</td>
<td>0.667</td>
<td>26.06</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word2 vs. Word1</td>
<td></td>
<td>-1.65</td>
<td>0.685</td>
<td>-2.41</td>
<td>0.016</td>
</tr>
<tr>
<td>Proficiency (Advanced vs. Intermediate)</td>
<td></td>
<td>-0.13</td>
<td>0.916</td>
<td>-0.14</td>
<td>0.888</td>
</tr>
<tr>
<td>Word2:Proficiency Advanced</td>
<td></td>
<td>0.21</td>
<td>0.941</td>
<td>0.22</td>
<td>0.823</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reduced Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td>17.31</td>
<td>0.457</td>
<td>37.83</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word2 vs. Word1</td>
<td></td>
<td>-1.54</td>
<td>0.469</td>
<td>-3.28</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Word3/Word1
The full model showed that the intermediate and advanced groups did not behave significantly differently ($p=0.757$) and the LRT revealed that proficiency was not a significant factor ($p=0.953$) in explaining the pitch relationship of Word3 to Word1. We are thus justified in combining the intermediate and advanced subjects. Both the reduced and full models showed that the average pitch of Word3 is significantly less than the average pitch of Word1 ($p<0.001$). Results are given in Table 6.

**Table 6.** Results of linear mixed models for the comparison of Word1 and Word3.

<table>
<thead>
<tr>
<th>Full Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>17.37</td>
<td>0.574</td>
<td>30.26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word3 vs. Word1</td>
<td>-3.79</td>
<td>0.644</td>
<td>-5.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Proficiency (Advanced vs. Intermediate)</td>
<td>-0.13</td>
<td>0.789</td>
<td>-0.16</td>
<td>0.87</td>
</tr>
<tr>
<td>Word3:Proficiency Advanced</td>
<td>0.27</td>
<td>0.885</td>
<td>0.31</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reduced Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>17.31</td>
<td>0.394</td>
<td>43.93</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word3 vs. Word1</td>
<td>-3.64</td>
<td>0.442</td>
<td>-8.24</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Word3/Word2
The full model showed that the intermediate and advanced groups did not behave significantly differently ($p=0.946$) and the LRT revealed that proficiency was not a significant factor ($p=0.978$) in explaining the pitch relationship of Word3 to Word2. We are thus justified in combining the intermediate and advanced subjects. Both the reduced and full models showed that the average pitch of Word3 is significantly less than the average pitch of Word2 ($p<0.001$, $p=0.002$ respectively). Results are given in Table 7.
Table 7. Results of linear mixed models for the comparison of Word2 and Word3.

<table>
<thead>
<tr>
<th></th>
<th>Full Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td>15.72</td>
<td>0.542</td>
<td>29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word3 vs. Word2</td>
<td></td>
<td>-2.14</td>
<td>0.687</td>
<td>-3.11</td>
<td>0.002</td>
</tr>
<tr>
<td>Proficiency (Advanced vs. Intermediate)</td>
<td></td>
<td>0.08</td>
<td>0.745</td>
<td>0.11</td>
<td>0.91</td>
</tr>
<tr>
<td>Word3:Proficiency Advanced</td>
<td></td>
<td>0.06</td>
<td>0.944</td>
<td>0.068</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reduced Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td>15.76</td>
<td>0.372</td>
<td>42.36</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word3 vs. Word2</td>
<td></td>
<td>-2.10</td>
<td>-.471</td>
<td>-4.46</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Summary

These data thus far reveal that (a) the intermediate and advanced groups are behaving the same when it comes to pitch patterns, and (b) that there is a significant decline in fundamental frequency from Word1 to Word2 and from Word2 to Word3. In other words, there are no prosodic boundaries between word One and word Four.

Interlanguage Grammars

These data suggest that the non-native speakers have acquired a grammar in which the phonological contiguity of the WH phrase and the question particle is maintained. This does have broader implications which we will return to in section 3.3 but for now we note that (a) this is not a property which could have been transferred from L1 English, and (b) that this is not a property which was taught in class. This latter point was established in the debriefing session of the subjects (some of the native speaker controls were the Japanese teachers). When we went over the goals of the research (after the data collection), all subjects and all teachers confirmed that this was not something covered in the textbooks or in the classroom instruction.

We will not go into statistical detail here, but we also probed to see if the non-native speakers were performing in a nativelike fashion with respect to the fundamental frequency of pitch descent. We used a 2-level proficiency variable: native or non-native. The results from the LRT comparing a full model with random effects for speaker and fixed effects for
word, proficiency, and the word-proficiency interaction to a reduced model without proficiency terms are shown in Table 8.

Table 8. Native vs. non-native pitch patterns.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Significance</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word2 vs Word1</td>
<td>p=0.171</td>
<td>NS=NNS</td>
</tr>
<tr>
<td>Word3 vs Word1</td>
<td>*p&lt;0.001</td>
<td>NS ≠ NNS</td>
</tr>
<tr>
<td>Word3 vs Word2</td>
<td>*p=0.001</td>
<td>NS ≠ NNS</td>
</tr>
</tbody>
</table>

This indicates that for the pitch drop from Word1 to Word2 the non-native speakers were behaving like the native speakers but for the other sequences (while there was, on average, a decline from Word2 to Word3 and from Word1 to Word3, these differences between groups were significant meaning that the non-native speakers were not implementing the decrease in fundamental frequency in a nativelike way.

It is worth noting that these two points are not incompatible: (1) the non-native speakers had significant categorical changes between all three target words, (2) the changes were not the same as the changes native speakers make. The L2ers have acquired Japanese prosodic phrasing which matches the Japanese syntax. But they are not nativelike in the gradient pitch patterns implementing these phrases. They have accents.

Let us consider this in the broader context of assessing second language speech. There is a long tradition (often couched in the Critical Period Hypothesis) to see if non-native speakers can perform a particular task (e.g. Voice Onset Time, vowel contrast) with the range of native speakers. It is clear that age of acquisition correlates highly with accentedness of L2 speech. However, as Munro (2008:194) notes, “native pronunciation in the L2 is not only uncommon but unnecessary.” The nativelikeness measurement bar is a high bar, and not the only bar. As Munro & Derwing (1995) have shown, heavily accented speech can be perfectly intelligible. So, our L2ers have acquired a targetlike phonology but have not acquired nativelike phonetics.

This is reminiscent of some of the work done in the acquisition of L2 stress
(Archibald, 1993; Kijak, 2009). Consider what an L2 stress error could look like. There are two possibilities. The first is that the L2er places the stress on the wrong syllable. For example, the word *aróma* being pronounced *ároma*. This is a phonological error. The second possibility is that the L2er places the stress on the correct syllable as a result of targetlike metrical settings but implements the stress prominence in a non-nativelike way. So, the L2er might have higher pitch on the second syllable of *aróma* but not have reduced vowels in the first and third syllable thus pronouncing the word [aróma] instead of [ərómə]. This would be a phonetic error.

What we are seeing in the performance of our subjects is targetlike phonology which achieves prosodic contiguity, but accented phonetic implementation.

### 3.3 Effects of Instruction/ UG-Constrained Interlanguage

These results suggest that the L2 learners *are* able to acquire the target grammar insofar as they have no pitch peaks within the WH domain. The interlanguage grammar licenses WH *in situ* by virtue of prosodic contiguity. We argue that the knowledge cannot be attributed to an effect of instruction but is rather the consequence of the interlanguage grammar being constrained by prosodic contiguity.

One reviewer has suggested that this targetlike performance is the result not of universal constraints on grammatical systems but rather the ‘tacit acquisition’ of Japanese phonology based on instructor input. While we would agree that the knowledge acquired is tacit in the sense that it is below conscious radar, we would not agree with the position that it arises from an implicit analysis of the sentence intonation. Crucially, there are some properties of Japanese speech that the L2 learners are *not* imitating correctly (e.g. the pitch boost on WH words). Our position explains why some properties are acquired while others are not. The ‘tacit acquisition’ argument cannot make such a distinction.

Furthermore, we would argue that this aspect of the Contiguity Principle could not have transferred from L1 English. Remember that English invokes the movement strategy to achieve linear contiguity *not* the strategy to achieve phonological contiguity. As the interlanguage grammars evidence phonological contiguity, the question is *what led the learners to adopt this strategy?* Our position is that the learners do not even consider
hypotheses that are not found in other natural languages; phonological contiguity is an option provided in the grammar-making toolbox that is UG.

3.4 Interfaces

Sorace (2011) argues that interfaces between syntax and pragmatics show optionality or indeterminate acquisition. Furthermore, she suggests that the syntax/lexicon interface shows optionality when involving encyclopedic knowledge or extralinguistic factors (external) but not with semantic roles (internal). Sorace (2011: 9) summarizes by saying:

"There is sufficient evidence for important developmental differences between linguistic structures that require conditions of a formal nature within the grammar, and structures that require the integration of contextual factors."

The structures under investigation here seem well-suited to contributing to the literature on the Interface Hypothesis in two ways. First of all, this is an area where the structures in question are narrowly grammatical (and do not involve pragmatic or other external factors). Secondly, in probing what actually causes the residual optionality observed in the L2 learners for structures involving external interfaces, Sorace et al. (2009) and Serratrice et al. (2009) argue that it is restricted input which drives the incomplete acquisition. The subjects discussed in this study are classroom learners of Japanese who have spent no extended periods in Japan. Their exposure to Japanese has not been extensive. And yet, the subjects have acquired the target structure of the phonological phrasing which licenses WH in situ. These results then, confirm the prediction of the interface hypothesis that these formal properties are, indeed, acquirable, and yet raise questions as to whether the acquisition is, in fact, input-driven. Rather, it seems to suggest more that the L2-learner grammars are governed by universal properties (such as contiguity) which are not directly read off the input.

3.5 Phonetics and Phonology

The empirical results of our study demonstrate a difference in ultimate attainment in the domains of phonetics and phonology. With respect to phonetics, the L2 learners were not
For Peer Review

implementing a phonetic boost on WH elements compared to DP elements (as has been reported in the literature (Hirotani, 2005; Ishihara, 2003; cited in Richards, 2010). However, they were encoding a representational system in which the grammars were constrained by universal principles; principles which cannot be directly inferred from the primary linguistic data, and which are not addressed explicitly via instruction.

Given such poverty of the stimulus effects (see Özcelik & Sprouse, 2016) in SLA, this has certain architectural implications which, to our minds, justify a well-defined phonetics/phonology demarcation. It is uncontroversial to note that phonology interfaces with other motoric and grammatical systems. The interface with morphology is crucial (Distributed Morphology, Embick, 2010; Prosodic Morphology, McCarthy & Prince, 1996; Zimmermann, 2017); Lexical Phonology (Rubach, 2008; Mohanon, 1986). The interface with syntax is critical (Selkirk, 2011). However, there is much work which blurs the line between phonetics and phonology (Flemming, 2017; Boersma & Pater, 2016). This view is in opposition to the stance of phonology as cognition (Reiss & Hale, 2008; Burton Roberts, et al., 1999). The roots of such a computational model of phonology may well stem from the tenets of minimalist architecture which view spell-out to Phonetic Form (PF) as largely a motoric interface. Even the work on the evolution of the language faculty (Berwick and Chomsky, 2016) views the central development to be Merge but still views phonology as largely a production device. But Richards (2010, 2016) reminds us that there is an earlier interaction between phonology and syntax in which phonological properties may license syntactic structures. Drawing on work such as Selkirk (2011) and Elfner (2015), we see evidence for syntactic constituents being mapped one-to-one onto prosodic domains (contra Nespor & Vogel, 1986). Thus phonological domains are also fundamentally recursive (see also Wagner, 2010).

Thus, we can view L2 phonology as cognition and recognize that a phonetics/phonology split actually helps to account for the knowledge and behavior of our subjects. They are more successful at the categorial phonological phenomena (i.e., groupings of prosodic phrases) than they are at gradient phonetic phenomena (i.e. pitch level of a particular syntactic category). This is consistent with aspects of the literature on
L2 phonology and nativelikeness. It is well-documented that it is uncommon for adult second language learners to fall within the range of native-speaker performance when it comes to gradient phonetic phenomena such as VOT or vowel quality (Flege, Munro & Mackay, 1995; Abrahamsson & Hyltenstam, 2009). We are not proposing a causal connection here (i.e., that these new contrasts are difficult because they are gradient). Indeed, it is well-established that some gradient contrasts are easier to acquire than others (Bohn, 1995; Polka & Bohn, 2011). What we are suggesting is that one way of interpreting the differential learner performance is that they have acquired the targetlike phonology but not target like phonetic implementation of Japanese WH questions.

3.6 Return to Learnability

As Yang (2017) reminds us, language acquisition is a fundamentally psychological activity, and our acquisition models must take into account both universal constraints on the system, and the type of input available to the learner.

We therefore note that the learnability of the two phonetic and phonological phenomena is quite different. Phonological representations are constrained by UG (Match Theory; Contiguity). We can conceptualize it along the following lines:

*IF WH remains in situ THEN no prosodic boundaries between WH and [Q] particle.*

The learners are unable to set up grammars which violate UG properties. This is consistent with other L2 UG properties (Özcelik & Sprouse, 2016; Archibald & Yousefi, 2018)\(^{23}\).

Language-specific phonetic implementation\(^{24}\), on the other hand, must be picked up only from the environment, and, in this case at least, the patterns are much subtler, and potentially masked. The L2 learner must note the fundamental frequency from a range of speakers (men, women, children) on different syntactic categories (WH- and non-WH DPs), keep track of the effects of intonation, and pragmatics factors such as presupposition...

\(^{23}\) A reviewer notes that in contrast to the wealth of literature addressing UG principles in L2 syntax, there has been much less attention to UG in L2 phonology. Perhaps this paper will trigger future research by others!

\(^{24}\) Motoric skills which are not governed by universal properties of the language faculty.
and focus, and from this determine that WH words should have higher pitch than DP objects. At the very least, we would expect it to take longer for the L2 learner to acquire this property of Japanese.

4.0 Conclusion

In this paper we have looked at some of the structural properties of L2 Japanese WH questions. In Japanese the WH words are licensed to remain in situ by virtue of the prosodic contiguity (Richards, 2016) properties of the phrases which have no prosodic boundaries between the WH word and the question particle. In a rehearsed-reading, sentence production task, we looked to see whether non-native speakers of Japanese who are learning the L2 in university classes (thus with relatively limited input compared to naturalistic learners) are able to acquire grammars which are constrained by such universal properties as Match Theory (Elfner, 2015; Selkirk, 2011) and Contiguity Theory (Richards, 2010, 2016). While LME analyses of the pitch contours reveal that the L2ers have not acquired the phonetic implementation of the documented pitch boost on WH words compared to non-WH DPs (i.e., they have foreign accents), our data show that the subjects have acquired the pitch compression patterns indicative of having no prosodic phrases intervening between the WH word and the question particle.

The classroom learners were given explicit instruction as to the word order of Japanese WH-questions. Their classroom environment appears to have been sufficient for them to acquire targetlike syntactic structure for this word order. This syntactic knowledge, in turn, generates a targetlike prosodic structure.

Thus, we argue, that the data are supportive of the position that interlanguage grammars are constrained by universal grammatical principles of Contiguity and Match theory which govern the prosodic contiguity of WH-phrase licensing. We also presented these results as being counter to the Shallow Structure Hypothesis (Clahsen & Felser, 2017).
References


http://mc.manuscriptcentral.com/SLR


### Appendix A. Reading Sentences

1. Satsuki ga depāto de ojōsan no tame ni ranpu o eranda.
   Satsuki-NOM dept.store-LOC daughter-GEN sake-DAT lamp-ACC chose.
   サツキが、デパートでお嬢さんのためにランプを選んだ。
   'Satsuki chose a lamp for her daughter at the department store.'

2. Naoya ga nanika o nomiya de nonda.
   Naoya-NOM something-ACC bar-LOC drank.
   ナオヤが、何かを飲み屋で飲んだ。
   'Naoya drank something at the bar.'

3. Naoya wa nani o nomiya de nonda no?
   Naoya-TOP what-ACC bar-LOC drank Q.
   ナオヤは、何を飲み屋で飲んだの？
   'What did Naoya drink at the bar?'

4. Umi ga depāto de dare no tame ni ranpu o eranda no?
   Umi-NOM dept.store-LOC who-GEN sake-DAT lamp-ACC chose Q.
   ウミが、デパートで誰のためにランプを選んだの？
   'Who did Umi choose a lamp for at the department store?'

5. Dare ga depāto de dare no tame ni ranpu o eranda no?
   Who-NOM dept.store who-DAT
   誰が、デパートで誰のためにランプを選んだの？
'Who chose a lamp for whom at the department store?'

6. Anata wa Haru ga nani o suki ka shitteimasu ka?
あなたは、ハルが何を好きか知っていますか?
'What do you know Haru likes?'

7. Anata wa dare ga Tatsu o suki ka shitteimasu ka?
あなたは、誰がタツオを好きか知っていますか?
'Who do you think likes Tatsuo?'

8. Dare ga nani o kaimasita ka?
誰が何を買いましたか?
'Who bought what?'

9. Seiko wa dare ga tsukutta sushi o tabeta no?
セイコは、誰が作った寿司を食べたの?
'Seiko ate sushi that who made?'

10. Shiro wa hon’ya de nani o katta no?
シロは、本屋で何を買ったの?
'What did Shiro buy at the bookstore?'

11. Kumiko wa doko de zasshi o katta no?
クミコは、どこで雑誌を買ったの?
'Where did Kumiko buy a magazine?'

12. Hayao ga senshū katta hon ga nakunatta
ハヤオが先週買った本が、無くなった。
'The book Hayao bought last week disappeared.'

13. Masahiro ga katta hon ga itsu nakunatta no?
マサヒロが買った本は、いつ無くなったの?
'When did the book Masahiro bought disappear?'

14. Ponyo wa kinō hon o kaimashita ka?
ポニョは、昨日本を買いましたか?
'Did Ponyo buy a book yesterday?'
15. Miki wa kinō nani o kai-mashita ka?
Miki-TOP yesterday what-ACC bought-POL Q
ミキは、昨日何を買いましたか?
'What did Miki buy yesterday?'

16. Miki wa kinō hon o kai-mashita.
Miki-TOP yesterday book-ACC bought-POL
ミキは、昨日本を買いました。
'Miki bought a book yesterday.'

17. Noboru wa piza o mottekitandesu ka?
Noburu-TOP pizza-ACC brought-POL Q
ノボルは、ピザを持って来たんですか?
'Did Noburu bring pizza?'

18. Noboru wa kinō piza o kai-mashita.
Noburu-TOP yesterday pizza-ACC bought-POL
ノボルは、昨日ピザを買いました。
'Nobouru brought pizza yesterday.'

19. Tarō wa nani o mottekitandesu ka?
Taro-TOP what-ACC bought-POL Q
タローは、何を持って来たんですか?
'What did Taro bring?'

Appendix B: Subject Profiles

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Sex</th>
<th>Proficiency</th>
<th>Either Parent Speaks Japanese</th>
<th>Mainly Instructed or Natural Input?</th>
<th>Length of Residency in Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>M</td>
<td>INT</td>
<td>N</td>
<td>Instructed</td>
<td>&gt;1 Year</td>
</tr>
<tr>
<td>JM</td>
<td>F</td>
<td>INT</td>
<td>N</td>
<td>Instructed</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>KR</td>
<td>F</td>
<td>INT</td>
<td>N</td>
<td>Instructed</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>CS</td>
<td>F</td>
<td>ADV</td>
<td>N</td>
<td>Instructed</td>
<td>&lt;1 month</td>
</tr>
<tr>
<td>A</td>
<td>M</td>
<td>ADV</td>
<td>N</td>
<td>Natural</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>RKV</td>
<td>M</td>
<td>INT</td>
<td>N</td>
<td>Instructed</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>KSD</td>
<td>F</td>
<td>ADV</td>
<td>N</td>
<td>Instructed</td>
<td>&lt;1 month</td>
</tr>
<tr>
<td>ZM</td>
<td>M</td>
<td>ADV</td>
<td>N</td>
<td>Natural</td>
<td>&lt;1 month</td>
</tr>
<tr>
<td>TF</td>
<td>F</td>
<td>ADV</td>
<td>N</td>
<td>I=N</td>
<td>&gt;1 year</td>
</tr>
<tr>
<td>CJ</td>
<td>M</td>
<td>INT</td>
<td>N</td>
<td>Both</td>
<td>&gt;1 year</td>
</tr>
<tr>
<td>OW</td>
<td>M</td>
<td>ADV</td>
<td>N</td>
<td>I=N</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>AK</td>
<td>F</td>
<td>INT</td>
<td>Y</td>
<td>Natural</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>JJ</td>
<td>F</td>
<td>INT</td>
<td>N</td>
<td>Natural</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>RC</td>
<td>M</td>
<td>INT</td>
<td>N</td>
<td>Natural</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td>SK1</td>
<td>F</td>
<td>INT</td>
<td>Y</td>
<td>I=N</td>
<td>&lt;1 year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>F</td>
<td>ADV</td>
<td>Y</td>
<td>Natural</td>
<td>&lt;1 month</td>
</tr>
<tr>
<td>SK2</td>
<td>F</td>
<td>ADV</td>
<td>M</td>
<td>Natural</td>
<td>&gt;1 year</td>
</tr>
</tbody>
</table>