Boersma, Paul

How to infer utterance-internal prosodic boundaries from those at the edges

In infant research there is some discussion whether children acquire word segmentation from the statistical correlations between successive syllables (Saffran, Aslin & Newport 1996) or from the correlations between syllable sequences and utterance edges (Johnson, Seidl & Tyler 2014). Both mechanisms could play a role, and to lend computational support to the latter mechanism, this talk provides a simulation of how learners could generalize from what happens at utterance *edges* to where syllable boundaries lie utterance-*internally*.

Let's look at a child who wants to figure out where the syllable boundaries are in the overt adult utterance [kwatro]. Suppose, for simplicity, that the child knows that there are two syllables (because there are two vowels) and that the child knows that there are syllable boundaries at the edges (Strict Layer Hypothesis; Selkirk 1984). The child then has three possible hypotheses for the phonological surface structure: /.kwa.tro./, /.kwat.ro./, and /.kwatr.o./.

I simulate the acquisition of this hidden syllable structure with a fairly standard type of virtual OT learner, combining Tesar & Smolensky's (2000) Robust Interpretive Parsing with the crucially frequency-sensitive Gradual Learning Algorithm (Boersma & Hayes 2001). The learner has three simple constraints for syllable starts, namely *.C ("NoOnset"), *.CC ("NoComplexOnset"), and *.V ("Onset"), and three constraints for syllable ends, namely *C. ("NoCoda"), *CC. ("NoComplexCoda"), and *V. ("Coda"). In addition, there is a faithfulness constraint. Note that the constraint set has no preference at all for any specific kind of syllabification.

The virtual learner considers 55 triplets of underlying form, surface structure, and overt form. Each of the five possible underlying forms (|VCV|, |CVCV|, |CVCVC|, |CCVCV|, |CVCCV|) comes with the same five possible overt forms ([VCV], [CVCV], [CVCVC], [CCVCV], [CVCCV]). Most of these overt forms have two possible surface interpretations, namely with the syllable boundary before and after the intervocalic C, and the overt form [CVCCV] has three possible interpretations. In total, each underlying form comes with the same 11 possible surface–overt pairs.

We feed the learner pairs of underlying and overt forms in which the overt form is segmentally identical to the underlying form, such as |CVCVC|~[CVCVC], without giving any information about syllable structure. Each underlying–overt pair comes with a different relative frequency in the input, as given in the table behind "Spanish":

	|VCV|~ [VCV]	|CVCV|~ [CVCV]	|CVCVC|~ [CVCVC]	|CCVCV|~ [CCVCV]	|CVCCV|~ [CVCCV]
"Spanish"	100	500	50	50	100
"Tagalog"	0	100	300	5	100

We can see that this toy version of Spanish has many words that start with CC. In the simulation, all seven constraints are initially ranked at 100, and learning proceeds with standard settings in Praat, i.e. an evaluation noise of 2.0 and a plasticity of 0.1. After 100,000 pieces of data, the structural constraints have settled themselves into equilibrium. The learner ends up with the ranking *CC. >> *C. >> *V. >> *.V >> *.C > *.CC, which would handle an underlying |CVCCV| by inserting a syllable boundary **before** the internal consonant cluster, as can be seen in Tableau 1. Hence, "Spanish" learners prefer to maximize their utterance-internal onsets, based on the abundance of complex clusters utterance-initially and the relative scarcity of utterance-final consonants.

Tableau 1: “Spanish”

	106.9	100.0	97.2	94.1	92.4	91.7	89.4
|CVCCV|	*CC.	Faith	*C.	*V.	*.V	*.C	*.CC
|CVCCV| / .V.CV. / [VCV]		*!*		**	*	*	
|CVCCV| / .VC.V. / [VCV]		*!*	*	*	**		
|CVCCV| / .CV.CV. / [CVCV]		*!		**		**	
|CVCCV| / .CVC.V. / [CVCV]		*!	*	*	*	*	
|CVCCV| / .CCV.CV. / [CCVCV]		*!*		**		*	*
|CVCCV| / .CCVC.V. / [CCVCV]		*!*	*	*	*		*
|CVCCV| / .CV.CVC. / [CVCVC]		*!*	*	*		**	
|CVCCV| / .CVC.VC. / [CVCVC]		*!*	**		*	*	
☞ |CVCCV| / .CV.CCV. / [CVCCV]				**		*	*
|CVCCV| / .CVC.CV. / [CVCCV]			*!	*		**	
|CVCCV| / .CVCC.V. / [CVCCV]	*!			*	*	*	

One can do the same for a toy version of Tagalog. This language has many words that end in C and very few that start with CC (also none that start with V, as initial glottal stops are phonemic [Schachter & Otanes 1972, French 1988], but that does not influence the result). The relative frequencies are in the table above. After learning, the ranking is such that the virtual child inserts the syllable boundary **between** the two medial consonants, as can be seen in Tableau 2. Hence, Tagalog likes to split up its intervocalic clusters, based on the abundance of utterance-final codas and the relative lack of utterance-initial consonant clusters.

Tableau 2: “Tagalog”

	112.4	101.1	100.2	100.0	98.7	94.1	93.5
|CVCCV|	*.V	*CC.	*V.	Faith	*C.	*.C	*.CC
|CVCCV| / .V.CV. / [VCV]	*!		**	**		*	
|CVCCV| / .VC.V. / [VCV]	*!*		*	**	*		
|CVCCV| / .CV.CV. / [CVCV]			**!	*		**	
|CVCCV| / .CVC.V. / [CVCV]	*!		*	*	*	*	
|CVCCV| / .CCV.CV. / [CCVCV]			**!	**		*	*
|CVCCV| / .CCVC.V. / [CCVCV]	*!		*	**	*		*
|CVCCV| / .CV.CVC. / [CVCVC]			*	*!*	*	**	
|CVCCV| / .CVC.VC. / [CVCVC]	*!			**	**	*	
|CVCCV| / .CV.CCV. / [CVCCV]			**!			*	*
☞ |CVCCV| / .CVC.CV. / [CVCCV]			*		*	**	
|CVCCV| / .CVCC.V. / [CVCCV]	*!	*	*			*	

A more comprehensive simulation would include segmental substance (e.g. Spanish /.kwa.tro./ but /.kwan.do./, based on the ubiquity of utterance-initial [tr] and the lack of utterance-initial [nd]) and it would probably also include multiple candidates for underlying forms (in the case of alternations), which is the other type of hidden structure.

What has been simulated here for syllable boundaries, will also go for the phonological structure of *word* edges. Therefore, the frequency-dependency of syllable structure observed in my simulations renders the influence of edges in word segmentation more feasible.