OCP study extended to Finnic Language

Syllabic Constraints between Indo-European and Finnic Languages

J. M. Purcell

Presented in partial fulfillment of the requirements of

Linguistics 692: Phonology II

Professor Steven Weinberger

George Mason University

14th May, 2020

1. INTRODUCTON: What does it mean to violate syllabic constraints among languages?

Discovered by E. C. Fudge, there is a hypothesized linguistic universal based on an observation in the English lexicon. Monomorphemic constructs with phonemic voiceless plosives in the onset and coda position cannot followed by the initial onset /s/. In this paper, I will explore the phonologic parameters that prevent certain syllabic constructs across Estonian language, defined by the syllable structure sCVC. In English, syllables with the onset and coda in the same place of articulation and noncoronal[1] are avoided (Fudge, 1969). In order to do a complete search of constructs in the target Finnic languages of my research I want to explore areas of more natural speech that would embrace human paramaters. In extension of the study by Davis (1989) and Paradis & Prunet (1991), the ability to access samples of spontaneous speech is expanded using over 60,000 data samples from the Universal Dependencies archive. Davis observes in addition to Fudge's study that nasals and liquids are not observed in the same sCVC structure in English. This expansion of my research applies this study to an unrelated language family, Finnic, where, I investigate places where OCP occurs in different methods(Frisch, 1997) and specifically in Estonian (Hayes, 1986). In this paper I question if Fudge or Davis' Obligatory Contour Principle (OCP) obstruent parameters are present in Estonian and Finnish while concurrently bound by moraic and stress parameters. One of the specific parameter of the OCP is the Same Place Avoidance principle (SPA), which is observed in the constraints of Davis and Fudge. Like English, Estonian observes the similar resistance to $C_1C_2VC_3$ syllables where clusters C_1C_2 clusters are permitted while C_2 and C_3 are in the same place of articulation.

Standard Estonian phonology is unique in this study because the inventory includes traits that are not present in previous studies. For instance, there are no voiced plosives in the Estonian

[1] Notes on differentiation of phones based on position, such as velarization and aspiration

to be slightly devoiced in the coda position like in the English inventory. Phonemes described in Frank's and Davis' papers are limited to place of articulation and less focused on the traits of each phone such as +/- voice or +/- quantity. Davis' study does not account for a language where the place of articulation is consistent while the phoneme has the same phonetic traits—this is not including the differentiation of onset aspiration and no release for coda. The Finnic languages are again significant to OCP and SPA parameters because there is no difference between the manner, place, and traits of articulation in the onset and coda position except for the limitations of clusters relative to the nucleus.

Estonian differentiates phones, obstruents and vowels based on length, known as gradation, arranged in quantities of Q1, Q2, and Q3, in order to provide some contrast in the same place of articulation. This research, in addition to validating the presence of OCP and SPA theories, I will provide additional information whether the place of articulation is more significant than phonetic gemination for contrastive purposes. In the following model of the syllabic structure I will need to be determine permissible what segments are in Estonian to better evaluate the presence and constraints of OCP; using the restraint format, $C_1C_2V{:}C_3$, where C_1 and C_2 are a valid cluster, the gemination of either C_2 or C_3 allows for C_2 and C_3 to be the same place of articulation without any resistance.

Concurrently, Estonian OCP is present in its moraic vowel system; vowel length must be contrastive in polysyllabic consonants of the same vowel, and preferred in non-primary stressed syllables (Hayes, 1986). While Estonian permits less clusters than English, CCVCC versus CCCVCC, so, the first steps of this analysis would be looking into the potential clusters of Estonian, then explore clusters in relation to stress, and finally provide insight as to what additional restraints are present in Estonian as a generalization to predict Finnic language rule

from OCP. Later in this this paper I will explore the patterns of stressed syllables in relation to consonant clusters.

Just like the contrast of geminated consonant clusters will be searched, the length of will be vowels as well. According to Hayes, (1989), the gemination of a vowel is reliant on the stress of the syllable--naturally, the stresses permit longer, geminated vowels. Given that there may be a singular effect of stress on the length of vowels and potentially clusters, is it then possible that clusters must be dependent on stress to be more salient and obscure than their environment. If the length of a vowel should be contrastive enough in the same way a diphthong alters, then the environment cannot permit it if and only if geminated consonants are also considered contrastive enough to be more significant than place of articulation, then there cannot be another segment with the same place of articulation adjacent. The Formula for this theory would be found as the following $C_1C_2V{:}C_3$ where C_1 and C_2 are a valid cluster; the gemination of $V{:}$ allows for C_2 and C_2 to be the same place of articulation.

Finally, The compared language is unique because of its phonetic inventory features and other present OCP factors that propose possible syllabic restraints which . To clarify, the first question is to discover whether languages in the Finnic family are also averse to the occurrence of $C_1C_2V{:}C_3$ structures and, like English, in what circumstances they may be permitted. This paper leaves potential to expand on the gemination of voiceless consonants in Finnic languages and their significance OCP and SPA phonologic theories. One question is to narrow down if this aversion to SPA is to prevent combinations of minimal, less-marked segments, or a matter of balancing stress in the presence of unbalanced obstruents and vowels in a language that limits such a ratio of clusters.

2. LITERATURE REVIEW

To approach a sense of reasoning why certain combinations are avoided is based on one of many restraints observed in human speech, some more easily tested than others[2], where the measurements of the following theories require more technical attention for every variety of language. That is, the constraints OCP vary per language, and there are specific phonologic and syllabic traits that differ in every language variety. Whereas our focus of analysis in is based on the place of articulation, other phonetic traits are present to increase dissimilarity between phones that would otherwise be allophonic among native speakers[3]. However, in order to discern which data that would be acceptable in a language variety would need to be done by native speakers. Therefore, in order to find proper analysis of these theories, it is best to have natives and non-native judgements of the segments in question with consideration of how they may be accepted into both the theory and the analyzed language. The theories which will be used to explain the anomaly of specific lexical absences are as follows:

The Obligatory Contour Principle (OCP)

The OCP is a hypothesis of human language where phonemes with similar traits tend to not occur in the same environment, whether that be directly adjacent (Fudge, 1969) or within the same syllable (Davis, 1989). The OCP presents itself in a variety of language families aside from English; to reference a few: South-Semetic, Sino Tibetan, Canaanite-Semitic, and Bantu ranging

[2] Nuances in every language require a person with linguistic ability to test allophony compared between native and non-native speakers. Only then could we provide an OT analysis of what is necessary.

[3] The presence onset aspiration and no release in the coda position.

from the phonetic level of traits to the resisted adjacency of root morphemes with homorganic segments (Frisch, Pierrehumbert, & Broe, 2004). The languages in OCP studies are subject to types of differentiation; in an Estonian study, vowels require differentiated adjacent quantities within a singular phrase (Hayes, 1989; Ehala, 2003). Since Estonian language adheres to OC Principles that constrain sonorants, the language could be subject to constrains of obstruents like those found in English (Davis, 1989). The place of articulation does not change in the gradation of a consonant from <g, k, kk,> phonetically /k., k, and k:/; these phones do not differ in aspiration, voice, or place of articulation. Languages where similar place of articulation are averse to gemination is referred to as OCP – Place. Frisch (1997) also considers the principle of OCP-Place as a gradient constraint that decreases with the number of intervening segments, rather than an absolute "on/off switch." Hence, there is the possibility of Hayes'(1989) study of vowel quantities creating an additional factor to consider while approaching Davis' (1989) work on the reach of OCP-place across syllables. Although the OC Principle can be generalized as "similarity avoidance," it's too vague for this study; the study of Finnic-Estonian OCP will analyze just same place of articulation avoidance within a syllable structure.

Same Placement Avoidance (SPA)

Pozdniakov and Serger's (2007) claim is that SPA, like the OCP, occurrence is not limited to any language family, and is in fact observed in most, if not all, languages in the world and therefore likely to be a universal property of human language. While the Same Placement Avoidance is clear in languages such as Japanese (Kawahara et al., 2005), Muna (Coetzee & Pater, 2006), the SPA test has not yet occurred within a study of the Finnic family although

studies of the syllabic structure and stop sequences have by native speakers as a foundation for external language study (Viitso, 2003: 14-35). Specific analysis of language by place of articulation has its exceptions, for instance nasal assimilation is an Optimality Theory (OT) occurrence that seems to be a superficial change—although OT suggests that all constraints are superficial, some assimilations impede comprehension like the sCVC construction while others are rules to language like nasal assimilation. Features in the Estonian inventory have exceptional traits which make it a language suited for SPA and OT analysis.

Optimality Theory (OT)

In each variety of language there are rules which characteristics of language abide, primarily in phonology, which can be deemed significant to other traits in language. Dependent on a theory of universal grammar, language families abide to a system of ranking. One may also suggest the ranking defines the language family (Kager, 1992a). For example, the following constraints of markedness: the stress in Estonian and Finnish must be placed on the initial syllable for native lexical items bringing NonFinality (NonFin) to the top of the hierarchy. while “heavy syllables” occur in stressless positions because the onset and coda are constrained in the initial position, making Weight-to-Stress(WSP) at the bottom of the hierarchy. Additionally, there are Faithfulness constraints, where items in Estonian are deleted in order to make adjacent morphemes compatible, subverts the MAX constraint to another significant constraint like IDENT(+quantity) is always present across morphemes (Prince, Alan, and Smolensky, 1993). Hayes and Viitso both reference OT as an appropriate analysis of syllabic stress where light syllables followed by heavy are generally avoided (Carlson, 1978; Kiparsky, 1991). “Every

polysyllable nevertheless begins with a bisyllabic foot, the strongest in the word, no matter what the composition of the first two syllables is." We will account for this in section 4. "The constraint EDGEMOST is crucially involved; specifically, EDGEMOST(head left), the member of the EDGEMOST family which requires the head foot of the prosodic word to be placed in initial position." In the prosodic-syllabic constraint of Estonian, the initial syllable is the most stressed, otherwise noted in section 5. "EDGEMOST(head left) evidently dominates whatever prominential or foot-shape constraint is responsible for the LH effect" (Prince, Alan, and Smolensky, 1993: 31).

3. FACTORS OF THE ESTONIAN INVENTORY

To compare the inventory of the original SPA studies featuring a generalized English (Fudge, 1969; Davis, 1986; Pozdniakov and Serger, 2007), the phonetic inventory of standard Estonian (Asu & Teras, 2009) differs from English in several ways that contribute to parts of this study. Here are the following inventories:

Table 1: A cross analysis of English and Estonian phonetic inventories.

CONSONANTS	Bilabial		Labiodental		Dental		Alveolar		Postalveolar		Palatal		Velar		Glottal	
Pulmonic																
Plosive	p	b	p̪	b̪	t̪	d̪	t	d			c	ɟ	k	g	ʔ	
Nasal	m̥	m	m̥̪	m̪	n̥̪	n̪	n̥	n			ɲ̊	ɲ	ŋ̊	ŋ		
Trill		ʙ					r̥	r								
Tap or Flap		ⱱ̟		ⱱ			ɾ̥	ɾ								
Fricative	ɸ	β	f	v	θ	ð	s	z	ʃ	ʒ	ç	ʝ	x	ɣ	h	ɦ
Lateral fricative							ɬ	ɮ								
Approximant		ʋ						ɹ			j̊	j		ɰ		
Lateral approximant								l				ʎ̝		ʟ		
Affricate	pf	bv	pf	b̪v	tθ	dð	ts	dz	tʃ	dʒ	cç	ɟʝ	kx	gɣ		

Red: English Blue: Estonian Pink: Both

We can observe, as aforementioned, there is a variety of plosives with different features that differ. Palatals in Estonian, like nasals, operate under the same pretense of OT assimilation, yet occur as separate phonemes. Voiced plosives do not occur in the Estonian inventory where they commonly do in the English inventory, except for in loaned words, where +voice is perceptually similar, as estimated from confusability, des not solely determine unfaithful production of non-native structures (Gallagher and Graff: Davidson and Shaw and Mielke, 2012: 109). To clarify between the graphemic <b> and /b/ in these languages, orthographic <b> in Estonian is [p.] and English loaned words in Estonian <b> is preserved as [b.] or devoiced as [p]. In Table 2, the graph are presented as the graphemic forms in Estonian <b> as [p.].

Table 2. Monomorphemic clusters of two consonants after the vowel of the initial syllable in native and adapted Standard Estonian stems (Viitso, 2003: 22)

.

	b	*d*	*g*	*p*	*t*	*k*	*s(s)*	*m*	*n*	*l*	*r*	*v*	*j*
b										*bl*	*br*		*bj*
d											*dr*	*dv*	*dj*
g									gn	*gl*	*gr*	*gv*	
p					*pt*		*ps*			*pl*	*pr*		*pj*
t						*tk*	*ts*			tl	*tr*	*tv*	*tj*
k							*ks*		kn	*kl*	*kr*		*kj*
s					*st*	*sk*			*sn*	*sl*	sr	*sv*	*sj*
h					*ht*	*hk*		*hm*	*hn*	*hl*	*hr*	*hv*	*hj*
m	*mb*			*mp*			*ms(s)*		mn	ml	mr		
n		*nd*	*ng*		*nt*	*nk*	*ns(s)*			nl	nr		
l	*lb*	*ld*	*lg*	*lp*	*lt*	*lk*	*ls(s)*	*lm*			lr	*lv*	*lj*
r	*rb*	*rd*	*rg*	*rp*	*rt*	*rk*	*rs(s)*	*rm*	*rn*	*rl*		*rv*	*rj*
v										vl	vr		

Key: The non-italicized clusters result from active morphophonological apocope in synchronically simple word stems, clusters occurring only on synchronical morpheme boundaries are not taken into account. The letter in parentheses reflect the writing of the corresponding clusters with ss in Q3, exceptional syllabants.

Table 3: Monomorphemic clusters of three consonants after the vowel of the initial syllable in native and adapted Standard Estonian stems (Viitso, 2003: 24).

								psl											
								ksl											
					spl	stl	skl							str					
						htl	hkl					hvl		htr	hkr			hvr
mps					mpl				mbl				mpr			mbr	ndr	
	nts	*nks*				ntl	nkl			ndl	ngl			ntr	nkr		ldr	
lps	*lts*	*lks*	*lst*											ltr		lbr		
	rts	*rks*	*rst*	*rsk*		rtl	rkl		rbl	rdl	rgl					rbr		

Key: The non-italicized clusters result from active morphophonological apocope in synchronically simple word stems, clusters occurring only on synchronocal morpheme boundaries are not taken into account.

These clusters are specifically noted as those following the initial onset. Observing this data, we can conclude that there is something significant about stress and the permission of clusters in a syllable. However, the ability to discern the place of articulation decreases in consonant clusters (Just, Michaels, & Suskick, 1979), meaning that in the initial stress the onset and coda must be preserved. Therefore, we may observe a disparity of clusters in this position despite the maximal onset and coda for Estonian language. This is evidence to support that stressed syllables of monosyllabic or polysyllabic lexemes affect the constraint of obstruent clusters. If we compare the occurrences of /s/ within clusters of 2 segments, and segments of 3, we can deduce that the /s/ clusters only occur as an onset cluster to the next syllable instead of coda to the previous because the affricate /ks/ does not occur in Estonian. Therefore, the syllabic constraint that occurs in the syllable is preferred to be C, and the coda CC. However, a counter analysis would suggest that the clusters can only be parsed by a native speaker and leftmost /s/ could be assimilated from coda to onset phonologically. Nevertheless, the occurrence of these clusters are common, and notably do not occur in the stressed position, creating a correlative relationship that suggests an OT constraint that that would prevent a lack of salience connecting the theories of OCP and OT.

4. GEMINATION OF SEGMENTS—the relation between OCP constraints and Stress

In addition, we can observe that there is another pattern, or lack of, in the data that should so far be able to support a restraint of geminated items; Gradation in Estonian includes a set of stem alternations that occur in stressed syllables or at the beginning of post-tonic syllables when the word is inflected:

(a) quantity alternation of long stressed syllables whereby Q3 in the strong grade alternates with Q2 in the weak grade.

(b) quality alternation that results from the weakening of former stops. (Viitso, 2003: 25)

In tables 1 & 2, the possibility of the previous clusters do not have any geminated items that could be in the onset or coda of a syllable. These phones are recognized between secondary morphemic and harmonic stress determined by the circumstance a lexical item would be created to fit any of the clusters. Below discusses the possibility of syllabic stress and gemination of syllables. The gemination of items does several things: it prevents the presence of other segments, and in doing so it provides salience between contrastive segments which would be permitted in unstressed[4] positions. /k:/,/p:/,and /t;/ cannot occur adjacent to /s/ or /s:/ in the same syllable. While there is the possibility of non-geminated segments in stressed clusters, /sC/ and /Cs/ does not occur across syllables where C is a voiceless plosive.

In addition to stress, the behavior of geminated items combines with grammatical case. In the table below, gemination occurs differently across primary and secondary stress. The occurrence of extended segments is not only related to stress, they are morphologically marked, further suggesting that stressed syllables are more marked with unique segments.

[4] I suggest the term commonly used as “unstressed” be used more appropriately as “stressless” due to the fact that stress is added instead of removed. Shifting stress is also a form of reapplication, adding stress.

Table 4: The grammaticality of gemination: Old and New Quantity alternations caused by stress pattern changes: *petlik* ‘deceptive’ and *õnnelik* ‘happy’

	Old	New	Old	New
Nominative Sg	*`pet`lik*	*`petlik*	*´õnne`lik*	*´õnnelik*
Genitive Sg	*`pet´liku*	*`petliku*	*´õnne´liku*	*´õnne´liku*
Partitive Sg	*`pet`likku*	*`pet`likku*	*´õnne`likku*	*´õnne`likku*
Genitive Pl	*`pet`likkude*	*`pet`likkude*	*´õnne`likkude*	*´õnne`likkude*

In the previous studies, contrastive voicing is a trait in plosives that differentiates place of adjacent place of articulation across syllables (Davis, 1989); again, voiced plosives are not present in the Estonian inventory. The range of sonority between phonemes ranges from voiceless plosives, fricatives, to nasals and syllabants. Note, in the event of a 3-consonant cluster, there is always a syllabant present. There is a gap of sonority between plosives and articulations above nasals. To summarize, I assert that the theory has been constructed, are able to support that in stressed syllables, the onset and coda are limited beyond that of others maximally CVCC. Furthermore this suggests that it is more optimal for the plosives in the stressed syllable be distinguished and the reduction of the cluster makes the phoneme more salient (Hudson, 1995). From this observation, I can further assert that the study of Berkeley (1994) states that maximally similar consonants in stressed syllables are less frequent in the onset position as well. Additionally, Frisch (1997) noted that initially stressed syllables, unrelated to the place of articulation, will avoid onset and coda pairs of “high sonority”—in the case for Estonian this includes /r, n, l, v, m/. Therefore, there must also be an expected range of both sonority and place of articulation in the first syllable compared to those following.

Table 5: Stress Patters in words up to five syllables

While we can account for the position of stress among word form, by the claim that there are different patterns only beyond the initial stress, and other stresses are predictable when the onset cluster is not present. Beckman (1998) claims that a segment with place of articulation may be replicated in a stressed syllable, while unstressed syllables do not permit co-articulation yet greater clusters. Clusters cannot be in the same place as the syllable, then there is support to additional claim that restrictions of clusters not only cannot occur on the onset where primary stress lies, but also on all other stresses that are otherwise stressless. A morphological restraint affects the pattern of stress rather than the troche-expected stress pattern; the secondary stress is allows adjacent to the primary stress and on the ultimate, penultimate syllable. From observing Table 3, the possibility of a secondary stress is allowed if it can be followed by an unstressed syllable. Inasmuch, the secondary stress is only present adjacent to the primary stress, given that final stresses can be either unstressed or rhythmic where syllables are less constrained. The rhythmic, troche, stress cannot be adjacent to morphologically conditioned stress, and this prevents clusters across the word form (Kager, 1992b).

Table 3: patterns of stress in Estonian

No. of syllables	Pattern	Example		Gloss
1	P	*`suu*	[*sû·*]	'mouth'
	U	*ma*	[*ma*]	'I'
2	PM	*`pet`lik*	[*pe·ɫ̀li:k̀*]	'deceptive'
	PU	*´mina*	[*mi·nà*]	'I'
3	PMU	*`joo´mine*	[*jô·mi:ne*]	'drinking'
	PUM	*´õnne`lik*	[*e̯·n̆neli:k̀*]	'happy'
	PUU	*´sadamad*	[*sa·ᴅamaᴅ*]	'harbours'
4	PMUU	*`joo´misega*	[*jô·mi:zeɢa*]	'with drinking (ComSg)'
	PUUM	*´elaja`lik*	[*e·lajali:k̀*]	'beastly (adj.)'
	PUMU	*`lükka´mine*	[*lü·k̀k̆ami:ne*]	'pushing (noun)'
	PUAU	*`lükka´sime*	[*lü·k̀k̆azi:me*]	'we pushed'
5	PMUAU	*`joo´mise´gagi*	[*jô·mi:zeɢa:ɢi*]	'even with drinking'
	PUUMU	*´kogele´mine*	[*ko·ɢelemi:ne*]	'stuttering (noun)'
	PUAUM	*´õpe´taja`lik*	[*e̯·p̆petta:jali:k̀*]	'teacher-like (adj.)'
	PUAUU	*´koge´lesime*	[*ko·ɢele:zime*]	'we stuttered'

Key: P- primarily stressed syllable, A- syllable with an automatic or rhythmical secondary stress, M- syllable with a morphologically conditioned secondary stress, U- unstressed syllable. Evidence of OT: Faithfulness: EDGEMOST.

Since the stressed syllables have been addressed as something regulated, what are the ways that the initial syllable is able to be presented in Estonian? From the data in Table 2, the permitted onsets are CCV, where the onsets are limited, only in a very few circumstances does the sibilant /s/ occur as the initial onset as a cluster—again, these are clusters between less-stressed items and initial /s/ is most likely /s/ as coda from the previous syllable while the adjacent segment is onset to the adjacent syllable. If we reduce the clusters that are present in CC

and CCC clusters, the presence of /sC/are sparse in comparison to all other permitted items, clusters are avoided in root morphemes.

So why is /s/ allowed in the onset position, but not in the coda? That is, why does /pops/ ot /tots/ exist across families of language but not /spops/. In order to explore this phenomenon, the we must investigate the effects of clusters and the effects of adjacent phonetic properties. Whereas it commonly known in English that significant aspiration occurs in voiceless plosives in the initial onset position and no release of the plosives with the same voiceless traits in the final coda position (Hudson, 1995). So far there is no mention of these phenomena in Estonian or Finnish, however, we can now suggest that the salience between phonemes in the initial syllable are more significant than the rest of the syllables in a given lexeme.

Furthermore, my claim can be explained using an Optimality Theory hierarchy. In addition to the *CLUSTER claim in the onset position, /s/ and /s:/ occur in a singular prevocalic position of prepositional post-fixations. Previously I stated that the /s/ and /t/ clusters cannot be adjacent in any given position in the onset within the same morpheme. This is elaborated upon by order of acceptance: that /st/ and /ts/are the least controlled coronal clusters, /ps/,/sp/ and /ks/,/sk/ are avoided clusters; geminated segments cannot be adjacent to anything, coronal or not. The lack of /sC-/ items is sparse especially for monomorphemic items, so much that to avoid this occurrence, the stress shifts to the second position and an additional nucleus as a type of transferred paragoge to fit into the target phonology. Next, an additional support for morphemic types suggests that there is less restriction of /s/ and /t/ adjacency across final morphemes, on the ultimate syllable.

In circumstances where an onset is present, the markedness is not significant to the lexeme. Like English, when the sibilant, or strident, occurs on the coda, the factor of avoidance

does not apply in the same way. But, why does the presence of /s/ as an onset differ from /s/ as a coda? In both languages, suffixation of -<s> carries additional semantic properties; in English sibilants are used in plurality, possession, and verb paradigms. In Estonian, <s> occurs in the presence of an adjectival, Inessive, Elative, Translative, and Illative prepositional case post-position fixations, relatively as significant in simple clauses with motion or relations to an object.

4.1 THE PHONOLOGY OF CASE ENDINGS. [a chapter]

We should consider that some samples of the most frequent morphemes in the post position must be compatible with any root, regardless of the syllable construct. In brief analysis, there are a total of 14 cases with multiple single and plural variants, totaling 47 possibilities. Of these, 41 have alveolar fricatives and/or stops, and 10 have a combination of <d[5]>,<t>,or <s>.

Below is a following of the data that edges close to the parameters of the study, that despite SPA, aversion by place of articulation submits to length features of Estonian:

Case	Singular	Plural	Difference
Partitive	--	-sid	manner
Illative	--	-desse/tesse[6]	manner
Innessive	--	-des/tes	manner
Elative	-st	-dest/-test, -ist	manner
Abessive	--	-deta/-teta	quantity
Translative	-ks	-deks/-teks	manner

[5] In phonetics, Estonian is probably best known for its three degrees of contrastive quantity: short (Q1), long (Q2) and overlong (Q3). Another reminder that <d> is not voiced, rather defined as a "Short plosive" Q1, whereas <t> is a voiceless unaspirated consonant Q2 /t/.

[6] -esse also occurs here when the coda /t/ or /d/ is present from the previous morpheme., same for -iks TRANS.PL

Note: all the possible constructs of these syllables prefer clusters in the coda. Across the nucleus, there is differentiation of phonetic features enough to satisfy OCP while violating the expectations of the SPA, where the limitations of clusters are still bound to the onset of the ultimate syllable. The OCP parameters taken account for the coda position where manner of place and quantity are "different enough" to be within the most popular morphemes. Evidence of OT: Faithfulness: MAX, where segments are deleted in the presence of others.

5: THE PROCESS OF DATA COLLECTION

Where I have discussed that the prevocalic sC cluster is already avoided in Estonian, the same occurrence as a cluster is complicated to devise, in fact does not occur as an onset from any findings in a morphological dictionary (Eesti Keele Instituut, 2018). To view data, please refer to Attachment I, for data gathered from the Universal Dependencies Archive, Attachment II for data gathered from the Morphological Dictionary of the Eesti Keele Instituut.

Spontaneous speech found in the Universal Dependencies Archive (Estonian UD, Finnish UD, 2014) represents diverse examples of syntax analysis with over 60,000 lexical samples. In order to optimize the search in these archives I successfully developed a search script to search through a text document with a coded ordering of letters representative of the conflicting combinations. In the script the sequence of letters are specifically searched for in order to be discoverable in a document of data; of the 4 initial scripts, few data we found, most of which were loan words, mentioned later in this section, are exceptions to stress. Each of the following pieces of code is spliced to explain the order to operations:

1: cat *.conllu. :prints out the content of all files with the .conllu extension. 2: egprep. Searches for all strings within a file that match the contents of the quotation marks. 3: cuts. cuts out the sections matching the items provided. 4: sort:sorts the entries provided in alphabetical order. 5: uniq -c. :removes repeated entries and provides a count of those repeated. 6: sort -k. :sorts the entries again from coda. 7: > (X)-blab.txt :prints the resulting entries into this (X)file

The script is defined under the following order: A Categorical instance within file type*-collu, one instance between representations of text that contain in order; starting /s/ adjacent to bilabials (p or pp or b or m) adjacent to (any vowel combination within the inventory) adjacent to the same categorical frame of bilabials(p or pp, or b, or m). That is, all the possibilities of letters (sounds) with the same place of articulation across a syllable that fit into the $C_1C_2VC_3$ formula are considered. i.e. <Lt>,<nk>,<pl>, all the combinations in Table 1. and substituted into respective search places of articulation. Then, the data is sorted into alphabetical order by the first two sorts and then by frequency. Finally, a file is created based on the text.

The following was used on the script for Estonian bilabial corpus data:

cat *.conllu | egrep "^[0-9]\t[ts](p|pp|b|m)+[aeiouöäüõ]+(p|pp|b|m)+" | cut -f2-4 -f6 | sort | uniq -c |sort -k2 > est-blab.txt

Due to the efficiency of this method, the search was expanded from the Estonian corpus to Finnic counterpart, Finnish; essentially the same codes were applied to the Universal Dependencies archive for the Finnish corpus. Similarly, none of these transfer items have been

altered in ways others have. Some speculation could be that they are so significant in cannot be changed in order to maintain differentiation of other lexical items.

The following was used on the script for Finnish bilabial corpus data:

```
cat *.conllu | egrep "^[0-9]\t[ts](p|pp|b|m)+[aeiouöäüõ]+(p|pp|b|m)+" | cut -f2-4 -f6 | sort | uniq -c |sort -k2 > fin-blab.txt
```

As predicted, the data found from the first round of searches were remarkably slim. Of the found data, the allowed clusters, even in these borrowed items, behaved differently than other borrowed English items in the language; onsets in these items are preserved instead of being subject to phonologic assimilation into either of the Finnic varieties. Some of the word forms that we can sample from are in the technology field, and we can observe that the initial onset is presented as CCVVC, which violates the CVVC maximum for stressed syllables, however these data are supported to be valid by the studies of sonority whereas in Finnish “the sonority hierarchy determines fixed stress in loanwords” (Kiparsky, 2004: 22). The following loan words can be first assessed by the onsets and then determine if the stress would be altered to the second syllable, which contradicts the possibility of an absolute shift for all onsets. So, the question is presented: do loan words expand the syllabic restrictions of Estonian, or does Estonian assimilate the OT: Faithfulness: MAX constraint by altering the stress to the second syllable while there must be a right most nucleus? In Appendix II, the loan words are polysyllabic-3.

Here is some of the following data that can be observed in their nominal forms

'printata',	'skannata'	hipsteri	videoblogi
EN: Print	scan	hipster	video-blog

(Wall, D., 2013)

So far observed, /s/ and <s> are significant because the cluster is preserved and yet causes less salience between the onset and coda. From the previous data, we can support the claim that stress in a syllable constrain allowed onset and coda. Broselow (1982) states that Finnic languages shift the primary stress to the right, to the next syllable for loan words. In the data presented in appendix A, the borrowed terms are monosyllabic: /spa:m/ and /stat/, so the stress would be placed onto the following CV syllables of Finnic origin. While the onset loses properties that differentiate it, the coda also is expected to be in the pre-position of the syllable. The split of CCC clusters in table 3 are typically separated as follows VCC:CV. The same occus for Table 2; CC is typically separated as VC:CV (Vittso, 2007: 26).

where the difference between /t/ and /t:/ may be treated different enough despite following the pattern of OT differentiations of the syllable rather as a matter of efficiency.

6: DISCUSSION AND SUMMARY

Reviewing the analysis that constrains syllables in the Finnic languages, we can assert that there are OT and OCP constrains documented that take similar precedence in Estonian and Finnish. The languages in the focus of this study have been well documented by natives, and their analyses into Universal Theories are still in process. I would like to note that the words and phrases that were found to breach the parameters of OCP-place and quantity were reviewed by

educated native speakers Estonian and verified that the parameters were indeed accurate. The pattern of $C_1C_2V{:}C_3$ where clusters are permitted to the maximal amount are not permitted across syllables in the onset position, but this is not limited in the coda position. Less constrains are left on the coda than the onset of syllables in both the stressed and less-stressed positions. However, we are confronted with a more naturalistic dilemma: speakers of human language can conceptualize and easily produce such OCP-conflicting, language avoiding syllables: /spup/,/stut/,/or/skuk/. These syllables are available for our use, yet they are avoided and not used unless in onomatopoeic or wordplay. These constructions are recognized as 'funny' and are used in a symbolic way, signifying that we have some innate knowledge about why they should not be in the lexicon. I assert that recognized novel lexemes in language readily violate these parameters and reinforce permission to violate OCP once they are normalized (Gaskell, M., Dumay, N. 2003). For instance, nearly every violation stated is documented with a narrow definition in The Urban Dictionary and a sample of the word form in use (1999-2020).

Political conflicts and language contact affect language structure, it is commonly known among linguists that language contact affects structure and inventory. By comparing the lexical change in the continuum of Estonian, we support to conclude that Fudge's hypothesis affects languages in contact with others over several generations. Estonian by far may be one of the worst languages for this study because of its consistent contact with other Indo-European languages. However, Fudge claims that the OCP is present in isolated languages, and the next step of Finnic OCP would be Veps or Mordvin, languages less-contacted by Indo-European varieties. Before 1940, Finnish and Estonian did NOT allow /sC/ clusters as an onset, as noted within an Estonian state dictionary. Once /sC/ became a cluster, then the rule applies.

But what about the distance across syllables? While Davis (1989) shows that OCP-place operates across boundaries, is there a gradient of the triphthong or potential V::V::V quantities that occur? The most-simple answer to this is no, because these, like the previously analyzed clusters only occur while there is no presence of an onset nor coda and are split syllables.

Lastly, we may conclude with another answer to the initial question proposed in this paper: “ Are the same restraints held on English also held in Finnic?” The answer is yes, but it is from different constraints prevent these syllables from occuring. These Finnic languages do not focus on how /s/ makes aspiration l with a plosive cluster less salient, because it does not simply occur. In Estonian and Finnish, however, stress-like constraints prevent these clusters, period, in any natural syllabic form.

FINISH THE CITATIONS {CTRL F)

Carlson, Lauri. 1978. Word stress in Finnish. Ms. MIT, Cambridge, MA

Davis, S. (1989). Cross-vowel phonotactic constraints. Computational Linguistics, 15, 109–110

Ehala, M. (2003). Estonian Quantity: Implications for Moraic Theory, *Generative Approaches to Finnic and Saami Linguistics*, 51-81. CSLI Publications.

Fudge, E. (1969). Syllables. *Journal of Linguistics, 5* (2), 253-286. doi:10.1017/S0022226700002267

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. Natural Language & Linguistic Theory, 22(1), 179–228

Gaskell, M., Dumay, N. (2003) Lexical Competition and the Acquisition of Novel Words, Cognition, 89(2), 105-132

Hayes, B. (1986) Compensatory Lengthening in Moraic Phonology, Department of Linguistics, UCLA 253-306

Lehiste, I. (1965). The Function of Quantity in Finnish and Estonian, *Language*, 41(3), 447-456.

Prince, A., & Smolensky, P. (1993). Optomality Theory, *Constraint Interaction in Generative Grammar*. 262.

Roach, P.J. (1979). "Laryngeal-oral coarticulation in glottalised English plosives", *Journal of the International Phonetic Association*.

Viitso, T-R. (2007). Structure of the Estonian Language: Phonology, Morphology, and Word Formation, *EstoF(nian Language*. (1)1-84

Wall, D. (2013) *Sounds familiar? It's because it's an English loan word.* Yle Uutiset. Retrieved May 13, 2020, from https://yle.fi/uutiset/osasto/news/sounds_familiar_its_because_its_an_english_loan_word/7021382

Eesti Keele Instituit. *ÕS Eesti õigekeelsussõnaraamat ÕS (2018).* Retrieved May 13, 2020, from http://www.eki.ee/dict/qs/index.cgi?Q=st*&F=M