ABSTRACT: Intuitive syllabifications of thirty seven 6-7 year-old and seventy one 9-10 year-old children were studied in an experimental setting. The children syllabified Bengali disyllabic nonwords with intervocalic consonant clusters that were presented orally to them. The aim was to find out to which (if any) of three linguistic theories of syllable structure these syllabifications correspond: the maximal onset principle, the legality principle or the sonority sequencing principle. Segmentation behaviours of the younger children showed a greater agreement with predictions based upon the maximal onset principle rather than sonority, subject to the constraint AVOID COMPLEX ONSET. Older children, however, also made categorical distinctions between phonotactically legal vs. illegal onset clusters. In addition, the analyses point to the nature of interactions between phonotactic knowledge and orthographic representations possibly impacting intervocalic segmentation.

KEYWORDS: intuitive syllabification, intervocalic consonant clusters, the maximal onset principle, phonotactic regularities, Bengali

0. INTRODUCTION

Syllables play an important role in the production and perception of speech (Elbers 1982; Oller 1986; Bertoncini & Mehler 1979). Research on phonological awareness has shown that children are able to segment speech into syllables at a very young age. One interesting aspect of these syllabifications is the use of knowledge of phonology (the sound system of a language) for establishing the position of syllable boundaries in polysyllabic sequences. Children’s choice of boundary positions can be used to achieve insight into the nature of their phonological
knowledge, more particularly the principles of syllabification that they bring to the task. It may also indicate to what extent syllabification is affected by knowledge of spelling. Such information about the location of syllable boundaries is needed to determine the nature of syllable processing in spoken and written language. Such analyses can also lead to pointers to how syllables as a phonological unit can be used in instruction, for example when teaching word decoding in general, and for spelling instruction specifically.

In order to discover the principles that govern children’s intuitive syllabification, one can present simple segmentation tasks to children. Examining changes in choice of syllable boundaries in children of different age groups can additionally throw light on developmental changes and the potential role of literacy on syllabification behaviour.

1. THEORETICAL FRAMEWORK

There are a number of principles related to syllabification. Three major ones, of interest to this writing, are discussed below:

1.1 The Maximal Onset Principle
The Maximum Onset Principle (MOP) states that the syllable boundary must be placed before the maximum, allowable legal onset. It requires that we should extend a syllable’s onset at the expense of the preceding syllable’s coda whenever it is legal to do so (Kahn 1976). In the case of word medial clusters that cannot appear at the beginning of a word/syllable, syllabification would be located between the consonants of the intervocalic cluster. MOP is therefore a basic phonological rule that functions under language-specific phonotactic constraints (Hooper 1972, Pulgram 1970; Treiman & Zukowski 1990). For the English word casket, MOP prefers [ca.sket] because it maximises the onset, and [sk] is also a legal onset in English.

1.2 The legality principle
The legality principle constrains the segments that can begin and end
syllables to those that appear at the beginning and end of words. In other words, a syllable tolerates in onsets only segments that are possible word initially and in coda only segments and sequences that are possible word finally (Steriade 1999; Goslin & Frauenfelder 2001). In Optimality Theory, this is one of the POSITIONAL FAITHFULNESS CONSTRAINTS. Thus, a word like admit must be syllabified as [ad.mit] because [dm] never appears word initially or word finally in English. However, when a segment can be a legal onset as well as coda, the principle does not imply just one syllabification. For instance, the word casket would be syllabified as [cas.ket], [ca.sket] or [cask.et] as all three employ legal onsets and codas.

1.3 The Sonority Sequencing Principle
The Sonority Sequencing Principle (SSP) provides a stricter definition of legality. The principle states that sonority should increase from the first phoneme of the onset to the syllable’s nucleus, and then fall off to the coda (Selkirk 1984). Languages prefer to build syllables with the most vowel-like sounds nearer the middle, and the least vowel like sounds plosives, voiceless fricatives near the edge(s). Syllable structured in this way are said to conform to the sonority profile (Clements 1990). Therefore, string CVC1C2 will typically be syllabified as CV.C1C2V where sonority is C1<C2 and as CVC1.C2V where sonority is C1>C2. CVC1C2.V may be a possible syllabification for the C1>C2 sequences, however in Bengali this parse is not preferred. For example, shasti ‘punishment’ is never syllabified as [shast.i].

The universal sonority sequencing which Bengali also follows is

\[\text{plosives < fricatives < nasal < laterals < flaps < glides < vowels}\]

(Kar 2009)

In this sequence, plosives are the least sonorous sounds and vowels are the most sonorous. When syllabifying a Bengali word like shasti ‘punishment’, the theory would imply a CVC.CV syllabification [shas.ti], since /t/ is lower on sonority scale than /s/ even though [st] is a legal onset in Bengali. The illustrative English example casket would therefore be syllabified as [cas.ket] or [cask.et] according to SSP.
1.4 Orthographic representation
Evidence from different writing systems suggests that the nature of the orthographic representation influences intuitive syllabification. In Dutch, literate subjects were found to make many ambisyllabic syllabifications (i.e. putting a single intervocalic consonant as both coda of the first syllable and onset of the succeeding syllable) when the vowel preceding a single intervocalic consonant was short (e.g. /ap\text{pel}/). This is because the Dutch spelling has a doubling of the consonant grapheme in this phonological context, [appel] and suggests to the subjects that the phoneme is shared by the two syllables. The nonliterate subjects, on the other hand, tended to syllabify [appel] as [a.pel] (Gillis & De Schutter 1996). English literate speakers, similarly, are more likely to place the critical intervocalic segment /m/ in both syllables when the word is written with a double letter (e.g. [com.ma] for the English word \textit{comma}) than when it is written with a single letter (e.g. [le.mon] in \textit{lemon}) (Treiman & Danis 1988). Among French literate speakers there is a significant increase in ambisyllabic responses when intervocalic clusters are represented by an opaque orthographic form (e.g., /ks\text{ in taxi} with [x]) than when there is a direct, one on one relationship between the phonology and orthography of the cluster /tr/ in the word \textit{citron} `lemon’ (Goslin & Floccia 2007). Further crosslinguistic evidence comes from Arabic where Derwing (1992) found that literates’ syllabification tended to reflect orthography while this was not the case for nonliterate’s syllabifications.

In the Brahmi derived scripts such as Hindi, syllables do not correspond neatly to akshara, and a CVCCV word can typically have three akshara (CV-C\textsuperscript{0}-CV) and two syllables (CVC.CV) (as in \textit{balti} ‘bucket’ which has three akshara [ba], [la] and [ti] but syllabified as [bal.ti] ) or with two akshara (as in \textit{baksa} ‘box’ with two akshara [ba] [ksa]) and two syllables (either CVC.CV [bak.s\text{a}] or CV.CCV [ba.ksa]). Hindi speakers tend to prefer the syllable boundary between the intervocalic complex akshara in the former more than in the latter suggesting that the nature of orthographic representation affects syllabification (M. Ohala 1999). In Bengali, also a Brahmi based script, the extent to which language specific orthography-phonology mappings play a role in intuitive syllabification is unclear.
2. BENGALI CONSONANT CLUSTERS: PHONOTACTIC CONSTRAINTS AND AKSHARA-SYLLABLE MAPPINGS

Bengali has 16 canonical syllable patterns (Sarkar 1986), which in descending order of frequency are CV, CVC, V, VC, VV, CVV, CCV, CCVC, CVVC, CCVV, CVCV, CCCV, CCCVC, VVC and CCCVV. Bengali shows partial sensitivity to syllable weight. Vowel length is not distinctive in Bengali, and closed syllables (CVC) count as heavy (bimoraic), with the coda consonant contributing a mora to the syllable. How the moraic structure affects word level stress is beyond the scope of this paper. The CV syllables constitute 54% of the whole language (Dan 1992), while CCCV, CCCVC and CCCVV are infrequent. The maximum syllabic structure in native Bengali is CVC, though words borrowed from Sanskrit have the syllable structure of [CCC]V[C] which shows that Bengali allows consonant clusters having up to three consonants word initially\(^1\) and only one consonant word finally. Bengali allows a whole set of consonant clusters in word medial positions particularly in mono-morphemic words (e.g. sopno, sotyo, surjo). Foreign borrowings from Urdu, Persian and English have however pushed up the syllable structure to [CCC]V[CCC].

Cluster combinations that are possible word medially in Bengali are plosive-plosive (/pt/, /p\textsuperscript{b}/), plosive-fricative (/ps/, /ks/), plosive-nasal (/pn/, /tn/, /k\textsuperscript{h}n/, /g\textsuperscript{n}/, /g\textsuperscript{h}n/), plosive-liquid (/pl/, /p\textsuperscript{b}l/, /kl/, /bl/, /gl/), nasal-plosive (/mp/, /mp\textsuperscript{h}/, /mb/, /mb\textsuperscript{h}/), liquid-plosive (/lp/, /lk/, /lb/, /lg/), fricative-nasal (/sn/, /sm/) and nasal-fricative (/ns/). Apart from this, Bengali has many geminate consonants (/kk/, /kk\textsuperscript{h}/, /pp/, /pp\textsuperscript{h}/, /bb/, /bb\textsuperscript{h}/, /ll/, /rr/, /tt/, /dd/, /nn/, /tt/, /dd/ /dd/) that can occur word medially (and never word initially) and contrast with single consonants. Apart from these pure geminates there are geminate aspirated sounds like /d\textsuperscript{h}d/ / tt\textsuperscript{h}/ and / bbbh/ as in /a\textsuperscript{h}d\textsuperscript{h}o / ‘a surname,’ / pott\textsuperscript{h}o / ‘diet,’ / obb\textsuperscript{h}e\textsuperscript{f} / ‘habit’ respectively.

The akshara is the basic orthographic unit of the Bengali orthography. The simplest akshara carry the inherent vowel /ɔ/ corresponding to [Ca]\(^2\) Bengali.

\(^1\) Although native Bengali words do not allow consonant clusters at the onset position, Sanskrit based words and foreign borrowings together contribute to 48-53% words with onset clusters (Kar 2009:52).

\(^2\) We represent the orthographic syllable as [Ca], the phonological representation for which would be /C\textsuperscript{a}/.
syllables, an orthographic feature that has been argued to be an elegant solution for marking minimal units of articulation (Pandey 2007). Other akshara units are CV, CCV and even CCCVs. However, certain systematic differences seem to exist between akshara and the syllable; though post-vocalic consonants are not accommodated within the akshara structure, they may be integral to the syllable structure. This has important implications if the akshara were to be used as an anchor for syllabification of words.

Table 1 below gives examples of the variety available. In the case of open syllables, akshara maps neatly onto syllables (poka, bichana), but in sukno, while the syllable includes the post-vocalic consonant /k/, it is orthographically represented with a single akshara. The word baksho has an intervocalic complex akshara, the first segment of which forms the coda of the first syllable, and the second the onset of the second syllable.

<table>
<thead>
<tr>
<th>Words</th>
<th>Syllables</th>
<th>Akshara</th>
</tr>
</thead>
<tbody>
<tr>
<td>pokā ‘insect’</td>
<td>po.ka po.ka cv.cv</td>
<td>po.ka po.ka cv.cv</td>
</tr>
<tr>
<td>bichana ‘bed’</td>
<td>bi.cha.na bi.cha.na cv.cv cvbi.cha.na bi.cha.na cv.cv</td>
<td>cv.cv cvbi.cha.na bi.cha.na cv.cv</td>
</tr>
<tr>
<td>sukno ‘dry’</td>
<td>su.k.no su.k.no cv.c.cv</td>
<td>su.k.no su.k.no cv.c.cv</td>
</tr>
<tr>
<td>baksho ‘box’</td>
<td>bak.sho ba.ksho cv.c.cv</td>
<td>ba.ksho ba.ksho cv.c.cv</td>
</tr>
</tbody>
</table>

Table 1: Syllable and akshara segmentation in Bengali words

It is the fourth type of akshara-syllable mismatch (*baksho* type) we are interested in, in this paper.

3. ITEM ANALYSIS

In this paper, we report children’s syllabification behaviour on a set of CVCCCV Bengali nonwords. The role of the sonority profile within consonant clusters and phonotactic regularities on syllable-based segmentation were investigated. We examined whether the sonority profile of the intervocalic consonant cluster would influence syllable
segmentation, and what was the role of the legality principle and the maximum onset principle (MOP). We also examined whether syllable boundary assignment would align with orthographic representation where, despite there being no default spelling rule for nonwords, segmentation might mirror spelling inferences made by the child. To examine how segmentation behaviour may change over the early years of literacy development, we compared the segmentation behaviour of 6-7 year old children with 9-10 year olds.

3.1 Test items
A list of ten nonwords having CVCCV structure was presented to children to syllabify. Since intuitive syllabification can be influenced by spelling rules (Treiman & Danis 1988; Gillis & De Schutter 1996), we tried to avoid or minimise this bias by using nonwords, and ensuring that children only heard the word. Since they were nonwords, children could not be influenced by morphemic boundaries in their syllabification. Since they were designed from words taken from language textbooks that are used in Grade 2 and Grade 3 (and those known to children according to our intuitions of child lexicon)³, some orthographic representation could have been construed by children from their analogies with real words. There is evidence of use of such phonological analogies from lexical items in children’s reading of nonwords (Sircar & Nag, in press). No diphthongs were used in the word list, and vowel length (which is not contrastive in Bengali) was not controlled for. One nonword leddhi (similar to buddhi ‘intelligence’ from early vocabularies) was dropped from the final analysis, since the segmentation for this nonword was primarily ambisyllabic (CVC.CCV). It is not clear whether in the oral presentation of this nonword by the examiner, there was an unintended gemination, which could have resulted in an ambisyllabic syllabification.

The test items were organized as a function of consonant type, combined from nasal (N), fricative (F), liquid (L), and plosive (P) consonants. A subset of seven was chosen for this item analysis (PL, FP, PN, PF, LF, LF).

³ Examples of words from language textbooks used in Grade 2 and Grade 3 that match the selected nonwords are: chakra ‘cross,’ tibro ‘fast,’ kabja ‘to capture,’ shasti, ‘punishment,’ pushpa ‘flower,’ bhorti ‘full,’ dorja ‘door,’ and bondo ‘closed’.
LP, and NP). The nonwords and their characteristics are provided in Table 2. Two nonwords were used as training items, so that children become familiar with the task demands and any difficulty in segmentation is not due to their poor understanding of the instructions.

<table>
<thead>
<tr>
<th>Nonword</th>
<th>Cluster category</th>
<th>Legality of the cluster</th>
<th>SSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>[mɑːкра], [libɾɑ]</td>
<td>PL</td>
<td>legal</td>
<td>respects</td>
</tr>
<tr>
<td>[baːстит]</td>
<td>FP</td>
<td>legal</td>
<td>violates</td>
</tr>
<tr>
<td>[kuʃpa]</td>
<td>FP</td>
<td>illegal</td>
<td>violates</td>
</tr>
<tr>
<td>[kɪтνə]</td>
<td>PN</td>
<td>illegal</td>
<td>respects</td>
</tr>
<tr>
<td>[нυбдя]</td>
<td>PF</td>
<td>illegal</td>
<td>respects</td>
</tr>
<tr>
<td>[mɑːрдь]</td>
<td>LF</td>
<td>illegal</td>
<td>violates</td>
</tr>
<tr>
<td>[ʃɔорті]</td>
<td>LP</td>
<td>illegal</td>
<td>violates</td>
</tr>
<tr>
<td>[пɔndo]</td>
<td>NP</td>
<td>illegal</td>
<td>violates</td>
</tr>
</tbody>
</table>

Table 2: The nonwords used in the task and their characteristics

3.2 Procedure
Each child was tested individually in a 15-minute single session in a quiet room. The nonwords were presented to the children orally. Children were given a pause insertion task, consisting of slow repetition of target nonwords with pauses inserted between each syllable of the nonwords. They were asked to add rhythm to their speech by clapping their hands if necessary. Each child was initially trained to clap once for each syllable in their name (e.g. shi-ba-ni)\(^4\), followed by two words as practice, one CVCV and one CVCCV. In the latter, we accepted CVC.CV, CV.CCV and CVC.CCV as possible answers. The word *syllable* was not used in the instructions. Children’s responses were recorded on a scoring sheet by the examiner. For each response, the position of the syllable boundary

\(^4\) Post hoc we realized that such practice with clapping for one’s own name might have demonstrated 1:1 akshara-syllable mapping (CV.CV.CV) for some children, while those with names like ‘Snigdha’ and ‘Pintu’, would have received additional demonstration of the complexities of syllable-boundary assignment when akshara do not correspond to syllables.
/CV.CCV/, /CVC.CV/ or /CVC.CCV/ and/or the category of atypical responses were noted down. The omission or the addition of phonemes not found in the stimuli was treated as atypical. No CVCC.V responses were attested since Bengali does not allow complex codas.

3.3 Subjects
Thirty 6-7 year olds (from Grade 2, Mean = 7;1, SD = 0.42) and seventy-one 9-10 year olds (from Grade 4, Mean = 9;07 SD = 0.56) participated in this survey. The younger group represented learners who could speak and understand the language but were at an earlier stage of literacy development. The older group represented children who had two to three years of exposure to reading, writing, and spelling. The children were tested in their school in Kolkata, West Bengal. All children were native speakers of Bengali. No hearing or speech problems were reported by their teachers.

4. RESULTS AND DISCUSSION

Children’s responses were taken together to examine the type of segmentation chosen for each item. Apart from the non responses, 30 out of a total of 1080 responses had a phoneme either omitted or substituted, and these responses were not included in the analysis reported here. The rest of the segmentations of the CVCCV nonwords were CV.CCV, CVC.CV and CVC.CCV. We were here interested to find out whether children could identify the number of syllables in the nonword. The issue of syllable boundary detection was done post hoc to find out what background principles were being used in intuitive syllabification. Certain sonority sequences would make children prefer a CV.CCV segmentation, while other sonority sequences would make children assign a syllable boundary as CVC.CV. From the perspective of orthographic mediation in intuitive syllabification, spelling options could be with 2 full akshara (as in sukho) or with a complex akshara (as in baksho). If the chosen spelling for the nonword is CV.CCV then it could be argued that an akshara-based syllabification has occurred, while a parse of CVC.CV would indicate an intra-akshara split. However since
there is no clear indication of what the spelling choices of the children might have been, we will keep this aspect aside for now and return to it later in the paper.

The segmentation responses for the younger and older groups of children are presented in Figures 1A and 1B. For each group the frequency of the three segmentation responses (CVC.CV, CV.CCV and CVC.CCV) were calculated separately, for each cluster category. The majority of responses by children in both age groups and across all the cluster categories was the CVC.CV segmentation. Recall that one of our aims was to test whether sonority sequencing in intervocalic clusters explains intuitive syllabification. If sonority did affect syllabification, then the PL, PN and PF clusters with the C1<C2 sonority in our item set should show a CV.CCV segmentation and the others a CVC.CV segmentation. Figure 1A shows that this was not the case; children segmented intervocalic consonant clusters to derive a single consonant onset for the second syllable (preferring CVC.CV). The sonority sequencing principle does not seem to account for intuitive syllabification of particularly the younger children. The syllabification was also irrespective of the legality of the onsets; for both legal onsets [kr] and illegal onsets [tn] the syllabification is CVC.CV. So, none of the three principles we had started out with – SSP, legality, or MOP, seem to account for syllabification of younger children.

An explanation for the pattern of syllabification in our data set may however be found in the construct of Complex Onset Constraint that is ranked higher than No Coda Constraint This ranking, however, is not inconsistent with the maximal onset principle. In ranking, a CC coda is more marked than single consonant coda, and a single consonant coda is more marked than a no consonant coda. Bengali respects the No Complex Coda Constraint, but allows single consonant coda. Another constraint in Optimality Theory pertains to the presence of onsets. All languages have simple onsets but many languages do not allow complex coda.

Therefore, Complex Onsets is a marked constraint. In Bengali, both simple and complex onsets are allowed, though it is only PL and FP clusters which can occur as onsets. Younger children in the study seem
to apply an **Avoid Complex Onset Constraint**\(^5\) (which is permissible in adult language) and this constraint is ranked higher than legality, MOP and SSP in the syllabification of CVCCV words. This is true of the older children as well, though we see other principles – legality, phonotactic constraints and orthographic representation – coming into play.

As Figure 1B shows, there is a definite trend for the older children to map their intuitive segmentations to a developing knowledge of phonotactic constraints in Bengali. We chose to compare two of the cluster strings, PL and FP, where the segmentation types are spread across all three possibilities (CV.CVC, CV.CCV and CVC.CCV). Group differences were statistically significant (PL, \(\chi^2(2) = 6.10, p<.05\); FP, \(\chi^2(2) = 6.19, p<.05\)). If examined from the perspective of the hypothetical spelling of CV.CCV, intuitive segmentations of many more older children would fall between the akshara (PL, 6% younger, 20% older; FP, 3% younger, 17% older), though the majority preference across both age groups was for what may be considered an intra-akshara split (CVC.CV).

---

\(^5\) We thank one of the reviewers of the paper for pointing out that the younger group appear to rank the **Avoid Complex Onset Constraint** over principles of onset maximization, sonority and legality.

\(^6\) Note that in Figure 1A and 1B responses for [kr] are included in PL and responses on [ʃp] are included in FP.
A small difference is seen between those consonant clusters which form legal onsets and those which do not. Table 3 gives the proportion of the CVC.CV segmentation, calculated as a proportion score. Three clusters – PL (/kr/, /br/); and FP (/st/) – induced small but noticeable preference for a CV.CCV segmentation among the older children. Note that Bengali phonotactic constraints allow these three clusters to occur as possible onsets. The principles of MOP, SSP and legality, and a spelling based influence, would all lead to the same segmentation pattern of CV.CCV for these clusters. At this point, with a small dataset, it would be difficult to categorically say which of these are specifically being applied by the older children. Younger children, we found, showed a stronger preference for CVC.CV segmentation for even those clusters which are legal onsets (e.g. makra, libra, baasti). We propose that the difference in the two age bands is because of growing knowledge of phonotactic constraints in Bengali, perhaps because of an expanding vocabulary. The further possibility of lexicalized phonological analogies being invoked for the spelling of nonwords cannot be ruled out as an additional mediator.

Among those clusters that induced some degree of CV.CCV parse, we had two FP clusters – /st/ and /fp/ – both of which disobey SSP. However, while /st/ is a legal onset, /fp/ is illegal. We see that these phonotactic constraints are evident in the two FP clusters in our item set among the older children: for baasti CVC.CV the average preference ratio is 0.61, but for kushpa it is 0.86, indicating that children appear to be influenced...
by phonotactic constraints rather than the sonority principle. However, the higher proportion of CVC.CV parse (0.96) for a legal cluster /st/ as well as /ʃp/ an illegal cluster (0.90) in younger children cannot be explained.

<table>
<thead>
<tr>
<th>Cluster category</th>
<th>Nonwords</th>
<th>Younger children</th>
<th>Older children</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PL</td>
<td>[mɑːkra]</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>PL</td>
<td>[lɪbra]</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>FP</td>
<td>[kuʃpa]</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>FP</td>
<td>[bɑːsti]</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>PN</td>
<td>[kɪtna]</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>PF</td>
<td>[nubfɔ]</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>LF</td>
<td>[mɑːtʃɔ]</td>
<td>0.96</td>
</tr>
<tr>
<td>8</td>
<td>LP</td>
<td>[ʃɔrti]</td>
<td>0.96</td>
</tr>
<tr>
<td>9</td>
<td>NP</td>
<td>[pəndo]</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Table 3: Proportion of CVC.CV responses

At this point, we need to account for a language specific phonological characteristic and its impact on intuitive syllabification. In Bengali, a consonant + liquid cluster [Cr/l] has a geminated effect on the obstruent (Dan 1992), so in pronunciation it is CCr/l. For example putro is pronounced /puttro/ and amlan /ɔmmlan/. In syllabification, therefore, the first member of the CCr/l sequence becomes the coda of the preceding syllable, and the second and the third members, i.e. Cr/l, together form a cluster which becomes the onset of the following syllable, thus giving rise to an ambisyllabic response. We found a noticeable number of ambisyllabic responses in the PL category in both younger (0.14) and older children (0.22) (see Figures 1A and 1B).

We also examined whether the level of knowledge of akshara had any role to play in intuitive syllabification. A trend seen in Figure 1A is that older children showed more phonotactic based segmentations in legal onsets (/kr/, /br/, /st/), the first two of which respect SSP, and the last /st/
does not. The segmentation is also consistent with a spelling based syllabification. Rather than making the analysis by age, we chose to reclassify children in our survey on the basis of their performance on a recognition task for CV, CVV and CCV akshara. Children were divided into two groups on the basis of accuracy score of below and above 80%: low akshara band (56 children) and high akshara band (52 children). The segmentation behaviour was then analyzed to see if higher knowledge of akshara contributed to a different segmentation pattern. The proportion of intra-akshara split was greater than the proportion of intuitive segmentations being fidel to akshara representations\(^7\). The PL and FP clusters had the largest so called akshara–wise CV.CCV responses (for lower akshara group 14% and 11% and for higher akshara group 21% and 7% respectively). Both groups, irrespective of level of knowledge of the akshara, showed similar segmentation patterns (PL, \(\chi^2(2) = 1.91, ns\); FP, \(\chi^2(2) = 1.09, ns\)). Therefore, within the band of akshara knowledge seen in our sample, differences in knowledge at the level of orthographic units do not explain trends in the data. For the hypothetical nonword spelling of CV.CCV both groups of children gave segmentation responses that would amount to syllabifying the complex CCV akshara, retrieving the phonemic values for each of the two consonants, and assigning these values to different syllables.

A last example is *pedo*, where the intervocalic cluster is both illegal and disobey sonority. Children show a considerable CV.CCV parse on this nonword when compared to other nonwords, in both age groups. Syllabification of NC (nasal+stop) clusters are known to behave differently and are affected by the universal principles of MOP, SSP and legality, and as well as by micro-phonetic details such as the duration of the nasal, its place of articulation, and the duration of the following stop which in turn depends on whether it is voiced or voiceless (Ohala & Ohala, 2011).

\(^7\) Proportion of intra-akshara split (CVC.CV) in segmentation was as follows: low akshara group, for PL (0.66), FP (0.81), PF (0.92), PN (1), LF (0.94), LP (0.96) and NP (1) and for high akshara group, for PL (0.58), FP (0.85), PF (0.90), PN (0.93), LF (0.92), LP (0.95), and NP (0.89).
5. CONCLUSION

The main aim of this item analysis was to investigate the phonological basis of children’s intuitive syllabification in Bengali. The clearest result is that intervocalic consonant clusters were syllabified with a single consonant onset. So the highest ranked constraint or the most preferred constraint is AVOID COMPLEX ONSET, that overrides legality and sonority of the consonant clusters, and akshara based representations.

One of our aims in this analysis was to throw light on the mental representation of syllables. Linguistic theories of syllable structure propose invariant syllabification behaviour on the basis of specific rules and principles. The particular syllabification principles explored here were the maximal onset principle and the sonority sequence principle. Our results showed that children’s intuitive syllabification was consistent on nonwords containing intervocalic consonant clusters that formed illegal onset clusters to the second syllable. This suggests that these particular tokens were being syllabified according to categorical phonological knowledge. Other judgments about word internal boundaries could also be explained through the same categorical knowledge: legal onset clusters may be syllabified differently. A pertinent point to make here is that sonority values of the intervocalic consonants and the representation of intervocalic clusters in orthography appeared to be factors with lesser explanatory power for the data. Segmentation behaviour of children therefore appears to go with the predictions of a Maximal Onset Principle which prefers a CV.CCV parse, but subject to the constraint AVOID COMPLEX ONSET which is ranked higher than MOP, SSP and legality. Therefore, children could violate the latter three, but the former is inviolable, particularly in younger children’s syllabification. With older children, the patterns show a growing tendency to violate AVOID COMPLEX ONSET CONSTRAINT in case of a legal onset that obeys SSP.

At the level of orthographic representations influencing intuitive syllabification, we concede that this dataset with nonlexical items is limited and the specific influences of orthography could not be tapped. Our analysis, however, suggests that a fruitful line of investigation would
be to compare intuitive syllabification in words exploiting an interesting orthographic duality available in Bengali (and Hindi, among other akshara languages). CVCCV words have two possible spellings, some words are spelled with a complex akshara, others use two full consonants.

A separate methodological issue has to do with order effects. Some of the trends in the data that we have reported may well have been due to the fixed order in which the items were presented, and we are unable to rule out whether earlier items in some way primed the responses in the later items. Counter-balancing the design, where different cohorts of children receive the same items but in different order, would offer more robust evidence about factors that influence intuitive syllabification.

Other caveats are also necessary. Children’s dialects can influence intuitive syllabification of nonwords. Syllable weight of the item set can influence trends in the data. And importantly, the task itself can influence syllabification judgments. Syllable boundary perception of subjects can be assessed through whole word repetition with pause insertion, a last syllable repetition task, a syllable reversal task, and a written slash-insertion task, and might lead to between and within-speaker variation in syllabification judgments even when identical stimuli are used. The syllable can change form (within limits) to facilitate the task at hand, and to fulfil the specific cognitive demands of the task. Different experimental paradigms do not give the same weight to the many factors involved in syllabification, and data collected using non-identical procedures may not be directly comparable. But despite the role of task demands and the methodological limitations of our study, certain general features of intuitive syllabification in Bengali are evident.
REFERENCES


Shruti Sircar
The EFL University
Hyderabad, India
shrutisircar@gmail.com

Sonali Nag
The Promise Foundation
Bangalore
sonalinag@t-p-f.org