Improving the Formalism of Phonological Parameter Hierarchies

Johanna Benz
University of Leipzig
benz.johanna@web.de

Shanti Ulfsbjorninn
UCL – London
s.ulfsbjorninn@ucl.ac.uk

1 Aims and Structure

The growing *Substance Free* phonological movement (Hale & Reiss 2008; Samuels 2017 and articles within) argues that markedness has no place in a formal theory of phonological competence. This is considered to be overlapping both with Blevins' program of eliminating formal phonological universals (2004, 2009, 2017), and Haspelmath's critique (2006). However, when it comes to syllable structure, we think this idea has been overextended. There are implicational universals that properly belong to the phonology. Moreover, in the correct formalism, markedness is more than just a 'metaphor for a cognitive state' (Haspelmath 2006). Following Ulfsbjorninn (2017), we will show that markedness is epiphenomenal but that it converts directly into linguistic categories, both in the complexity of a representation ($\mathfrak{R}$), and the corresponding number of positive setting of parameters ($\mathfrak{P}$) required to describe/permit it.

Taking consonant clusters (CCs) as a case study, we will defend the view that the formal typology of CCs is explained by the decision tree in (2).[1] Markedness does not have any *active* role to play in the grammar, however, it does directly correspond to the number of <yes> parameter settings. These, in turn, directly correspond to the complexity of the related phonological form. Moreover, because the parameters are arranged in a hierarchy, a parameter's <yes> setting may be contingent on the <yes> setting of other parameters. The hierarchy explains the implicational universals.

At the end, we will concede that the parameter hierarchy itself needs to be explained. Why is the hierarchy is this way, and not some other way? For two of the hierarchical steps, the solution appears to be quite general: (a) Filled categories are stronger than weak categories, and (b) Non-locality implies locality. Similar implications are expected in other parts of the grammar and could be captured in the same way, by arranging the corresponding parameters into a hierarchy. The third implicational step (so far) resists obvious explanation in general terms, (c) Final nuclei are stronger than Medial nuclei (cf. Charette 1990; 1992).

2 Background

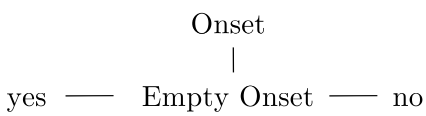
Ulfsbjorninn (2017) shows that a number of implicational universals (related to syllable structure) are best modelled by parameter hierarchies. Consider the simple case of an onset at the beginning of the word. There are no languages where syllables may *only* begin with vowels. For any given language, vowel-initial syllables presuppose the presence of consonant-initial syllables.[2] This

[1] Counterexamples from Malagasy, Thai and Gbe/Kru languages will be discounted on the basis of phonological behaviour and loanword adaptation.

[2] Arrernte is sometimes cited as a counterexample but 25% of the Arrernte lexicon begins with a consonant (cf. the re-analyses of Kiparsky 2013, Topintzi & Nevins 2017).

typological fact is explained by the decision tree in (1). According to this schema, vowel-initial words require **Empty Onset** to be set to <yes>. However, the choice of setting **Empty Onset** is contingent on the presence of **Onset** (which allows consonant-initial words).

(1)

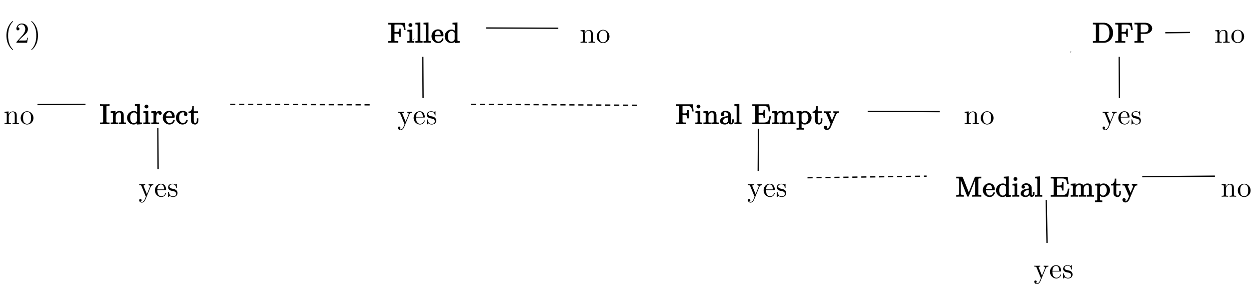


The structure in (1) also shows the formal difference between a Principle like **Onset** and a Parameter like **Empty Onset**: a Parameter is a Principle with a decision point.[3] The hierarchy in (1) gives a satisfying formal account of the typology but it also makes markedness epiphenomenal on the representation (if we assume Strict CV (Lowenstamm 1996; Scheer 2004)). In Strict CV, there is universal template for syllable structure. This can be defined in terms of precedence: $\# \rightarrow C$, $C \rightarrow V$, $V \rightarrow \%$ or C (ibid.; Faust & Ulfsbjorninn to appear). In the default/unmarked state, each of these C and V positions is filled with featural content. This means that for every representation that is not $CVCV^{(n+1)}$, every move away from the default will require an empty node as part of its structural description and every empty node requires a <yes> setting to permit it.

3 Proposal

Taking the typology of CCs as a case study, we accept Charette's (1990, 1992) two key insights: (a) CCs are licensed by the nucleus that follows them, and (b) the following implicational universals hold: **Medial** implies **Final**, **Empty** implies **Filled**, and **Indirect** implies **Direct**.[4]

(2)



We propose the following formal innovations. Parameters are instructions of how to move away from the CVCV baseline so: (1) the default is always a <no> setting. (2) the <yes> setting is dominated by its parameter (which is on the same level as the <no> default setting). Being dominated takes you to a lower level in the parameter hierarchy. (3) Only <yes> settings allow

[3] Unlike OT, no re-ranking is possible in this system and there are no violable principles.

[4] In addition to these related parameters, the DFP (domain-final parameter) is also required. Since this parameter is independent of the others, it is represented in its own treelet.

you to descend a level. A parameter that is contingent on a higher <yes> setting is linked at the same level as that <yes> setting. (4) Principles (such as **Direct**) are not part of the tree, the tree now only shows decisions (parameters). (5) Markedness is a description of the <yes> settings only and each <yes> setting corresponds directly to a linguistic category or state in the representation.

In addition to providing a formal and precise definition of markedness, this model also allows reference to a second formal property, namely depth in the hierarchy, which can be determined for any given parameter. While some parameters are accessible immediately (such as Filled and DFP), others are embedded to a varying degree. Parameter Hierarchies therefore formalise the difference between Principles and Parameters (by excluding the former due to their lack of a decision point) as well as a difference between parameters, which are grouped into different accessibility classes.

References

Blevins, J. 2004. *Evolutionary phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press. **Blevins, J.** 2009. Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics*, 48: 264-73. **Blevins, J.** 2017. What are grammars made of? In: Samuels, B. 2017. (ed.), *Beyond Markedness in Formal Phonology*, John Benjamins Publishing Company. **Charette, M.** 1990. License to govern. *Phonology*, 7:233-253. **Charette, M.** 1992. Polish and Mongolian meet Government Licensing. *SOAS Working Papers in Linguistics and Phonetics*, 2:275-291. **Faust, N. & S. Ulfsbjorninn.** to appear. Arabic stress without moras, extrametricality or syllables. In: *Linguistic Review.* **Hale, M., and C. Reiss.** 2008. *The Phonological Enterprise.* Oxford University Press. **Haspelmath, M.** 2006. Against markedness (and what to replace it with). *Journal of Linguistics*, 42(1):25-70. **Kiparsky, P.** 2013. On the empirical bases of phonological typology. Talk @ Workshop on Phonological Typology, Somerville College, University of Oxford. **Lowenstamm, J.** 1996. CV as the only syllable type. In: J. Durand and B. Laks (Eds.), *Current Trends in Phonology: Models and Methods.* 419-441. Salford, Manchester: ESRI. **Samuels, B.** 2017. (ed.) *Beyond Markedness in Formal Phonology*, Amsterdam: John Benjamins Publishing Company. **Scheer, T.** 2004. *A Lateral Theory of Phonology. Vol 1: What is CVCV, and why should it be?* Berlin: Mouton de Gruyter. **Topintzi, N. & A. Nevins.** 2017. Moraic onsets in Arrernte. *Phonology*, 34(3):615-650. **Ulfsbjorninn, S.** 2017. Markedness and formalising phonological representations. In Samuels, B. (ed.), *Beyond Markedness in Formal Phonology.* Amsterdam: John Benjamins Publishing Company.

What can we learn from implementing Optimality Theory?

Tamás Biró
ELTE Eötvös Loránd University

Lauri Karttunen (2006) argued more than a decade ago for "the insufficiency of paper-and-pencil linguistics". He showed that even a leading phonologist could make a mistake when developing a complicated analysis, namely, by omitting relevant candidates from an OT tableau. Karttunen therefore suggested that the best recipe to avoid such a problem is to implement the grammar computationally, such as by using a finite-state realization thereof.

Already before Karttunen's paper, and also since then, numerous software tools have been developed to support the linguists working with Optimality Theory and Harmonic Grammar. In fact, many of us are not as well-versed in programming as Karttunen is, not to mention the restrictions posed by finite-state technology on Optimality Theory. Consequently, those OT tools are very useful to phonologists who would like to check their analyses for possible mistakes, but would not be able to do so without them.

Now the problem is that OT allows a very broad spectrum of "objects": not only (underlying and surface) linguistic representations, but also Gen and the constraints come in all shapes and sizes. While they are postulated to be universal across languages, they immensely differ per linguist and per article. Many OT tools simply expect the user to enter a tableau manually; so the software immediately work with the violation levels (number of stars) in the cells, without caring for either the representations (the leftmost column) or the constraints (the uppermost row). These tools, in turn, can only implement grammars that contain not simply a finite number of candidates, but a "reasonable" number of them. Unless the linguists are able to generate the tableau automatically, they have to create it themselves by hand, which would be too laborious, were the tableau realistically complex.

The toolkit developed by the present author comes with predefined sets of forms, candidates, Gen and constraints. Some of them are motivated by contemporary phonology, and others by the simplicity of the formalism. Moreover, the toolkit also permits the combination of constraints, such as their addition, multiplication or logical combinations. In short, the toolkit invites the linguist to build up the modules of their analysis in a novel, creative way.

In my talk, I shall present the conceptual framework behind this toolkit, and demonstrate how colleagues can use it. I argue that a novel perspective on OT's well-known building blocks not only makes it possible for the phonologists to double-check their analyses, but it is also provides a new conceptual understanding. I compare the situation to the famous Gestalt picture that can be perceived either as a young lady looking backwards, or as an old woman looking downwards. Similarly, the building blocks of an OT grammar can also be perceived it two ways: either as linguistic concepts, or as mathematical objects. In turn, this second perception may benefit the linguist's creativity.

Reference

Karttunen, Lauri 2006. The insufficiency of paper-and-pencil linguistics: the case of Finnish prosody. *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. 287–300.

Unifying digital phonology with an analogue brain

Joe Collins (UiT)

One frequently encountered criticism of phonological formalisms is that their rigidly discrete nature is fundamentally incompatible with the standard view of the brain as an analogue system, as well as the gradience observed in the phonetic realisation of phonological processes (Port and Leary 2005). This talk will demonstrate that this perceived incompatibility is incorrect. Discrete phonological formalisms are compatible with certain classes of gradient systems which exhibit a high degree of non-ergodicity. The information encoded in such a system will always be highly unevenly distributed, meaning that the system can be characterized at a macro-level abstraction by discrete symbols. An argument that the brain is such a system is presented in the form of a neural network *attractor* model, which demonstrates the emergence of discrete categories from an underlyingly gradient system. It will also be shown how this model can account for incomplete devoicing, which has been argued to be an example of phonetic gradience that discrete phonological models cannot explain (c.f. van Oostendorp 2008). Finally, it will be shown that the formal phonological analysis of devoicing has a higher *Effective Information* (EI), in the sense of Hoel (2017), than the neural model. Thereby demonstrating that not only are phonological formalisms compatible with a gradient view of the brain, but that they are *causally emergent* (*ibid.*), and therefore necessary if we wish to have a complete explanation of natural language grammar.

An attractor network can be defined as a type of dynamical system whose behaviour will always tend asymptotically towards one of a smaller set of networks states, which are referred to as attractors (Hopfield 1982). Their basic structure is typical of most neural networks: a number of identical units connected by “synapses” of varying efficacy. For this reason, they can be employed as an effective model of neural dynamics. One relevant property of attractor networks is that they can function as a way of storing discrete, content addressable memories, which makes them capable of representing the kinds of discrete symbols employed in formal phonological theories.

This fact is demonstrated by the implementation of a toy phonological grammar consisting of 6 possible phones – 3 places of articulation, each with a voiced and voiceless variant – and the capacity to distinguish coda and non-coda positions. The 6 phones are encoded as attractor states in the network, without any process of supervised learning, while information about syllable structure is supplied to the network as a simple inhibitory signal to the network – intended to approximate slow speed neural oscillations (c.f. Ding et al. 2016). What will be demonstrated is that the network can self-organize to voiceless memory, in the case when it is supplied with a voiced input in the context of the inhibitory coda signal. Crucially, those voiceless outputs which are derived from a voiced input, can vary fractionally from those voiceless outputs which are underlyingly voiceless. This small variation is easily interpretable as a small, but consistent, difference in the VOT of the phone during realization. In this way, this simple model is a proof of concept for how a discrete phonological system, when implemented in an underlyingly continuous system, can exhibit the sorts of gradience observed in phenomena such as incomplete devoicing.

Moreover, what can be demonstrated using Hoel’s measure of EI, is that the formal, macro-level characterization of the system is more informative regarding causal structure, than the micro-level characterization of the attractor network.

At the macro-level, the toy grammar can be understood as a system having *n=12* possible states *S*={[b]#, [d]#, [g]#, [b], [d], [g], [p]#, [t]#, [k]#, [p], [t], [k]}. The dynamics of the system can be understood as an intervention over each state s_i, at time=*t,* and a resulting effect at time=*t+1.* We can then determine two probability distributions, *Intervension Distribution* (I_D) and *Effect Distribution* (E_D), which can then be used to calculate the *effectiveness* of the system (normalized EI). Following Hoel, the I_D is considered in the maximum entropy case, where $I_D(i)=n^{-1}$. and the E_D is calculated by observing the effects of the interventions at time=*t+1* (see table). These values can then be used to determine the *degeneracy* of the system:

I_D at time=t	t+1	E_D
<do(b#)>=$\frac{1}{12}$	[p]#	<b#>=0
<do(d#)>=$\frac{1}{12}$	[t]#	<d#>=0
<do(g#)>=$\frac{1}{12}$	[k]#	<g#>=0
<do(p#)>=$\frac{1}{12}$	[p]#	<p#>=$\frac{2}{12}$
<do(t#)>=$\frac{1}{12}$	[t]#	<t#>=$\frac{2}{12}$
<do(k#)>=$\frac{1}{12}$	[k]#	<k#>=$\frac{2}{12}$
<do(b)>=$\frac{1}{12}$	[b]	<b>=$\frac{1}{12}$
<do(d)>=$\frac{1}{12}$	[d]	<d>=$\frac{1}{12}$
<do(g)>=$\frac{1}{12}$	[g]	<g>=$\frac{1}{12}$
<do(p)>=$\frac{1}{12}$	[p]	<p>=$\frac{1}{12}$
<do(t)>=$\frac{1}{12}$	[t]	<t>=$\frac{1}{12}$
<do(k)>=$\frac{1}{12}$	[k]	<k>=$\frac{1}{12}$

$$degeneracy = \frac{D_{KL}(E_D|I_D)}{log_2(n)} = log_n(2)\sum_i E_D(i)log_2\frac{E_D(i)}{I_D(i)}$$

Which will then allow us to calculate the $effectiveness = [determinism] - degeneracy$. Since our toy grammar is strictly deterministic, the *determinism* is equal to 1. Crunching the numbers gives our toy grammar $eff = \sim 0.93$

Now we can turn to the case of our attractor model. Because the storage capacity of an attractor network can be well defined, relative to the size of the network (i.e. the number units), we can quantify exactly the total number of states in the system. This allows us, in principle, to calculate the lower bound for *degeneracy* in an attractor implementation of our toy grammar. In practice, the state-space of our attractor is so large that it precludes a brute force computation over each and every state. A trinary node implementation of 12 attractors would require 50 units (Ziong and Zhao 2010), giving a system size of $n=3^{50}$. However, demonstrating the high *degeneracy* can be accomplished by averaging over the system. Since each state which is not an attractor will ultimately lead to an attractor, each of the 12 attractor states can be reached by an average of $\frac{3^{50}}{12}$ states, the average value of $\frac{E_D(i)}{I_D(i)} \cong \frac{12^{-1}}{3^{-50}}$, which will cause the *degeneracy* to tend heavily towards 1, resulting in a very low *effectiveness*. Therefore, even when our discrete phonological representations are taken as emergent phonemena from an underlyingly gradient system, such as an attractor network, it is in fact the phonological model which has the highest *effectiveness*, rather than the neurological model.

Finally, it can also be demonstrated that, if the micro-level characterization is a discrete system of the sort repudiated by Port & Leary, then *causal emergence* does not occur. This suggests that formal phonological models are, counter intuitively, **more** valuable in the case where discrete symbols are emergent rather than primitive (pace Port & Leary).

Ding, R., Melloni, L., Zhang, H., Tian, X., Poeppel, D. (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19, 158–164

Hoel, E. P. (2017) When the Map Is Better Than the Territory. *Entropy* 19:188.

van Oostendorp, M. (2008) Incomplete devoicing in formal phonology, In Lingua, Volume 118, Issue 9, Pages 1362-1374

Port and Leary (2005) Against formal phonology. *Language*, 81 pp. 927-964

Xiong, D., Zhao, H. (2010) Estimates of storage capacity in the q-state Potts-glass neural network. *J. Phys. A: Math. Theor.* 43 445001

Formalizing contrast and redundancy in phonological representations
Daniel Currie Hall, Saint Mary's University

The motivation for a formal theory of representations As discussed in depth by Anderson (1985), theories of phonology have divided their attention in various ways between phonological representations and the rules (or, more recently, constraints) that operate on them. Any fully formalized theory necessarily encompasses both—but, as noted in the call for papers for this workshop, it is not always practical or useful to attempt to formalize all aspects of a theory at once. This does not mean that those aspects that are not made explicit remain unconstrained. In particular, a formal theory of representations will restrict the kinds of things that rules and constraints can do. To take a colourful example from Hale & Reiss (2008), a theory in which bananas are not licit linguistic objects will have no need of a NoBanana constraint, nor can it have rules that insert, delete, or slice bananas. More generally, and more realistically, theories that restrict the information content of representations thereby also restrict the power of rules and constraints: the phonological computation can only work with what it is given. There is thus a clear methodological reason to pursue parsimonious theories of representations, in that they can easily be falsified by the discovery that the rule system needs access to more information than they provide.

Application to contrast It has often been observed that contrastive features—ones that serve to mark phonemic distinctions—appear to have some kind of special status in phonology; specifically, there are at least some phonological patterns that refer to contrastive features but ignore redundant ones. Broadly speaking, theories of representations have responded to this observation in two kinds of ways: either by positing that redundant features are unavailable to some or all of the phonological computation (e.g., Archangeli 1988; Dresher 2009; Mackenzie 2013), or by positing that both kinds of features are phonologically visible, but that the computation is able to distinguish between them (e.g., Calabrese 1995; Halle et al. 2000; Nevins 2010). In other words, the special status of contrastive features can be encoded either by subtracting information from phonological representations (excluding redundant features), or by adding information (marking specifications as contrastive or redundant). In the additive approach, one might posit that contrastive feature specifications have a special 'colour' (akin to the colours used by van Oostendorp 2007 to mark morphological affiliation) that makes them visible to a superset of the rules that can see redundant features.

A case in point: Uyghur vowels Halle et al. (2000) take an additive approach to the formalization of contrast, but their account of Uyghur vowel harmony is striking in that it suggests that a feature cannot be coloured contrastive once and for all; rather, they claim, its status must be re-assessed during the course of the derivation.

Uyghur has the vowel inventory shown in Table 1, which is also that of Finnish. As in Finnish, the vowels /i/ and /e/, which have no minimally different back counterparts, are transparent to vowel place harmony. Harmony spreads [±back] rightward to alternating suffixes such as the plural *-lær/-lar,* as in (1) and (2).

	FRONT		BACK	
	UNRND	ROUND	UNRND	ROUND
HIGH	i	y		u
MID	e	ø		o
LOW	æ		ɑ	

Table 1: Vowel inventory of Uyghur

(1) a. [jyz] [jyz-lær] ‘face(s)’
b. [køl] [køl-lær] ‘lake(s)’
c. [xæt] [xæt-lær] ‘letter(s)’

(2) a. [pul] [pul-lɑr] ‘money’/‘monies’
b. [jol] [jol-lɑr] ‘road(s)’
c. [ɑt] [ɑt-lɑr] ‘horse(s)’

The transparency of /i/ is illustrated in (3).

(3) a. [køl-imiz-gæ] ‘lake-our-DATIVE’
b. [jol-imiz-ʁɑ] ‘road-our-DATIVE’

There are also non-alternating suffixes such as *-ʧæ,* which not only remains [−back] after [+back] stems, but can also transmit [−back] to a subsequent suffix:

(4) a. [tyrk-ʧæ] ‘(in the) Turkish (manner/language)’
b. [ujʁur-ʧæ] ‘(in the) Uyghur (manner/language)’
c. [kitɑp-ʧæ] ‘booklet’
d. [kitɑp-ʧæ-m-dæ] ‘in my booklet’

Low vowels in medial open syllables are raised to [i], and strikingly, this causes them to become transparent to harmony:

(5) a. [bɑlɑ] ‘child’ [bɑli-lɑr] ‘children’
b. [iʃæk] ‘donkey’ [iʃiɣ-i] ‘his/her/its donkey’
c. [næj-ʧi-dæ] ‘child+*ʧæ*+LOCATIVE’
d. [kitɑp-ʧi-dɑ] ‘book+*ʧæ*+LOCATIVE’

In Halle et al.’s (2000) account, all features are specified, but harmony spreads, and can be blocked by, only contrastive values of [±back]. In their account, then, the transparency of an [i] derived from /æ/ means that its [−back] specification must become non-contrastive as soon as it becomes high.

However, an alternative account is possible within the more restrictive information-subtracting approach to contrast. Suppose that segments are assigned only contrastive features as designated by a contrastive hierarchy (Dresher 2009). A partial such hierarchy for Uyghur vowels is shown in Fig. 1. Underlying /e i/ are transparent to harmony because they have no value for [±back]. The process that changes low vowels to [i] is not merely raising, but reduction, both in the sense that it involves a decrease in sonority and in the sense that it involves the deletion of marked structure. Note that this process, as shown in (5), neutralizes the place contrast between underlying /æ/ and /ɑ/. In the underspecification account, the neutralization and concomitant harmonic neutrality are neatly captured by saying that reduction involves deletion of [±back], rather than changing the status of the feature from contrastive to redundant (and, in the case of /ɑ/, its value from + to −).

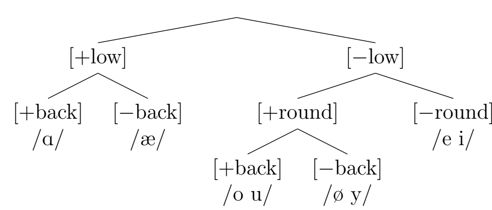


Figure 1: Partial contrastive hierarchy for Uyghur vowels

Formalizing the difference between contrastive and redundant features by saying that the latter are simply absent from phonological representations is both conceptually more elegant and methodologically more useful than formalizing it by painting the two types of features different colours. In Uyghur, it also yields a more satisfactory account of the interaction between reduction and harmony.

References

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations.* Chicago: University of Chicago Press.

Archangeli, Diana. 1988. Underspecification in phonology. *Phonology* 5.2: 183–207.

Calabrese, Andrea. 1995. A constraint-based theory of phonological markedness and simplification procedures. *Linguistic Inquiry* 26.3: 373–463.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Hale, Mark & Charles Reiss. 2008. *The phonological enterprise.* Oxford: Oxford University Press.

Halle, Morris, Bert Vaux & Andrew Wolf. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31.3: 387–444.

Mackenzie, Sara. 2013. Laryngeal co-occurrence restrictions in Aymara: Contrastive representations and constraint interaction. *Phonology* 30.2: 297–345.

Nevins, Andrew Ira. 2010. *Locality in vowel harmony.* Cambridge, Mass.: MIT Press.

van Oostendorp, Marc. 2007. Derived environment effects and consistency of exponence. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.), *Freedom of analysis?*, 123–148. Berlin: Mouton de Gruyter.

One-Relation Representation for a Simpler-Than-Strings Phonology

Maxime Papillon - University of Maryland

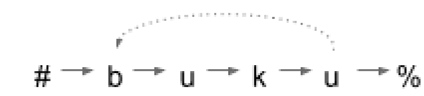
This presentation aims to argue in favor of a phonological representation derived from that of Raimy (2000) by exploring the additional power conferred by it. In this respect this presentation is in the lines of of Reiss's (2012) bottom-up approach to phonological typology.

One goal is reducing the complexity of the representation by going from strings to directed graphs. Raimy (2000) suggested to replace strings such as (1) with graphs like (2), where # and % are explicit beginning- and end-of-word markers respectively. The difference being that in such a representation single segments can immediately precede or follow multiple others. This allowed him to analyse reduplication as "loops" in the directed graph. Indonesian plural reduplication as in (3) is linearized as (4). This structure then allowed Raimy to tame a number of complicated patterns of over- and under-application of phonological rules interacting with reduplication and infixes

(1) kæt

(2) # → k → æ → t → %

(3)



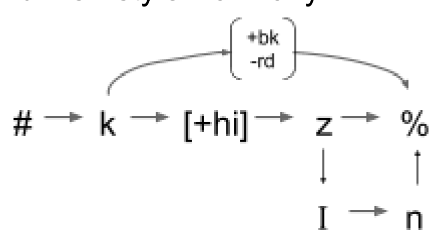
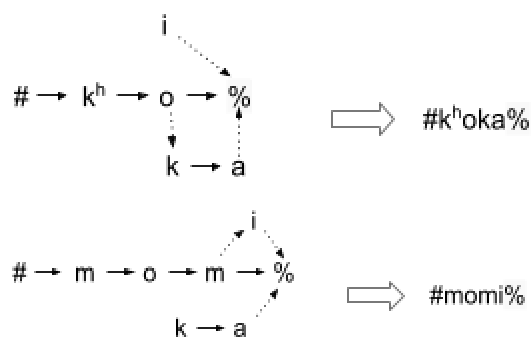
(4) #bukubuku% "books" (Indonesian)

It is important to distinguish a) the fact that directed graphs like (2) are simpler than strings like (1) in the way mentioned above from b) the graphical conventions of our writing system that allows us to easily represent strings as left-to-right sequences of symbols, and c) the richer set of structures allowed in graphs than than in strings. a) is a mathematical fact about the properties of the data structure. Strings are graphs with a total order relation, namely the relation is total, transitive, asymmetric, and irreflexive. A graph without these properties is a simpler mathematical object with fewer assumptions. And with Autosegmental Phonology, feature geometry, and prosodic structures, and correspondence theory almost everyone is already doing phonology on graphs that contain multiple types of relations, obviously a higher amount of complexity than the one-relation graphs I am arguing for here.

It is true that thanks to b) we can represent strings in fewer symbols than graphs as in (1), but that's because our writing system implicitly packages all these assumptions, representing linear order in time as linear left-to-rightness on a the page. Let's not allow this purely pictorial fact of writing influence your impression about mathematical complexity. Strings do require a logical relation of precedence between pairs of elements as well as a notion of start and end, the graph notation is simply actually writing those down. The presence of arrows and special START and END symbols in (Aa) is explicitness, not complexity.

And as said in c) it is true that we can do more with generic graphs than with strings (since all strings are graphs but not vice versa). As seen above graphs can contain branchings, loops, and parallel paths that strings cannot represent. But this is representational power, not complexity. My claim is that talking about generic graphs rather than strings is making fewer assumptions about what phonology is and therefore it is a simpler claim. But one would be correct to think that that it is powerful. Simpler things often offer more possibilities. The point of this paper is to argue that this power matches very well with all the less string-like phenomena of phonology and morphology that are attested. This simpler-than-string phonology is powerful, and it seems to contain just the right power to do morphology.

In addition to reduplication and infixation which were already explored by Raimy (2000), other phenomena that turn out to immediately follow from the additional power of this representation include at least tonal and harmony phenomena (5), allomorphy (6), ineffability and morphological unproductivity, and intonational phenomena.

<table>
<tr>
<td>(5)</td>
<td>Turkish-style Harmony

</td>
<td>(6)</td>
<td>Korean phonologically-conditioned allomorphy

</td>
</tr>
</table>

Reiss (2012) argued that in the realm of segment inventories, the best option is to start from the simplest model -unrestricted feature sets- and crucially to accept it in all its combinatorics power. There can be many reasons for a possible pattern to be unattested, ranging from biases of language change, to the small sample size of studied languages, other than strict limitations of our grammar. Tuning our theories to all and only the attested languages will inevitably lead to overfitting, and phonological theory must therefore care more about the simplicity of the system than about how tightly it predicts all and only what is attested.

A phonological representation based on graphs is simpler than one based on strings, and its space of possibilities is not a cause for concern. To the contrary it will lead to simpler explanation of morphological phenomena.

References

Raimy, E. (2000). *The phonology and morphology of reduplication* (Vol. 52). Walter de Gruyter.

Reiss, C. (2012). Towards a bottom-up approach to phonological typology. *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces. John Benjamins Publishing Company, Amsterdam*, 169-191.