



**HAL**  
open science

# Representation vs derivation: The case for a modular view of phonology

Joaquim Brandão De Carvalho

► **To cite this version:**

Joaquim Brandão De Carvalho. Representation vs derivation: The case for a modular view of phonology. *Radical: A journal of Phonology*, 2020, 1, pp.292-327. halshs-03154028v2

**HAL Id: halshs-03154028**

**<https://shs.hal.science/halshs-03154028v2>**

Submitted on 23 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**volume 1, 2021**

**Brandão de Carvalho, Joaquim**

**Representation vs  
derivation:**

**The case for a modular  
view of phonology**

Brandão de Carvalho, Joaquim. 2020. "Representation vs derivation : The case for a modular view of phonology". *Radical: A Journal of Phonology*, 1, 292-327.

Editor: *Quentin Dabouis*  
Reviewers: *Heather Newell, Tobias Scheer, Marc Van Oostendorp*

# REPRESENTATION VS DERIVATION: THE CASE FOR A MODULAR VIEW OF PHONOLOGY<sup>1</sup>

JOAQUIM BRANDÃO DE CARVALHO (UNIVERSITÉ PARIS 8, UMR 7023)

Abstract. Plurilinear representations and constraint-based derivation have been the main tools used in the last 40 years by autosegmental approaches and OT respectively. It is argued here that the border between representation and derivation depends on another key division in phonological theory: in phonology proper, there are no such things as processes involving feature propagation or delinking; only by virtue of morphophonological alternations, that is of phonology/morphosyntax interaction, may an object be said to change into another one. Thus, representational theories better describe core phonology, seen as an autonomous module of grammar, while constraint-based models are necessary to deal with computation associated with phonology's upper interfaces.

Keywords: Phonological representation, phonological derivation, morphophonological alternation, modular grammar.

---

1 This squib resumes and elaborates two papers that were presented at the 1<sup>st</sup> meeting of the Phonological Theory Agora (Lublin, 2015) and at the 40<sup>th</sup> anniversary conference of the Centro de Linguística da Universidade do Porto (Porto, 2016). I am grateful to Noam Faust and Francesc Torres Tamarit for their input on a first draft of this article, and to Heather Newell, Marc van Oostendorp and (with special thanks) Tobias Scheer, as well as to three members of *Radical's* editorial team for their reviews of a previous version.

## 1 THE DUAL LEGACY OF *SPE*

### 1.1 FROM RULES TO CONFIGURATIONS

It is instructive to look back upon the last forty years of research in phonology. The crisis caused by Chomsky & Halle's (1968) book (henceforth *SPE*) came to a provisional end in the mid-seventies, when Goldsmith (1976) introduced a break whose effects are felt to this day and are commonly labeled under the term "autosegmental phonology". It was, according to Encrevé (1988:146), the most "spectacular innovation" in the history of phonological representations since the invention of the alphabet. From then on, it has no longer been possible to view words as mere sequences of letter-like tokens: phonologists have come to the conclusion that a distinction has to be made between the phonemes of a morpheme and the positions they occupy, that the phoneme has an internal structure, and that representations are based on how melodies and positions are synchronized, features behaving much in the same manner as only tones were previously supposed to do.

While the number of possible phonological operations has been drastically reduced to two (spreading and delinking), this milestone naturally made representations much more complex – and possibly more abstract as well: for example, there is nothing more abstract than a pure timing slot. It should be noted that abstraction is not, in itself, a reason for criticism or rejection. A theory should not be dismissed because it is too abstract; it should, however, if it is *arbitrary*. Arbitrariness was precisely the "serious flaw" of their own proposal that Chomsky & Halle emphasized in the last chapter of their book: rules are arbitrary as they are not able to distinguish a commonplace phenomenon from one that is unlikely or even impossible. By contrast, within autosegmental approaches, processes are supposed to be *motivated* if they obey a small number of principles, i.e. well-formedness conditions, imposed to autosegmental configurations (OCP, No Line-crossing, etc.), contrary to other processes which are therefore said to be

ungrammatical. Thus, autosegmentalism, whose main representative is currently Government phonology, appeared to certain scholars (see, e.g., Encrevé 1988, Kaye *et al.* 1990, Goldsmith 1993) as a project where, thanks to representational devices, the expressive power of *SPE*-type computation could be both restricted and motivated. In particular, the serial nature of early generative phonology was hardly tolerable for many people in the 1980s, and was eventually abandoned in (classic) Optimality theory.

## 1.2 FROM RULES TO CONSTRAINTS

It is also instructive to compare the contribution of autosegmental phonology with the second great break in our field, which dates back to Prince & Smolensky (1993), and is now mainstream: Optimality theory (OT). As is well-known, three main points characterize this approach in relation to the *SPE* original framework:

- (1) a. Violable constraints, which, unlike *SPE* rules, are supposed to be universal, only their ranking (or weight) being language-specific.
- b. Dialectic tension between markedness and faithfulness constraints, the former applying to the surface representations (SR), and the latter to the relation between SRs and underlying representations (UR); in *SPE* faithfulness effects were obtained by default, through the absence (or underapplication) of rules.
- c. Parallel derivation, which replaces *SPE* serialism (at least in the classic version of OT).

With respect to representations, Prince & Smolensky argued that their model builds on Autosegmental Phonology. Freedom of Analysis, the first of the three principles underlying the theory of GEN, stipulates that “any amount of structure may be posited”: “GEN may supply candidates with syllabic, moraic, or other prosodic structure, with association lines, and with additional segmental material, ranging from empty root nodes

through fully specified vowels or consonants.” McCarthy & Prince (1993: §2.3)

The comparison between autosegmental and OT accounts is extremely interesting as it provides two different ways of answering the following question: what should a rule-free phonology be like? So as to get rid of serial rules, the autosegmental approach focuses on the structure of *representations*, OT, by assessing candidate outputs partly against their inputs, explicitly focuses on *derivation*, i.e. on processes whereby an object is changed into another one. The question then arises: should we try to overcome this opposition by unifying the two theories? Assessing the respective roles of representation and computation is, as Anderson (1985) had already sensed, one of the most challenging epistemological problems that emerge from the phonology of these last forty years.

### 1.3 IS UNIFICATION POSSIBLE?

To begin with, is this unification desirable? I think it is, because the two approaches show major and specific drawbacks. Both involve the same flaw: some kind of circularity.

On the representational side, Government phonology is unable to provide a straightforward account of variation.<sup>2</sup> Variability is assumed to follow from parameters associated to the principles mentioned above. Yet, parameter settings through licensing relations look like ad hoc stipulations: consider, for example, the variable behaviour of word-final empty nuclei from language to language according to whether they can or cannot dispense licensing (Charette 1991:132-142). In contrast to this “legalist” approach (Boltanski 1999:153), the same markedness constraints can be said to be differently ordered cross-linguistically in OT according to their ranking vis-à-vis faithfulness constraints; in other words, unlike parameter settings, the ranking of markedness con-

<sup>2</sup> Like one of my reviewers (p. 310), many, if not most, proponents of GP may argue that variation is extra-phonological and therefore should not be accounted for within a theory of phonology. I have a different stance: grammars generally model one arbitrarily chosen variety of a language; since the same speakers always employ more than one variety, at least a certain type of variation should be viewed as an integral part of grammar. At any rate, the line between phonology and variation is not of the same nature as those between phonology and morphology or between phonology and phonetics; there is no sociolinguistic module.

straints has an *extrinsic* motivation. More generally, recent developments of OT have led to the most interesting formal accounts of free variation since Labov's times (see e.g. Boersma 1997; Anttila 1997, 2007; Hayes 2000; Hayes & Wilson 2008).

As to OT, markedness constraints can also be said to be ad hoc, as they generally lack formal motivation.<sup>3</sup> For example, why are CV syllables unmarked *vis-à-vis* CVC (with one additional element) and V (with one element less)? ONSET and NOCODA simply record typological and acquisitional evidence; they do not explain anything. By contrast, Strict CV allows a unified account of CVCv and cV markedness, assuming that empty positions (v, c) are marked.

There are several cases of theoretical competition and eventual unification in the history of science. Physics offers at least three different examples. The first and the simplest one is when a new and more global approach annexes an earlier and “narrower” theory, the latter becoming a particular case of the former. This is the case of general relativity and classical mechanics. I do not think that this applies to the phonological frameworks at stake: neither of them can be shown to be more global than the other.

The second example is when two competing theories account for different empirical domains. This is the case of the theory of relativity *versus* quantum theory. Here a real unification should occur (and is still expected), both theories becoming particular cases of a third more global one. This case does not seem more relevant than the former, since the domains of configurational and constraint-based approaches largely overlap.

The third example may be the most appropriate for our purpose. Until the twenties of the last century, certain natural phenomena – light for example – seemed to exhibit a mysterious dualism: some data supported the view that they had a corpuscular nature,

---

3 For most phonologists working in the OT framework, formal motivation might be unnecessary as markedness constraints are supposed to be “phonetically grounded” (see Archangeli & Pulleyblank 1994, Hayes & Steriade 2004, Bermúdez-Otero & Börjars 2006, de Lacy 2006, Kingston 2007). However, not only is this groundedness problematic in many cases, but phonologists are far away from a consensus on the relationship between phonology and phonetic substance (see, e.g., Iosad 2017 and Reiss 2017).

and that a particle should therefore be sought; other data pointed towards a wave theory. This continued until Louis de Broglie succeeded in explaining that light was *simultaneously* a particle and a wave, according to the “perspective”: one is the other’s limit. I have long forgotten what de Broglie’s perspectives are, but I think there might be a similar way to answer the question that arises from the dual nature of phonology: if representation and derivation are both part of our field, *where should we draw the line between them?* Interestingly, the response I will argue for squares with repeated claims for the autonomy of phonology in relation to upper modules (see in particular Scheer 2011).

## 2 BACK TO BASICS

### 2.1 THE DUAL NATURE OF RULES

Let us remember the scope of *SPE* rules. They are a legacy of structural phonology, and, as such, they are dependent on its two main historical trends: Prague School and American structuralism. Due to the former source, rules should affect phonemes that are strictly defined through distributional analysis. For example, as shown in (2a), Spanish has three nasal phonemes – a labial /m/, a coronal /n/ and a palatal /ɲ/ – that contrast in onset position, while, as shown in (2b), only an underspecified nasal can occur in coda position – an “archiphoneme”, whose place feature is imposed by the following onset, if any.

(2) Trubetzkoy's "(archi)phonemes"

a. Spanish /m/ ~ /n/ ~ /ɲ/

['kama] “bed”	['kana] “rattan, stick”	['kaɲa] “reed”
[so'mar] “to sum”	[so'nar] “to sound”	[so'ɲar] “to dream”

b. Spanish /N/ = {nasal}

['kampo] “countryside”	['kanto] “I sing”	['baŋko] “bank”
['aɲfo] “large”	['pan/ŋ] “bread”	['bjen/ŋ] “well, asset”



It follows that Spanish will be said to have the “allophonic rule” in (3).

- (3) a. Sp. /N/ → [m, n, ɲ, ŋ] / \_LAB, COR, PAL, VEL  
 b. Sp. /N/ → [n/ŋ] (according to the variety) / \_#

However, in accordance with a tradition that dates back to Bloomfield (1933), the input of *SPE* rules is not only based on distribution; it can also follow from morphophonological alternations. Consider the data from Somali in (4). Distributional evidence would have led to posit /la:N, siN, da:N, sa:N/ in the singular and definite forms for the same reasons as in Spanish: no place contrast is allowed in coda. Yet, as this would make the plural form unpredictable, the /m/ ~ /n/ contrast is assumed to exist in all positions underlyingly.

- (4) (Neo-)Bloomfieldian "morpho-phonemes"

	<i>singular</i>	<i>definite</i>	<i>plural</i>	
a. Somali /m/	[la:n] [sin]	[la:nta] [sinta]	[la:mo] [simo]	“branch” “hip”
b. Somali /n/	[da:n] [sa:n]	[da:nta] [sa:nta]	[da:no] [sa:no]	“shore” “hiding place”

Thus, the allophonic rule yielding homorganicity no longer affects one archiphoneme as in (3), but two specified nasal phonemes, in particular /m/ which, as shown in (5), undergoes a real change by losing its labial feature before coronals and word-finally.<sup>4</sup>

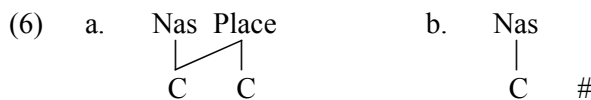
- (5) Som. /m/ → [n] / \_COR, #

4 Conversely, /n/, putatively unmarked as to place, is supposed to acquire [Labial] before labials (although I could not find examples from the scarce literature on Somali phonology).

The rules in (3) and (5) differ considerably due to the motivation of their inputs. As said above, (3a,b) are dictated by distributional analysis, while (5) is supposed to account for the *morphological* knowledge of speakers: the plural of [la:n] is [la:mo], not \*[la:no]; therefore, its [n] is not a “real” /n/ but an /m/ that changes into [n]. Despite that difference, due to Chomsky & Halle's denial of an intermediate phonemic level, (3) and (5) have long been assumed to be formally identical instances of phonological derivation, whereby an object is converted into another one. Does this hypothesis hold? It might be the case that the difference between (3) and (5) dispenses with the need for derivation in one of the two examples. We should thus ask if derivation is necessary in both cases.

## 2.2 DERIVATION IS A MORPHOSYNTACTIC EFFECT

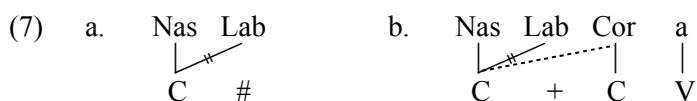
Is the derivation in (3a) necessary? I contend that it is not. The concept /N/ follows from the classic view of phonemic representations where features do not overlap (hence /N+p/). However, strict linearity at the underlying level – causing place spreading at the phonetic level (hence [mp]) – results from unidimensionality. The autosegmental approach does not require feature spreading or delinking in such cases: Spanish homorganic NC clusters should simply be viewed as underlying geminates, in which linearity (i.e. the sequence Nasal+place features) is expressed *in the melodic tier*, as shown in (6a). No one has ever claimed that lexical geminates or long vowels result from spreading.<sup>5</sup>



5 The same holds, say, for languages where front vowels palatalize consonants beyond what is expected universally (i.e. at Bermúdez-Otero & Trousdale's (2012) “phonologisation” level): the l-element is underlyingly associated with two positions as well. Likewise, vowel harmony simply requires many-to-one association, as in (6a); only affixation allows us to refer to derivation, as in Turkish /ip-um/ → [ipin] ‘rope-gen.’, /ip-lar/ → [ipler] ‘rope-pl.’

As to (3b), this rule involves no phonological change either, as no feature is added nor deleted: [n] (or [ŋ]) is the unmarked realization of the archiphoneme /N/ in (6b). Derivation only occurs by virtue of morpheme adjunction in Spanish, where cross-morpheme NCs behave like their Somali cognates, with the underspecified coda receiving the place feature of the following onset: *Jua*[n] but *Jua*[ŋ] *Carlos*, *do*[n] *Antonio* but *do*[m] *Pedro*.

Let us now consider the case of Somali. Like Spanish, it has homorganic NC clusters, with the nasal coda lacking a place feature of its own, including intramorphemic ones from which no alternation can be subsumed: see Gabbard (2010: 33). However, unlike Spanish, Somali *has* the rule in (5). This is because (i) as in Sp. *Jua*[ŋ] *Carlos*, *morphemic information* (the suffix /-ta/ but also the boundary #) is introduced; (ii) the two languages have the same (nasal) phonology *but different lexicons*: Somali, unlike Spanish, has radicals ending with /m/, hence the derivation in (7a) for [(la:)n], where the feature [Labial] is deleted, and the one in (7b) for [(la:)nta], where [Labial] is replaced with [Coronal].



Spanish may be said to have roots ending with any nasal consonant (e.g. *cam-a* 'bed', *man-o* 'hand', *cañ-a* 'rod'), but these consonants, except [n], are never found word-finally because the Spanish lexicon stores stems with theme vowels, not roots with inflectional class features (Bermúdez-Otero 2013). As a result, there are no m-n alternations parallel to Somali [(la:mo)-[(la:n)]. The only exceptions come from a ridiculously small number of loanwords like *álbum* or *islam*. These might be pronounced with a final [n] in low varieties. However, even the existence of a couple of alternations like [islámiko]-[islán] or [álbumes]-[álbun] supporting an underlying /m/ in such styles is highly dubi-

ous given the erudite nature of the adjective and the rarity of the plural form.<sup>6</sup>

In sum, the facts discussed above suggest the following answer to the question asked in §1.3 about the border between representation and derivation:

- (8) a. There is no derivation in morpheme-internal phonology: the phonological processes of a language do not apply to items belonging to the same morpheme.
- b. Only when it comes to morphophonological alternations – due to affix concatenation or to sandhi – may derivation emerge, lexicon permitting.<sup>7</sup>

### 3 WHAT IS MEANT HERE

Before concluding, four points are worth noting to make clear what my proposal is and what it is not. First, as the reader may have understood, the divide I suggest between morpheme-internal phonology and its interfaces does not parallel the one between the so-called “lexical” and “post-lexical” levels posited by Lexical phonology and its heirs, nor can one object that the same work is done twice, once by phonology, a second time by phonology’s upper interfaces since Spanish and Somali share the same (nasal) phonology. What is at stake here is not a matter of levels but of perspectives (cf. §1.3). The rules (3) and (5) can be viewed as different instances of the same phenomenon seen from different angles: both imply underspecified nasal codas; unlike (5), however, (3) is not a derivational process insofar as it neither adds nor deletes any phono-

6 In any case, if the reader wants an example of a language where these alternations are totally absent, I suggest Japanese, which has homorganic NC clusters, but whose nasal coda never alternates with [m] before morpheme-initial vowel.

7 Note that non-derivational alternatives to alternation-based rules have long been proposed: e.g. the “via-rules” of Natural Generative Phonology (Hooper 1976), whereby (5) could be replaced with a correspondence between /m/ and the archiphoneme /N/. For my part, I think that both views are potential options: /la:m/ (and the resulting derivation) should entail word-final [m] before vowel *even in sandhi contexts*, while /la:N/, with a lexicalized archiphoneme (= (6a)), precludes [m] in sandhi – as in Somali (Barillot, p. c.). As an example of the latter case, Korean *lenis* /p, t.../, aspirated /p<sup>h</sup>, t<sup>h</sup>.../ and *fortis* /p\*, t\*.../ consonants merge word-finally into voiceless plosives ([p̥, t̥...]) that undergo the same voicing process in sandhi before word-initial vowels as the word-internal unmarked *lenis* between vowels. Despite regular allomorphic alternations with marked phonemes, these final plosives can thus be argued to be underlyingly archiphonemes that have lost all connection with their historical sources.

logical feature, low level phenomena remaining outside phonology, like all morpheme-internal processes.

Secondly, the assumed difference between (3) and (5) does not fit in with the eternal debate regarding the trade-off between the lexicon and computation in phonology – see thereon the recent discussion by Vaux & Samuels (2018): is it preferable to have a more complicated lexicon or a more complicated computation? My purpose is not to play down the importance of computation as a matter of principle. I simply aim to suggest a reason for the division of labour between the two facets of phonology, and I contend that the unique trigger of phonological computation is morpheme concatenation; hence, computation operates in the phonology-morphosyntax interface, since phonology, on its own, does not concatenate such items. Nonetheless, the claim “same phonology but different lexicon” departs from OT, where there are no lexical differences, and everything is put into the grammar to a point that the examples above would be explained in a quite paradoxical manner. It is the simplest case, Spanish, the one with place neutralization in coda position and no morphophonological alternations, that would require the most complex treatment, where the principles of Richness of the Base and Lexicon Optimization must be considered for banning specified /m, n/ inputs in coda position. I have argued instead for a strong version of Consistency of Exponence in the morpheme-internal situation, and only in this case, where “no changes in the exponence of a phonologically-specified morpheme are permitted” (McCarthy & Prince 1993: §2.3) simply because the output is fully faithful to the input in the lexicon.

Thirdly, the proposed division between morpheme-internal phonology and its interfaces does not coincide with the distinction between non-derived and derived environments (Kiparsky 1993) either, for two reasons. On the one hand, derived environments may be due to morphology (affix adjunction), as in my approach, but also to phonology, with new material being the result of another phonological process (which is typically

the case with opacity effects).<sup>8</sup> On the other hand, in Non-derived Environment Blocking, rules apply in one case and do not in the other, hence producing different and (apparently) contradictory outputs. However, there is no such conflict in our Spanish-Somali example: both languages attest homorganicity, which is lexical in one case, and derived in the other.

Lastly, it should be emphasized that even though the terms “computation” and “derivation” have been employed here as synonyms, the former may be assigned a much broader meaning than the latter. I have used computation/derivation in a particular sense, referring to processes that convert a phonological string into a different one, by manipulating (i.e. adding or deleting) equally phonological objects like features. However, morpheme-internal homorganic NC sequences do not come out of the blue. On the one hand, representations are subject to well-formedness principles (No Line-crossing, OCP, etc.) and implicational markedness scales. On the other hand, languages can make different choices: German, for example, has lexical *non*-homorganic NC clusters, as in *Hemd* “shirt”. At this level, thus, homorganic and non-homorganic NC sequences can be said to compete against each other in the lexicon, which requires some kind of computation, since the same phonetic string, say [nt], can result either from /mt/ or from a (6a)-like underlying geminate depending on the language.<sup>9</sup> Nonetheless, such computation is not derivation; it is a lexicalization process whereby “L1 learners or adults transform the acoustic signal into a stored representation” (Scheer forth.), that is a cognitive category. Interestingly, computation, albeit non-derivational, appears once again to be associated with another interface, in this case between phonetics and phonology.

8 Morpheme-internal opacity effects pose an interesting challenge to the claim defended in this paper. However, as it is part of the general problem caused by opacity (and by the issue of whether derivation requires a serial approach), this topic will be left for further discussion.

9 That is from entirely different representations, /mt/ implying an empty nucleus in between the consonants (see Carvalho 2017).

## CONCLUSION

I hope that the idea expressed in this article may contribute to the ongoing discussions about the work sharing between representational and computational approaches to our field. If I am right, the dual legacy of *SPE* should not be seen as contradictory: representations and derivation actually do partly complementary jobs. Representational approaches account for the internal workings of the phonological module, and might help to motivate the markedness constraints required by phonological computation. Constraint-based approaches are required by the latter, that is by derivation which emerges when phonology interacts with other modules such as morphology, syntax, and the lexicon. Thereby, the respective roles of representation and derivation strongly support modularist theories of grammar, and the autonomy of the phonological module.

## REFERENCES

- Anderson, Stephen R. (1985) *Phonology in the twentieth century: theories of rules and theories of representations*. Chicago: University of Chicago Press.
- Anttila, Arto (1997) *Variation in Finnish phonology and morphology*. Ph.D. diss., Stanford University.
- Anttila, Arto (2007) 'Variation and optimality', in P. de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: CUP, pp. 519-536.
- Archangeli, Diana & Pulleyblank, Douglas (1994) *Grounded phonology*. Cambridge, MA: The MIT Press.
- Bermúdez-Otero, Ricardo (2013) 'The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features', *Probus*, 25, pp. 3–103.
- Bermúdez-Otero, Ricardo & Börjars, Kersti (2006) 'Markedness in phonology and in syntax: the problem of grounding', in P. Honeybone & R. Bermúdez-Otero (eds), *Linguistic knowledge: perspectives from phonology and from syntax*. *Lingua*, 116 (4), pp. 710–756.

- Bermúdez-Otero, Ricardo & Trousdale, Graeme (2012) ‘Cycles and continua: On unidirectionality and gradualness in language change’, in T. Nevalainen & E. C. Traugott (eds), *The Oxford handbook of the history of English*. New York: OUP, pp. 691–720.
- Bloomfield, Leonard (1933) *Language*. New York: Henry Holt.
- Boersma, Paul (1997) ‘How we learn variation, optionality, and probability’, in *Proceedings of the Institute of Phonetic Sciences of the Univ. of Amsterdam*, 21, pp. 43–58.
- Boltanski, Jean-Elie (1999) *Nouvelles directions en phonologie*. Paris: Presses Universitaires de France.
- Carvalho, Joaquim Brandão de (2017) ‘Deriving sonority from the structure, not the other way round: A Strict CV approach to consonant clusters’, *The Linguistic Review*, 34, pp. 589–614.
- Charette, Monik (1991) *Conditions on phonological government*. Cambridge: CUP.
- Chomsky, Noam & Halle, Morris (1968) *The sound pattern of English*. New York: Harper & Row.
- de Lacy, Paul (2006) *Markedness: Reduction and preservation in phonology*. Cambridge: CUP.
- Encrevé, Pierre (1988) *La liaison avec et sans enchaînement: phonologie tridimensionnelle et usages du français*. Paris: Seuil.
- Gabbard, Kevin M. (2010) *A phonological analysis of Somali and the guttural consonants*. BA Linguistics Honors Thesis, The Ohio State University.
- Goldsmith, John A. (1976) ‘An overview of autosegmental phonology’, *Linguistic analysis*, 2, pp. 23–68.
- Goldsmith, John A. (ed.) (1993) *The last phonological rule: reflections on constraints and derivations*. Chicago: University of Chicago Press.
- Hayes, Bruce (2000) ‘Gradient well-formedness in optimality theory’, in J. Dekkers, F. van der Leeuw & J. van de Weijer (eds), *Optimality Theory: Phonology, Syntax, and Acquisi-*



tion. Oxford: OUP, pp. 88–120.

Hayes, Bruce & Steriade, Donca (2004) ‘The phonetic bases of phonological markedness’, in B. Hayes, R. Kirchner & D. Steriade (eds), *Phonetically Based Phonology*. Cambridge: CUP, pp. 1–33.

Hayes, Bruce & Wilson, Colin (2008) ‘A maximum entropy model of phonotactics and phonotactic learning’, *Linguistic Inquiry*, 39, pp. 379–440.

Hooper, Joan B. (1976) *An introduction to natural generative phonology*. New York: Academic Press.

Iosad, Pavel (2017) *A substance-free framework for phonology: An analysis of the Breton dialect of Bothoa*. Edinburgh: EUP.

Kaye, Jonathan, Lowenstamm, Jean & Vergnaud, Jean-Roger (1990) ‘Constituent structure and government in phonology’, *Phonology*, 7, 193–231.

Kingston, John (2007) ‘The phonetics-phonology interface’, in P. de Lacy (ed.), *The Cambridge handbook of phonology*. Cambridge: CUP, pp. 401–434.

Kiparsky, Paul (1993) ‘Blocking in Non-derived environments’, in S. Hargus & E. Kaisse (eds), *Phonetics and Phonology 4: Studies in Lexical Phonology*. San Diego: Academic Press, pp. 277–313.

McCarthy, John J. & Prince, Alan (2001 [1993]) ‘Prosodic morphology: Constraint Interaction and Satisfaction’. *Linguistics Department Faculty Publication Series*, 14.

Prince, Alan & Smolensky, Paul (2004 [1993]) *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell.

Reiss, Charles (2017) ‘Substance-free phonology’, in S. J. Hannahs & A. R. K. Bosch (eds), *The Routledge handbook of phonological theory*. Oxford: Routledge, pp. 425–452.

Scheer, Tobias (2011) *A guide to morphosyntax-phonology interface theories: How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsingale*. Berlin:

Mouton de Gruyter.

Scheer, Tobias (forth.) '3xPhonology', unpublished ms.

Vaux, Bert & Samuels, Bridget (2018) 'Abstract Underlying Representations in Prosodic Structure', in D. Brentari & J. Lee (eds), *Shaping phonology*. Chicago: University of Chicago Press, pp. 146–181.

---

## DISCUSSION WITH HEATHER NEWELL

(UQAM)

---

Newell, Heather. 2020. discussion in: Brandão de Carvalho, Joaquim (auth.) “Representation vs derivation: The case for a modular view of phonology”. *Radical: A Journal of Phonology*, 1, 308-312.

### COMMENTS

This paper argues that the work done by representational theories (autosegmental, ex. Government Phonology (GP)) and derivational theories (Optimality Theory (OT)) should be shared in a unified theory of phonological computation. Below I outline questions that arose as I read the article. The main topics covered below are *arbitrariness*, *phonology vs morphology/phonetics/sociolinguistics* (modularity), and *variation*.

**1. Arbitrariness and the status of representations.** De Carvalho begins the discussion (p. 293) by bringing up the arbitrary nature of phonological rules in SPE, as opposed to the motivated (or at least limited) operations available to Autosegmental Phonology (AP) (delinking, spreading). This opposition clearly favours the predictability offered by an Autosegmental account, and I believe is not too controversial across phonological subfields. The one question I have concerning this specific comparison is the status of the ‘configurations’ that are proposed to bring about phonological operations. While the No-Line-Crossing constraint is (virtually) universally upheld in autosegmental analyses, the OCP is appealed to on a very ad-hoc basis. De Carvalho’s conclusion to this section is that GP and other autosegmental theories have worked to do away with arbitrary rules of the SPE variety, which I believe is uncontroversial.

At the beginning of Section 1.2 De Carvalho then contrasts SPE and OT. There is no specific theory of phonological representation that necessarily ‘goes with’ OT, and OT has not taken firm stands on questions of autosegmentality (although autosegmental

structure is used sometimes in OT analyses of segmental phenomena. Ex. Shona Vowel Harmony (Beckman 1997)). The author includes a quote from Prince & Smolensky (1993) that OT allows that “any amount of structure may be posited” (p. 294). This is true, but I still believe that the OT literature has not engaged adequately with structural questions, and that this weakness impedes our ability to discuss comparisons/thoughts on unification. Another point on this section is that, in the point-by-point comparison of OT and SPE in (1), there is only a direct comparison of the two theories in (1c). In (1a) and (1b) only the OT stance is mentioned. This comparison should be made explicit for all points, and it should be clarified how these points are pertinent to the proposition that OT is more concerned with the derivational side of phonological operations.

At the end of this first section De Carvalho suggests that the representational and derivational theories might come together to create a unified theory of phonology. I think this is an interesting conversation to have. But, I wonder if it is premature to discuss unification before discussing where the theories diverge and where they overlap. Will unification lead to a doubling of some types of analyses? If so, which? Also, before we can consider a potential unification, we have to be clear on what the predictions of each theory are. Do we know the predictions of OT? Do we know the predictions of Autosegmental Phonology (AP)? Are these terms (OT and AP) too broad to be useful? How can we reconcile the fracturing of both camps (many OTs (Harmonic Serialism, Stochastic OT, Stratal OT...), many autosegmental theories (Government Phonology, Government Phonology 2.0, CVCV, Dependency Phonology...))? Do we need unification of these sub-theories before we try to overcome the 'great divide'? Why or why not? I am not suggesting that all of these questions need to be dealt with in this article, but as the question is raised (and is a central point of the paper) I think that the complexity of the problem should be given more ink.

**2. Variation.** In De Carvalho’s section 1.3 he moves on to discussing specific advantages and disadvantages of OT and AP/GP. The first point to be discussed is *variation*.

De Carvalho states that GP “is unable to provide a straight-forward account of variation” (p. 295). It is unclear to me whether parameter settings are more or less ad-hoc than constraint-rankings. A more in-depth discussion would be nice here. Carvalho states that “...recent developments in OT (see Coetzee & Pater 2011) have led to the most interesting formal accounts of phonological variation since Labov’s times” [*the references cited have been amended in the latest version of the article. —Ed.*] (p. 296). Again, this is put forth as a statement of fact with no discussion. I agree that GP has not focused on variation, but I also believe that most proponents of GP would take the stance that variation is extra-phonological and therefore should not be accounted for within a theory of phonology (as is now mentioned in fn. 2). This is the only point that the author brings forth to support OT’s derivational prowess. Given the question of whether variation is a problem for Phonology-proper, I feel it needs either (i) more discussion, or (ii) to be replaced with something more unanimously phonological. One of the great problems of phonological theory is defining its domain of study (where is the line between phonology and morphology, between phonology and phonetics, etc?). See Zwicky & Pullum 1986, for example, on variability and the view that it has nothing to do with phonology. Also, the examples discussed in Coetzee & Pater (2011)<sup>10</sup> are not

10 Coetzee & Pater spend a lot of time talking about post-consonantal t/d deletion in English. They mention the morphological conditioning of this deletion (it is more likely when it is not a representation of the past tense marker) but then only talk about cross-linguistic variation based on the following environment (C, V, pause), a pattern that may still arguably be phonetic (not relevant for phonology). The fact that t/d is pronounced more often when it is a separate morpheme could also be due to ‘executive override’ of social/pragmatic cues, which again are not necessarily phonology, but could rather be the factors that lead to choosing one register over the other (Coetzee and Pater say “It is hard to see how such a model could capture the observed differences in frequency of deletion observed across languages/dialects, as in the English t/d-deletion case.” In regards to van Oostendorp’s (1997) suggestion that variation is due to social factors that choose between registers. I have not attempted to work through the problem, and Coetzee & Pater don’t discuss van Oostendorp’s proposal in any detail. They do discuss Harmonic Grammar proposals that incorporate register (style) into the factors that determine constraint rankings, but again, here we are incorporating something that is clearly non-phonological into the phonology and we have to ask ourselves if this is truly motivated.) In any case, the sensitivity to t/d *as a morpheme* is not the same as sensitivity to *morphological context*, which is what is relevant for the ‘early phonology’ mentioned in Coetzee & Pater as relevant for the question of whether variation is limited to ‘early’ or ‘late’ phonology.

The discussion of variability in loanwords with regards to Lyman’s Law and geminate devoicing is more clearly phonological, but it is unclear how such an analysis is less ‘circular’ than parameter settings that determine the characteristics of FENs in GP (De Carvalho compares GP’s analysis of FENs with OT’s ability to account for variation). When we talk about the arbitrary nature of parameters we could also be talking about constraints, both of which are theoretical postulates.

clearly directly comparable to the question of parameter settings with regards to Final Empty Nuclei, and so it is unclear what point this comparison is making. I get what the author wants to say, but I don't think the examples used are parallel.

**3. Unification.** The author then goes on to compare the unification of GP and OT with the unification of theories of physics. I think less time could be spent on this (or rather that there are more important phonological topics to address in this paper, if one is going to make a choice). But the question posed at the bottom of page 297 : “...*where should we draw the line between [the role of representation and derivation]...*” is a great one. This also needs to be further discussed internal to each theory, and is therefore not limited to the comparison of GP and OT, or to questions of their unification.

**4. Back to Basics.** Here there is a nice discussion of how much work we want derivation to be doing in our system. In cases where we see no alternations (specifically in the case of homorganic nasals) what does the speaker propose as an underlying representation? Is it the most specified form that accords with the surface form, or is it an underspecified form that is derived in the same way that alternations at morphological boundaries are? The author argues for the former analysis.

**5. Conclusion.** In conclusion, this paper brings up the discussion of whether all phonologists are working on the same problems, and whether our work can be seen to complement each other's in a way that could lead to a grand unified theory of phonology. I think this is a very interesting discussion to have and encourage the author to go into more depth in order to clarify his points.

---

Notably, also, no one has offered evidence (that I know of) for the type of variability seen in t/d deletion for FENs. Does this exist? Are these comparable phenomena?

It is possible that the discussion of variability is pertinent here in a way I'm not seeing, but it would be nice for the author to be clear about what is being put forth as a problem here. It is clearly something that we want to account for, and, as Coetzee & Pater note, lexical, register, frequency, and phonological/phonetic environmental effects are seen, but I think we need to be careful to be certain that variability is something we expect to find (here with regards to FENs) before we say it is problematic for a theory to not be able to account for it.

**Beckman**, J.N., 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology*, 14(1), pp.1-46. **Coetzee**, Andries W. & **Pater**, Joe (2011) 'The place of variation in phonological theory', in J. A. Goldsmith, J. Riggle & A. Yu (eds), *The handbook of phonological theory*, 2nd ed. Oxford: Blackwell-Wiley, pp. 401-431. **Oostendorp**, Marc van. 1997. Style Levels in Conflict Resolution. In Frans Hinskens, Roeland van Hout and Leo Wetters, eds. *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins. p. 207-229. **Prince**, Alan & **Smolensky**, Paul (1993) *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell. **Zwicky**, A.M. and **Pullum**, G.K., 1986. The principle of phonology-free syntax: introductory remarks. In Newmeyer (Ed.) *Linguistics, the Cambridge survey, vol 1*. CUP.

#### REPLY

See reply p. 320.

---

## DISCUSSION WITH TOBIAS SCHEER

(UNIVERSITÉ CÔTE D'AZUR)

---

Scheer, Tobias. 2020. discussion in: Brandão de Carvalho, Joaquim (auth.) “Representation vs derivation: The case for a modular view of phonology”. *Radical: A Journal of Phonology*, 1, 313-318.

### COMMENTS

Based on the claim that monomorphemic strings are never subject to any phonological computation when they are by themselves (i.e. in absence of other morphemes, or of morpheme boundaries), the author addresses the question how labour is divided between representations and computation. That is, what is representation competent for, and what is the purview of computation? This is the red line running through Anderson's (1985) book on the history of phonology, which Anderson rightfully described as a see-saw movement between representationally and computationally oriented theories. Writing at the peak of autosegmental representationalism, he predicted another round of computationalism to come. The fact that this prediction was borne out a few years later when OT dwarfed SPE in computational radicalism suggests that Anderson stood on solid analytic ground.

The author now ambitions to halt this see-saw movement: he claims that a correct line of division between representations and computation exists (having this or that phenomenon analyzed in either terms is not a matter of taste), and that it may be identified: when phonology is just by itself (monomorphemic strings), i.e. when it does not interact with anything else (morpho-syntax, phonetics), it is representational in kind. But phonology is computational when morpho-syntactic concatenation has produced a plurimorphemic string which undergoes modification in speech production, and also upon lexicalization, i.e. when L1 learners and adults transform the acoustic signal into a stored representation.



The author views this identification of representational and computational areas of competence as a unification in the sense of the insight in physics that light is not either a wave or made of particles, but both at the same time. In the same way, he argues, representations and computation should not be seen as competing instantiations of phonology, but rather as two aspects of the same thing.

That much we knew: like all other natural things (biology, chemistry, physics etc.), phonology is made of objects (stored in long term memory) which are subjected to forces that modify them. The question is how much of the cake (the set of phenomena to be explained) is representational and how much is computational. All theories claim that they have found the correct division of labour. The author introduces his own take on the matter, which has the virtue of drawing a clear line, but which remains unmotivated: the reader is not presented with any argument why the red line should run where the author says it does, rather than somewhere else. The author does provide a rationale for monomorphemic strings to be uncomputable, but assuming that this is indeed the case, I cannot see how this answers the question how much of the plurimorphemic cake is representational, and how much of it is computational. For the ban on computation of monomorphemes decides the competition for these strings, but does not speak to plurimorphemic items. Here the situation is thus as before: for a given alternation there will be competing representational and computational accounts.

The idea that monomorphemic strings are uncomputable revives Kiparsky's derived environment effects, embodied successively in the Alternation Condition (Kiparsky 1968-73: 14ff), the Revised Alternation Condition (Kiparsky 1982: 152) and the Strict Cycle Condition (SCC, Kiparsky 1982: 153f).<sup>11</sup> Kiparsky's idea was that phonological computation can only apply to strings that are already derived. The author is right in pointing out that there were two ways for a string to count as derived in the SCC: when containing more than one morpheme (morphologically derived) and when it was already

---

11 A more detailed discussion of the historical development regarding derived environment effects appears in Scheer (2011: §§183-200).

subject to a phonological process (phonologically derived). True, the uncomputability of monomorphemic strings that the author argues for is thus only half of the SCC, but the idea is the same and was present as such since the original Alternation Condition.

All versions of the derived environment idea met empirical challenges that the contemporary literature discussed in detail. Kiparsky's motivation for declaring monomorphemic strings uncomputable was the observation that processes such as trisyllabic shortening are triggered by (class 1) suffixes as in *sane* [ɛj] - *s[æ]n-ity*, but never within a root, although the triggering conditions may be satisfied: the first vowel of *ivory* and *nightingale* is pronounced [aj], not [ɪ] (as in *Christ* [aj] - *christ-ian* [ɪ]). But if processes do not apply to monomorphemic strings, what about nasal cluster simplification *gn* → *n* in root-final position when the root occurs alone as in *sign* [n] (compare with *si[gn]-ature*)? Or the loss of *g* in /ɲg/ as in English *long* [ŋ] (as opposed to *long-er* [ŋg])? And what about aspiration of *p,t,k* in English, which occurs word-initially (*p<sup>h</sup>olitics*, *t<sup>h</sup>ime*) and before stressed vowels (*p<sup>h</sup>olit<sup>h</sup>ic-ian*), i.e. also in monomorphemic strings? And how should stress placement be considered, given that monomorphemic strings of course do receive stress?

The numerous counter-examples where a monomorphemic string is transformed by phonological computation are acknowledged since Kiparsky (1968-73: 18) and Kiparsky (1973: 64), who concedes that "if a form appears in a constant shape, its underlying form is that shape, except for what can be attributed to low-level, automatic phonetic processes" (Kiparsky 1968-73: 18). So this is a yoker: computation cannot affect monomorphemes, except in case it is "low-level", "automatic" or "phonetic". Kiparsky (1968-73: 18) explains what that means: "these can be defined as processes which do not cause neutralization of distinct representations."

The full-bodied statement about monomorphemes as such is thus narrowed down quite a bit. But after having run the proposal against a more substantial body of empirical evidence, Kenstowicz & Kisseberth (1977) are pessimistic even regarding the weakened version:

"Kiparsky's principle may be too strong in that some rules of nonautomatic neutralization apply in nonderived contexts. If so, it is not immediately clear that there is a way to predict which rules will apply only in derived contexts and which will apply in nonderived contexts as well." Kenstowicz & Kisseberth (1977: 214)

Halle (1978: 132) agrees: "[t]o conclude it would appear that the special conditions on rules discovered by Kiparsky are unconnected with the automatic or nonautomatic character of the rule." More patches were added over the years as the SCC met more and more adversity, until Kiparsky (1993) himself declared its bankruptcy because it is empirically wrong. Bermúdez-Otero (2018: 102) believes that maintaining the SCC against all odds for so long has contributed a good deal to the discredit of Lexical Phonology, whose modern version (Stratal OT) needs to be cleared of this kind of historical burden.

The author may have wanted to address this historical and empirical record which played an important role in the development of the discipline. Discussing just one single example (NC homorganicity) where computation of a monomorphemic string can be avoided by inscribing homorganicity in the lexicon may be a bit short a motivation for the kind of overarching claim that is made. But there are two interesting properties of the particular version of the derived environment idea that the author exposes.

In both the Spanish and the Somali pattern discussed, a modification of word-final lexical nasals occurs according to the analysis of the author: in Spanish a nasal unspecified for place, /N/, appears as [n] (or [ŋ] depending on dialect) word-finally, while in Somali word-final /m/ is pronounced [n]. How is that possible in monomorphemic strings such as Spanish *Jua/N/* → *Jua[n]* and Somali */laam/* → *[laan]*? The author explains that morpheme boundaries such as the end of the word # constitute independent morphological information. That is, *laam* + # is not a monomorphemic string, but rather morphologically complex and therefore a legitimate target of phonological computation.

This is a new interpretation of what counts as a derived environment: in the classical literature on derived environments mentioned, only plurimorphemic strings are morphologically complex; a single morpheme with its boundaries is morphologically underived. The question that the author needs to answer, then, is what kind of situation could count as morphologically simplex at all when boundaries make a morpheme morphologically complex. For morphemes cannot occur without boundaries. Isn't it the case, thus, that there is no such thing as morphologically underived strings? What would be an example of such an item? If there are no, the article would lose its object of inquiry: monomorphemic strings. I am afraid the move made by the author to count boundaries as morphemes may obliterate the subject matter of the article and reduce its empirical content to nothing.

The other interesting property of the specific version of derived environments defended by the author is the difference in kind that is made between Spanish /N#/ → [n] (or [ŋ]) and Somali /m#/ → [n]. The author explains that while the latter operation is true phonological computation, the former "is the unmarked realization of the archiphoneme /N/" (p. 300). There is thus some operation distinct from phonological computation that turns stored /N/ into its unmarked realization [n]. The author does not dwell on the nature of this operation, but in a modular environment the obvious candidate is spell-out, i.e. the conversion of the output of phonological computation into phonetics items (Scheer 2014). This conversion is list-based: rather than an input being turned into an output by an algorithm (computation), there is a list of correspondences like in a dictionary where pairs of phonological and phonetic items are defined. Just like past tense ↔ -ed in English is a dictionary entry of the spell-out that converts morpho-syntactic into to phonological units, N ↔ [n] is an entry of the Spanish spell-out dictionary that associates phonological and phonetic items.

Spell-out may thus be a candidate for the "low-level", "automatic" or "phonetic" processes that the original derived environment literature has identified as contravening evidence. At the time modularity and hence spell-out were not an issue, but today the

clear distinction between modular (phonological) computation and the conversion of its output into the vocabulary of the following module may make sense of the two types of modification experienced by stored items.

**Anderson, S.** (1985) *Phonology in the Twentieth Century*. Chicago: University of Chicago Press. **Bermúdez-Otero, R.** (2018). ‘Stratal Phonology’, in Hannahs, S. J. & Bosch, A. R. K. (eds.) *The Routledge handbook of phonological theory*. Abingdon: Routledge, pp. 100-134. **Halle, M.** (1978). ‘Formal vs. functional considerations in phonology’. *Studies in the Linguistic Sciences*, 8, pp. 123-134. Reprinted in Brogyányi, Bela (ed.) 1979. *Studies in diachronic, synchronic and typological linguistics: Festschrift for O. Szemerényi*, 325-341. Amsterdam: Benjamins. **Kenstowicz, M. & Kisseberth, C.** (1977) *Topics in Phonological Theory*. New York: Academic Press. **Kiparsky, P.** (1968-1973). ‘How abstract is phonology?’, in Fujimura, O. (ed.) *Manuscript circulated since 1968 and published 1973 in: Three Dimensions of Linguistic Theory*. Tokyo: TEC, pp. 5-56. **Kiparsky, P.** (1973). ‘Abstractness, opacity and global rules’, in Fujimura, O. (ed.) *Three Dimensions in Phonological Theory*. Tokyo: TEC, pp. 57-86. **Kiparsky, P.** (1982). ‘From Cyclic Phonology to Lexical Phonology’, in Hulst, H. v. d. & Smith, N. (eds.) *The structure of phonological representations I*. Dordrecht: Foris, pp. 131-175. **WEB. Kiparsky, P.** (1993). ‘Blocking in nonderived environments’, in Hargus, S. & Kaisse, E. (eds.) *Studies in Lexical Phonology*. San Diego: Academic Press, pp. 277-313. **Scheer, T.** (2011) *A Guide to Morphosyntax-Phonology Interface Theories. How Extra-Phonological Information is Treated in Phonology since Trubetzkoy's Grenzsignale*. Berlin: Mouton de Gruyter. **Scheer, T.** (2014). ‘Spell-Out, Post-Phonological’, in Cyran, E. & Szypra-Kozłowska, J. (eds.) *Crossing Phonetics-Phonology Lines*. Newcastle upon Tyne: Cambridge Scholars, pp. 255-275.

#### REPLY

See reply p. 320.

---

**DISCUSSION WITH MARC VAN OOSTENDORP**

(MEERTENS INSTITUUT, RADBOD UNIVERSITY)

---

Van Oostendorp, Marc. 2020. discussion in: Brandão de Carvalho, Joaquim (auth.) “Representation vs derivation: The case for a modular view of phonology”. *Radical: A Journal of Phonology*, 1, 319-327.

**COMMENTS**

This paper seems to be a short essay on some of the differences between Optimality Theory on the one hand and Autosegmental Phonology (or Government Phonology) on the other. This is a very big topic, maybe one of the core issues in phonological theory of all times (who does not remember Stephen Anderson's history of phonology where this history is sketched as a pendulum between those two aspects), but given the sketchiness, I am not completely sure that I find the result very compelling.

First, it should be noted that not all scholars have treated OT and AP or even GP as incompatible or incommensurable; the author makes the claim that :

With some exceptions (like tonology or metrical phonology), the UR's used in OT-based research are identical to those of SPE, that is mere sequences of letters. [*This quote appears in the first draft of the manuscript and has not been retained in the final version of the article.* —Ed.]

But this seems false, or at least not true for all of the relevant literature. As a matter of fact, in the original Prince and Smolensky text, it is argued quite explicitly that OT builds on Autosegmental Phonology, and at various times there have been people showing that we can also combine OT and GP. One could say that these proposals were wrong, but I do not think one can ignore them. In the end, also the author seems to admit that we need derivation (at least for morphologically complex forms), and then the

question arises how those should be dealt with: maybe in terms of OT after all? If not, why not?

One problem one would face immediately, however, is a problem that OT was set out to solve, viz. that of 'doubling'. One can claim that within a morpheme derivations are not necessary, but one of OT's observations is that *the same* constraints often seem to hold morpheme-internally as in complex forms. One does not want to, and does not need to, stipulate those constraints twice. Again, one can say that this is the same strategy of Optimality Theorists, but one cannot ignore that this has been one of the central motivations for OT to begin with.

Another, related issue, is that OT and GP differ in one important way, viz. the treatment of variation, which in the former case is seen as constraint ranking, and in the latter often stipulated as 'parameter-based'. However, it seems to me that the latter only seems to seriously apply to the 'morpheme-internal' phonology. If for derivations we have simple rules, it means that there is a new form of variation, viz. which rules do and do not apply within a given language (e.g. Turkish has a roundedness harmony rule but French does not). That seems another type of harmony that seems rather ugly.

Yet, on top of all this, the author seems to propose a new form of doubling, by stating that maybe OT and GP are true at the same time, just like light behaves as particles and as waves at the same time. However, in the case of physics, this double nature is well described mathematically and we know in what sense we can expect particle-behaviour and when light looks more as organized in terms of waves. This precision is missing here, and the reader is actually left in the dark as to in what way OT is supposed to be right.

#### REPLY

I wish to thank the reviewers for their careful reading of the manuscript and their constructive remarks. Most comments revolve around three main topics: 1) the central sub-

ject of my squib: the distinction between representation and derivation; 2) the empirical challenges that my proposal faces; 3) my description/interpretation of phonological theories (structuralism, SPE, Autosegmental phonology/Government phonology – henceforth AP and GP –, OT), and the historical perspective drawn. I will address these topics in the same order.

## 1. Representation and derivation.

### 1.1. There is no competition between them.

*"(...) the ban on computation of monomorphemes decides the competition for these strings, but does not speak to plurimorphemic items. Here the situation is thus as before: for a given alternation there will be competing representational and computational accounts."* (Tobias Scheer, p. 314)

There is no competition whatsoever. There is no modification of underlying forms (hereafter UR) in the morpheme-internal situation. Spanish NC sequences involve underlying geminates: the consonant's P.A. is bipositional in the underlying form; it does not result from spreading. The same holds, say, for palatalization: in languages where front vowels palatalize adjacent consonants beyond what is expected universally (that is, at what Bermúdez-Otero calls the "phonologisation" level), the I-element is underlyingly associated with two positions as well. In other words, and more generally, there are no such things as "allophonic rules" (see abstract, p. 292), which actually result from the superseded view according to which URs are sequences of letter-like tokens. (And this is to be regarded as a crucial, albeit largely unexploited, outcome of AP.)

By contrast, cross-morpheme phonology does involve derivation: /m/ in Somali "becomes" [n] before coronal and word-finally. However, there is no competition either in



this case: only a computational account *explains* the selected input, representations coming into play only to the extent that they motivate the markedness constraints required by ruling out such constraints as NOBANANA or NOBLUEVOWEL!

Admittedly, if there are no allophonic rules in morpheme-internal phonology, it might be said, though, that URs involve computation, as they result from constraints which, for example, outlaw non-homorganic NC sequences in both Spanish and Somali; hence "a new form of doubling" since "THE SAME constraints [...] seem to hold morpheme-internally as in complex forms" (p. 320) (Marc van Oostendorp, also Heather Newell). I agree that representations do not come out of the blue; rather, they are ruled by constraints such as implicational markedness scales. What I meant is simply that since they do not derive from a more abstract underlying form, there is no "repair" to be considered. Markedness does not relate to optimality or well-formedness, but to formal complexity. Both unmarked and marked segments, syllables, etc. are "well-formed" insofar as an adequate grammar generates "not the best form, but just the form it generates" (Reiss 2017: p. 432). Returning to the Spanish case, as there is no underlying /NC/ (with a placeless N), homorganic and non-homorganic NC clusters do not compete against each other: the latter does not occur simply because Spanish doesn't admit fake clusters, which involve an entirely different representation.

On the other hand, as Spanish and Somali share "the same phonology", there is no "doubling" in morpheme-internal and cross-morpheme phonology. In both languages, what may happen in cross-morpheme computation follows from what is allowed morpheme-internally; one does not need to stipulate the same constraint twice (Marc van Oostendorp).

## 1.2. The ambiguity of boundaries.

*"In both the Spanish and the Somali pattern discussed, a modification of word-final lexical nasals occurs according to the analysis of the author: in Spanish a nasal unspecified for place, /N/, appears as [n] (or [ŋ] depending on dialect) word-finally, while in Somali word-final /m/ is pronounced [n]. How is that possible in monomorphemic strings such as Spanish Jua/N/ → Jua[n] and Somali /laam/ → [laan]? The author explains that morpheme boundaries such as the end of the word # constitute independent morphological information. That is, laam + # is not a monomorphemic string, but rather morphologically complex and therefore a legitimate target of phonological computation.*

*This is a new interpretation of what counts as a derived environment: in the classical literature on derived environments mentioned, only plurimorphemic strings are morphologically complex; a single morpheme with its boundaries is morphologically underived. The question that the author needs to answer, then, is what kind of situation could count as morphologically simplex at all when boundaries make a morpheme morphologically complex. For morphemes cannot occur without boundaries." (Tobias Scheer, p. 316-317)*

There is a misunderstanding about the word "boundary", which stems from the ambiguous usage of the symbol "#". True, morphemes have a beginning and an end. Do this beginning and this end constitute in themselves "*independent morphological information*", though? I do not think so. Spanish has a *morphologically underived string* /θjeN/ "hundred". It can be transcribed as /#θjeN#/ in order to note its limits. Nevertheless, the final # has no particular effect on /N/ unless this morpheme is in final position, where /N/ → [n] /\_# (*tengo cien*); otherwise, this "boundary", albeit still present, does not prevent /N/

from being pronounced [m] (*cien piedras*), [ɲ] (*cien yeguas*) or [ŋ] (*cien gallos*). In my view, only the hash sign required by the rule conveys "morphological information", not the one with which the generative tradition dresses up monomorphemic inputs. The same should hold for Somali /laam/, whose morpheme-final /m/ loses its labial feature under the effect of a "#" of the first type: /laam/ is underived; only [laan#] is derived.

## 2. Empirical challenges.

*"All versions of the derived environment idea met empirical challenges that the contemporary literature discussed in detail. Kiparsky's motivation for declaring monomorphemic strings uncomputable was the observation that processes such as trisyllabic shortening are triggered by (class 1) suffixes as in sane [ej] - s[æ]n-ity, but never within a root, although the triggering conditions may be satisfied: the first vowel of ivory and nightingale is pronounced [aj], not [ɪ] (as in Christ [aj] - christ-ian [ɪ]). But if processes do not apply to monomorphemic strings, what about nasal cluster simplification  $gn \rightarrow n$  in root-final position when the root occurs alone as in sign [n] (compare with si[gn]-ature)? Or the loss of g in /ŋg/ as in English long [ŋ] (as opposed to long-er [ŋg])? And what about aspiration of p,t,k in English, which occurs word-initially ( $p^h$ olitics,  $t^h$ ime) and before stressed vowels ( $p^h$ olit<sup>h</sup>ic-ian), i.e. also in monomorphemic strings? And how should stress placement be considered, given that monomorphemic strings of course do receive stress?" (Tobias Scheer, p. 315)*

I agree that the only example of NC sequences is far from sufficient to support the idea that there is no monomorphemic computation. What I found interesting enough for a squib is the difference between Spanish and Somali. But I admit that further research is needed. In particular, I think that all the counterarguments that followed Kiparsky's

conjecture should be re-examined under a dual concern: on the one hand, we should distinguish the facts that depend on morphology (which are the object of my paper) from those that are purely phonological (that is all the opacity-related examples that Kiparsky includes in his NDEB); on the other hand, and above all, we should track the false counterexamples.

These are of at least three sorts. First, there are all the low-level processes. I believe that I made myself clear when I explained why /m/ → [n] /\_# in Somali is a derivation, while /N/ → [n] /\_# in Spanish is not: the former process involves the loss of a distinctive feature, while the latter does not. Secondly, there are "accidental" alternations, in the sense that the variants can hardly be shown to follow from one single UR. *Pace* Tobias Scheer, I very much doubt whether *si*[n] – *si*[gn]ature is phonologically-conditioned. Thirdly, there are alternations that are mere artefacts of the classic unilinear representations. For example, the "loss" of /g/ in English /loNg/ → [loŋ] is, above all, that of a letter; "/Ng/" is nothing else than a melodic sequence associated with one single position – just as [mb], [nd], etc. in Spanish involve one single (place) feature associated with two positions; there is no derivation whatsoever. On the contrary, there is derivation, caused by the suffix *-er*, in *longer*, where /g/ is resyllabified onto an empty onset. I am afraid, once again (cf. §1.1 of this reply), that many criticisms of Kiparsky's idea have sometimes forgotten the lessons of autosegmental approaches.

### 3. Phonological frameworks and historical perspective.

Several topics concern the comparison of OT with earlier frameworks.

#### 3.1 Parameters vs ranking, and variation.

a. Is constraint ranking really better than parameter settings? Parameters are unattractive in my view (i) because they weaken the principles to which they are linked, and (ii) because the settings seem *ad hoc*. By comparison, the same markedness constraints can be

said to be differently ranked from language to language according to their ranking vis-à-vis faithfulness constraints; in other words, unlike parameter settings, the ranking of markedness constraints has an *extrinsic* motivation.

**b.** I do not understand Marc van Oostendorp when he writes that "[parameter-based variation] only seems to seriously apply to the 'morpheme-internal' phonology. If for derivations we have simple rules, it means that there is a new form of variation, viz. which rules do and do not apply within a given language" (p. 320). If I agree that representational approaches are more suitable for morpheme-internal phonology, this does not mean (i) that I am supporting the adoption of "principles and parameters" theories *as they stand*, and (ii) that derivation should involve *rules* accordingly.

**c.** I understand Heather Newell's view according to which "variation is extra-phonological and therefore should not be accounted for within a theory of phonology" (p. 310). I feel, however, that this remains an opinion, and I have a different stance: grammars generally model one arbitrarily chosen variety of a language; since the same speakers always employ more than one variety, (at least a certain type of) variation should be viewed as an integral part of grammar. At any rate, the line between phonology and variation is not of the same nature as those between phonology and morphology or between phonology and phonetics: there is no sociolinguistic module. I acknowledge that this sounds much like a profession of faith, but not more than Zwicky & Pullum's (1986) opposite claim. This was made explicit in the revised version.

**d.** Heather Newell says that "the examples discussed in Coetzee & Pater (2011) are not clearly directly comparable to the question of parameter settings with regards to Final Empty Nuclei, and so it is unclear what point this comparison is making" (fn. 10). I agree, I chose the wrong examples. If Coetzee & Pater's article is a good survey about the place of variation in OT, their examples, mostly based on the morphological condi-

tioning of phonological processes, are not comparable to the parameter setting on FENs in GP. This has been changed in the second version. Anttila's work, for example, seems much more relevant for the intended purpose.

**3.2. URs used in OT.** As noted by Heather Newell, "OT has not taken firm stands on questions of autosegmentality" (p. 308). True, there is no specific theory of phonological representation that 'goes with' OT, and I have replaced the issue of the URs used in OT with the problem of constraint motivation. That being said, usage matters, even if "in the original Prince and Smolensky text, it is argued quite explicitly that OT builds on AP, and at various times there have been people showing that we can also combine OT and GP" (p. 319) (Marc van Oostendorp). A typical instance of this is the rather systematic denial of any underlying status to the syllable, in line with SPE (and most structuralism).

**3.3. Unification.** True, as Heather Newell says, the question of unification is probably premature in practical terms: before overcoming the divide between AP/GP and OT, both camps have to be unified, as there are several representational and constraint-based frameworks. However, should this be given more ink in the paper? I am not quite sure: this squib was intended to be a simple essay on the complementarity of AP and OT beyond the fracturing of both approaches.

As Marc van Oostendorp observes, while "in the case of physics, this double nature [of light] is well described mathematically" (p. 320), this does not hold for phonology. True again, but phonology is in my view at the age of pre-Galilean physics, and it should be recalled that the dualism of light was not "well described mathematically" when it was just one of de Broglie's hunches.

**Reiss, Charles** (2017) 'Substance-free phonology', in S. J. Hannahs & A. R. K. Bosch (eds), *The Routledge handbook of phonological theory*. Oxford: Routledge, pp. 425–452.