A theory of the theory of vowels

Edoardo Cavirani & Marc van Oostendorp

Draft. May 19 2017.

Abstract

We represent the most common vowel contrasts in a theory that allows only (recursive) embedding of sets, including the empty set. Such a theory needs neither features nor elements. We show that from such a theory we can actually derive some common properties of the element set |A, I, U|: why are there only three of them? And why does |A| behave differently from the other two? Furthermore, the theory also gives a natural place to both schwa and the completely empty nucleus. We also show how this theory is related to some earlier proposals in the literature.

Sets usually play some role in the representation of phonological segments, since these are seen as sets of elements or features (Bale et al. submitted). Feature geometry theories can be seen as a refinement of this: a segment is a set which consists of a subset of place features, a subset of aperture features, etc. Especially within Element Theory, there has been a tendency to furthermore reduce the number of representational primitives. A radical example of this is so-called Radical CV Phonology by Van der Hulst (1988, 1994, 1996, 2015), which claims that there are only two such primitives, called C and V, which can be interpreted differently in different parts of the tree.

Where would we go if we would reduce the number of primitives even further? And what would the ultimate reduction be? Would there be a possibility of giving a representation for segments that would not include *any* elements? We know that set theory is a rather strong mathematical tool – so strong that there have been hopes of founding all of mathematics on it. Those projects failed, but could we not achieve the much less ambitious project of building a theory of vowels on them? We think we can, and the following can be seen as a complete definition of vowel representations in natural language motivated by the *desiderata* above. It is the purpose of this paper to sketch what such a theory would look like.

1. Vowels without features of elements. A proposal

We first define the vowel recursively:

(1) a. The empty set ({}, also written as ∅) is a vowel.
b. If σ is vowel, {σ} is also a vowel.
c. If σ, τ are vowels, and {σ, τ} is balanced, {σ, τ} is also a vowel.

This definition relies on a definition of 'balance' that is familiar from search algorithms, but which we will ignore for now. The definition in (1) gives us in principle an infinite number of V-sets, of which the following are some simple examples:

(2) a. ∅ (per (1a))

b.	{∅}	((1b) applied once to (1a))
ci.	{ ∅, ∅ }, or	((1c) applied once to two instances of (1a))
cii.	{ {∅}, ∅ }, or	((1c) applied once to (1b) and (1a)
ciii.	{ {∅}, {∅} }	((1c) applied once to two instances of (1b))

This thus means that we claim that all representations in (2) are vowels. We think it is easier to represent them as treelets. The following are thus exactly equivalent to the ones in (2):

```
(3)   a.      ∅

      b.      |
              ∅

      ci.    / \
            ∅   ∅

      cii.   / \
            |   ∅
            ∅

      ciii.  / \
            |   |
            ∅   ∅
```

Note that linear order is irrelevant, so that (cii) could also be represented as:

```
(4)      / \
        ∅   |
            ∅
```

Further trees can be formed by replacing {∅} in any of these treelets by another treelet, for instance:

```
(5)    / \    +   / \    >   / \
      |   |      ∅   ∅      |  / \
      ∅   ∅                 ∅ ∅   ∅
```

Since the original definitions are recursive, we can go on doing this forever (as long as we abide by the balance requirement in (1c)? which is actually a condition imposed on ‘free’ recursion). The treelets in (3) have a special status: they are as it were the most primitive structures that exist in the theory, as they are the only ones which can be formed by applying the definitions in (1) at most once. The result of the operation in (5) is more complicated.

We propose that each of these ‘primitive’ treelets has a simple phonetic interpretation, that is reminiscent of classical Element Theory. Except for the empty set which has no phonetic interpretation (and thus counts as an ‘empty nucleus’[1]), each simplex treelet in (3) is like an Element:

```
(6)      |     corresponds to |@|
         ∅

           / \     corresponds to |U|
         ∅  ∅

          /  \     corresponds to |I|
         ∅   |
             ∅

           / \   corresponds to |A|
          |   |
          ∅ ∅
```

Logic tells us that, since linear order is irrelevant (4), there can be only tree binary treelets, corresponding exactly to the typical number of ‘normal’ elements, and one unary treelet for the schwa interpretation. In section 1.2, we will go into the question why we think each individual treelet has the interpretation it does.

When we combine treelets (per (1ciii)), the resulting tree is again a representation of a vowel, with an interpretation that is familiar from Element Theory, or to be more precise, of Particle Phonology (Schane 1984), because there is no headedness in our representations, and the same treelet can occur more than once. This is true in particular for the |A| treelet. E.g. in (5) above we see |A| combined with |U|, so that the resulting tree is a representation for /o/, and the following would be a representation for /ɔ/:

```
(7)       / \      +     / \                  >   / \
         |   |           |  / \                   |  / \
         ∅ ∅          ∅   ∅∅                 ∅  |   / \
                                                   ∅ ∅  ∅

         |A|              /o/                         /ɔ/
```

One does not have to have the ‘derivational history’ of a V-set to understand its phonetics; by reading a tree one layer at a time, one also gets the desired interpretation:

```
(8)     /  \                          |A|
       |  / \                     @      |A|
      ∅ |  / \                        @     |U|
         ∅ ∅  ∅
```

The top layer is a binary treelet with two non-empty daughters; this is interpreted as an |A| element. The second layer consists of a unary set (a schwa) plus another binary treelet with two

[1] Note that this implies an ontological difference between the segmental and the prosodic planes,

non-empty daughters, so this is again an |A| element, of which again one of the daughters is a schwa; the other is a treelet with two empty daughters, which is an |U|. This representation can thus be understood as |A| . @ . |A| . @ . |U|, or (given that @ is an element that does not add any value other than background noise; Harris 1994) as |A|.|A|.|U|, which is an /ɔ/ in Particle Phonology.

1.1. Balance and the limits of recursion

The |A| treelet allows for recursion in this way – limited, we propose, only by extralinguistic factors such as the ability of humans to distinguish vowel heights from each other articulatorily or acoustically. Note that the |U| treelet does not allow any kind of further embedding, because each of its daughters should be empty. |I| does allow for embedding of |U|, giving us /y/:

```
(9)                 / \                |I|
                   ∅  / \                 |U|
                      ∅ ∅
```

We propose that other types embeddings are not allowed, because of what we call the 'balance property' of phonological representations. This property is most easily defined informally on trees, but an equivalent (and more formal) definition could of course be given for sets:

(10) The number of embeddings N(T) in a tree T, is the number of steps it takes to go from the root of T to the most deeply embedded leaf.

(11) A binary tree {A, B} is **balanced**, if N(A) – N(B) ≤ |1| (Adelson-Velsky & Landis 1962)

If we embed something else than a @ treelet or a |U| treelet into an |I| treelet, we get a tree that is not balanced. This for instance would be the result of putting an |I| inside another |I|:

```
(12)                / \                |I|
                   ∅ / \                  |I|
                     ∅  |                   @
                        ∅
```

The first daughter has a number of embedding of 0 (because it is an empty set), the second daughter (the embedded |I|) has a number of embedding of 2, so the difference between the two nodes becomes too big to be balanced. The embedding of |U| in |I| is balanced, however (see (9)): the first daughter of |I| still has an embedding of 0, but the second one (the |U|) has an embedding of 1. This difference is within the limits put forward by (11).

Note that this implies that |A| can embed twice into itself, because each daughter is non-empty, and has therefore a number of embedding of at least 1. The representation of /ɔ/ above (7) is therefore balanced. We cannot embed three |A|'s into each other in this way, however, as then the top 'simple' (@) daughter becomes too simple (note that this also derives the (nearly absolute; Crothers 1978) maximal 4-degrees height of vowel systems). We can do more embedding, but in that case each of the daughters needs more internal structure.

Note that we have tacitly assumed so far that schwa formation is not recursive: we cannot embed {∅} into another single set ({{∅}}, {{{∅}}}, etc. are not feasible representations). We assume that the reason for this is that such structure is redundant: a schwa embedded in a schwa would still be phonetically interpreted as a schwa. But for this reason, schwa embedding is also not allowed to 'balance' treelets that would otherwise be unbalanced.

1.2 Some examples of vowel sets

Using the kinds of representations just outlined, we can now define a number of well known vowel sets. For instance a three-vowel set [i, u, a] has the following elements (we use familiar square brackets to denote vowel sets, since ontologically such 'sets' are obviously of a different type than individual segments):

(13)	a. [{∅, ∅},	([u])
	b. {∅, {∅}},	([i])
	c. {{∅}, {∅}}]	([a])

We can call (13a) the |U| set, (13b) the |I| set and (13c) the |A| set. Such a vowel inventory can be described in the following way:

(14) a. All vowels are binary sets.
b. The sets that are members of vowels sets have at most one member.
c. All sets are balanced.

In a four language vowel set that also includes schwa, the requirement in (a) is canceled. If we instead cancel (13b) from (14) and add the assumption of (15), we get a typical five vowel set (assuming that an |A| set that dominates another |A| set is still also an [a] because of 'redundancy'):

(15) Vowel sets have an embedding depth of at most 2.

(16)	[{∅, ∅},	([u])
	{∅, {∅}},	([i])
	{{∅}, {∅}},	([a])
	{{∅}, {∅, ∅}},	([o])
	{{∅}, {∅, {∅}}}	([e])
	{{∅}, {{∅},{∅}}}]	([a])

Extending the embedding depth to level 3 instead of level 2, we of course also extend the vowel set even further:

(17)	[{∅, ∅},	([u])
	{∅, {∅}},	([i])
	{{∅}, {∅}},	([a])
	{{∅}, {∅, ∅}},	([o])
	{{∅}, {∅, {∅}}}	([e])
	{{∅}, {{∅},{∅}}}	([a])
	{{∅}, {{∅},{∅, ∅}}}	([ɔ])

{{∅}, {{∅},{∅, {∅}}}}   ([ɛ])
{{∅}, {{∅},{{∅}, {∅}}}}]   ([a])

We can understand lowering processes by embedding a vowel set in an |A| set. Removing the ‘balanced’ requirement in (14c) gives us inventories with front rounded vowels. In other words, the calculus for at least the most common vowel inventory types can be described with a small set of possible restrictions on sets. Notice that the view on vowel structure which we thus get is not incompatible with autosegmental views of frontness or roundness harmony: these sets can behave as autosegmental elements and spread. Lowering harmony would need to have a different representation, on the other hand.

1.3 Extensions to larger segmental inventories

Obviously, segmental inventories do not just consist of vowels – and even within vocalic phonology there are many distinctions we have not made yet – we will need representations for nasality, for tone, and many other distinctions. Space does not permit to go into these details, but note that we can make use of the kinds of ideas developed in other Element-based frameworks (such as RcvP; see below) in which the same elements can have different phonetic interpretations, depending on their position in the tree. Making segmental trees bigger (having ever more recursion) will expand our space of possibilities.

2 Substance reduction and set theory: some precedents

2.1 Early precursors

Ours is not the first proposal for reducing the number of representational primitives. As hinted at in section 1, several proposals have been put forward with a similar aim. Since our work represents a rather extreme move along the same line of research, we think it is important to recapitulate the most relevant stages of this research line. More importantly, this will also give us the room to stress the differences and similarities between our proposal and the preceding ones. One of the first of this type of measures to reduce the number of representational primitives is Feature Geometry (Clements 1985), which (implicitly and informally) applies the notion of a set to the unordered bundles of features of Chomsky and Halle (1968). As a result, segments become sets of subsets of features, which, thus, are formally conceived of as organized in groups headed by nodes in a (segmental) tree[2]. Crucially, the geometric restructuring of the featural content of segments allows for generalizations which target subsets, i.e. representational nodes[3]. As a matter of fact, this framework reduces the computational components of phonology (e.g. both the

[2] According to Clements (1985: 230), this embodies the view according to which “the varying degrees of independence among phonetic features can be expressed by a hierarchical grouping such that higher-branching categories tend to be more independent than low-branching categories. More exactly, the relative independence of any two features of feature classes is correlated with the number of nodes that separate them”.
[3] More recently, Bale et al. (submitted: 1) resort to set theory in a more explicit fashion: “taking […] feature bundles to be sets [and natural classes sets of sets] allows us to apply ideas from set theory to phonology”. This allows them to propose the reconceptualization of a fully underspecified segment as empty set, which, in turn, “can be used to define a natural class over all segments”.

structural description and the structural change of a given rule can now just refer to the relevant parent node), rather than the representational one. Even if representations are still as rich as they were before, though, with Feature Geometry, trees enter the subsegmental scene.
As a matter of fact, trees were already on the phonological market since a few years (see section 2.5 below). Indeed, assuming the Structural Analogy Hypothesis, whereby both morphosyntactic and phonological structures are represented as dependency relations holding between representational primitives, Anderson & Jones (1974) developed Dependency Phonology (henceforth DP; see also Anderson 1985, 1992; Anderson & Ewen 1987, van der Hulst 2006 and van der Hulst 2011). Within such a model, the organization of features essentially parallels the one proposed by Clements (1985), with major nodes corresponding to laryngeal, manner and place categories. Differently from Feature Geometry, the representational primitives co-occurring under the relevant nodes are arranged according to a variable head-dependent schema: given two features α and β, the relationship they enter into can be either α- or β-headed, each corresponding to a (potentially contrastive) phonological expression. As a matter of fact, dependency relations are suggested to hold also between nodes and sub-nodes. However, no restrictive theory constraining the various combinatorial possibilities has been developed, resulting in overgeneration; see van der Hulst 2006 and 2011 for a brief discussion.
One difference between Feature Geometry and DP concerns the representational primitives, which are binary in the former case and unary in the latter. Furthermore, the primitives proposed by DP are "(in an Aristotelian sense) 'substances' in themselves rather than properties of substances'. Whereas mainstream binary features are arguably properties of segments, DP-primes are segments themselves. Indeed such primes can occur independently as fully pronounceable phonological segments" (van der Hulst 2006: 455). Traditionally, these primes have been referred to as components, their primary phonetic interpretation being acoustic (e.g. "|V|, a component which can be defined as 'relatively periodic', and |C|, a component of 'periodic energy reduction'" (Anderson and Ewen 1987: 151)).
Many DP proposals were further elaborated within e.g. Radical CV Phonology (henceforth RcvP; van der Hulst 1988, 1994, 1996, 2015) and Government Phonology (henceforth GP; Kaye *et al* 1990; Charette 1991; Lowenstamm 1996 and Scheer 2004). Both RcvP and GP maintain a similar conception of primes, which are unary, 'substantial' and combinable in head-dependent structures. RcvP and GP, though, attempt to solve the overgeneration problem DP suffered because of the lack of a constrained theory of primes (and their combinatorial possibilities). These frameworks proposed two different solutions, which are briefly described in what follows.

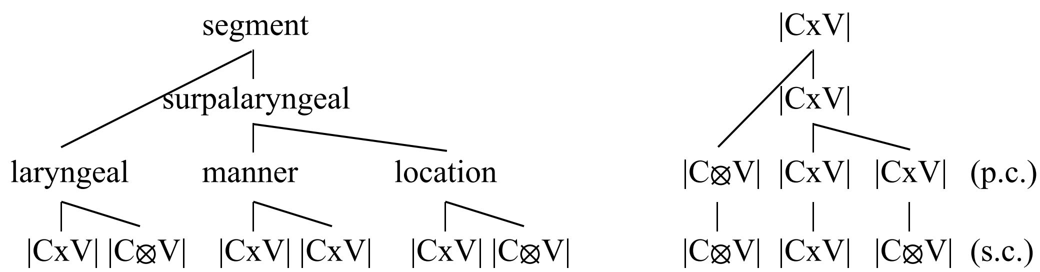
2.2 RcvP

In order to limit the generative power of the system developed within DP, RcvP capitalizes on a suggestion already present in Anderson & Ewen (1987), according to which a given component can occur under different nodes of the segmental tree. This is the case, for instance, for the |i| and |u| components, which are interpreted as high and low tone, respectively, when occurring under the tonological node (Anderson & Ewen 1987: 273)[4]. The possibility for the same component to occur in various structural positions, in turn, allows for the formalization of similarities (same component) and differences (different structural position) among (the phonetic interpretation of)

[4] Interestingly, Anderson & Ewan (1987: 215) argues for the identity of |a| and |V|. As we will see below, this alleged identity is in line with our proposal, as well as with those proposed e.g. by Rennison (1998) within the GP camp.

segments. Together with the head-dependent asymmetry DP shares with Feature Geometry, this possibility allows for a further reduction of the number of components: a contrast previously formalized by e.g. two features can now be conveyed by one and the same component occurring in a head or dependent guise, or in different structural positions. For instance, the |V| component can translate [sonorant] and [voice] depending on its head *vs* dependent status, or it can identify sonorants, vowels, [low] and [open place] depending on its structural position.
RcvP exploits these possibilities to their maximal extent by constraining the typology of structures to head-dependent configurations of just two primes: |C| and |V| (whereas DP resorted to |C|, |V|, |O| |G|, |K|, |i|, |u|, |a|, |@|, |α|, |l|, |t|, |d|, |r|, |L| and |n|), which are organized in an arboreal structure such as the following, where '|CxV|' means that |C| and |V| can combine, '|C⊗V|' that they cannot, and DP gestural labeled nodes (on the left; Clements 1985) are "defined in purely structural terms" (on the right), 'p.c.' and 's.c.' indicating, respectively, the primary and secondary component; van der Hulst 2017):

(18) RcvP translation of DP segmental tree



Notice that, even though RcvP, like DP, assumes that components have a default (acoustic) phonetic interpretation, |C| and |V| are assigned specific interpretations depending on their a) syllabic position (onset head *vs* onset dependent *vs* rhyme head *vs* rhyme dependent), b) class (manner *vs* location *vs* laryngeal), c) component status (primary *vs* secondary) and d) element status (head *vs* dependent). RcvP therefore comes as close to completely reducing the role of differences between elements as one can get without really abandoning the whole concept of elements completely.

2.3 GP & GP 2.0

As mentioned above, DP-style representational primitives are kept in both RcvP and GP, inasmuch as they all resort to unary primes which are acoustically grounded and combinable in head-dependent structure. In the GP literature, these primes are known as elements. Clearly rooted in DP (and Particle Phonology; Schane 1984), elements are introduced by Kaye *et al* (1985) and further developed along various directions, which differ in the element number and/or in the way elements can combine (see Backley 2012 for an overview and a brief discussion of the variants on the market).
In its standard form (Backley 2011), there are six elements, which are extensionally equivalent to the objects defined within RcvP by means of elements and (unlabeled) gestural nodes:

(19) GP *vs* RcvP (van der Hulst 2017)

GP elements	
|ʔ|	|A|
|I|	|U|
|H|	|L|

RcvP elements		
Aperture	C	V
Location	C	V
Laryngeal (Phonation/Tone)	C	V

As discussed in the preceding section, the resort to an arboreal structure enriched with gestural nodes allows RcvP to shrink the number of elements to a binary set. Notice, however, that the gestural nodes constitute a sort of representational primitives themselves, even if of a different nature than components/elements. As a consequence, GP and RcvP display the same number of elements, the difference between the two theories consisting mostly in the presence *vs* absence, within the representational toolbox, of the gestural nodes. In other words, GP and RcvP differ in the relative balance between structure and substance: whereas GP decides to minimize structure and maximize substance (*viz* many elements on small trees), RcvP gets rid of most substance by maximally exploiting the structural dimension (*viz* a few elements on bigger trees).

Among the directions GP evolved into, variants can be found that try to reduce substance in a similar fashion[5]. One of the targets of Occam's razor is the |A| element, which is repeatedly shown to behave differently from other vocalic/resonance elements such as |I| and |U|. For instance, |A| is argued to be more syllabic than |I| and |U|, thereby showing a preference for occupying the head position of nuclei while avoiding the nuclear dependent position[6]. Furthermore, |A| is shown to interact with nasalization and length[7].

As discussed above, the price to pay for substance reduction is structure enrichment. As a consequence, |A| is replaced by structure. For instance, in order to formalize the preference for |A| syllabicity, Rennison (1998) proposes to associate the phonetic counterpart previously related to |A| (i.e. a centrally converging F1–F2 acoustic pattern) with the presence of a nuclear position lacking any elemental specification (whose unmarked status is thus representationally encoded; see Cavirani & van Oostendorp *in press* for a similar proposal).

An even more extreme development of GP towards substance reduction is represented by what came to be known as GP 2.0 (Pöchtrager 2006; Živanović & Pöchtrager 2010; Kaye & Pöchtrager 2013; Schwartz 2010), which eliminates |H|, |ʔ| and |A| by resorting to structures and

[5] While discussing the variants which resort to more elaborate arboreal structure to get rid of elements, Backley (2012: 75) warns that the standard theory "manages to strike a useful balance between the two, providing a restrictive model of phonological knowledge in which elements are abstract enough to function as cognitive units of linguistic structure yet concrete enough to be realized phonetically without the need for explicit rules of phonetic interpretation".

[6] As discussed in Backley (2012), this could be the reason why, for instance, diphthongs such as [ai] and [au] are typologically less marked than [ia] and [ua], which is in turn possibly related to the fact that only the latter diphthongs are reinterpreted as glide-vowel sequences. This shows that, whereas |I| and |U| may be (re)interpreted as belonging to the onset preceding the |A| nucleus, the same does not hold for |A|, which keep on projecting to its nuclear node.

[7] For instance in French, where only |A| nuclei can be lengthened and nasalized (Ploch 1995).

mechanisms inspired by syntactic analogues, such as control[8], m-command[9] and head-adjunction. Let's focus now on |A| (referring the reader to Kaye & Pöchtrager 2013 for |H| and |ʔ|).

As just mentioned, |A| is argued to display a special interaction with length: "more specifically, |A| seemed to make bigger structure possible" (Pöchtrager 2015: 261). As a matter of fact, what is traditionally referred to as |A| is formalized as pure structure, where the extra structure is guaranteed via head adjunction: "in the case of head adjunction, the head xN projects to another level but remains the same type, i.e. an xN" (Pöchtrager 2015: 261). This is shown in (20), where the arrow between xN and its sister (in [a]) represents control:

(20) GP 2.0 vocalic elements

```
 [ɪ]     [ʊ]     [ɨ]          [ə]        [a]
 xN      xN      xN           xN         xN
  |       |                   / \        / \
 |I|     |U|                xN   x     xN→x

      no head adjunction          head adjunction
```

As shown in (20), both [ə] and [a] are represented as pure structure. The only difference is the presence in the latter of control, which is thus deemed the responsible for the [a]-interpretation of such an empty structure, otherwise sounding [ə]. Notice that "the control relationship also expresses that within [a] both positions are used up, while in [ə] there is one position (the non-head) available. In some sense, [ə] takes up less room than [a]. This neatly capture Lowenstamm's (1996) observation that [ə] is the shorter version of [a][10]" (Pöchtrager 2015: 261). Furthermore, the lack of control in [ə] is considered to be the reason why it can be coloured by adjacent melody: the absence of control leaves "one position […] available", which can thus host elements spreading from adjacent structures (e.g. in the analysis of Putonghua proposed in Živanovič & Pöchtrager 2010).

2.5 Comparison

As shown in the preceding sections, a research line can be identified within the DP-inspired tradition which aims at reducing substance by exploiting structure. This is particularly evident in the case of RcvP, which attempts at minimizing substance (only |C| and |V| are left) by enriching the structural dimension, whereas GP minimizes structure and maximizes substance (|ʔ|, |H|, |L|,

[8] In Pöchtrager (2006:77) control is described as, "[an] unannotated x in a non-maximal onset projection must be controlled by its xO [*viz* the onset head]." In GP2.0 control is generalised to structures occurring in nuclear projections. Its general effect is that of making the controlled point inaccessible.

[9] Živanović & Pöchtrager (2010) define m-command as a sort of licensing necessary for phonetic interpretation, whereby terminals, i.e. elements or empty structural position, can be interpreted only if m-commanded. In the case of an empty structural position, m-command has the same effect of spreading, the commanded receiving the same phonetic interpretation of the commander.

[10] Notice, also, that control is somehow analogous to standard GP headedness, whereby controllers head controlees. See Cavirani & van Oostendorp (in press) for a slightly different proposal on the structural relatedness of [a], [ə] and empty nuclei.

|A|, |I| and |U|). Even within the latter, though, variants have been proposed that prefer to pay a little structural price to get rid of elements (which reduce to |I|, |U| and |L|).
Focusing on the vocalic half of the phonological world, we try to go even further by resorting to set theory to eliminate all the substantial content. More precisely, following the path initiated by DP and RcvP, we resort to the possibility for a given primitive to occur in different structural positions, the difference laying in the fact that, instead of components, we replace ("Aristotelian") substance with pure structure, namely with sets which are recursively built on top of the empty set.
This move echoes the attempts we mentioned above to reduce e.g. |a| to |V| (Anderson & Ewan 1987: 215) or |A| to empty nuclear positions (Rennison 1998). In a similar fashion, we propose a representational account of markedness whereby, differently from Rennison (1998), [ə] is represented as the set containing the empty set (see also Cavirani & van Oostendorp *in press* for a similar proposal).
Furthermore, assuming that, as proposed e.g. by Lowenstamm (1996), Pöchtrager (2006) and Živanovič & Pöchtrager (2010), [ə] is the shorter version of [a] and that "|A| seemed to make bigger structure possible" (Pöchtrager 2015: 261), we represent [a] as the set that contains two sets containing an empty set. In prose, this entails that [a] is consists of two other vowels (typically [ə]). This allow us to get rid of the control mechanism introduced by Pöchtrager (2006; 2015) and Živanovič & Pöchtrager (2010) the account for the difference between two sounds, [a] and [ə], that are otherwise represented in an identical fashion (see (20) above). We keep, though, something similar to head-adjunction (even if, as a matter of fact, we do not need to make any head-dependent distinction).
Together with control, we can also get rid of the c-command solution proposed by Pöchtrager (2015) to solve the problems raised by the Harris' (1990) Complexity Condition[11]. In a nutshell, Harris' (1990) and Pöchtrager (2015) concerns relate to the preference for complex elemental structures to occupy the head position of diphthongs. Or, in other words, they relate to the preference for diphthongs' heads to contain |A|.
According to Pöchtrager (2015: XXX), the "problems [of Harris' (1990) account] stem from a failure to take into account the individual nature of elements, their individual character". As a consequence, Pöchtrager (2015) proposed the structures given in 20) together with c-control, a mechanism evidently (though 'unfaithfully'[12]) borrowed from syntactic theory. In the present paper, rather than introducing c-command, we derive the same effect from the structural properties of 'elements', namely from their "individual nature": complex structures are preferably built on top of [a] because its representation consist of two non-empty sets that can

[11] "a. Let α and β be segments occupying the positions A and B respectively. Then, if A governs B, β must not be more complex than α;
b. The complexity value of a segment is simply calculated by determining the number of elements of which it is composed" (Harris 1990: 274).
[12] As recognized by Pöchtrager (2015: 270) himself, "syntactic binding is about co-reference, while phonological bounding [is about] distributional restrictions on melody". There seems to be other problems with binding and c-command as well, as c-command (alone) is not enough: it needs an extra mechanism to constrain its application domain (for this reason, c-command has been 'expanded' into the c++command in Živanovič & Pöchtrager 2010). However, "the locality of binding cannot be captured by this alone, but also by no other command relationship I can envision. How this boundedness is formally captured is another question" (Pöchtrager 2015: fn 11).

thus be further expanded by adjoining additional vocalic structures (with the limitations discussed in section 1.2 above).
Similarly, the representations we propose for |I| and |U|, whereby only the former present non-empty expandable nodes (see (6)), might account for their asymmetrical behavior. This asymmetry is also discussed in Pöchtrager (2015), who claims that "the English vowel system never allows combinations of |I| and |U| within some phonological expression. [This] is true for monophthongs [and] diphthongs" (Pöchtrager 2015: 258)[13]. As in the case of |A|, we encode this asymmetry in the representations we propose for |I| and |U|, rather than resorting to c-command ("I can bind U, but U must not bind I", where "α binds β iff α c-commands β" (Pöchtrager 2015: 263)).
With respect to standard element theory (Backley 2011), a crucial difference concerns phonetic interpretation. As we mentioned above, the standard theory provides "a restrictive model of phonological knowledge in which elements are abstract enough to function as cognitive units of linguistic structure yet concrete enough to be realized phonetically without the need for explicit rules of phonetic interpretation". Thus, the more elements/substance we replace by structure, the more complex the phonetic interpretation procedure. Assuming strict modularity, though, whereby phonology and phonetics are two different realms and the former is translated into the latter in a lexical access fashion (Scheer 2014), this problem is perhaps not that dangerous. If anything, there would be an issue of learnability, which is arguably much simpler assuming, as per the element standard theory, that primes are at the same time cognitive *and* concrete units.

3. Conclusion: Substance as structure

From one fairly simple definition in (1), we can thus derive a theory that seems to mimic the results of known Element Theories. Notice, however that this theory has two properties that seem to its disadvantage. First, it does not seem to add a lot of new empirical coverage to what we know from Element Theory, also because the latter is fairly successful. Secondly, although the theory we present itself is extremely simple, it leads to representations of even simple vowels like /ɔ/ that are rather difficult to read (such as {{{}},{{{}},{{},{}}}}).
In other words, our set theoretic definition gives a notational variant of existing theories in an obtuse notation. So why would it be helpful to consider this? We believe that in most practical analyses it will be more useful to write vowels in terms of familiar |I|, |A|, |U| representations, but the set theoretic definition gives us insight into the internal structure of these elements: why there are three of them, why only one of them can fully embed, whereas the other two are heavily deficient in this respect and, possibly, why we have at most 4 degrees of vowel height. We thus get a deeper insight into the reason why elements function a certain way that would not be available if we treat them as completely primitive, atomic elements. At the same time, for studying e.g. the vowel set of a particular language, we may not always need to know why elements function in some way. This is of course familiar from most kinds of (linguistic) analysis: for an analysis of stress in a certain language, we typically also do not need to give the full internal structure of all vowels involved: we use the 'higher-order' representation of IPA symbols with the understanding that these stand for combinations of elements. On the other

[13] Other asymmetries listed by Pöchtrager (2015), which are thus amenable to be accounted for account by our representational proposal, are front rounded vowels headedness, Turkish two /e/'s against one /o/ and the fact that Turkish |U|-harmony is more restricted than |I|-harmony.

hand, sometimes certain features may be relevant for the assignment of stress (like height features, or tone).
Our theory is similarly a theory of vocalic elements; it aims to explain the properties of these elements, but in the everyday business of phonological analysis, it may not be necessary to refer to them all the time.
Because our system has definite properties, it is also not compatible with any interpretation of element theory. For instance, it is impossible to introduce a notion like ‘headedness’ into the system without making crucial changes to it. The kind of asymmetry that headedness applies can only come about by an extra theoretical device that is not available in the current theory.

References

Adelson-Velsky, G.; Evgenii Landis (1962). An algorithm for the organization of information. Proceedings of the USSR Academy of Sciences (in Russian). 146: 263–266 [English translation by Myron J. Ricci in Soviet Math. Doklady, 3:1259–1263 (1962)].

Anderson, J. M., & Ewen, C. J. (1987). *Principles of dependency phonology* (No. 47). Cambridge University Press.

Anderson, J., & Jones, C. (1974). Three theses concerning phonological representations. *Journal of linguistics*, *10*(01), 1-26.

Backley P. (2011). *An introduction to element theory*. Edinburgh University Press.

Backley P. (2012). Variation in Element Theory. *Linguistic Variation*, 12.1, pp. 57-102.

Bale A., C. Reiss and D. Ta-Chun Shen (submitted). *Sets, rules and natural classes: { } vs. [ ]*.

Botma E.D. & M. van Oostendorp (2012). A propos of the Dutch vowel system 21 years on, 22 years on. In Botma E.D. & Noske R. (eds.), *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: De Gruyter.

Cavirani, E. & van Oostendorp, M. (in press), The markedness of silence. In Samuels B. (ed.). *Beyond Markedness in Formal Phonology*. Amsterdam: John Benjamins.

Charette, M. (1991): Conditions on phonological government. Cambridge: Cambridge University Press.

Chomsky N. & Halle M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*(01), 225-252.

Crothers, J. (1978). Typology and universals of vowel systems in phonology. In Greenberg, J. H., C. A. Ferguson & E. A. Moravcsik (eds.), *Universals of human language: phonology. Vol. 2*. Stanford University Press, pp. 95-152.

Harris, J. (1990). Segmental complexity and phonological government. *Phonology 7*, pp. 255-300.

Harris J. & M. Lindsey (2000). Vowel patterns in mind and sound. In Burton-Roberts N., Carr P. & Docherty G. (eds), *Phonological knowledge: conceptual and empirical issues*. Oxford: OUP, pp. 185-205.

Hulst, H.G. van der (1988). The geometry of vocalic features. In: H. van der Hulst & N. Smith (eds.). *Features, segmental structure and harmony processes*. Dordrecht: Foris, 77-126

Hulst, H.G. van der (1994). An introduction to Radical CV Phonology. In: S. Shore & M. Vilkuna (eds.). *SKY 1994: Yearbook of the linguistic association of Finland*. Helsinki, 23-56.

Hulst, H.G. van der (1996). Radical CV Phonology: The segment – syllable connection. In: J. Durand & B. Laks (eds.). *Current trends in phonology: Models and methods. Vol 1*. CNRS/ESRI Paris X, 333-363

Hulst, H.G. van der (2006). Dependency phonology. In Keith Brown (ed.). *The encyclopedia of language and linguistics. 2nd edition. Volume 3*. Oxford: Elsevier, 451-458.

Hulst, H.G. van der (2011). Dependency-based phonologies. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory, 2nd edition*. Malden, MA: Wiley-Blackwell, pp. 533–570.

Hulst, H.G. van der (2015). The Opponent Principle in RcvP: Binarity in a Unary System. In: Eric Raimy and Charles Cairns (eds.). *The Segment in Phonetics and Phonology*. Wiley-Blackwell.

Hulst, H.G. van der (2017). The integration of segmental and syllabic structure in Radical CV Phonology. Talk presented at the workshop *The Interface Within*. *What relations hold between prosody and melody?*, 13th March 2017, Meertens Institute, Amsterdam, The Netherlands.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1985). The internal structure of phonological representations: a theory of charm and government. *Phonology Yearbook* 2, pp. 305-328.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1990). Constituent structure and government in phonology. *Phonology* 7, pp. 193–231.

Kaye J. and M. Pöchtrager (2009). *GP 2.0*. Paper presented at the "Government Phonolog Round Table 2009", Piliscsaba/Hungary.

Kaye, J. & M. Pöchtrager (2013). GP 2.0. *SOAS Working Papers in Linguistics*, 16, pp. 51-64.

Lowenstamm, J. (1996) CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current Trends in Phonology: Models and Methods*. European Studies Research Institute, University of Salford Publications. 419-442.

Nasukawa K. (2015). Recursion in the lexical structure of morphemes. In van Oostendorp M. & van Riemsdijk H. (eds.), *Representing Structure in Phonology and Syntax*. Berlin: Mouton de Gruyter, pp. 211-238.

Ploch, S. (1995). French Nasal Vowels - A First Approach. *SOAS Working Papers in Linguistics*, 5, pp. 91-106.

Pöchtrager, M. (2006). *The structure of length*. Ph.D. dissertation, University of Vienna.

Pöchtrager M. (2012). *Deconstructing A*. Paper presented at the "MFM Fringe Meeting on Segmental Architecture", Manchester/UK.

Pöchtrager M. (2015). Binding in Phonology. van Oostendorp M. & van Riemsdijk H. (eds.), *Representing Structure in Phonology and Syntax*. Berlin: Mouton de Gruyter, pp. 255–275.

Reiss C. (2017). Substance Free Phonology. In Hannahs S. J. and Bosch A. R. K. (eds), *Handbook of Phonological Theory*. Routledge.

Rennison, J. (1998). Contour segments without subsegmental structures". In Eugeniusz Cyran (ed.), *Structure and Interpretation: Studies in Phonology*. Lublin: Folium, pp. 227–245.

Schane, S. A. (1984). The fundamentals of particle phonology. *Phonology Yearbook* 1, pp. 129-155.

Scheer, T. (2004). *A lateral theory of phonology*. Berlin: Mouton de Gruyter.

Scheer, T. (2012). *Melody-free syntax and two Phonologies*. Paper presented at the annual conference of the Réseau Français de Phonologie (RFP), Paris 25-27 June.

Scheer, T. (2014). Spell-Out, Post-Phonological. Eugeniusz Cyran & Jolanta Szpyra-Kozlowska (eds.), *Crossing Phonetics-Phonology Lines*. Newcastle upon Tyne: Cambridge Scholars, pp. 255-275.

Schwartz, G. (2010). Auditory representations and the structures of GP 2.0. *Acta Linguistica Hungarica* 57:4, pp. 381-397.