

Two theories on Polish syllable structure

Armin Schmidt <armin@ikiro.de>

Torsten Marek <shlomme@gmx.de>

1. Introduction

In our work, we try to compare two different theories about Polish syllable structure. The obvious problem every researcher has to cope with if working on this topic are the seemingly long consonant clusters. At the beginning of our work, we read both „Syllable structure assignment“ from Rubach & Booij and „Polish Syllables – The role of prosody in phonology and morphology“ by Christina Bethin with the plan to compare both theories and find data where we can actually prove one of them to be wrong. After some consideration, we finally concluded that both theories have certain weaknesses and ugly workarounds for certain problems (mostly concerning the sonority hierarchy), but do explain the syllable structure well enough. Moreover, finding weaknesses has been made more difficult because we could not find out what the actual theory by Bethin was. Thus, the option of comparing these their two theories was not feasible.

After that, we thought about comparing two different phonological rules where one is based on syllable structure and the other is not, as proposed in the working assignment. While this is actually possible for Polish (Rubach & Booij have given some examples for this), the phonological processes have, due to their cyclic nature, proven to be too advanced for our understanding. Moreover, we could not have copied the examples from the literature, but propose our own ones, which would have made them attempt even more complex.

Finally, we decided to compare the more concise theory of Rubach & Booij with a little theory of our own and show where our theory, which hopefully is not too naive, fails in explaining Polish syllable structure.

Therefore, we will briefly display the „golden“ theory first, then lay out our own one and in the last step, compare them both.

2. Difficulties for syllable structure assignment in Polish

Polish has a large number of onsets with more than one segment. Bethin92 speaks of ~250 two-segment onset clusters, and ~170 three segment ones, where the actual number depends on the notion of segments to phonemes mapping.

Several studies have shown that long onsets are much more common in Polish than long codas, a fact that is vital to both theories presented in this paper. In terms of frequency, Bethin92 itself cites a survey by Sawicka, where the distribution for long initial clusters is 86,7% and 13,3% for long final ones. Although there are some restriction on these clusters, have experience a broad variety of violations of the Sonority Sequencing Generalisation (SSG):

SSG: the sonority of segments must decrease towards the edges of a syllable

The sonority hierarchy used by the SSG is defined as:

Vowels > liquids > nasals > fricatives > stops

In the history of work on Polish syllable structure, different approaches have been made, one of them is to completely ignore sonority for the organization of Polish consonants. Bethin ('92) suggests to split up long onsets into smaller units that locally obey the sonority hierarchy, a theory that inspired us when designing our own approach.

3. Syllable structure assignment algorithm for Polish (Rubach & Booij)

In this article, the authors try to give a definition of a „syllabification algorithm“ for Polish. This algorithm applies cyclically due to the nature of Polish phonology/morphology where processes know about one another and make use of the structure of their input,.

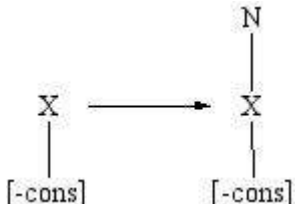
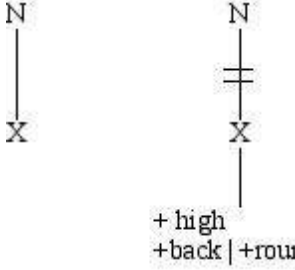
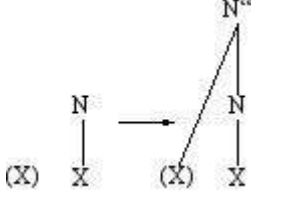
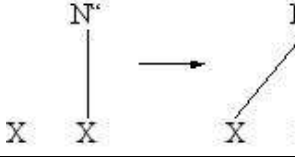
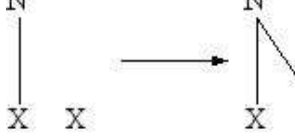
The algorithm

In order not to drop the SSG for Polish completely, the authors state the „the consonant at an edge of a word ... does not count from the point of view of the SSG“, which has been proposed several times before and takes care of a large number of otherwise invalid constructions. The reanalysis of all questionable words with the work hypothesis also leads to the result most of the SSG violations are now wrong orders among obstruents. This suggests that it is desirable to merge fricatives and stops into one sonority class of obstruents.

Before working on the algorithm, the authors give a couple of more general rules:

- Obstruent Sequencing Constraint (OSC)
 Geminate sequences are always split between two syllables (VC₁C₁V is always syllabified as VC₁-C₁V), otherwise there is no need of sonority distance between two distinct consonants from the same sonority class
- Coda constraint (CC)
 The coda may not contain two sonorants
- High vowel constraint (HWC)
 If [i] occurs before [u], it does not receive an N slot and therefore is changed into a glide

The rules of syllabification will now be specified. All of them have to satisfy the SSG, the OSC, the CC and the HWC as well (all drawings taken from Rubach & Booij (1990)):

	<i>Name</i>	<i>Obligatory</i>	<i>Description</i>	<i>Picture</i>
1	N-Placement	yes	For every non-consonantal sound at the melody tier, an N is placed at the syllable tier	
2	Progressive gliding	yes	If [i] or u occur after a vowel, their slot does not receive an N but they are changed into the glides [j] or [w]	
3	CV-Rule	yes	If there is something to the left of an N, it is included into the onset. In any case, a N'' node is created.	
4	Polish onset	no	Put another consonant into the onset (may apply several times)	
5	Polish coda	no	Insert a consonant from the right into the coda	

6	Complex coda	no	Insert one more consonant from the right into the coda	
---	--------------	----	--	--

These rules lead to sequences of VCV always being syllabified as V-CV, while for VCCV, V-CCV (Rules: N-Placement – CV Rule – Polish onset) or VC-CV (Rules: N-Placement – CV Rule – Polish coda) are equally possible. To cope with this kind of ambiguity, the authors have conducted field study with Polish native speakers. Their results suggest that longer onsets are preferred over long codas, justifying the order of the formalisms.

Application, reapplication and restructuring:

Since we are not interested in the cyclic structure of Polish phonology here, we just subsume the further findings of the authors, stating that their syllabification algorithm itself is cyclic and is reapplied to the prosodic word structure after every phonological rule. Furthermore, they say that the codas from the last run of the algorithm are erased at the beginning of a new algorithm application. Since this, too, has to do with cyclic phonology, we simply ignore it for the sake of brevity and only work with their syllabification rules.

4. The Semi-Syllable Approach (Marek&Schmidt)

Our hypothesis presumes some of the major constraints in Polish phonology that have already been stated introducing Rubach and Booij's model and which seem to be consensus among phonologists working on Polish syllable structure. These are specifically: the Sonority Sequencing Generalization (SSG), the Obstruent Sequencing Constraint (OSC), and the High Vowel Constraint (HWC).

Furthermore, we assume that every vowel functions as a nucleus and thus agree on N-placement as a first step in the syllabification process. In addition, we accept the finding of Rubach & Booij that long onsets are preferred over long codas, which will prove to be quite central in our model.

As for the sonority hierarchy used, we will actually include glides and distinguish between fricatives and stops, thus arriving at the following sonority scale:

vowels > glides > liquids > nasals > fricatives > stops

The basic or "core" syllable we assume to be CV that can be extended to CVC. In addition, we will introduce a syllable missing a vowel and thus having no nucleus, a concept that shall be called semi-syllable (\$) in the ongoing discussion.

The idea here is to try to show that SSG does apply in Polish and can be retained even in the case of consonant clusters that appear to violate it heavily from the first sight. These consonant clusters which most of the times occur word initially can, as is to be seen, be split into smaller units that do confirm the SSG. These semi-syllables consist of only consonants whereas, due to the lack of a vowel, one of the consonants takes the nucleus role. In a final step, they are concatenated with the onset of the following "full" syllable in an iterative manner.

The algorithm executed on the example krtáń is as follows:

<i>Step</i>	<i>Description</i>	<i>Picture</i>
1	N-placement: for every non-consonantal sound at the melody tier, a N is placed at the syllable tier	

2	Take as much of the consonants on the left of the V and put them into the onset as long as they are in rising sonority (SSG) and do the same for the coda and elements on the right of the nucleus:	<pre> O N C C V C k r t a ń </pre>
3	Semi-syllabify the remaining consonants into as many semi-syllables as needed. In our example it's only one	<pre> \$ σ / \ / \ C C O N C k r t a ń </pre>
4	The word <i>krtáń</i> is syllabified into only one syllable by Polish speakers, so our suggestion is that in Polish, all full syllables need at least one vowel, semi-syllables merge into the onset of the next full syllable	<pre> σ / \ O N C / \ k r t a ń </pre>

5. Comparison Syllable Structure Algorithm (SSA) by Rubach&Booij vs. Semi-Syllable Approach (SM) by Marek&Schmidt

The SSA as well as the SM both explain most of the complications concerning the retainment of the SSG in Polish well enough.

The SSA does, to our knowledge, actually approach all cases and suggests solutions for them, but it is also found that it exhibits certain difficulties explaining, for example (a) initial obstruent-liquid-obstruent clusters and (b) splits of consonant clusters in multisyllabic words. These problems are dealt with in a rather unsatisfying manner, introducing complicated extra-rules with an unclear base: for (a) the iteratively applying Polish Onsets Rule, and for (b) the observation that onsets are preferred over codas. In terms of predictability, these struggle with many cases where several alternatives are possible. Nevertheless, the algorithm provides solutions for (almost?) all cases.

The SM leads to an algorithm that is a lot simpler and does not introduce any extra rules. Moreover, it explains how the problem of seeming violation of the SSG can be dealt with in a theory-based approach. However, there are two major weaknesses with SM. For one, the algorithm heavily overgenerates. According to Bethin ('92), it is not arbitrary what consonants do appear in Polish consonant clusters. On the opposite, field study shows that certain sequences prove to occur very often, whereas others appear to be disallowed totally with an occurrence rate of zero. The SM algorithm, however, would allow, in theory, any consonant sequence in a semi-syllable as long as it is not ruled out by the OSC. In terms of predictability, SM would not help a lot.

Furthermore, there is a large set of consonant clusters that seemingly violate the SSG and occur word finally. These naturally cannot be merged into a following onset (there is none) and would thus be disallowed according to SM. This happens even in quite short and very frequent clusters consisting of an obstruent-sonorant sequence, for example in *teatr* 'theatre' (stop followed by liquid). This single point of failure suggests the discard of the algorithm or at least some extension, which cannot be provided here.

6. Conclusion

Accounting the analysis of the two different models given above, we find that our own model does not suffice in a great number of cases, which Rubach and Booij's SSA actually can handle, although in a quite unaesthetical manner. Still we understand that trying syllabification on sample data with both algorithms clearly speaks in favour of Rubach&Booij.

Literature

¹Bethin, Christina Y.. **Polish Syllables: The Role of Prosody in Phonology and Morphology**. *Publisher: Slavica. Place of Publication: Columbus, Year of Publication: 1992.*

²Rubach, Jerzy. Booij, Geert. **Syllable Structure Assignment in Polish**. *Phonology. 7 (1): 121-158. 1990.*