



# Prosodic versus Sequential Constraints in Phonotactics : A Case Study from Japanese

Pinter, Gabor

---

(Citation)

神戸言語学論叢, 5:193-206

(Issue Date)

2007-12

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81001534>

(URL)

<https://hdl.handle.net/20.500.14094/81001534>



# PROSODIC VERSUS SEQUENTIAL CONSTRAINTS IN PHONOTACTICS A CASE STUDY FROM JAPANESE\*

Gábor Pintér

*Kobe University*

## 1. INTRODUCTION

The aim of phonotactics by definition is to describe those sequential arrangements of phonological units which can occur in a language (Crystal 1997:292). Although this definition does not overtly specify any methodology, the vast majority of phonotactic studies base their analyses on prosodic constraints. Branching onsets and nuclei, minimal sonority distance restriction, sonority sequencing principle are all familiar concepts. However there is a growing number of studies that propagate a different direction. In their view the inventorial characteristics and distributional peculiarities of sounds are traced back to principles of articulatory and perceptual phonetics. The present paper aims to add a further piece to this body of research by providing an analysis for some co-occurrence restrictions in Japanese consonant-vowel sequences. It shows that the ungrammaticality of \*[si], \*[ni] and \*[hi] sequences in Japanese originates in the fact that the high vowel masks perceptual cues that mark their distinctions from [ci] [ni] and [çi] respectively.

In order to emphasize the divergence between prosodic and phonetic views of phonotactics the present article approaches the problem of Japanese CV restrictions from a somewhat wider perspective. Section 2 outlines the basic mechanism of prosodic and sequential/phonetic approaches. Section 3 highlights two fundamental flaws in the logic of prosodic paradigms. Section 4 discusses the problem of Japanese CV restrictions. This section compares prosodic and sequential/phonetic analyses to the problem and demonstrates the superiority of the phonetic approach. Section 5 summarizes the article and concludes that instead of striving to exclusively define one framework to cover every aspect of phonotactics, it would be more promising to look for a division of labor between prosodic and linear approaches in a coherent system.

## 2. PROSODIC VERSUS SEQUENTIAL LICENSING

The following sections sketch out the most important characteristics of prosodic and sequential approaches to phonotactics for further evaluations.

### *2.1 Prosodic licensing*

Prosodic approaches to phonotactics share the assumption that segments can be arranged recursively into higher units. These units or constituents and their particular

way of hierarchical arrangements make up the syllable structure. Currently, there are two major views on how to represent this structure. According to the classical Onset-Rhyme approach, the syllable node branches into onset and rhyme constituents, yielding a structure as depicted in *Figure 1*. In contrast, the Moraic Theory puts emphasis on the representation of syllable weight and uses moras – that is weight units – as immediate constituents under the syllable node *Figure 2*. In this latter view the so-called “onset segments” are the ones which are directly associated with the syllable node. Needless to say, there are many variations of these to two basic structures; however their comparison or discussion is not the topic of the present research (but see Ewen & van der Hulst 2001, Hayes 1989).

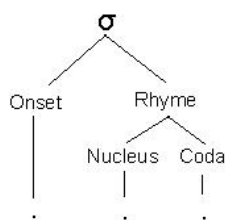


*Figure 1*

*Syllable structure in Onset-Rhyme Theory*

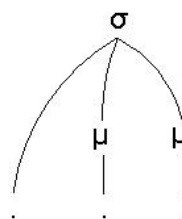


*Figure 2*

*Syllable structure in Moraic Theory*

In prosodic approaches phonotactic well-formedness is defined by reference to syllable structure. First of all, every segment in a word must be associated with a syllable position. Second, the association between syllable nodes and segments – or features – must follow some general principles. All of these components – including some structural parameters (e.g. for branching onsets) – are language-specific, nevertheless showing universal patterns. These universal patterns are defined in terms of implicational universals based on the Sonority Hierarchy. For example, onset position shows a preference for low sonority segments (e.g. voiceless stops), while nucleus opts for high sonority ones (e.g. low vowels). Every language specifies the range of segments that can occupy certain syllabic positions. This rather straightforward way of controlling association between segments and syllable nodes is called Prosodic Licensing (Ito 1988). For example, in many languages velar nasals are not licensed by onset position (e.g. English, French, Spanish... etc). The restricted licensing potential of codas is also a well-known fact: codas often cannot accommodate voiced obstruents (e.g. in German, Russian) or features for place of articulation (e.g. in Japanese see Ito 1988).

A less straightforward way of constraining segment-to-position associations is expressed by the Sonority Sequencing Principle (Selkirk 1984). This principle requires that syllable peaks are occupied by segments of the highest sonority (i.e. most often by vowels) and the magnitude of sonority decreases towards the margins of the syllable. It is easy to see that Prosodic Licensing coupled with the Sonority Sequencing Principle can greatly restrict the number of possible syllables.

A further restriction is expressed by co-occurrence restrictions between segments that belong to the same constituent. For example, a word-initial stop-nasal cluster satisfies the requirement for the pre-nuclear rising sonority profile; however this onset cluster is not allowed in many languages (e.g. English, French, Spanish, Italian). Its absence is attributed to the Minimal Sonority Principle which defines the minimum

amount of sonority distance that onset clusters should provide. It is important to note, that the same sound sequence is well-formed if it stretches over a syllable boundary.

(1) Co-occurrence restrictions in Onset clusters

- |    |                    |               |                |
|----|--------------------|---------------|----------------|
| a. | onset (*kn-)       | ‘know’        | [n]ow, *[kn]ow |
| b. | coda-onset (-k.n-) | ‘acknowledge’ | a[k.n]owledge  |

In sum, the common features of prosodic theories can be outlined as follows. First, the most basic component in prosodic theory is a well-defined syllable structure. Second, different licensing potentials are attributed to syllable positions. Third, the associations between segments and positions are controlled by a requirement for a rising-falling sonority profile and co-occurrence restrictions within syllable-constituents. The specific configuration for all of these components are determined language specifically, nevertheless they follow sonority driven implicational universals.

## 2.2 Phonetic/Sequential Licensing

Using sequential constraints without reference to syllable structure is not a novel idea in phonology. *The Sound Pattern of English* by Chomsky & Halle (1968) proved that it is possible to substitute syllables by linear descriptions. Even Theo Venneman, who is often credited with (re)introducing the syllable into generative phonology admits the following:

- (2) "[A]ll phonological processes which can be stated in a general way with the use of syllable boundaries can also be stated without them, simply including the environments of the syllabification rules."

(Venneman 1972:2)

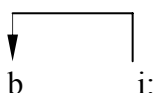
Of course, mechanically translating prosodic rules and restrictions into sequential constraints can not raise much interest. The ultimate question is not whether phonotactics can be cultivated without syllables – because it can – but how to justify such an approach.

The sequential approach being developed here has its grounds in perceptual and articulatory phonetics and relies heavily on works of Steriade (1999, 2001), Blevins (2003, 2004), Ohala (1992) and Browman & Goldstein (1986, 1990). Sequential Licensing claims that a sequence is grammatical if all of its segments are phonetically licensed. The concept of phonetic licensing involves both articulatory and perceptual aspects; however, due to spatial limitations only the latter one is going to be spelled out here in details. Briefly, Articulatory Licensing expresses an implicational hierarchy of speech sounds based on the articulatory effort their pronunciation requires.

The concept of Perceptual Licensing has its roots in the achievements of speech perception studies (Pisoni & Remez 2005, Raphael 2005). It claims that a segment is perceptually licensed if its perceptual cues can be recovered from the acoustic signal (Wright 2001). Cues can manifest either within the temporal domain of the segment or on neighboring sounds. A segment is *self-licensed* in the former case, that is, if the cues that are necessary for its perception do not extend to adjacent sounds. For example, vowels are primarily cued by their formant values, and they can be perceived without any neighboring sounds. In contrast, the most salient cues for stop consonants are

release bursts and second formant transitions. These features typically occur on adjacent segments. That is why stops usually need perceptual licensors, that is, other segments that accommodate these perceptual cues. Bursts and formant transitions can manifest most saliently on vowels, so it is nothing but surprising that, for example, consonant letters in the alphabet are spelled together with vowels (e.g. ‘B’ [bi:], ‘K’ [kei]).

### (3) Perceptual licensing



Licensing is not a categorical but a gradual concept: some contexts are better licensors for a feature or a segment than others. In terms of phonetics this means that some environments prove better perceptibility for certain cues than others. For example, the best perceptual licensing environment for a stop is intervocalic position, and the worst one is between other stops. Word boundary on the right side is not as bad a context as one may think, as it lets the release burst manifest. Environments can be ranked according to their perceptual licensing potential.

### (4) Perceptual licensing potential of contexts

| <i>Left context</i> |   | <i>Right Context</i> |  |
|---------------------|---|----------------------|--|
| C                   | – | C                    |  |
| #                   | – | C                    |  |
| R                   | – | C                    |  |
| R                   | – | #                    |  |
| V                   | – | V                    |  |

↑  
increasing  
perceptual  
difficulty

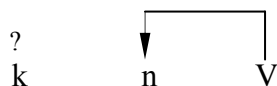
*(based on Blevins 2003:387)*

Perceptibility rankings can be interpreted as implicational universals. If a segment or feature occurs in a certain environment it is also expected to occur in a perceptually more enhancing environment.

In order to demonstrate how Perceptual Licensing works, consider the above mentioned problem of /kn/ sequences in English. Prosodic Licensing claims that /kn/ cannot occur as an onset cluster because the sonority distance between the two segments is not sufficient. First of all, let us translate the prosodic terminology into sequential description. Being an onset cluster means that the sequence is preceded by a word boundary or a consonant, and followed by a vowel. Accordingly, the following forms are under discussion: /#knV/ and /CknV/. A perceptual approach to this problem would focus on the acoustic cues for the velar stop (k). The absence of /k/ in these contexts is due to the fact that neither the word boundary/consonant on the left side nor a nasal on the right side offers proper licensing for /k/.<sup>1</sup> Formant transitions and release bursts are not sufficiently retrievable in these environments for English listeners. On the other hand, the well-formedness of the coda-onset configuration (linearly Vk.nV) is attributable to the fact that neighboring vowels can properly license /k/.

### (5) Perceptual Licensing of /kn/ sequences ①

(a) In onset position: \*#kn- ‘know’ → [nou] \*[knou]



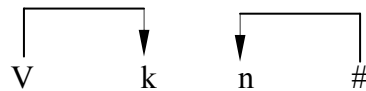
(5) (b) In coda-onset position: -k.n- ‘acknowledge’



In casual speech /kn/ can also appear word finally in such forms as ‘broken’ [brəʊk(ə)n] or ‘taken’ [teɪk(ə)n]. This is not surprising if we accept that nasals can be licensed by word final position. Note that word-final stop-nasal sequences are considered to be exceptional in the prosodic paradigm and treated by introducing the idea of ‘syllabic consonants’ into the theory. Nevertheless, this idiosyncratic extension of the syllable theory does not provide any insight to the problem.

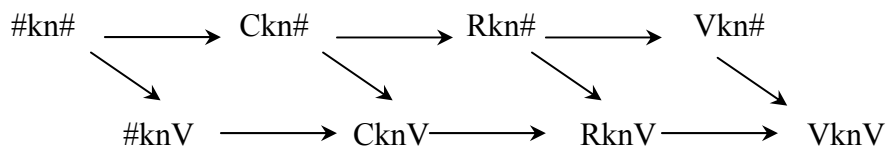
(6) Perceptual Licensing of /kn/ sequences ②

In coda: -kn# ‘broken’



Of course, confining ourselves to a few cases can not make justice between prosodic and perceptual approaches. The really attractive aspect of Perceptual Licensing is that it offers a straightforward and intuitive way of making implications about phonotactic grammaticality of sounds in various contexts. In case of /kn/, we know that the perceptual salience of /k/ - concentrating only its left side - is increasing in the following order: #k < Ck < Rk < Vk; where R is a sonorant sound and C is an obstruent. As for /n/, it is more salient in front of a vowel than word finally: n# < nV. For sake of brevity I discard other cases here. The combination of these two scales yields in an implicational hierarchy depicted in Figure (6).

(7) Perceptual Implicational hierarchy of /kn/ as a function of context



Occurrences of /kn/ in a language should form a non-interrupted chain in the hierarchy. The fact that in English not only /VknV/ but also /RknV/ and /Vkn#/ forms are attested is in accordance with the implicational hierarchy. Of course more typological data is necessary to fully justify this approach.

(8) Attested /kn/ forms in English

- Vkn# : broken, taken
- VknV : acknowledge, thickness
- RknV : darkness, banknote, frankness

Since an in-depth explanation of Sequential Licensing with discussions of both perceptual and articulatory licensing would far exceed the scope of the present article, let us stop after this brief introduction and summarize the most fundamental features of the theory. Sequential Licensing aims to explain phonotactic regularities in languages based on phonetic characteristics of speech sounds. As opposed to the Prosodic Licensing approach, it does not use the syllable as an explanatory device rather it evaluates perceptual and articulatory viability of segments as a function of their contexts. A sequence is grammatical if all of its segments are perceptually and articulatorily properly licensed.

It is important to note here that this paradigm does not aim to banish the concept of syllable from the phonological dictionary for good. It only claims that phonetically grounded linear explanations can offer better insight to phonotactical problems than insufficiently motivated prosodic approaches. Undoubtedly, the syllable has important roles in some fields of phonology (e.g. in weight related phenomena), however segmental phonotactics can benefit more from a phonetically motivated linear approach.

### 3. ONTOLOGICAL ARGUMENTS

The present section aims to criticize two fundamental flaws of prosodic approaches: first, the controversial status of Sonority Hierarchy and second, the circularity of argumentations in phonotactic descriptions.

#### *3.1 Sonority Hierarchy and its lack of grounding*

The use of the Sonority Hierarchy in phonological research is so widespread and roots so deeply in practice that its adequacy in phonotactic description is seldom questioned (but see Ohala 1992, Steriade 1999, Coté 2000, Blevins 2003). Nevertheless linguistic theories are not to be evaluated on the basis of popularity, and widespread use in itself does not add up to justification. It is not unlikely that sonority, the backbone of the syllable theory, may easily turn out to be its Achilles' heel.

The idea of sonority hierarchy, which is a unitary segment scale to describe phonotactic patterns, existed well-before structuralism, but its phonetic background was not addressed in research until the 20th century (for reviews see Ohala 1992; Blevins 1995). Otto Jespersen was one of the first modern linguists who tried to give an objective phonetic justification for such a scale. In a perceptual experiment he measured the distances from where phonemic oppositions ceased to be distinguishable (Jespersen 1932:191, reported by Nathan 1989:62). Linking sonority to loudness by ranking the greatest distances that guaranteed perceptibility was an approach with promising results: vowels had the greatest sonority while voiceless stops the smallest. The classroom definition of sonority is based on this phonetic interpretation.

#### (9) Sonority Hierarchy by Jespersen (1932)

voiceless stops < voiced stops < voiced fricatives < nasals < liquids < glides < high vowels < mid vowels < low vowels

While Jespersen was confined to relatively unsophisticated techniques in his efforts to give sonority an objective measure, technical developments in the post-war period

enabled speech scientists to investigate acoustic aspects of speech sounds in a direct way. Laboratory instruments made it possible to measure intensity, that is, the acoustic correlate of loudness. Ladefoged (1975) reports the following sonority ranking according to his estimations of intensity.

(10) Relative sonority scale for English by Ladefoged (1975:220)

$k < t < d < ʃ < s < v < z < m < n < l < i < u < ɪ < ε < æ < a$

Using similar instrumental methods Fry (1979) establishes a more elaborated scale. At some points his results violate some generally believed tendencies. For example, fricatives are to be more sonorous than stops, but we can see  $[θ] < [p]$  and  $[v] < [g]$ .

(11) Relative Intensity Scale by Fry (1979) (partial)

$θ < p f < b d < v ð < t g k < z s < ʒ ʒ < n < ts < m < ŋ < sh < r j l < w \dots$

Lindblom (1983) tries to pin down sonority on jaw replacement measured in continuous speech. His experiment denotes a smooth shift from perceptual to articulatory orientation, with some remarkable differences compared to previous scales.

(12) Jaw replacement scale for Swedish by Lindblom (1983)(partial)

CLOSED <  $s ʃ$  <  $\begin{matrix} p t k f \\ b d g \end{matrix}$  <  $\begin{matrix} m n ŋ \\ j v \end{matrix}$  <  $r l$  <  $a$  < OPEN

The great number of alternatives highlights the futility of attempts to cover all phonotactic problems with an absolute phonetically grounded scale. Indeed, all of these hierarchies have their preferred range of applications beyond which they face serious difficulties. For example, the idiosyncratic behavior of /s/ as a prefix to syllables can be grasped easily with Lindblom's scale; however, the same scale would predict /hC/ onset clusters as well.

Facing this dilemma, it seems to be unreasonable to insist on the idea of an absolute and unitary ranking of segments in order to explain several remotely related phenomena. Indeed, many phonological works noticed this problem and aimed to remedy it. For example, proposals have been made to use Sonority Hierarchy in a language-dependent parametric fashion (Malsch & Fulcher 1989), or to interpret it as a complex prototypical category in which articulatory and perceptual measures together decide sonority rankings (Nathan 1989). Some research even proposed to use sonority in a context sensitive manner (Larson 1992). None of these proposals managed to take root. Of course, this controversy does not prevent the researcher from using sonority as an explanatory device – while explicitly postponing the justification for it to some future research.

### 3.2 Circularity

Phonology – just as theoretical linguistics in general – provides fertile ground for producing novel rules, constraints and principles. Although there is an increasing need for establishing independent grounds in phonological explanations (Archangeli & Pulleyblank 1993, Hayes et al. 2004, Hume & Johnson 2001, but see Hale & Reiss 2000a, 2000b for criticism), many explanatory devices and techniques take off from

purely theoretical considerations. The point I am making here is not the apparent problem of proliferation of ad hoc constraints or rules (McMahon 2000), rather the unconscious practice of declaring generalizations to be full-fledged explanations, and leaving the real driving factors unknown. Prosodic views of phonotactics provide some examples of this type.

The Minimal Sonority Distance Principle defines the minimum amount of sonority distance that must hold within onset segments. In many Indo-European languages (e.g. English, French, Italian) there is a borderline between attested stop-liquid clusters and non-attested stop-nasal ones (e.g. /kr/ vs. \*/kn/). It was these coinciding observations that allowed the analyst to assign numeric indices to segments alongside the sonority scale (Selkirk 1984). Afterwards it was enough to say that stops and sonorants have to be at least 5 units away.

(13) Sonority indices by Selkirk (1984)

|       |       |   |     |   |     |   |   |   |     |    |
|-------|-------|---|-----|---|-----|---|---|---|-----|----|
| p t k | b d g | f | v z | s | m n | l | r | u | e o | a  |
| 0.5   | 1     | 2 | 3   | 4 | 5   | 6 | 7 | 8 | 9   | 10 |

Impressed by the clarity of this explanation, subsequent studies invoke the Minimal Sonority Distance Principle to give account for onset cluster patterns in other languages. What is usually left unnoticed is that this principle does not give any substantial explanations for the observed data, which leads to circularity. Without independent evidence this device is nothing more than a tool for generalizations. Consequently, its application in phonotactics is equal to matching individual cases to a general pattern.

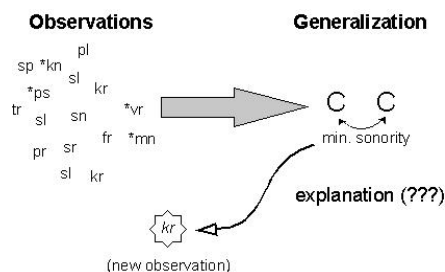


Figure 3 Circular Arguments

The absurdity of referring to the Minimal Sonority Distance Principle as an explanation can be aptly illustrated by the following example. Let us think of the variation in average air temperature as a function of time. From our observations we know that changes in the temperature follow a roughly a 365 day-long cycle with one relatively warm and one relatively cold season. With some corrections this pattern seems to hold in the majority of living spaces. This is a *generalization*. Let us assume that we discover a new territory which surprisingly shows a similar 365-day cycle. Can we say that our previous generalization is *the explanation* for the pattern in the newly discovered territory? Obviously not. This would be a circular argument. The real explanation relies on the astrological fact that it takes 365 days for the Earth to orbit around the Sun (hence the 365-day cycle), and the temperature/seasonal variation is caused by a 23.5° tilt in the axis of our planet. Thus every territory is expected to show this variation on the surface of the Earth. The Minimal Sonority Distance Principle in its present form is only a generalization just like the observation about the 365 day cycle. The independent

motivation - corresponding to the astrological facts in the metaphor - that leads us to the real explanation has yet to be found.

Accumulating cases, collecting data that fits into general pattern is a necessary and important part of the scientific activity. It can expose the problem from several points of view and give us implications about the nature of the underlying causes. However the generalization - no matter how important it is - should not be mistakenly understood as the cause itself. Sonority based principles of the prosodic phonology together with their controversies denote the problems and not solutions in phonotactics.

#### 4. JAPANESE CV RESTRICTIONS

In this section I turn to the problem of CV restrictions in Japanese and compare how prosodic and sequential approaches score on this task.

##### 4.1 Japanese CV restrictions

The strongly restricted phonological system of Japanese is an often cited topic in phonological studies. While the small syllable inventory of Japanese is its most frequently invoked feature, its limitations on possible consonant-vowel sequences are also deserves attention. It is not uncommon in Japanese that in front of high vowels some phonemes can not occur. For example, alveolar voiceless fricatives cannot occur before the front high vowels (\*[si]). Since auditory [si] is usually perceived as /sh/, the problem can be interpreted as the neutralization of the existing /s/-/sh/ opposition before /i/. The following table enlists some similar cases.

##### (14) CV restrictions in Japanese

| <u>Absent</u> | <u>Attested</u> | <u>Opposition</u> | <u>Minimal Pairs</u> |                      |
|---------------|-----------------|-------------------|----------------------|----------------------|
| *hi           | çi              | h-ç               | haku - çaku          | 'vomit' - '100'      |
| *ni           | ji              | n-ɲ               | noo - joo            | 'brain' - 'urine'    |
| *si           | çi              | s-ç               | sakai - çakai        | 'border' - 'society' |

As the right column shows, these sounds are distinct phonemes and it is not difficult to find minimal pairs with them in contexts other than /i/.

##### 4.2 Prosodic Approach

Analyzing consonant-vowel restrictions seems to be an ill-suited problem for prosodic approaches for two main reasons. First, co-occurrence restrictions are usually expressed in terms of sonority, and looking at the data (\*[hu], \*[hi], \*[ni], \*[si]) it is questionable whether or not any sonority-based constraint would be able to generalize over these cases. Apparently, place of articulation is going to be the key to the problem; however this does not translate well into sonority restrictions.

Second, co-occurrence restrictions are believed to hold between constituents, however CV sequences parse into onsets and nuclei, which usually do not form constituents in the syllable structure. As for the onset-rhyme theory, one of the major motivations behind splitting syllables into onsets and rhymes is that they can freely combine. This very principle seems to be broken in Japanese, where the number of legitimate onset plus nucleus combinations is strongly limited. The Moraic Theory in its classical form faces similar problems, however it is not impossible to change the

original configuration into one in which consonants and subsequent vowels hang from the same moraic node as it is shown in *Figure 5*. Among others Kubozono (1985, 1989, 1996) and Katada (1990) argue for such moraic representation for Japanese on basis of speech errors, blend formations and language games. While this representation offers structural motivation for Japanese CV restrictions, it leaves open many questions. Although a thorough discussion of sub-syllabic structures will not be included here, it is worth noting that for example consonant-glide-vowel sequences (e.g. /kyo/) according to this view have flat structure, which simply leads us back to linearity.

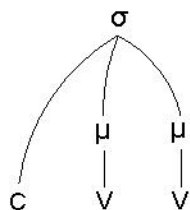


Figure 4  
Traditional moraic structure

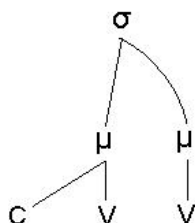


Figure 5  
An alternative internal structure

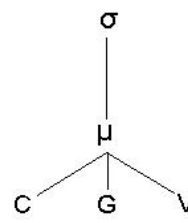


Figure 6  
Representation of CGV

Even if we can put consonant-vowel sequences into a single constituent, it is not obvious what kind of constraints we shall define in order to rule out the unattested forms. For example, the disfavor against \*[si] sequences is difficult to trace back to general prosodic or sonority constraints. The voiceless alveolar fricative [s] is not particularly close to or distant from the high front vowel [i]. Thus referring to minimal or maximal sonority distance would not work. Compared to [si], a stop-vowel combination (e.g. [pi]) has greater sonority distance, while a sonorant-vowel sequence (e.g. [ri]) has smaller. Turning to place of articulation does not help either. Although there is a tendency in languages to avoid shared place of articulations (e.g. English \*m<sup>w</sup>), [si] is not such a case. On the contrary, as Table (14) testifies, it is exactly the forms with shared palatal place of articulations that are welcome (e.g. [ci], [ni], [çi]).

#### 4.3 Phonetic/Sequential Approach

The phonetic licensing approach to the problem of Japanese CV restrictions builds on some basic characteristics of human perception. In perception the listener's task is to successfully identify the incoming acoustic signs. This task can be broken down into two parts: identification and discrimination. Identification involves the correct perception of acoustic cues in speech sounds. As it was explained at length in Section 2.2, the salience of an acoustic cue can greatly vary according to context. This context-sensitivity can be responsible for distributional peculiarities of a segment.

Another – so far not discussed – aspect of perception is the discrimination of speech sounds from each other. Speech sounds are not independent entities whose perception can be seen as individual processes. Speech sounds make up segment inventories in which each sound can be defined in relation to other sounds. Namely, the sound inventory can be recognized as a system of oppositions. Accordingly acoustic cues can be interpreted as markers for oppositions.

The choice of acoustic cues to mark an opposition (or a feature) can widely vary across languages (Sebastián-Gallés 2005). For example, the /s/-/sh/ opposition is

marked in many languages mainly by differences in the spectral envelope (Harris 1958, Jassem 1965). Since /sh/ (but not /s/) has an /i/-like formant transition part, the transitional cues can also be used to tell /sh/ apart from /s/ (Whalen 1981). Some research showed that Japanese listeners tend to rely more on transitional than spectral cues (Hirai et al. 2005), however the weighting of these cues is subject to individual variation.

At this point we can return to the problem of /si/ and /shi/ in Japanese. Since the /s/-/sh/ opposition relies heavily on transitional cues the perceptual contrast is expected to be the most salient in front of /a/, where the presence or the absence of an /i/ like formant portion is the most apparent. So a contextual /a/ is a perfect licenser for the /s/-/sh/ contrast. This can not be said about contextual /i/ as the differences in formant transitions are masked in this environment. Therefore, the same perceptual strategy is unreliable before /i/. Nevertheless, the trained ear still can tell the difference here by relying on the distinct spectral characteristics /s/ and /sh/. However, Japanese listeners do not seem to resort to this strategy. It is probably due to the fact that consonants are always followed by vowels<sup>2</sup>, so Japanese listeners can entirely rely on the transitional cues in sibilants.

The disadvantageous environment of /i/ in the discrimination of /s/ and /sh/ is a well-documented fact justified by experimental data (Lambacher et al. 2001, Pintér 2007). The role of formant transition can be also supported by historical facts, as up until Early Modern Japanese (16<sup>th</sup> century) the /s/-/sh/ opposition was absent not only before /i/ but before /e/ as well. This is not surprising since the second worst licenser for the sibilant contrast – due to its closeness to /i/ – is /e/.

The similar line of argumentation can explain the absence of /ni/≠/ɲi/ and /hi/≠/çi/ contrasts. In these cases – similarly to /si/ – the palatal consonant is cued by an /i/-like formant transition, which is an unreliable cue in front of /i/. The careful reader may point out here that while this perceptual approach can explain the absence of the oppositions before /i/, it does not give an account for the fact that it is always the non-palatal consonant that is missing (\*[si], \*[hi], \*[ni]). Indeed, this fact has rather an articulatory than a perceptual explanation. Namely, the place of articulation for palatal consonants and high front vowels are almost identical. Thus from an articulatory point of view they are more convenient than their pairs with non-palatal consonants.

#### 4.4 Conclusion

Restrictions on the co-occurrences of consonants and vowels are not uncommon in the languages of the world (Kawaski 1982, Flemming 1995). The above introduced examples explain the restricted occurrences of consonantal contrasts in Japanese CV sequences based on perceptual factors. Similar phonetic account can be given to cases where not consonantal but vocalic oppositions are neutralized. For example in Korean, labial consonants provide poor licensing to /u/≠/ɯ/ contrast and cause their neutralization into /u/. Another variation on this theme is when not CV but rather VC or CVC templates are concerned. In German the palatal [ç] and velar fricative [x] are used to be in allophonic relation [ç] occurring only after high-vowels ([iç], [ax] but \*[aç], \*[ix]). As a result of an ongoing phonologization process, at present [ç] can also occur after non-front vowels (e.g. [frauçən] ‘female dog master’)<sup>3</sup>. However, high-vowels in VC sequences still do not license the /ç/≠/x/ contrast. An example for co-occurrence

restrictions in CVC templates can be found in Cantonese. If both consonants in CVC are coronal, the vowel can not be a non-front high vowel ([tit], [tyt] but \*[tut]). That is /i/≠/u/ and /y/≠/u/ oppositions are not licensed within coronals.

All of these phonotactical patterns can be explained by relying on perceptual and articulatory factors in a similar way as was discussed in dealing with Japanese. The underlying generalization can be summarized by claiming that phonemic contrasts are subject to neutralization if the perceptual cues that are necessary for their discrimination are masked by the phonetic context. It seems impossible to arrive at the same level of generalization by prosodic means. First, structural motivation is difficult to find for all of these CV, VC and CVC cases. Second, even with such motivations, prosodic analyses eventually have to refer simply to the linear order of sounds and their perceptual and articulatory characteristics.

## 5. SUMMARY

The aim of this paper was to draw attention to the viability of phonetically grounded alternatives to prosodic phonotactics. The phonetic orientation to a large extent requires linear or sequential interpretations of observed distribution patterns with no reference to such syllabic constituents as onset, nucleus or rhyme. Still, this approach does not deny the existence of the syllable, it only points out that building analyses on such theoretical constructs as Sonority Sequencing Principle or Minimal Sonority Distance does not lead to insightful explanations.

Presumably, both sequential and prosodic constraints have their place in phonotactic descriptions. However their symbiosis is expected to be more intimate than their mere co-existence in a loosely defined theory. Ohala (1992) believes in the priority of sequential constraints in phonotactics and treats the syllable as a derived concept. That is the phonotactic system of a language is shaped by articulatory and perceptual factors and the concept of syllable is conceptualized as a generalization over the attested patterns. Ohala leaves us with this provoking idea, but without any further direction about how syllables can be derived. Nevertheless many years later a growing number of research on frequency effects and phonotactic probability seems to point toward a promising direction. For example, Elman (1990) shows how neural networks can learn phonotactic generalizations on the basis of segment sequences. Chen et al. (2004) demonstrates how the syllable can emerge from the statistical properties of phoneme distributions in Chinese. The overall message of these results is that researchers should not insist on the exclusive superiority of either the sequential or the prosodic views for all problems, but rather they should aim to find the means to which these two paradigms can co-exist in a coherent system.

## NOTES

\* I would like to thank Haruo Kubozono, Yoshihiro Nishimitsu, Takayuki Tohno and Andrea Tews for their help and assistance in the preparation of this paper.

1. A word boundary can be a decent licenser for a stop provided it follows and not precedes it. The silence – associated with the boundary – allows stop bursts to be released and to be recovered perceptually. Indeed, words ending in /nk#/ are not uncommon in English ('monk', 'sink', 'punk'... etc.).

2. Vowel devoicing can produce consonant clusters or word-final consonants where /s/ is not followed by vowel. However these alternatives are not discussed here.
3. A minimal pair with palatal and velar fricatives: [ku:çən] -[ku:xən] 'little cow' – 'cake'.

## REFERENCES

- Archangeli, D. & Pulleybank, D. 1993. *Grounded Phonology*. Cambridge, Mass. : MIT Press.
- Blevins, J. 1995. The syllable in phonological theory In Goldsmith, J. (ed.) *The Handbook of Phonological Theory*. Oxford: Blackwell. 206-244.
- Blevins, J. 2003. The Independent Nature of Phonotactic Constraints. In Féry, C. and R. van de Vijver (eds.) *The syllable in Optimality Theory*. Cambridge: University Press. 375-403.
- Blevins, J. 2004. *Evolutionary Phonology : the emergence of sound patterns*. Cambridge ; New York : Cambridge University Press.
- Browman, C. & Goldstein, L. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.
- Browman, C. & Goldstein, L. 1990. Tiers in articulatory phonology. with some implications for casual speech. In Kingston, J. and M. E. Beckman (eds.) *Papers in laboratory phonology I. Between the Grammar and Physics of Speech*. 341-376.
- Burton-Roberts, N. & Carr, P. & Docherty, G. 2000. *Phonological Knowledge - Conceptual and Empirical Issues*. New York: Oxford University Press.
- Chen, T-M. & Dell, G. & Chen, J-Y. 2004. A Cross-Linguistic Study of Phonological Units: Syllables Emerge from the Statistics of Mandarin Chinese, but not from the Statistics of English. *Paper presented at the meeting of the 26th Annual Meeting of the Cognitive Science Society, Chicago*.
- Coté, M. 2000. *Consonant cluster phonotactics: a perceptual approach*. PhD thesis. MIT.
- Chomsky, N. & Halle, M. 1968. *The sound pattern of English*. New York ; Evanston ; London : Harper & Row.
- Crystal, D. 1997. *A dictionary of linguistics and phonetics*. Oxford : Blackwell.
- Elman, J. L. 1990. Finding Structure in Time. *Cognitive Science* 14, 179-211.
- Ewen, C. J. & van der Hulst, H. 2001. *The phonological structure of words*. Cambridge : University Press.
- Flemming, E. 1995. *Auditory representation in phonology*. PhD thesis. University of California. Los Angeles.
- Fry, D. 1979. *The physics of speech*. Cambridge: Cambridge University Press.
- Hale, M. & Reiss, C. 2000a. Phonology as Cognition. In Burton-Roberts, N. et al. (eds.) *Phonological Knowledge - Conceptual and Empirical Issues*. 161-184.
- Hale, M. & Reiss, C. 2000b. Substance abuse and dysfunctionism : current trends in phonology. *Linguistic Inquiry* 31, 157-169.
- Harris, K. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1, 1-7.
- Hayes, B. 1989. Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry* 202, 253-306.
- Hayes, B., Kirchner, R. & Steriade, D. ed. 2004. *Phonetically Based Phonology*. Cambridge : Cambridge University Press.
- Hirai, S.; Yasu, K.; Arai, T. & Iitaka, K. 2004. Acoustic Cues in Fricative Perception for Japanese Native Speakers. *IEICE technical report. Speech* 104, 25-30.
- Hume, E. & Johnson, K.. ed. 2001. *The role of speech perception in phonology*. Tokyo : Academic Press.
- Ito, J. 1988. *Syllable Theory in Prosodic Phonology*. New York: Garland.
- Jassem, W. 1965. The formants of fricative consonants. *Language and Speech* 8, 1-16.
- Jespersen, O. 1932. *Lehrbuch der Phonetik*. Leipzig: B.G. Teubner.

- Kawasaki, H. 1982. *An acoustical basis for universal constraints on sound sequences*. PhD thesis, University of California, Berkeley.
- Katada, F. 1990. On the representation of moras : evidence from a language game. *Linguistic Inquiry* 21:4, 641-645.
- Kubozono, H. 1985. *A theory of phonological weight*. Dordrecht : Floris.
- Kubozono, H. 1989. The mora and syllable structure in Japanese. *Language and Speech* 32:3, 249-278.
- Kubozono, H. 1996. Speech segmentation and phonological structure. In Otake, T. and A. Cutler (eds.) *Phonological Structure and Language Processing*. 77-94.
- Ladefoged, P. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Lambacher, S.; Martens, W.; Nelson, B. & Berman, J. 2001. Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *Acoustical Science and Technology* 22:5, 334--343.
- Larson, G. N. 1992. Local computational networks and the distribution of segments in the Spanish syllable. *CLS* 26, 257-272.
- Lindblom, B. 1983. Economy of Speech Gestures. In MacNeilage, P. F. (ed.) *The Production of Speech*. New York: Springer-Verlag. 217-245.
- McMahon, A. M. S. 2000. *Change. chance. and optimality*. New York : Oxford University Press.
- Malsch, D. L. & Fulcher, R. 1989. Categorizing phonological segments: the inadequacy of the sonority hierarchy. In Corrigan, R. et al. (eds.) *Current Issues in Linguistic Theory*. Amsterdam: John Benjamins. 69-80.
- Nathan, G. S. 1989. Preliminaries to a theory of phonological substance: the substance of sonority. In Corrigan, R. et al. (eds.) *Current Issues in Linguistic Theory*. Amsterdam: John Benjamins. 55-67.
- Ohala, J. J. 1992. Alternatives to the Sonority Hierarchy for Explaining Segmental Sequential Constraints. *CLS* 26:2, 319-338.
- Pintér, G. 2007. The interaction of contextual vowels and formant transitions in Japanese fricative perception, talk at *315th Meeting of the Phonetic Society of Japan*.
- Pisoni, D. B. & Remez, R. E.. ed. 2005. *The handbook of speech perception*. Malden. Mass. : Blackwell.
- Raphael, L. J. 2005. Acoustic cues to the perception of segmental phonemes In Pisoni, D. B. and R. E. Remez (eds.) *The handbook of speech perception*. 182-206.
- Sebastián-Gallés, N. 2005. Cross-Language Speech Perception. In Pisoni, D. B. and R. E. Remez (eds.) *The handbook of speech perception*. 546-566.
- Selkirk, E. 1984. The major class features and syllable theory. In Aronoff, M. et al. (eds.) *Language Sound Structure*. Cambridge, Mass.: MIT Press.
- Steriade, D. 1999. Alternatives To Syllable-Based Accounts Of Consonantal Phonotactics. *Proceedings of the 1998 Linguistics and Phonetics Conference*. The Karolinum Press. 205-242.
- Steriade, D. 2001. Directional asymmetries in place assimilation. In Hume, E., and K. Johnson (ed.) *The role of speech perception in phonology*. San Diego : Academic Press.
- Venneman, T. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18, 1-18.
- Whalen, D. H. 1981. Effects of vocalic formant transition and vowel quality on the English [s]-[sh] boundary. *The Journal of the Acoustic Society of America* 69, 275-282.
- Wright, R. 2001. Perceptual cues in contrast maintenance. In Hume, E. and K. Johnson (eds.) *The role of speech perception in phonology*. 251-277.