radical | A Journal of Phonology
ISSN 2592-656X

volume 2, 2020

Volenec, Veno
Reiss, Charles

Formal Generative Phonology

Volenec, Veno & Reiss, Charles. 2020. "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 1-148.

Editor: *Shanti Ulfsbjorninn*
Reviewers: *Iris Berent, Joaquim Brandão de Carvalho, Alex Chabot, Guillaume Enguehard, Gillian Gallagher, John Goldsmith*

PAPER

FORMAL GENERATIVE PHONOLOGY*

VENO, VOLENEC (CONCORDIA UNIVERSITY)
REISS, CHARLES (CONCORDIA UNIVERSITY)

This paper outlines a program for the study of phonology as a branch of cognitive science. Building on the legacy of classical generative phonology and biolinguistics, it provides a theoretical framework that strictly differentiates phonological competence from aspects of articulation, acoustics and perception. We argue that phonological competence is to be characterized as a formal—that is, explicit, logically precise, and substance-free—manipulation of abstract symbols. We propose that a productive way to execute this program is to adopt a model called *Logical Phonology*, where phonological competence is described and explained by a system that maps between phonological data structures like strings via rules constructed from basic set-theoretic operations. We show the merits of this model by applying it to Turkish, Hungarian and English data. The remote and complex relationship between phonological competence and speech is elucidated by *Cognitive Phonetics*, which proposes that the outputs of phonology are transduced via two algorithms into temporally distributed neuromuscular activities. Taken together, *Logical Phonology* and *Cognitive Phonetics* aim to explain the nature of what is loosely referred to in the literature as 'the externalization of language' and to delineate its components.

generative phonology, I-language, substance-free phonology, rule, feature

* We are grateful to Johanna Benz, Noam Chomsky, Dana Isac and Björn Köhnlein for their constructive feedback on earlier versions of the paper. We do not imply that their views are fully aligned with ours, and we assume responsibility for whatever errors remain.

INTRODUCTION

The externalization of language—the conversion of the output of a syntactic derivation to a spoken utterance—is a process of great complexity. Because of the assumed autonomy of syntax from the components of this conversion (Chomsky et al. 2017: 18), the syntactic literature understandably lumps the whole process under a variety of interchangeable and somewhat vague labels such as 'AP' for the 'articulatory-perceptual system'; 'EXT' for 'externalization'; 'SM' for the 'sensorimotor system'; 'PF' for either 'phonological form' or 'phonetic form'; or a catch-all 'PHON'.[1] Syntacticians assume that it is the job of the 'sound' researchers, phonologists and phoneticians, to explain what is going on between the output of syntax and the mouth. 'Sound' researchers usually do agree that both phonetics and phonology are needed to account for the externalization of syntax, but there is very little agreement about how to characterize these domains and where to place the boundary between them. Furthermore, the theoretical assumptions of various scholarly communities—syntacticians, phonologists and phoneticians—are so mutually incompatible as to preclude a coherent account of 'EXT'. For example, nativist assumptions based on the Argument from the Poverty of the Stimulus and other considerations widely accepted in the syntactic literature have been increasingly ignored or rejected by phonologists (e.g., Archangeli & Pulleyblank 2015), despite parallel arguments in the two domains. We also find a chasm between phonologists, who tend to accept the necessity of feature-based analyses (whether or not they are nativists with respect to features), and scholars in phonetics and speech processing, who largely reject or ignore phonological features. This situation is comparable to

1 This claim should not be taken to imply that there cannot be any interaction between phonology and morphosyntax, but rather that from the syntactic output to the acoustic output of the body, there are complexities that are sometimes obscured in the syntactic literature, for example by calling everything from the output of syntax to phonetics 'the SM system' or 'the phonetic form'. To constrain the exposition in this paper, we will abstract away from possible interactions between these grammatical systems, interactions which are covered, for example, in *The Sound Pattern of English* (Chomsky & Halle 1968), more recently by Distributed Morphology (Halle & Marantz 1993; Harley & Noyer 1999), and by other work.

Pylyshyn's description of the complexities that arise when different research communities try to arrive at unifiable models:

> It turns out that the cognitive theorist enjoys considerable latitude in deciding which functions to attribute to the transducer and which to the symbolic processor. The boundary between transducer and symbolic processor can be shifted about, but this shifting carries with it profound implications for cognitive theory. It is possible to assign the deepest mysteries of perception and cognition to different places, and consequently to make the part one is concentrating on at the moment appear either surprisingly simple or unfathomably complex. [...] Cognitive processes can also be simplified by attributing complexity to the functioning of the transducer, but a price must be paid for this shift. The price [...] is that one can trivially explain anything that way. Unless what counts as transduction is constrained in a principled manner, the simplification of the problems of perception gained by postulating certain transducers has, as Bertrand Russell is reported to have said once, all the virtues of theft over honest toil. (Pylyshyn 1984: 148)

There is currently a high degree of fragmentation in the phonology community and the rapid proliferation of new theories is matched only by their mutual incommensurability. Even theories labeled as 'generative' tend to be completely incompatible. As we will argue below, we think that the main factor that contributes to such a fragmentation is a serious confusion about the object of study in phonology, i.e., a misconception about what it is that phonological theory should explain.[2] Of course, such a state of affairs makes it much more difficult to say something coherent about the relationship between phonol-

2 In the introduction to *The Routledge Handbook of Phonological Theory* (2018), Hannahs and Bosch offer some more specific reasons for this fragmentation: "Perhaps the most obvious proximate cause of disagreement in generative phonology is the advent of Optimality Theory (OT). [...] The popularity of OT among a large group of phonologists, alongside its failure to persuade or engage others, may be said to be responsible for some of the ensuing fragmentation within the field of phonology even among those phonologists who consider themselves generativists, in the sense of Chomsky (1957, 1965)" (p. 2). Additionally, they cite issues related to abstractness, increased interest in the sociolinguistic aspects of phonology, and the resurgence of empiricist methods as further contributing factors.

ogy and domains related to it, particularly phonetics, thus precluding a principled account of 'EXT'. The present paper aims to mitigate this fragmentation, first by defining the object of phonological study in a way that is consistent with the philosophical foundations of generative linguistics (Chomsky 1968; 1986; 2000), and then by outlining a research program for the study of that object.[3]

In the next section, we explore the theoretical and methodological consequences of adopting the position that the object of phonological study is completely internal to an individual's mind/brain. In the rest of the paper we then establish that what is loosely called 'externalization' in the literature consists of two parts: (1) *phonology* – a grammatical subsystem governed by rules; (2) *universal phonetics* – an extra-grammatical, invariant component that interprets the output of phonology and produces speech. (Its invariance entails that there is no phonetic system specific to, say, French vs. Swahili.) In other words, we decompose 'EXT' into a 'symbolic processor' and a 'transducer', to use Pylyshyn's (1984) terms. Section 2 sketches *Substance Free Phonology*, a formal theoretical framework that strictly differentiates phonological competence from all aspects of articulation, acoustics and perception. In section 3, we outline a working model for substance-free phonology called *Logical Phonology*, which uses basic set-theoretical notions to describe the representational and computational aspects of phonological competence. *Cognitive Phonetics*, our conception of the phonology-phonetics interface, is presented in section 4. Building on the work of Lenneberg, Marr, Pylyshyn, Halle and others, it proposes a universal transduction system, whose inputs, crucially, are the feature-based surface representations of generative phonology. Summary and conclusions are given in section 5.

3 The main rationale for such a broad goal is that arguing about more specific phonological problems and their explanations is highly unproductive unless we at least agree about what constitutes 'a phonological problem' (as opposed to, say, a phonetic one). For example, we think it is not particularly useful to argue about phonological gradience unless we first agree on *where and in what form* gradience is supposed to reside. In the mind? In the movement of the articulators? In the acoustic signal? In all of these domains? Somewhere else? Once we have clearly defined our object of study, we will be in a better position to address such specific problems productively (for example, see section 1.1 where we return to the issue of gradience after defining the object of our study).

While the present study shares its theoretical orientation with substance-free phonology of Hale & Reiss (2008) and related work, particularly in the form of looking at phonology from the I-language perspective, it goes beyond that work and offers the following novel ideas:

- we use the cognitive neuroscience framework outlined in Gallistel & King (2010) to look at phonological features in a neurobiologically realistic way (§2);
- we draw a parallel with amodal completion in vision studies (Michotte & Burke 1951) to apply the Argument from the Poverty of the Stimulus to phonology and argue for the innateness of features (§2);
- we provide arguments against recent influential work (e.g., Mielke 2008; Archangeli & Pulleyblank 2015) that makes problematic claims about the nature of features and rules (§2);
- we use set theory to model phonological representations and computations, exploring the use of priority union in derivations, an idea we apply to natural language data from Hungarian and Turkish (§3);
- we apply David Marr's (1982) insights on the nature of information-processing systems to construct a phonology-phonetics interface theory (§4);
- we argue that our 'Logical Phonology + Cognitive Phonetics' combination eliminates the need for a language-specific phonetics module in the grammar, thus simplifying the sequence of conceptual steps needed to account for the externalization of language (§4).[4]

4 This paper is clearly programmatic in nature, and as such, it does not argue *directly* against approaches to phonology, phonetics or their interface that differ from those that are proposed here. The reader should also keep in mind that some of our arguments—for example, for the innateness of features—may be relevant to selecting between, say, competing versions of Optimality Theory or Government Phonology. Our claim is not that the argument we give leads inexorably to our 'Logical Phonology + Cognitive Phonetics' model, but rather that we have made some progress in defining a coherent split between phonetics and phonology and also provided a sketch of a promising model of their interface.

1 PHONOLOGY AS A BRANCH OF COGNITIVE SCIENCE

1.1 THE OBJECT OF STUDY IN PHONOLOGY: I-LANGUAGE

George Miller (1998) noted that the most profound consequence of the cognitive revolution of the late 1950s was the realization that in any branch of cognitive psychology, including linguistics, observable behavior, such as spoken utterances, is evidence, and not the object of study. The actual objects of study are the cognitive systems that underlie observable behavior. The utterances are merely one kind of evidence for the cognitive systems.[5] To put it in the bluntest terms, then, phonology is *not* the study of speech, but speech provides evidence for the study of phonological competence.

Phonology was included in the cognitive revolution right from the outset, in recognition of the crucial distinction between its object of study (phonological competence) and its main source of evidence (speech production and perception) (Chomsky 1951, 1957a, 1957b, 1964; Chomsky *et al.* 1956; Halle 1959, 1962, 1964). However, despite the clarity of this early work, phonology has not been consistently treated as a branch of cognitive science. Instead, there have always been strands of work that adopt the formal trappings of generative phonological theory without adopting its mentalist commitments. This attitude seems to stem from both a failure to engage with the philosophical foundations of the cognitive revolution, and also from a related preoccupation with describing linguistic *behavior* and not the unobservable mental competence that makes that behavior possible. This preoccupation is undeniably gaining in popularity in recent years,[6] as

5 Miller's (1998) observation should not be understood as suggesting that cognitive science is different from other sciences in this respect, although there is some confusion about this matter. For example, Peter Norvig, former Director of Research at Google, has specifically critiqued the fact that in generative linguistics the object of study is *not* a set of observations: "I can't imagine Laplace saying that observations of the planets cannot constitute the subject-matter of orbital mechanics, or Maxwell saying that observations of electrical charge cannot constitute the subject-matter of electromagnetism" (quoted from http://norvig.com/chomsky.html). We think, on the contrary, that neither Laplace nor Maxwell would describe their own *observations* as the object of study. Observations are subject to margins of error, faulty equipment, observer bias, and so on. The laws of nature are not.

6 This trend has recently been characterized as "[t]he present regrettable return of neo-empiricism, in the shape of Bayesian learning models, statistical regularities, collective exchanges and the extraction

witnessed by the pervasiveness of output-driven models of phonology and by the prevalent orientation towards 'the surface' throughout the field. What muddles this issue even further is the fact that a vast amount of phonological work is either indifferent or agnostic about whether phonology is concerned with the mind/brain or with sound patterns that just exist 'out there'.

Since the reasons for the paradigm shift in the late 1950s were to our knowledge never disputed, we see no reason *not to* adopt strict mentalism in phonology and accept *all* accompanying philosophical and methodological commitments that come with treating phonology as a branch of cognitive science. In the rest of this section we will therefore provide a mentalistic definition of our object of study and explore the methodological implications that such a definition has for phonology.

Our object of study, *phonological competence*, is an aspect of *I-language* (Chomsky 1986). Here, the prefix *I-* stands for several key properties that characterize this technical notion of language. First, language is *internal* to the human mind. Since the mind is just the functioning of the brain viewed from a sufficiently abstract point of view, language can ultimately be regarded as a biological object—a certain subsystem of the brain. Consistent with the conceptual shift Miller referred to, the object of our study is fully internal to a speaker and not something 'out in the world', belonging to a 'speech community' in some vague sense. Second, following directly from the first, language is a property of an *individual*. Since no two individuals have exactly the same experiences during the period of language acquisition, it is not surprising that they end up with I-languages that differ from each other. To the extent that they are similar, language-based communication between individuals occurs with more or less facility. The question also arises of why, despite these differences of experience, individuals may acquire very similar I-languages. In other words, by recognizing I-language as the object of study, we can start to ask what factors determine the plasticity of the human language faculty. Third, language is an *intensional* system, which means that it is properly characterized

of generic cognitive patterns, recurrent networks, and similar notions." (Piattelli-Palmarini 2018: 211)

by means of a finite set of precise statements that describe the generative capacity of that system. Language characterizes, in terms of rules or functions, the members of an unbounded set of structured expressions; these members are not listed in long term memory. In every I-language there is an infinity of members (sentences), so defining the system extensionally, by listing all of its members, is impossible.

Additionally, language is a kind of *implicit* or *tacit knowledge*. Its content is predominately below the level of consciousness and beyond the reach of introspection. The main task of linguistics is to provide an explicit characterization of this implicit knowledge system in terms of a framework that can account for all possible I-languages.[7] The speakers' judgments about their own linguistic output can be considered an important source of evidence for discovering the nature of this knowledge, but there is no reason to think that a speaker can report perfectly (i.e., completely reliably) on the contents of their I-language.

Crucially, the term *I-language* does not include the notion *innate*, although *nativism*, the existence of a non-trivial genetically determined component of the human language faculty, has also been a tenet of most proponents of the I-language perspective. In brief, one could imagine a cognitive psychologist who accepts that humans have I-languages encoded in their brains based on some kind of all-purpose pattern discovery capacity, without any genetically determined substrate that is specific to language. For such a psychologist, an I-language would be parallel to a person's knowledge of the rules of chess. So, I-language does *not* imply nativism. However, there *is* an implication in the other direction: if one accepts the existence of an innate aspect of the language faculty, then one also accepts, at a minimum, the internalist part of the I-language perspective, because whatever is innate must also be internal to an individual's brain. It follows, then, that a rejection of internalism entails a rejection of nativism. This logic will be important as we demonstrate, first, that nativism is not universally accepted by phonolo-

7 "The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, with no specific reference to particular languages." (Chomsky 1957: 11)

gists, and second, that the rejection of nativism can be tied to a more or less explicit rejection of a crucial aspect of internalism (see sections 2 and 3).

It is of paramount importance to distinguish—both in theory and in practice—between a cognitive system and the use of that system for a particular purpose on a particular occasion. That is, we must distinguish between I-language and the *use* of I-language. To be able to do that clearly, linguistics provides the well-known dichotomy between *linguistic competence* (an I-language) and *linguistic performance* (the use of I-language in, say, communication). Linguistic competence is always a core factor in any act of linguistic performance, interacting with many other linguistically irrelevant factors such as a speaker's level of fatigue, creativity, having a sore throat, free will, or being distracted by a mobile phone. Competence is present even when it cannot be used in acts of performance, as demonstrated, for example, by numerous studies documenting recovery from post-stroke aphasia (Pinker 1994; Hillis 2010; Tippett & Hillis 2016), and by infant comprehension studies which show that children understand speech at a level that far exceeds their ability to produce it (Hirsh-Pasek & Golinkoff 1996; Lust 2006; Smith 2010; Bergelson & Swingley 2012). Since competence is always at the core of every performance act, but crucially not vice versa, we can say that competence has *epistemological priority* over performance: We must first learn about the nature of competence in order to be able to undertake a scientific study of performance.

Language (competence, I-language) is a module of the human mind.[8] A *module* is an informationally encapsulated cognitive system governed by principles not completely shared by any other module. In line with the computational-representational conception of the mind, each module consists of a finite set of primitive, atomic symbols (basic representations) and a finite set of functions that manipulate those symbols (basic computations). As we will show in more detail in section 2, we can also make concrete hypotheses about the symbols and functions of phonological competence.

Different modules communicate via *interfaces*: configurations in which the outputs of one module serve as the inputs to another module. Manipulation (e.g., reordering, regrouping, deletion, addition) of representational elements within a module, and without a change in the representational format is called *computation*. Computation occurs within a single module. Conversion of an element in one form into a distinct form, i.e., a mapping between dissimilar formats is called *transduction*. Transduction occurs at the interface between different modules, and it is a way in which modules communicate

8 Language is modular in two ways: It is functionally separable from other, non-linguistic aspects of the mind, and it itself consists of (sub)modules. The main source of evidence for the modular nature of a cognitive capacity is *double dissociation*: If it can be demonstrated that system A can be impaired without affecting the functioning of system B, and if, on a separate occasion, it can be demonstrated that system B can be impaired without affecting the functioning system A, then systems A and B constitute distinct modules. There is a massive amount of empirical evidence of this kind pertaining to the modularity of language. Language is doubly dissociated from acoustic processing (Zatorre et al. 1992; Burton et al. 2000; Phillips et al. 2000; Dehaene-Lambertz & Pena 2001; Dehaene-Lambertz et al. 2002), from attention (Pulvermüller et al. 2008; MacGregor et al. 2012; Blank et al. 2014), from visual and spatial cognition (Emmorey et al. 1993; Wang et al. 1995; McGuire et al. 1997; Emmorey 2002), from 'general intelligence' (Lenneberg 1967; Rondal 1995; Curtiss 1982, 1988a, 1988b, 1995, 2011; Smith & Tsimpli 1995; Smith et al. 2010); from non-linguistic communication (Corina et al. 1992; Willems et al. 2009; Grosvald et al. 2012); from arithmetic competence (Schaller 1995; Grinstead et al. 1997; Brannon 2005; Curtiss 2011); from social cognition (Karmiloff-Smith et al. 1995; Smith 2003; Smith et al. 2010; Gernsbacher et al. 2016), from pragmatics (Bottini et al. 1994; Champagne-Lavau & Joanette 2009). In all of these cases, a double dissociation between language and other cognitive systems has been experimentally demonstrated. Furthermore, there is ample evidence for the separability of systems *within* linguistic competence: for the lexicon (Hart et al. 1985; Caramazza 1988; Jodzio et al. 2008); for semantics (Binder et al. 2009; Skeide et al. 2014; Patterson & Ralph 2016); for phonology (Phillips et al. 2000; Leonard 2017); and for syntax (Grodzinsky 1986; Grodzinsky & Finkel 1998; Friedmann et al. 2006; Bastiaanse & Zonnefeld 1998; Bastiaanse & Thompson 2003; Buchert et al. 2008; Friederici 2017). In light of this evidence, it seems difficult to maintain a non-modular conception of the mind.

with each other. So, modules compute, interfaces transduce. Computation within the phonological module is described in sections 2 and 3, while transduction of information between the phonological module and the system for speech production is described in section 4 on Cognitive Phonetics.

Since language is a module of the mind and therefore a subsystem of the brain, the maturation of language follows from the general maturation of the brain (Lenneberg 1967; Friederici 2017). At least two important states in language growth can be discerned: the initial state and the attained mature state (Chomsky 1976: 3). The initial state is the totality of the biologically predetermined (innate) language-related units and operations. The attained state is an I-language. As in the maturation of any other biological system, three factors play a role in the growth of language: innate factors (genetic endowment), experience (environment), and more general laws of nature (see Chomsky 2005 for details). Since different brains will be exposed to different linguistic experience during maturation, the attained states can be different (hence linguistic variation), but they can only vary within the range determined by the innate factors. An explicit characterization of the initial state of language growth is called *Universal Grammar* (UG). 'Universal' here refers to innate, biologically given cognitive capacities, not to surface-true generalizations about languages (understood non-technically) in the sense of Greenberg (1966) and related work. In other words, the main task of Universal Grammar is to precisely characterize the notion of a *possible human language*, without accounting for what is frequent, common, attested or currently attestable. An explicit characterization of the mature state of a language is called a *generative grammar*. Since phonological competence is a part of the mature state of language (see sections 2 and 3 for details), its description should be included in a generative grammar.

Because it is an aspect of the mind, an I-language is not directly observable.[9] Therefore, the main source of evidence for I-language is linguistic performance or the *use* of I-language (e.g., in speaking or understanding speech). Again, it is important to recognize clearly the difference between the *object of study* in linguistics (in all of its branches) and the *sources of evidence* for that study. The object of phonological study is phonological competence, which is a part of an I-language. The primary source of evidence bearing upon that object of study are spoken utterances. But spoken utterances, or speech more generally, are *not* the object of phonological study because by the time the human body outputs a spoken utterance, phonological competence has been obscured by numerous factors external to linguistic competence (e.g., transduction procedures, rate of speaking, wearing dental braces etc.). Not discriminating between the object of study and evidence for that object leads to a serious confusion that hinders progress in both phonology and phonetics.

An example of this confusion can frequently be observed in discussions of gradience in phonology. The following is a fairly standard definition of 'categoricality' vs. 'gradience', and by emphasizing certain words in it, we wish to draw the reader's attention to the conceptual level at which the definition is given:

> [C]ategorical *sounds* (...) are stable and represent clear distinct phonological categories (e.g. *sounds* showing all characteristics of voiced segments *throughout* their *realizations*) (...); gradient *sounds* (...) may change *during* their *realization* and may simultaneously represent different phonological categories (e.g. *sounds* that *start* as voiced and *end* as voiceless). (Ernestus 2011: 2115)

9 The only sense in which an I-language could in principle be directly observed is in terms of brain activity. Since our current understanding of the neurobiological substrate of all modules of cognition, including I-language, is still very limited, the insight about the nature of I-language that can be gained on these grounds is also very limited.

From the emphasized words it is obvious that this characterization of categoricality and gradience is relevant to the domain of speech (performance), not grammar (competence). A problem arises when phonetic data are used to make inferences about phonology *directly*, as if every idiosyncratic datum recorded in speech or found in a corpus is relevant for phonology, without acknowledging the distance between competence and performance. Consider another passage from Ernestus (2011: 2118):

> Ellis and Hardcastle (2002) found that four of their eight English speakers showed categorical place assimilation of /n/ to following velars in all tokens, two speakers showed either no or categorical assimilation, and two speakers showed gradient assimilation. Together, the data show that place assimilation processes (...) may be gradient in nature. These processes cannot simply be accounted for by the categorical spreading of a phonological feature from one segment to another.

The cited results, showing inter- and intra-speaker variation, as well as both discrete and gradient effects, may constitute a salient illustration of the ubiquitous lack of uniformity in the *behavior* of members of a speech community, but it is not in the purview of phonology to provide an explanation of such phenomena. The fact that such variation "cannot simply be accounted for by the categorical spreading of a phonological feature from one segment to another" (*ibid.*), a claim most certainly true, does not mean there is something wrong with phonology conceived as categorical symbol manipulation. A plausible explanation of these results is that gradience is introduced into the data by the post-phonological transduction procedure that determines the temporal coordination of muscular activity (see section 4), and not by phonological computation itself. Notice the references to time highlighted in the above quote from Ernestus (2011: 2118), for example, "during" and "start as... end as". Gradience involves change over time. If we think of phonology as involving a representational system (features and the like) and a computational system that can be regarded as a complex function of, say, composed rules

(see sections 2 and 3), then there is no temporal aspect to phonology.[10] Questions about gradience in phonology are like questions about how fast a *wh*-element moves in syntax; both reflect a category error. The only kind of conclusion a phonologist can draw from the Ellis & Hardcastle experiment cited by Ernestus is that the I-language of some English speakers contains a feature-based rule that changes coronal nasals to velar nasals before velar consonants. If a featural assimilation rule correctly models a part of a speaker's phonological competence, a phonetician can then posit hypotheses as to why such a pattern exists, why there is variability in externalization of this knowledge, what are the limits of its variation, whether the variation is purely biomechanical or partly/mostly/solely cognitive, and so on. Adopting such a perspective not only preserves a clear distinction between competence and performance, a necessity on many different grounds, but it also facilitates disentangling phonological conclusions from phonetic conclusions even though both are drawn from the same data.

1.2 CONSEQUENCES OF THE I-LANGUAGE PERSPECTIVE

Due to the nature of its object of study, generative linguistics assumes internalism, mentalism, realism, naturalism, and rationalism. These '-isms' are philosophical stances that ultimately determine an overarching research method for the study of language that has proved fruitful.

As we have seen, *internalism* refers to the idea that the object of linguistic study is internal to the mind/brain, and that there is no coherent, scientifically useful notion of language external to the human mind/brain. Surely, there is no phonologist who thinks that language is completely outside of people's brains. But just believing that there is something in the brain is not enough to make a linguist or a psychologist an internalist. An internalist claims that language is *only* in the mind/brain. As radical as this sounds, it is trivial to illustrate with a basic linguistic notion such as *word*. Inspection of the wave-

10 Which is *not* to say that precedence relations, and (multi)linear ordering of phonological elements more generally, are irrelevant for phonology (see Idsardi & Raimy 2013; Papillon 2020).

form of utterances demonstrates that there is no consistent correlation between the perceived boundary between words and the silent parts of the signal. The waveform of a spoken word like *stand* has silence between [s] and the release of [t], and a phrase like *There's Mary* has no silence before the [m] of *Mary*. Syntactician Howard Lasnik (2000: 3) has made this point forcefully:

> The list of behaviors of which knowledge of language purportedly consists has to rely on notions like 'utterance' and 'word'. But what is a word? What is an utterance? These notions are already quite abstract. Even more abstract is the notion 'sentence'. Chomsky has been and continues to be criticized for positing such abstract notions as transformations and structures, but the big leap is what everyone takes for granted. It's widely assumed that the big step is going from sentence to transformation, but this in fact isn't a significant leap. The big step is going from 'noise' to 'word'.

As Ray Jackendoff (1990: 164) puts it, "Language, as far as I can tell, is all construction." In the domain of phonology, the point has been made before: "It should be perfectly obvious by now that segments do not exist outside the human mind." (Hammarberg 1976). And even earlier, Edward Sapir (1933/1949: 46) recognized that the same principle holds for all domains of human cognition, stating that "no entity in human experience can be adequately defined as the mechanical sum or product of its physical properties." The internalist perspective thus maintains that the objects of cognition are better understood as our parses of the mind-external word, rather than as the mind-independent 'furniture' of that world (cf. Chomsky 1980: 218f). The related nativist perspective is that without the innate categories with which to parse the input, no learning can take place: "In any computational theory, 'learning' can consist only of creating novel combination of primitives already innately available" (Jackendoff 1990: 40, echoing Fodor).

Accordingly, linguistics is *mentalistic* in the sense that rather than being concerned with observable verbal behavior, it is concerned with 'mental aspects of the world' which stand alongside its mechanical, chemical, optical, and other aspects. The fact that the boundaries between mental and non-mental phenomena are not always completely clear is normal and unproblematic. As the French philosopher Pierre Jacob (2010: 219) put it, "no philosopher of physical sciences believes that he is expected to offer a criterion for what constitutes mechanical, optical, electrical or chemical phenomena." Why, then, should mental phenomena be different?

Realism is just the reflection of the fact that language is a real aspect of the world: It is a certain subsystem of the human brain, and when the functioning of the brain is viewed from a sufficiently abstract point of view, language can be regarded as a mental object. Realism commits us to admit that the apparent dualism between mind and brain is merely a result of our current inability to explain how an activity of neurons can give rise to thoughts, not a result of the belief that there indeed are two categorically different entities—the mind *is* the brain (cf. Mountcastle 1998: 1).

Since its object of study is a part of the natural world on a par with its mechanical, chemical, optical and other parts, linguistics can be regarded as a natural science. In his seminal book *Biological Foundations of Language* (1967), Eric Lenneberg, one of the founders of biology of language, noted that "we may regard language as a natural phenomenon—an aspect of our biological nature, to be studied in the same manner as, for instance, our anatomy" (p. vii). This *naturalism*, then, has certain methodological consequences, particularly in suggesting the validity of the use of standard research tools of the natural sciences such as simplification, isolation and idealization in linguistic inquiry. Thus, just as in physics, it is perfectly acceptable, and in fact necessary, to study language and its various subsystems in isolation and in a highly idealized setting in order to first gain an understanding of its fundamental nature. Only once that has been achieved to a reasonable degree can it be useful to introduce further complexity by

studying the interaction of language with other systems. This view concerning the isolability of grammatical components follows that of philosophers of science such as Lawrence Sklar (2000: 54–55), who has even made the point that without such isolability, science itself would probably be impossible:

> [W]ithout a sufficient degree of isolability of systems we could never arrive at any lawlike regularities for describing the world at all. For unless systems were sufficiently independent of one another in their behavior, the understanding of the evolution of even the smallest part of the universe would mean keeping track of the behavior of all of its constituents. It is hard to see how the means for prediction and explanation could ever be found in such a world [...] It can be argued that unless such idealization of isolability were sufficiently legitimate in a sufficiently dominant domain of cases, we could not have any science at all.

The standard *rationalist* stance is that the human mind contains innate knowledge—units and operations that exist in the mind independently of experience—and that this innate knowledge is a prerequisite for any kind of learning to take place: "the dimensions in terms of which experience is encoded cannot themselves be learned" (Jackendoff 2015: 187). We have already alluded to two arguments for nativism in phonology (see sections 2 and 3 for further discussion). First, there is the logical argument. Suppose that learning a given language involved learning that noun phrases like *dogs* come after adpositions like *with*. Such a rule can only be learned if the learner has access to the categories 'noun phrase' and 'adposition'. Similarly, a phonological rule like 'delete a consonant before another consonant (but not before a vowel or at the end of a word)' can only be learned if the category 'consonant' can be applied to the data. Second, as suggested above, there is no guarantee that the categories of language have any consistent articulatory or acoustic correlates that enable those categories to be applied or triggered. For example, there are no obvious physical properties that correspond to the set

of speech sounds [s, p, g, r, m, l] but not [i, u, a] and the end of a word: a *completely* unbiased scientist would not be able to derive the category 'consonant' from the signal. The degree of constriction in the vocal tract might be invoked as the relevant criterion here, but the cut-off point between the categories 'consonant' and 'vowel' is arbitrary and must be predetermined for the parsing to begin. For a human speaker, the innate category is applied to a certain input, thus determining that consonants constitute a 'natural class'. This problem of a *lack of invariance*—an absence of invariant properties that correlate with linguistic categories—is well known in the speech perception literature. Philosopher Irene Appelbaum (2006) suggests that there has been no progress in its solution despite half a century of work because the wrong question is being asked (see section 4).

The rationalist approach to the study of language is also based on the view that the data that we use to reason about language, e.g. the data that come from linguistic performance, represents the object of study in an *indirect* way. The reasons for this indirectness are clear: on the one hand, performance yields data that are necessarily intertwined with data about other cognitive systems and sensorimotor processes, and on the other hand, the data are qualitatively impoverished in comparison to the competence that underlies them (in interaction with other systems). It is therefore not at all obvious what constitutes 'data' for linguistics since (a) its object of study is not something directly observable, and (b) data do not come with a label that says that they are relevant for a particular object of study. For a phonologist qua cognitive scientist, it is necessary to understand that not all data from speech are relevant for phonology, and that it is therefore necessary to peel off the various complications that were introduced in the process of externalization from the underlying system of linguistic knowledge. In every scientific domain, it is the theory that determines whether something counts as relevant data or not. In other words, "we have to remember that what we observe is not nature in itself but nature exposed to our method of questioning" (Heisenberg 1962: 58). We have already seen the repercussions of this fact in the discussion of gradience in the preceding

section. Thus, while linguistics *is* an empirical science, it gives epistemological precedence to theoretical constructs of great explanatory depth over the observables. Again, this is a perfectly normal state of affairs in the natural sciences. For example, the existence of black holes was maintained on purely theoretical grounds for at least a hundred years, before the first ever direct image of a black hole and its vicinity was published on April 10th, 2019 (Akiyama et al. 2019).

To summarize, the object of study in phonology is phonological competence, a part of I-language and a module of the mind, which is to be studied by the classic, well-developed rationalist methods of natural science. Speech is merely one source of evidence for the study of that object, and it is always necessary to disentangle the phonologically relevant data from irrelevant artefacts of linguistic performance by giving priority to coherent theory over (putatively) unfiltered observation.

2 THE THEORETICAL FRAMEWORK: SUBSTANCE FREE PHONOLOGY

Phonological competence is an aspect of mind that consists of elementary units and operations not completely shared by any other module of the mind, linguistic or otherwise. Phonological units are primitives from which the representational aspect of phonology is built, while the operations constitute the computational aspect of phonology. The conceptual minimum needed to describe phonological competence consists of two levels of representation and an operational part that connects these two levels, yielding a derivational scheme that is characteristic of most work in modern phonology:

underlying phonological representation
↓
operational part of phonology
↓
surface phonological representation

} Phonological competence

In a fairly standard conception of the architecture of linguistic competence, the lexicon provides the syntactic component with minimal meaning-bearing linguistic units, morphemes, from which hierarchical structures are built. At the syntax-phonology interface, these hierarchical structures are flattened into linear strings of morphemes. A linearized string consisting only of the abstract forms of morphemes corresponds to the *underlying phonological representation* (*UR*). The *surface phonological representation* (*SR*) is a data structure which the phonetic implementational system converts into movements of the speech organs. Since they are part of the same module, URs and SRs are composed of the same type of units. These units can thus be said to comprise the *universal representational alphabet* of phonology from which all I-languages draw. The phonological representational alphabet consists of *phonological features*, and other distinctive, irreducible units of phonology such as *tones*, *moras* and *stresses*, i.e., elements of *prosody*. To constrain the exposition, we will primarily focus here on the nature of features, leaving the prosodic units aside.

The nature of phonological features can be elucidated by adopting the cognitive-neuroscience framework proposed by Gallistel & King (2010). There, what we have so far informally referred to as *units* are called *symbols*. Symbols are "physical entities in a physically realized representational system" (Gallistel & King 2010: 72), where the physical system in the case of phonological symbols, and all other cognitive symbols, is the human brain. Thus, phonological *features* are symbols realized in the human brain. The common properties of all neurobiological symbols are (at least) distinguishability, combinability and efficacy.

The standard assumption in cognitive neuroscience is that different symbols are *distinguished* by place coding of neural activity, rate coding, time coding, or, most likely, some combination of those. Of course, we are still far from being able to state how exactly features qua neural symbols are realized in the brain, but promising work such as Phillips et al. (2000), Scharinger et al. (2012), Hickok (2012), Bouchard et al. (2013), Mesgarani et al. (2014), Magrassi et al. (2015) and Scharinger et al. (2016) is emphasiz-

ing the importance of neural activity in the superior-most part of the superior temporal gyrus and sulcus, and Brodmann areas 44 and 6.

Features also meet the criterion of *combinability*: the hallmark of modern phonology is the notion that features can be grouped to construct higher-level, non-atomic data structures. An unordered, unstructured set of features constitutes a phonological *segment* (see section 3 for details), while a particular organization of segments constitutes a data structure of a higher taxonomic level, such as a *syllable*. This combinability of features allows phonology to construct complex symbols from an inventory of simple parts, and provides an explanation for the so-called *natural class behavior*—different structures can behave alike because they contain identical substructures.

Features are also an *efficacious* way of coding information since their combining leads to combinatoric explosion (Reiss 2012; Matamoros & Reiss 2016). For example, if we assume that UG provides the brain with only 30 binary features with the possibility of underspecification, then from this small set of primitive symbols we can construct 206 trillion (3^{30}) different segments. With a number of the same order of magnitude, say fifty, the number of definable segments rises to 3^{50} (about $7x10^{23}$). As the UG feature inventory grows linearly, the set of intensionally definable segments grows exponentially. Of course, the richness that arises from feature combinability should not be taken to imply that any particular I-language should come close to exploiting the full range of possibilities. Instead, what we expect to find in particular I-languages is in line with the traditional view of feature combination: “no language has as many phonemes as there are possible combinations of the utilized distinctive features” (Halle 1954). The same point is made in the context of genetics by biologist Peter Medawar (1960: 37):

> The potential variation of human beings is enormously greater than their actual variation; to put it another way, the ratio of possible men to actual men is overwhelmingly large.

A corollary of this combinatoric explosion is that such richness goes a long way towards eliminating the need for a phonetics module specific to each language (e.g., Keating 1985; 1990), which simplifies the sequence of conceptual steps needed to account for the externalization of language.

Our outlook on the nature of phonological features is an elaboration of a long tradition in generative linguistics, where the general research agenda is to build biolinguistically plausible cognitive models of language. It is therefore not surprising that Morris Halle already in the 1980s proposed that features be understood in terms of human neurobiology: "[T]he distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems" (Halle 1983/2002: 109). Even though some progress has been made in making Halle's proposal concrete (e.g. Mesgarani et al. 2014), we are still far from being able to refer to phonological features (or any other cognitive category) by stating their neurobiological substrate and therefore have to resort to using (linguists') symbols to refer to (mental) symbols. So when we write, for example, LABIAL, we use a sequence of letters to form a symbol for a particular feature, which in turn is also a symbol, just in the brain. To reiterate, LABIAL is a (non-neural) symbol for a (neural) symbol. We, the researchers, need these phonetic labels to know what we are talking about, the brain does not. The brain does not need such phonetic labels because the transduction algorithms at the phonology-phonetics interface (see section 4) interpret the identity of a feature by the place of the neural activity (or a combination of the activity's place and firing rate) in the brain. This is similar to how a computer does not retrieve the identity of a symbol solely on the basis of its form (1s and 0s), but rather by combining the information about the form with the location and context in the memory (Gallistel & King 2010: 73). Possibly, the actual form of all features is the same—a neural spike (i.e., an action potential). But more importantly, the unique location of the spike and/or the rate of its repetition is how the transducer determines the identity of the feature and 'knows' which neuromuscular schema (e.g., labiality and not, say, nasality) to assign to it. It can

of course be debated whether it is misleading or not to use phonetic labels such as LABIAL to refer to features qua neural symbols and whether there is a better solution to this. But a decision on this issue has no bearing on the actual nature of features: the neural symbol is, of course, the same irrespective of whether we refer to it as LABIAL or by using a non-phonetic label like Universal Feature 17.

Features are manipulated by phonological rules, which should be understood as neural functions in the sense of Gallistel & King (2010: §3). The totality of the primitive phonological symbols and functions that manipulate them constitute the phonological module of the language faculty. Also, features are interpreted by particular transduction algorithms at the interface between the phonological module and the sensorimotor system. The Cognitive Phonetics theory (Volenec & Reiss 2017; 2019), described in section 4, spells out how this transduction proceeds and provides hypotheses about how it is realized neurobiologically.

The symbolic nature of something is that it stands for something else, something that is not the same as the symbol. That for which a symbol stands, that which it represents, is variably called its *referent*, *correlate*, or *the represented*. Phonological features are symbols which refer to aspects of speech. At this point, it is of utmost importance not to "make the common mistake of confusing the symbol with what it represents" (Gallistel & King 2010: 56); "the tendency to confuse symbols with the things they refer to is so pervasive that it must be continually cautioned against" (p. 62). There is a connection between phonological features and speech, but this connection is highly complex and indirect (see section 4), and features *do not* encode speech-related information in any straightforward way. This was the crucial characteristic of Distinctive Feature Theory from its inception:

> Such considerations [that languages do not make free use of acoustic values or articulatory properties] were much in our minds thirty years ago when Jakobson, Fant and I were working on *Preliminaries to Speech Analysis*, and it was these considerations that led us to draw a sharp distinction between distinctive features, which were abstract phonological entities, and their concrete articulatory and acoustic implementation. Thus, in *Preliminaries* we spoke not of 'articulatory features' or of 'acoustic features', but of 'articulatory' and/or 'acoustic correlates' of particular distinctive features. (Halle 1983/2002: 108–109)

In linguistics, information related to speech is called *phonetic substance*. It is the totality of the articulatory, acoustic and auditory properties and processes that constitute speech. For example, properties and processes of speech such as movements of the tongue, values of formants, loudness, duration expressed in milliseconds etc. fall under the rubric of substance.

Since features are symbols physically realized in the brain, they cannot contain phonetic substance. In other words, features are *substance-free*. Believing that features 'are' substance or that they 'contain' substance is just an instance of the aforementioned mistake of confusing the symbol with what it represents.

The relation between any given feature and its correlate is not random or arbitrary. If it were, then any feature could in principle be realized by any possible human articulation, similarly to how the concept/signified 'DOG' can in principle be assigned to any possible signifier. If such Saussurean arbitrariness were applicable to the realization of features, then it should be possible that +ROUND in one language gets realized as a lowering of the velum, and in another as a raising of the tongue dorsum, and so on. Instead, there is a non-arbitrary, lawful relation between features and their correlates. The nature of this relation is described in more detail in section 4, but at this point it is important to emphasize that the lawful relation between features and their correlates is *phonologically* irrelevant. That means that phonological computation treats features as invariant

categories, manipulating them in an arbitrary way irrespective of the variability in their realization in speech and irrespective of phonetic substance in general.

This fact can be clearly observed in the case of Turkish vowel harmony, where certain suffixes will contain different vowels depending on the preceding vowels. In suffixes containing high vowels, the +ROUND root vowels [u] and [o] give rise to a suffix form with [u], whereas the corresponding –ROUND vowels [ɯ] and [ɑ] in a root give rise to a suffix form with [ɯ]. The phonetic correlate of +ROUND for [u] is systematically different from the phonetic correlate of +ROUND for [o]. Because the lower jaw is lowered to a greater extent in the articulation of [o] than in the articulation of [u], the lip rounding for [o] will always take a different shape than the lip rounding for [u]. Crucially, the operation of vowel harmony will always ignore this obvious phonetic difference, and will treat +ROUND as an invariant category—by virtue of containing the substance-free invariant category +ROUND, both [u] and [o] will trigger the operation of vowel harmony in the same way, giving rise to a suffix with [u]. A parallel situation holds for the vowels [ɯ] and [ɑ] which differ from each other phonetically in many ways, but are *phonologically* just –ROUND. The relation between features and phonological computation is thus completely independent from the relation between features and phonetic substance, as can be seen in this diagram:

feature (neural symbol) ····· *arbitrary relation* ····· computation (neural function)
⋮
non-arbitrary relation
⋮
phonetic substance

The diagram summarizes the fundamental tenet of *Substance Free Phonology*: Even though there is a lawful relation between features and phonetic substance, both the rep-

resentational and the computational aspect of phonology are devoid of phonetic substance (cf. Hale & Reiss 2008).

We maintain, as the null hypothesis, the view that Universal Grammar provides a finite set of discrete phonological features. This position currently appears to be out of favor among many phonologists, although anyone who accepts the innateness of constraints that refer to specific features (e.g., IDENT-IO$_{[\text{VOICED}]}$) in various versions of Optimality Theory (OT), including the original proposals of Prince & Smolensky (1993/2004) and McCarthy & Prince (1995), must be committed to the innateness of features.

Work that denies the existence of innate features suffers from at least two failings. First, as a whole, this literature (e.g., Carr 2006, Mielke 2008, Archangeli & Pulleyblank 2015) ignores a logical argument presented in various forms, at least since Kant concluded in the eighteenth century that "all objects of experience must necessarily conform" to cognition:

> Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them a priori through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence, let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an a priori cognition of them, which is to establish something about objects before they are given to us. (Kant, cited in Rohlf 2016)

This conclusion has been reiterated and explicated over decades in the generative literature (e.g., Fodor 1980; Hale & Reiss 2003). For example, Jackendoff (2015: 187) points out that:

> If one is studying item-based or exemplar-based learning—if one thinks that learning consists of encoding all the instances of whatever a person has experienced in some domain, and that understanding novel inputs is a matter of analogy or interpolation among instances [references deleted]—then one ought to be asking: Exactly how does the mind/brain encode instances it has encountered, and what are the dimensions available for encoding them? A crucial constraint on such a theory is that the dimensions in terms of which experience is encoded cannot themselves be learned. They form the basis for learning; they are what enable learning to take place at all. So they have to somehow be wired into the brain in advance of learning.

Until this logical argument for the innateness of features is addressed, so-called 'emergentist' theories cannot be taken seriously. It is noteworthy that even some researchers in the AI machine learning domain have recently recognized the need to build 'priors' into their systems (Versace et al. 2018).

A second, oft-repeated argument against innate features is the putative existence of rules that do not refer to sets of segments that constitute a natural class (Mielke 2008, *inter alia*). However, work that denies an innate feature set in this way does not provide a general and explicit theory of rules, so such claims cannot even be evaluated. In Logical Phonology (see section 3), rules refer to natural classes by definition: a statement that cannot be formulated in terms of natural classes is not a rule. An example that can clarify that point can be drawn from Halle's paper *Knowledge Untaught and Unlearned* (1978/2002). Halle explores the productivity of English plural formation and essentially presents a phonological Argument form the Poverty of the Stimulus (PoS), although he does not use that term explicitly. His main point is that rules refer to natural classes defined by shared features, and not to sets of atomic segments. The argument relies on predicting that an English speaker exposed to the pronunciation of *Bach* as [bax], with a voiceless velar fricative, can form the plural [baxs] despite never having encountered

the devoicing of underlying /z/ after /x/. The implicit reasoning is that a speaker has learned a rule that devoices /z/ after /p, t, k, θ, f/, and that /x/ falls into any natural class that contains those segments. So, the pronunciation [baxs] is just normal rule application, if one thinks at the level of features. In other words, the rule refers to a natural class that is defined *intensionally*, but the *extension* of the set of relevant trigger segments changes when [x] is introduced. It turns out that Halle's PoS argument from the plural can be pushed further, and the relevance of the intensional characterization of rules can be made even more apparent.

The English plural form [-ɨz] occurs in a natural class of environments, as expected, because the vowel insertion is triggered 'between coronal stridents'. The form [-z] does *not* occur in a natural class of environments, because it is the 'elsewhere' case, identical to the underlying form. However, the form [-s] arises by a devoicing rule, but it does *not* occur in a natural class of environments *on the surface*—the voiceless coronal stridents [s, ʃ, t͡ʃ] are in the natural class that contains [p, t, k, f, θ][11]. The well-known solution is that the vowel insertion rule precedes and bleeds the devoicing rule. Intensionally stated, devoicing applies to a natural class of segments (all the voiceless obstruents), but the data available to the learner contradicts this. We therefore argue for an interpretation of PoS that not only *allows* the learner to posit rules in terms of natural classes despite contradictory surface patterns, but actually *forces* them to do so.

Much of the literature says that rules 'typically' refer to natural classes, and as noted above Mielke (2008) proposes that the existence of rules that do not rely on natural classes constitutes an argument against an innate feature set. In contrast, we propose that natural classes are defining properties of rules: *If a statement cannot be made in terms of natural classes, then that statement is not a rule*. We argue that it is only possible for linguists to study the range of possible rules if one stipulates that rules are based

11 This works whether we assume that English [f] is +STRIDENT, as suggested by languages like Ewe that distinguish [f] from [ɸ], or that English [f] is unspecified for STRIDENT, in which case each of the segments that trigger the change of /z/ to [s] would be either unspecified, like [f], or else -STRIDENT, like [p, t, k, θ]. Whichever assumption we adopt, [p, t, k, f, θ] don't form a natural class, and thus by hypothesis can't define a rule environment.

on natural classes. By defining rules in terms of natural classes, one severely restricts the hypothesis space for an acquirer—only processes definable in terms of natural classes can be part of the target phonological grammar. As pointed out above, the [-s] form of the plural does not occur in a natural class of environments, but the devoicing rule intensionally applies in a natural class of environments. If UG determines that rules must be stated in terms of natural classes, then a learner formulates such rules even in light of contradictory evidence—this is a PoS argument.

The view of natural classes developed here parallels the phenomenon of *amodal completion* (Michotte & Burke 1951) in visual perception and other 'masking' effects. Figure 1 is interpreted by the visual system as two objects, a magenta one partially occluding a blue one, despite the lack of continuity between the two blue regions. The visual system infers the part of the large blue rectangle missing from the signal. Similarly, in acquiring 'English', the language acquisition device infers a single voiceless obstruent class as a trigger of devoicing, even in the absence of direct evidence. As the IPA symbols in Figure 1 show, we can think of the magenta region as representing the natural class of coronal stridents, which trigger vowel insertion. Because of the bleeding order, this rule 'occludes' some members (s, ʃ, t͡ʃ) of the intensionally defined natural class which triggers devoicing, the set of all voiceless obstruents. Just as our visual systems infer a single blue object, our phonological acquisition systems infer a single voiceless obstruent natural class. If learners could make a separate rule for each segment that devoices /z/, then Halle's (1978/2002) claim about devoicing after /x/ would not be justified. We propose that even in the face of contradictory evidence (the partial 'occlusion' by coronal stridents of the natural class of voiceless obstruents), a learner constructs a single rule for devoicing. Without this insight—

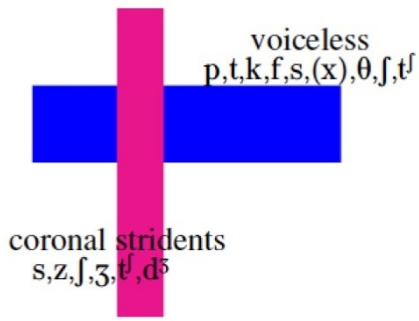

Fig.1: Amodal completion

the centrality of feature-based natural classes, despite the poverty of the stimulus—it is impossible to develop a universal theory of rule formulation and rule interaction.[12]

In light of these logical and methodological considerations, we maintain that UG provides a set of innate features. Since there are sound logical and empirical reasons to doubt that features can be learned, their innateness is the best null hypothesis. Numerous legitimate challenges remain for the idea that features are innate, including the question of whether sign languages can plausibly be encoded with the same set of features: Is a bilingual native speaker/signer of English and ASL using the same set of features for both languages, ones that we just happen to refer to by names like 'ROUND' and 'VOICED'? How many features are there? Is there any reason to think that the twenty or so features assumed in much of the literature are even close to sufficient?[13]

We mentioned in section 1.1 that there is a logical relationship between internalism and nativism: a denial of internalism entails a denial of nativism, since nativism is a claim about innate *internal* knowledge. We have argued that phonology should be treated like syntax with respect to issues like internalism and nativism. The following are some statements showing that this view is by no means universally held in the field:

12 The masking effect we just saw in vision is paralleled in auditory perception (Bregman 1990). If our parallel to phonological learning is on the right track, we can take this as further corroboration of the cognitive science approach to phonology advocated in this paper, since the notion of masking seems relevant to these three domains. Conversely, we might think of amodal completion in vision and audition as versions of the PoS argument, since they involve inference of structures for which there is no direct evidence in the signal.

13 Gunnar Fant (1969) considered the feature system presented in the *SPE* to be *"an improvement over that of Jakobson, Fant & Halle (1952)"*. However, he pointed out that *"there is a danger that the impact of the theoretical frame with its apparent merits of operational efficiency will give some readers the impression that the set of features is once for all established and that their phonetic basis has been thoroughly investigated. This is not so."* Of course, Fant's conclusion still holds. Research on features is not complete, any more than research on genes is complete. There is no reason to expect any contemporary phonology textbook or article to list the correct set of features. Even the validity of appeals to underspecification—the assumption that features can be valued '+', '–' or be absent from a segmental representation—is open for discussion.

> [H]owever much poverty of the stimulus exists for language in general, there is none of it in the domain of the structure of words, the unit of communication I am most concerned with. Infants hear all the words they expect to produce. Thus, the main proving ground for UG does not include phonology. (MacNeilage 2008: 41)

Peter MacNeilage is not a phonologist, but rather a psycholinguist who works on speech, so it is not surprising that his perspective is far from that of the orthodoxy of generative linguistics, especially syntax. Jeff Mielke is a linguist by training, and the clarity with which he rejects the generative tradition is refreshing:

> Many of the arguments for UG in other domains do not hold for phonology. For example, there is little evidence of a learnability problem in phonology (see Blevins 2004 for discussion). (Mielke 2008: 33)

> Most of the evidence for UG is not related to phonology, and phonology has more of a guilt-by-association status with respect to innateness. (Mielke 2008: 34)

Archangeli and Pulleyblank are also quite clear in rejecting the generative tradition and the idea that phonology is like syntax, not least by the title of their paper *Phonology without universal grammar*:

> See Mielke (2004) on why features cannot be innately defined, but must be learned. (Archangeli & Pulleyblank 2015: 2)

> [Children face] the challenge of isolating specific sounds from the sound stream. (*ibid*.)

> [T]he predictions of Emergent Grammar fit the data better than do the predictions of UG (*op. cit.*: 4)

In the first quotation, Archangeli and Pulleyblank reject innate features. In the second one they suggest, contrary to our discussion of Jackendoff, Sapir and Hammarberg, that segments exist in the sound stream. In the third one they claim that UG is empirically a worse hypothesis than their alternative Emergent Grammar. Phonological PoS and the innateness of features is also rejected by Philip Carr:

> Phonological objects and relations are internalisable: there is no poverty of the stimulus argument in phonology. No phonological knowledge is given by UG. (Carr 2006: 654)

Finally, Juliette Blevins rejects PoS by suggesting that the auditory input to the learner is sufficient to capture all relevant generalizations:

> Within the domain of sounds, there is no poverty of the stimulus. (Blevins 2004: 235)

> [I offer] general arguments against the 'poverty of stimulus' in phonology. [...] To my knowledge, there is no argument in the literature that phoneme inventories, stress patterns, tone patterns, phonotactics, and regular phonological alternations cannot be acquired on the basis of generalizations gleaned directly from auditory input. (*op. cit.*: 66)

All of these scholars seem to be ignoring standard views about what it takes to develop an acquisition theory:

> [T]he development of the theory of grammar [UG], and intensive application of this theory, is a necessary prerequisite to any serious study of the problem of language acquisition and many other problems of immediate psychological significance. This theory must contain a precise specification of the class of formalized grammars; it must provide a precise definition of this notion, that is, an explicit schema and notation for grammars and a specification of the conditions that a set of rules, stated in these terms, must meet in order to qualify as a grammar of some natural language. (Chomsky 1962: 534)

Without an explicit theory of phonological grammar, Blevins, Mielke and the others cannot, by fiat, declare that there is no learnability problem for phonology. Contrary to Blevins' (2004: 66) claim that "there is no argument in the literature" that aspects of phonology "cannot be acquired on the basis of generalizations gleaned directly from auditory input", what we find in various forms throughout the history of modern phonology, for example in the work of Sapir, Jakobson, Halle, Hammarberg and others, is that exact argument, as we have shown in the preceding sections.

It is of interest that Archangeli & Pulleyblank and Blevins cite approvingly the work of Terence Deacon, which includes the following passages:

> I think Chomsky and his followers have articulated a central conundrum about language learning, but they offer an answer that inverts cause and effect. They assert that the source of prior support for language acquisition must originate from *inside* the brain, on the unstated assumption that there is no other possible source. But there is another alternative: that the extra support for language learning is vested neither in the brain of the child nor in the brains of parents of teachers, but outside brains, in language itself. (Deacon 1997: 105)

> Over countless generations languages should have become better and better adapted to people, so that people need to make minimal adjustments to adapt to them. (*op. cit.*: 107)

> Grammatical universals exist, but I want to suggest that their existence does not imply that they are prefigured in the brain like frozen evolutionary accidents. In fact, I suspect that *universal rules or implicit axioms of grammar aren't really stored or located anywhere*, and in an important sense, they are not determined at all. (*op. cit.*: 115–116; emphasis added)

Of course, none of this makes sense from the I-language perspective. Despite the fact that Deacon (1997) studies the evolution of the brain and discusses, for example, the "loss of language abilities due to brain damage (termed *aphasia*)" (p. 280), he seems to locate language out in the world, not in the head. He rejects internalism. If we assume that the fact that Archangeli & Pulleyblank and Blevins cite Deacon can be read as an acceptance of his rejection of internalism, we can now see why they are driven to reject PoS and nativism. Without internalism, there is no nativism.

To reiterate how our proposals differ from these, let us summarize some of the lessons of this section. Words, segments and features are not 'in' the acoustic signal. These are categories that our minds impose on the signal. Without an innate inventory of such categories there can be no learning. Even granting the existence of such categories, the surface patterns of language do not reflect in a straightforward manner the nature of phonological computation, since even natural class behavior of segments can be 'occluded' by factors like rule-ordering, as we see in the English plural. These considerations support the idea that "a necessary prerequisite to any serious study of the problem of language acquisition [is] a precise specification of the class of formalized grammars" (Chomsky 1962: 534). In other words, we cannot even start to explain acquisition without an explicit theory of rules and their interaction. In the next section, we

will begin to flesh out a model of rules built on operations that apply to formally specified natural classes of segments. Phonetic substance will play no part in this formal model.

3 A WORKING MODEL FOR PHONOLOGY: LOGICAL PHONOLOGY

Much of the phonological literature since *The Sound Pattern of English* has been concerned with accounting for *markedness*, inspired by the possibility that the model developed in the first eight chapters of the book was "overly formal":

> The problem is that our approach to features, to rules and to evaluation has been overly formal. Suppose, for example, that we were systematically to interchange features or to replace [αF] by [−αF] (where α = +, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made use of the fact that the features have intrinsic content. (Chomsky & Halle 1968: 400)

For example, work in classical OT, which relies on two classes of constraints, one of which is known as the *markedness* or *well-formedness constraints*, cautions against extreme formalism:

> We urge a reassessment of [the] essentially formalist position. If phonology is separated from the principles of well-formedness (the "laws") that drive it, the resulting loss of constraint and theoretical depth will mark a major defeat for the enterprise. (Prince & Smolensky 1993/2004: 234)

Rather than following up on the initial worries expressed at the beginning of *Chapter 9* of the *SPE*, Logical Phonology maintains the ultimately strict formalist position, expressed later in the same chapter:

> It does not seem likely that an elaboration of the theory along the lines just reviewed will allow us to dispense with phonological processes that change features fairly freely. The second stage of the Velar Softening Rule of English (40) and of the Second Velar Palatalization of Slavic strongly suggests that the phonological component requires wide latitude in the freedom to change features, along the lines of the rules discussed in the body of this book. (Chomsky & Halle 1968: 428)

In other words, phonological theory should define a combinatoric space defined by representational and computational primitives. The characterization of which particular regions of this phonological combinatoric space are actually or potentially attested, or are more or less common, is within the purview not only of the diachronic linguist, the phonetician and the acquisitionist, but also of the historian and archaeologist who studies patterns of migration, settlement, conquest and genocide that have produced the current distribution of language families in the world. For the construction of a general mentalistic phonological theory, the frequency of occurrence of a particular phonological phenomenon is irrelevant, since that frequency is determined by political, social, economic, geographical and other extra-linguistic factors. As we have seen in section 2, the combinatoric space provided even by a modest set of features is exceedingly rich, and there is plenty of work for phonologists to do to further elaborate a formal model of computation and representation.

Logical Phonology (LP) maintains the *SPE* view of the phonology of a language as a complex function that maps input representations to output representations. This function is the composition (in a specific order) of more basic functions, typically called the *rules* of the language. As a first approximation, each rule maps strings of phonological

segments to other strings of segments. As Gallistel & King (2010) point out, functions are operations that take variables as input, and this property applies to phonological rules, too. For example, a traditionally expressed rule like 't → s /_ i' maps members of the class of strings such that *ati* maps to *asi*, *okutilo* maps to *okusilo*, *aka* maps to *aka* and *mokitapi* maps to *mokitapi*. These mappings are not arbitrary—the form of the input determines the form of the output. In this simple example, rules apply not to segments, but to strings, which are a more complex data structure. Phonologists know that even segments have an internal feature composition, and that a more complete model of the nature of rules available for phonological computation requires an elaboration of the taxonomy of phonological symbols. We need to get to the point where we can formalize processes such as 'coronal stops become their corresponding fricative before any high front vowel'.

Symbols in computation must be physically realized entities that are manipulated by the programming and architecture of a computational system. Because the brain is physically limited in power and space, and because scientific reasoning leads us to propose the simplest possible hypothesis, a desirable model is one with a small number of primitive symbols and basic processes for combining them. By combining primitive symbols, more complex symbols can be created when new stimuli require new representations. Gallistel & King (2010: 79–80) themselves use a linguistic example to illustrate the idea of a basic symbol inventory in both engineered and natural systems:

> No language in the world has a word for the message, "After the circus, I'm going to the store to get a quart of skim milk." (…) Minimizing the number of atomic data is desirable in any symbol system as it reduces the complexity of the machinery required to distinguish these data. This is why computers use just two atomic data and likely why nucleotide sequences use only four.

The atomic data enter into combinations that lead to more complex symbols, according to the following taxonomy:

(1) Gallistel & King's (2010: 79) symbol taxonomy
- atomic data
- data strings
- nominal symbols
- encoding symbols
- data structures

For example, atomic data can be grouped into sets, sets can be ordered in strings, and strings can be arranged into complex data structures like trees with precedence and containment relations among strings. Logical Phonology adopts a model of representation consistent with the symbol taxonomy in (1).

Vowels and consonants are *segments*. That is, they exist at the level of organization of language that we commonly represent with the symbols of the International Phonetic Alphabet. However, each segment symbol in LP is actually an abbreviation for a *set* of valued features.[14] These sets are *consistent*, which means that they cannot contain incompatible values;[15] for a given feature F, a segment cannot contain both –F and +F. We can thus think of the vowels in the words *beet* and *boot* as sets of valued features, as shown in (2).

14 "The phonetic symbols [p], [t], [e], [i], [u], etc., are simply informal abbreviations for certain feature complexes; each such symbol, then, stands for a column of a matrix of the sort just described." (*SPE*, p. 5)

15 For the sake of simplicity, we are abstracting away from the so-called *contour segments*, such as affricates and diphthongs.

(2) Specification of sets of valued features corresponding to the vowels /i/ and /u/

/i/ = {–BACK, –ROUND, +HIGH}
/u/ = {+ROUND, +HIGH, +BACK}

These sets of features are part of what Poeppel (2012: 35) calls "the infrastructure of linguistics [that] provides a body of concepts that permit linguists and psychologists to make a wide range of precise generalizations about knowledge of language". Realism (see section 1.2) commits us to accept that these sets of features are also part of language itself, not just tools that linguists use to describe language: "Since linguistics is an empirical science, the only justification for developing a theoretical framework is that the framework mirrors relationships among the facts of natural languages" (Halle 1975: 532–533; see also Hammarberg 1976: 355). Table 1 illustrates more fully the data structures relevant to phonology (Matamoros & Reiss 2016).

TABLE 1. *A HIERARCHICAL TAXONOMY IN LANGUAGE.*

Level	Notation	Name
Atomic	{+, – }, {HIGH, BACK, ROUND, *etc.* }	coefficients and features
Ordered pairs	<+ROUND>	(valued) features
Sets of ordered pairs	{ <+ROUND>, < –HIGH>, <+BACK>} abbreviated [k], [i], [θ], [p] *etc.*	segments
Strings of sets of ordered pairs	/kæt/	strings (phonology of morphemes)
Tripartite structures	/kæt/ w/ syntactic & semantic info	morphemes
Lists of morphemes	(a complex structured list)	mental lexicons

The atomic value and feature symbols are combined into ordered pairs in the process of parsing speech. Such ordered pairs are grouped into sets corresponding to segments. Minimal units of meaning, called *morphemes*, are stored with information related to

meaning and syntactic category (VERB vs. NOUN, for example). Of course, the segments that constitute the phonological representation of a morpheme must be stored in a specific order since the same three segments [æ, k, t] can be ordered to express more than one meaning, as in *act*, *tack*, *cat*.

Given this infrastructure for symbols, we can start to formalize the treatment of these symbols by a computational system. LP defines *possible rules* as functions that map sequences characterized by natural classes of segments to other sequences. Since each segment is treated as a set of valued features (which in turn are ordered pairs), we have available a simple method of formalizing the sometimes intuitively stated notion *natural class*:

Definition: The smallest natural class containing a set of segments S = $\{s_1...s_n\}$ is defined by the generalized intersection of S. The generalized intersection of a set of sets R = $\{r_1...r_m\}$ is just the set of elements that are in every set r_i.

This can be illustrated using the vowels /i/ and /u/ specified in (2). The generalized intersection ∩{i, u} = ∩{{–BACK, –ROUND, +HIGH}, {+ROUND, +HIGH, +BACK}}= {+HIGH}. The segments in the smallest natural class containing /i/ and /u/ are thus all those that are supersets of the set {+HIGH}. This natural class is described intensionally using standard set theory formalism thus: {x : x ⊇ {+HIGH}}. For ease of exposition, we will denote such a natural class using the traditional square brackets for the classes that constitute targets and environments of phonological rules, thus [+HIGH]. It follows, then, that if a language has a rule that refers to the vowels /i/ and /u/, then the rule must apply equally to the vowel /y/ = {+ROUND, +HIGH, –BACK}, if this vowel also occurs in the language, because /y/ is also a superset of the generalized intersection ∩{i, u}. This result formalizes the notion of *generalizing beyond the data*, although viewed at the correct level of abstraction, there is no generalization, just application of the rule to whatever value has been assigned to that variable serving as its input symbol. It also fol-

lows that if in some language a putative process affects /i/ and /u/ but not /y/, then that process cannot correspond to a single phonological rule. Adopting a simple, but explicit model like this allows us to leverage the well-established results of set theory, and also leads us to understand apparent exceptions as arising from, say, the masking effects of rule ordering, as in the discussion of the English plural in section 2. In other words, by having a theory of the scope of rules, we have a theory of the limits of rules—certain patterns that might be considered the result of a single rule in an informal analysis, have to be accounted for with two rules, when the theory progresses and becomes more explicit.

The approach we describe is an example of more general principles that have been advocated in linguistics for decades:

> By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. (Chomsky 1957: 5)

Note that Mielke's (2008) method for rejecting natural class reasoning is to scan the literature for ‘rules’ that do not apply to natural classes. One problem is that he counts as a rule anything that has been called a rule by his sources. In the absence of a formal theory of rules, just about any descriptive statement can be considered a rule; of course, science cannot be done in such loose terms. Another problem is that this technique should come up with the English devoicing rule as an example that does not use natural classes, since on the surface, it is not triggered by a natural class of segments (see section 2). In the case of English, it is obvious how to save a natural class solution through

a bleeding rule relationship, but for the many less-studied languages he cites, from Abujmaria to Zina Kotoko, it is unlikely that his sources are as well-documented and sophisticated as what we have for English.

Now that we have a better understanding of natural classes, we turn to considering another aspect of phonological computation, the nature of rules. Following Bale et al. (2014), we suggest that it is necessary to deconstruct the traditional arrow of phonological rules, because there are in fact several distinct operations at work in phonological computation. Here we focus on one operation, *priority union*, as a candidate for inclusion in the computational toolkit of phonological UG (see Reiss 2021 for further discussion).

Regular set union combines the elements of two sets into a set whose members are all the elements that are in at least one of the input sets. So, given sets A and B, $A \cup B = C$, where each element of C is in either A or B or both: $\{w, x, y\} \cup \{w, y, z\} = \{w, x, y, z\}$.

Applying union to a segment (which is a set of valued features) and another set of valued features potentially leads to an inconsistent output (3):

(3) Union can yield an inconsistent output

$$\{+\text{COR}, -\text{SON}, +\text{VOICED}\} \cup \{-\text{VOICED}\} = \{+\text{COR}, -\text{SON}, +\text{VOICED}, -\text{VOICED}\}$$

Because the output contains both +VOICED and −VOICED as members, it is inconsistent. In order to maintain the requirement that segments be consistent, an operation different from simple union is needed for phonological rules. Adapting a definition from Kaplan (1987), we make use of *priority union*, which is defined in (4). In this definition, the valued features '+F' and '−F' are referred to as *opposites*.

(4) Definition of priority union

The priority union of two sets of valued features, A and B, is the union of the set of valued features that are in A, and the set of valued features in B whose opposites are not in A.

Unlike simple set union, priority union is not commutative—the priority union of A with B is not necessarily the same as the priority union of B with A. Paraphrasing Dahl (2002), priority union (also called *default unification*) takes two feature sets, one of which is identified as *strict*, while the other one *defeasible*, and combines the information in both such that the information in the strict set takes priority over that in the defeasible set. We denote priority union by the normal union symbol with a subscripted arrow: '$\cup_{\rightarrow}$' or '$\cup_{\leftarrow}$' (for reasons that will become apparent below). The arrow points from the strict argument to the defeasible argument, as shown in (5).

(5) Priority union examples

a. $\{+F\} \cup_{\rightarrow} \{+G\} = \{+F, +G\}$
b. $\{+F, -G\} \cup_{\rightarrow} \{+G\} = \{+F, \mathbf{-G}\}$
c. $\{+F, -G\} \cup_{\rightarrow} \{+G, +H\} = \{+F, \mathbf{-G}, +H\}$
d. $\{+F, -G\} \cup_{\leftarrow} \{+G, +H\} = \{+F, \mathbf{+G}, +H\}$

In (5a), priority union gives the same result as normal union. In (5b), the valued feature −G of the strict argument takes priority over the +G of the defeasible argument. The same situation holds in (5c), except that the defeasible argument contributes +H, since there is no conflict with the strict argument. In (5d), the arrow has been reversed, so that +G is a member of the strict argument and −G is in the defeasible argument.

Inkelas (1995) analyses Turkish as having a three-way underlying consonant contrast that can be used to illustrate how priority union can capture *feature-filling* phonological rules. The segment /d/ is a voiced coronal stop, /t/ is a voiceless coronal stop, and /D/ is a coronal stop with no specification for voicing.[16] The fully specified stops do not alternate in the relevant environments: /t/ and /d/ surface unchanged in onsets and codas, but /D/ surfaces as [d] in onsets and as [t] in codas.

Given the set-theoretic formalization, there is no way to define a natural class that (intensionally) targets just /D/, but there is a way to make a rule whose effect *appears* to target just /D/ by using priority union. Rule (6) turns /D/ to [t] in codas.

(6) $[+\text{COR}, -\text{CONT}, -\text{SON}] \cup_{\rightarrow} \{-\text{VOICED}\}$ / in Coda position

Recall that this target specification refers to all segments that are supersets of the set of valued features {+COR, −CONT, −SON}, so the target natural class contains the three segments /t, D, d/. The target and the individual segments are defined as in (7).

16 The exact analysis of Turkish is controversial, but there exist other phenomena that share the same structure as the presentation here, so we can abstract away from phonetic details—we are just concerned with three underlying segments a, b, c, such that b is the intersection of a and c. In other words, b is unspecified for a feature with respect to which a and c have opposite values. Allowing underspecification makes for a more 'stripped-down' theory, because it does away with the stipulation that segments must be complete, i.e., that they must have a value for each feature. Removing the stipulation of completeness is parallel to removing the stipulation against Internal Merge in syntax, and thus unifying Merge and Move (Chomsky 2007). Additionally, underspecification gives the model more descriptive power for a given number of features, since the number of definable segments with *n* features in UG then increases from 2^n to 3^n. This increase in descriptive power has at least two benefits. First, it means that it is easier to resist the temptation of positing a language-specific phonetic module, as discussed below in section 4. Second, it is not necessary to posit as many features in UG to achieve a given level of descriptive power. Obviously, it is desirable to reduce the amount of information that we need to attribute to the genetic basis of language.

(7) A natural class with an underspecified member[17]

Description of class:

[+COR, −CONT, −SON...] = {x : x ⊇ {+COR, −CONT, −SON...}

Members of the class:

a. /t/ = {−VOICED, +COR, −CONT, −SON...}

b. /d/ = {+VOICED, +COR, −CONT, −SON...}

c. /D/ = {+COR, −CONT, −SON... (no VOICED feature)}

Priority union of each of the Turkish stops with the set {–VOICED} is shown in (8).

(8) Priority union for Turkish codas

a. /t/ $\cup_{\rightarrow}$ {−VOICED} = [t]

b. /D/ $\cup_{\rightarrow}$ {−VOICED} = [t]

c. /d/ $\cup_{\rightarrow}$ {−VOICED} = [d]

In (8a), priority union is applied vacuously, because /t/ is −VOICED. Therefore, the output is [t]. In (8b), priority union adds the valued feature −VOICED to /D/, yielding [t]. In (8c), priority union yields [d] because /d/, which is +VOICED, is the strict argument, and −VOICED is a member of the defeasible argument. Priority union thus allows us to formulate a rule that changes representations containing underlying /D/ without affecting those containing /d/ or /t/.

To account for *feature-changing* processes using priority union, we must distinguish two aspects of rule notation. In a rule like '*a* → *b* / __ *c*' the first element *a* is called the

17 The ellipsis '...' denotes whatever other features may be relevant to characterizing the segments in question, in this case anything except VOICED.

target, and the rule is understood to apply to strings that contain the subsequence *ac*, turning that subsequence into *bc*. The *structural change* of the rule is represented by the element *b*. The issue of which argument of priority union is the strict one and which is the defeasible one is independent of the *target* vs. *structural change* distinction which corresponds to position in rule syntax. In contrast, the *strict* vs. *defeasible* distinction is stipulated by the direction of the subscript arrow in the priority union symbol. Consider the priority union rules in (9).

(9) Rule target vs. strict argument

i. $a \cup_{\rightarrow} b$ / __c

a is the target; *a* is the strict argument

ii. $a \cup_{\leftarrow} b$ / __c

a is the target; *b* is the strict argument

The element *a* is both the target and the strict argument of rule (9i). However, in (9ii), although *a* is again the target, it is now *b* that is the strict argument of priority union. In both rules, the environment in both rules is the same, defined by the target *a* being followed by *c*.

An illustration of feature changing rules does not require an underlyingly underspecified segment such as /D/, therefore a language with just /t/ and /d/ such as Hungarian is appropriate. Hungarian manifests reciprocal neutralization whereby /t/ becomes [d] before a voiced obstruent and /d/ becomes [t] before a voiceless obstruent. All pairs of voiced and voiceless obstruents, such as /p/ and /b/, participate in parallel alternations (aside from some complications we can ignore). For example, the data in (10) shows stem-final /p, t/ mapping to [b, d] before a suffix that begins with voiced obstruent /b/; and /b, d/ mapping to [p, t], before a suffix that begins with voiceless obstruent /t/.

(10) Hungarian feature-changing voicing assimilation

Lexical	Bare Noun	in Noun /-ban/	from Noun /-to:l/	to Noun /-nak/	gloss
/ku:t/	ku:t	ku:dban	ku:tto:l	ku:tnak	'well'
/ka:d/	ka:d	ka:dban	ka:tto:l	ka:dnak	'tub'
/kalap/	kalap	kalabban	kalapto:l	kalapnak	'hat'
/rab/	rab	rabban	rapto:l	rabnak	'prisoner'

We assume that Greek letter variables are also part of the phonological UG toolkit (see Bale & Reiss 2018 for arguments that such variables are not merely part of phonologists' metalanguage), and so we can combine α-notation with priority union into a single feature-changing rule for Hungarian, as shown in (11).

(11) Hungarian voicing assimilation

Examples:

i. p, t → b, d / __b

ii. b, d → p, t / __t

Rule: [–SON] $\cup_{\leftarrow}$ {αVOICED}/ __ [αVOICED, –SON]

The target of the rule in (11) is the natural class of segments that are specified −SON, or more formally, the set of segments that are supersets of the set {−SON}, so this specification is listed leftmost in the rule, within square brackets. Next comes the priority union symbol, but with the arrow pointing from the right to the left, showing that the target's values are defeasible. Notice that the feature of the structural change 'comes from' (via α-notation) the segment that defines the triggering environment. This is what it means to be an assimilation rule. However, not every rule involves assimilation, and

the structural change can just be stipulated, without a Greek letter variable, and without an apparent segmental source. For example, a rule that devoices obstruents word-finally, as in Catalan and many other languages, would stipulate that the structural change involves priority union with the strict argument {−VOICED} as in (12), but this change is not bound to any segment in the triggering environment.

(12) Feature-changing final devoicing

$$[-\text{SON}] \cup_{\leftarrow} \{-\text{VOICED}\} \;/\; \_\_ \;\%$$

Here in (12) the voicing values on the targets are defeasible, so the rule is feature-changing. Compare this to the devoicing in (6) where the value on the structural change is defeasible, making that a feature-filling rule.

This example demonstrates the ongoing concern of Logical Phonology with 'deconstructing' the arrow of traditional phonological rules to better understand what kinds of functions phonological competence can contain. In other work (e.g., Bale et al. 2014, Volenec 2019 etc.), we have treated feature-filling using unification, which is a partial operation that, like priority union, is also related to set union. There, following a long tradition dating at least to Harris (1984) and adopted by many other scholars (such as Poser, 1993, 2004; Wiese 2000; Samuels 2011; McCarthy 2007) feature changing rules are modeled as a two-step process of deletion, accomplished with set subtraction, followed by insertion (i.e., feature-filling), accomplished with unification. The exploration of an alternative based on priority union reflects the methods and goals of Logical Phonology to understand the formal infrastructure of the language faculty. The choice between various formal approaches to phonological computation can be addressed in future research only after we recognize that there are several possibilities.

4 THE PHONOLOGY-PHONETICS INTERFACE: COGNITIVE PHONETICS

From the preceding discussion it is apparent that our conception of phonology is vastly different from phonetics. Phonology consists of abstract, symbolic, discrete, timeless units and formal operations, while phonetics is characterized by gradient, temporally arranged articulatory movements and continuously varying sound waves. Since humans do speak, it is a logical necessity that phonological competence successfully relays information to the sensorimotor (SM) system,[18] which is in charge of speech production. The communication between these two systems is the *phonology-phonetics interface* (PPI).

Cognitive Phonetics (CP; Volenec & Reiss 2017; 2019) proposes that the PPI consists of two transduction procedures that convert the substance-free output of phonology into a representational format that contains substantive information fed to the SM system which may (or may not) then externalize language through speech. The inputs to CP are the outputs of phonology, i.e., surface phonological representations (SRs). SRs are strings of segments, each of which is a set of features (still abstracting away from prosodic units). Each feature of SRs is transduced and subsequently receives interpretation by the SM system (cf. Lenneberg 1967: §3). This transduction is carried out by two algorithms (cf. Marr 1982/2010: 23–24). The *paradigmatic transduction algorithm* (PTA) takes a feature (a symbol in the brain) and relates it to a motor program which specifies the muscles that need to be contracted in order to produce an appropriate acoustic effect. The *syntagmatic transduction algorithm* (STA) determines the temporal organization of the neuromuscular activity specified by the PTA. In simpler terms, PTA assigns muscle activity to each feature, STA distributes that activity temporally. These transduction algorithms yield an output representation of CP, which then feeds the SM

18 Since the term *sensorimotor system* is used ambiguously in the linguistic literature, as was mentioned in the introductory section, it should be noted that we use that term to denote a large-scale brain network that includes the pre-central and post-central gyrus and the supplementary motor area.

system. The output of CP is called the *phonetic representation* (PR),[19] and it can be defined as a complex array of temporally coordinated neuromuscular commands that activate muscles involved in speech production. The standard schema of phonological competence presented in section 2 can now be expanded to accommodate the transduction performed by CP:

underlying phonological representation
↓
operational part of phonology (rules as logical functions)
↓
surface phonological representation
↓
cognitive phonetics (PTA & STA)
↓
phonetic representation
↓
the sensorimotor system

The gray parts of the schema represent phonological competence, while the black parts correspond to the initial *phonetic* steps in speech production. That is, the difference in shading parallels the competence/performance dichotomy: phonology is competence, cognitive phonetics is (one component of) performance.

To clarify the effects of PTA and STA, we can explore in some detail the transduction of a few hypothetical SRs (see Volenec & Reiss 2019 for more examples). Interestingly, as will be shown, PTA and STA have considerable implications: They open the possibility of elegantly accounting for subtle yet systematic interactions of two kinds of coartic-

19 In generative linguistics literature, the output of phonological computation, what we call a *surface phonological representation*, is sometimes called a *phonetic representation*. Note that we use the term very differently—our phonetic representations do not contain the phonological features that are in URs and SRs.

ulatory effects, which is only possible if we assume that the basic units of speech production are transduced phonological features. Suppose that a hypothetical I-language contains SRs [lok] and [luk]. Each segment is a set of features, and vowels [o] and [u] both contain the valued feature +ROUND, on which we will focus. One thing that should be noticed is that [o] and [u] are different in terms of height: [o] is –HIGH, [u] is +HIGH. The PTA takes a segment, scans its feature composition and determines the required muscular activity for the realization of every feature. Roughly, for +ROUND the PTA activates at least four muscles—*orbicularis oris*, *buccinator*, *mentalis*, and *levator labii superioris* (Seikel et al. 2009: 719–720)—which leads to lip rounding. The difference in PTA's effect on –HIGH and +HIGH is that for the latter the algorithm raises the tongue body and the jaw, while it does not for the former. While transducing +ROUND, the PTA takes into account the specification for HIGH and assigns a slightly different lip rounding configuration for [o] than for [u]. Let us refer to a *transduced feature*, which we take to be the basic unit of speech production, as $PR_{[F]}$, where 'PR' stands for 'phonetic representation' and 'F' stands for an individual valued feature. So, $PR_{[+ROUND]}$ is the transduced feature +ROUND. We can now say that $PR_{[+ROUND]}$ will be different for [o] because of its interaction with $PR_{[–HIGH]}$ than for [u] because of its interaction with $PR_{[+HIGH]}$. Since these interactions involve transduced features within a single segment, [o] or [u], we can refer to these effects as ***intra****segmental coarticulation*. The PTA accounts for intrasegmental coarticulation by assigning a different neuromuscular schema depending on the specification of features from the same segment.

Let us suppose further that, while determining the durational properties of transduced features, the STA temporally extends $PR_{[+ROUND]}$ from the vowel onto the preceding consonant, i.e., in the anticipatory direction. This amounts to the more familiar ***inter****segmental coarticulation*, where transduced features from *different* segments interact. Returning to SRs [lok] and [luk], it is now apparent that (a) $PR_{[+ROUND]}$ is different for [o] than for [u] due to its intrasegmental coarticulation with $PR_{[HIGH]}$; (b) [l]'s inherent $PR_{[–ROUND]}$ is now temporally overlapping with the $PR_{[+ROUND]}$ from the adjacent vowels be-

cause of intersegmental coarticulation. It is important to note that the difference in $PR_{[+ROUND]}$ from [o] and $PR_{[+ROUND]}$ from [u] will be reflected on the preceding consonant: [l] in [lok] will be articulated differently with respect to lip rounding than [l] in [luk]. Thus, [l] simultaneously bears the effect of both intra- and intersegmental coarticulation. CP allows us to account for such subtle yet systematic phonetic variations in an explicit and straightforward way—they follow automatically from PTA and STA, which are independently motivated by the need for transduction.

CP's transduction is *deterministic*, which means that it assigns the same neuromuscular schema to each feature every time that feature is transduced. This also includes all cases of feature combinations that lead to intra- and intersegmental coarticulation. CP thus makes another empirically testable prediction: In principle, given a full and correct list of features, it should be possible to exhaustively describe all possible intra- and intersegmental coarticulatory effects just by using the two algorithms proposed by CP.

It should be stressed that CP's outputs, phonetic representations, should not be equated with actual articulatory movements or with the acoustic output of the human body. What is actually pronounced is further complicated in the process of externalization by a great number of factors. Transduction is followed by other performance factors that have no bearing on either phonology or transduction, factors like muscle fatigue, degree of enunciation, interruptions due to sneezing, and many other situational effects, all of which will have an effect on the final output of the body, and will therefore make (co)articulatory variation seem greater. For that reason, it is *not* the case that the articulatory and the concomitant acoustic substance will always be identical for each feature or feature combination. The apparent *lack of invariance* in the realization of a cognitively invariant category is not a matter of transduction. In fact, lack of invariance is only problematic for those scholars of language and speech who assume a misguided *physicalist* stance (as opposed to a mentalist stance) towards their respective objects of study. Philosopher Irene Appelbaum (2006) has pointed out that "nearly half a century of sustained effort in a variety of theoretical perspectives has failed to solve the prob-

lem" of the lack of invariance in speech perception—in the output of the human body there are no constant correlates of particular segments or features across contexts. Appelbaum suggests that scholars attempting to find such invariant properties may be asking the wrong questions, and we find, in the study of vision for example, that the assumption that perception is stimulus bound must be rejected (Marr 1982/2010). This was recognized in the study of language and speech already by Sapir (1933/1949: 46): "No entity in human experience can be adequately defined as the mechanical sum or product of its physical properties''. In contrast, assuming a mentalist stance towards the study of language and speech, and appreciating the fact that it is phonetics that is grounded in phonology and not the other way around, leads us to perceive the lack of invariance not as a 'problem' but rather as an expected prediction of our theory (i.e., it follows naturally from the competence/performance dichotomy). Performance will always introduce into the acoustic record some degree of completely random and chaotic component, thus even in acoustic phonetics a certain degree of abstraction is necessary to arrive at law-like conclusions.

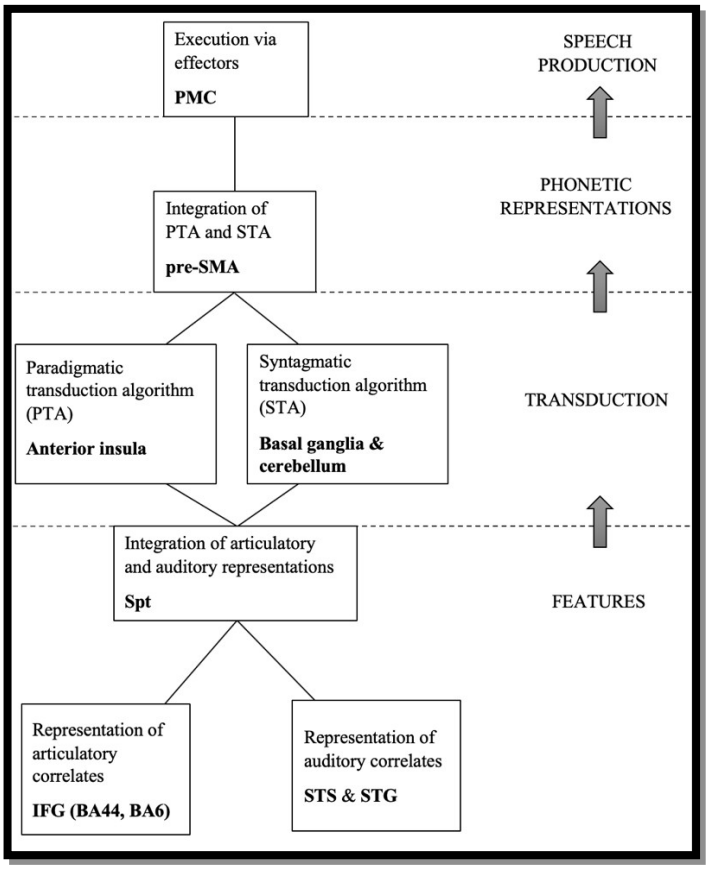

FIGURE 2. *NEURAL SUBSTRATE OF PHONOLOGICAL FEATURES*

We want to reiterate that CP differs from most other models of speech processing in the importance it places on linking *feature*-based phonological representations with speech. This stance is supported by recent neuroscience evidence which is consistent with the idea that CP transduces features, understood as symbols in the brain, into temporally distributed neuromuscular activities (elements of PRs), relating phonological competence to the vastly different SM system (Dronkers 1999; Hickok & Poeppel 2007; Guenther et al. 2006; Eickoff et al. 2009, Hickok 2012).[20] The activity in parts of the inferior frontal gyrus (IFG) corresponds to the representations of the articulatory correlates of features, while the activity in parts of the superior temporal gyrus (STG) and sulcus (STS) corresponds to the representations of the auditory correlates of features. An area in the Sylvian fissure at the boundary between the parietal and the temporal lobe (Spt) unifies these two aspects into a complete symbol, a feature. The symbols are sent to the anterior insula where the PTA is carried out, and to the cerebellum and the basal ganglia where the STA is carried out. The PTA and the STA are inte-

20 See Volenec & Reiss (2017: §4.3) for a more elaborate exposition of the neural basis of Cognitive Phonetics, which also includes extensive referencing to previous neuroscience work that corroborates it.

grated in the anterior part of the supplementary motor area (pre-SMA) to form the phonetic representation, which is a set of neural signals that the primary motor cortex (PMC) sends to the effectors that produce speech. These neural processes are graphically represented in Figure 2.

Since they are part of performance and not of competence, the transduction algorithms cannot be specific to any particular I-language. However, although they are biologically invariant in humans, CP will still show 'output variability' due to the PTA and the STA being applied to SRs that are featurally distinct. It is therefore important to distinguish between the outputs of the transduction system (phonetic representations) and the system itself (CP). The outputs of the system are, trivially, I-language-dependent and surface-representation-dependent since CP will automatically transduce any SR that is fed to it and any particular I-language will generate its own distinct set of SRs. For example, English URs and SRs will not have click segments in them, but some !Xhosa URs and SRs will have clicks. For CP, it is irrelevant whether an SR comes from an English, !Xhosa or Telugu speaker. The CP system itself is part of the human biological make-up and can in this sense be regarded as a universal property of the human species. The universality[21] of CP is merely a reflection of the fact that there exists a biological object we may informally refer to as *the phonetic implementational system*, of which CP is one part and the SM system another.

Consequently, we deny the existence of 'language-specific phonetics'. This is simply a result of consistently adhering to the competence/performance dichotomy. Something can be language-specific only by virtue of belonging to an individual's linguistic competence, i.e., to their I-language. Something that is not part of an I-language cannot be specific to an I-language, i.e., it cannot be language-specific. Since speech, the object of study of phonetics, is not language (competence, I-language), no part of it can be language-specific. The idea of 'language-specific phonetics' is thus a contradiction in terms

21 Even though CP is biologically universal, it is *not* part of Universal Grammar (UG), the innate component of the language faculty. CP is *not* part of UG simply because it is not part of the human language faculty (competence), just like, say, the visual or the respiratory system.

just as, say, the idea of 'language-specific vision', because neither vision nor speech are part of linguistic competence.[22] Adherence to *'universal' phonetics* means that many of the alleged 'language-specific phonetic patterns' should to be reanalyzed by capturing them through phonological features and other phonological primitives,[23] while at the same time strictly resisting the temptation of lumping distinct I-languages under socio-politically determined labels (e.g. 'English language', 'General American', 'German dialects'). That this feat is indeed possible is reflected in the fact that 30 binary features (slightly more than is usually assumed) with underspecification can yield over 206 trillion distinct segments, which is more than enough to capture the purported examples of 'language-specific phonetics' such as the systematic difference in the high front unrounded vowel between English and German speakers. In our framework, a German speaker's [i] and an English speaker's [i] must actually correspond to different feature sets because there is no language-specific phonetics which could have introduced a difference. Given the combinatoric explosion discussed in section 2, it is to be expected that there are many different feature sets (segments) that inhabit the 'high front unrounded vowel' space.

22 Note that both the visual system and the articulatory-auditory system *interface* and *communicate* with linguistic competence—vision in reading, the articulatory-auditory system in speaking—but neither is *part of* competence (see Volenec & Reiss 2017 for details).

23 Of course, the features that are currently assumed by most phonologists, presented with some variation in modern textbooks such as Hayes (2009), Odden (2013), Gussenhoven & Jacobs (2017) etc., might not be enough to achieve this task and a slight enrichment of the universal feature set might be necessary. See Hale et al. (2007) for an elaboration along this line.

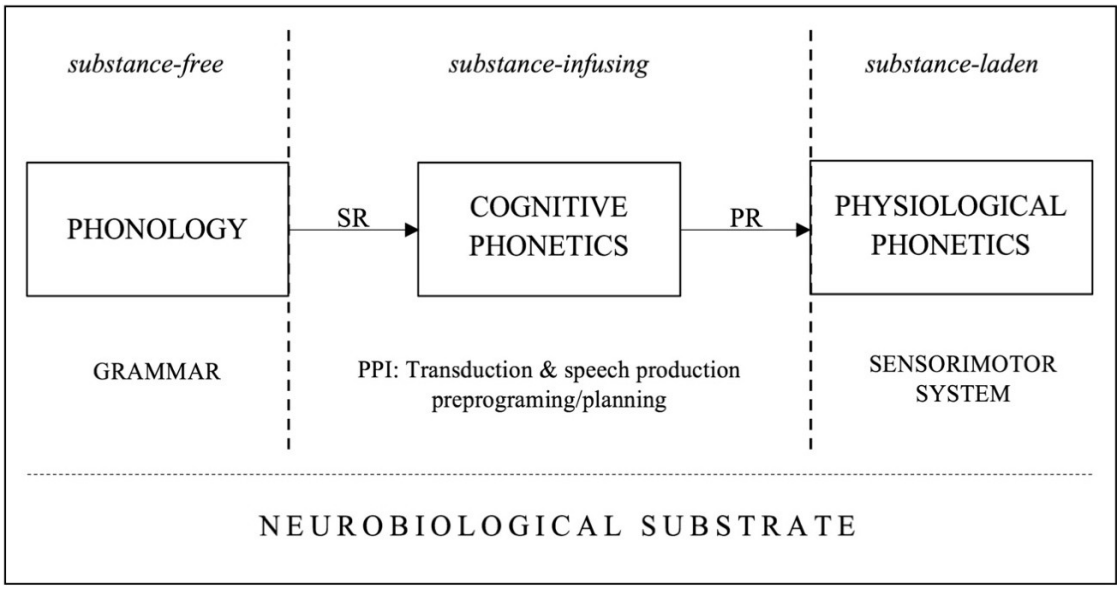

FIGURE 3. *THE ARCHITECTURE OF THE PHONOLOGY-PHONETICS INTERFACE AND THE PLACE OF COGNITIVE PHONETICS WITHIN IT.*

Figure 3 summarizes the general architecture of the phonology-phonetics interface (PPI) in the theory of Cognitive Phonetics (CP). To connect substance-free phonology with the substance-laden physiological phonetics, CP takes features of phonological surface representations (SRs) and relates them to neuromuscular activity (PTA) and arranges that activity temporally (STA), thus generating phonetic representations (PRs) that are directly interpretable by the sensorimotor (SM) system.

CONCLUSION

The object of study in linguistics is the biologically determined human language faculty which gives rise to each I-language, a module of an individual's mind/brain that generates an unbounded array of discrete hierarchical expressions. Each I-language is further divided into distinct submodules, one of which is phonology (phonological competence). Phonological representations are built from certain atomic symbols, such as the

features (VOICED, ROUND, etc.) and values (+, –); and phonological computation involves rules built from logical functions, potentially including priority union. The entirety of phonology resides in the human mind/brain, and its connection to phonetic substance (e.g., movements of the articulators during speech) is remote and complex. Given its mind-internal nature, phonology is to be studied consistently as a branch of cognitive science, seeking ultimate unification with neurobiology, with all the corresponding theoretical and methodological commitments that follow from this mentalist perspective.

In this paper, we have provided a general program for the study of phonology as a branch of cognitive science. Resting on the philosophical foundations of generative linguistics (Chomsky 1968; 1986; 2000), this program can be called *Formal Generative Phonology*. Its main tenet is that phonological competence is to be characterized as a formal—that is, explicit, logically precise, and substance-free—manipulation of abstract symbols. We have suggested that a productive way to execute this program is to adopt a model of formal generative phonology called *Logical Phonology* (building on work like Bale & Reiss 2018). There, phonological competence is described and explained by the application of set-theoretic functions, such as set subtraction, unification, generalized intersection, etc., to abstract representational data structures such as strings of segments. The remote and complex relationship between phonological competence and speech was then elucidated by *Cognitive Phonetics* (Volenec & Reiss 2017; 2019), which proposes that the outputs of phonology are transduced via two algorithms into temporally distributed neuromuscular activities.

A foundational pillar of linguistics as a modern science is the idea that language and speech are two different but connected entities. Over one hundred years ago, Ferdinand de Saussure likened their relationship to two different sides of a single sheet of paper, arguing persuasively that the formal underlying system (*langue*) must be strictly distin-

guished, both in theory-construction and in analytical practice, from the substance-based use of that system in communication (*parole*):[24]

> *Language is a form and not a substance*. This truth could not be overstressed, for all the mistakes in our terminology, all our incorrect ways of naming things that pertain to language, stem from the involuntary supposition that the linguistic phenomenon must have substance. (Saussure 1916/1959: 122)

A century later, the most serious confusions in phonology still stem from the supposition that phonological phenomena must have phonetic substance. If one thinks that phonology is language and not speech, then one is compelled also to accept that phonology is form and not substance. This line of reasoning provides "a principled manner" for determining "the boundary between transducer and symbolic processor" (Pylyshyn 1984: 148), or in more familiar terms, between phonetics and phonology. And with this conceptual boundary in place, we are in a better position to further develop an *explicit* theory of phonological representation and computation. We have provided a general program for such a phonological enterprise, a program that is grounded in cognitive neuroscience and formal logic, and not in phonetic substance and intuition-based claims. Is such a program on the right track? "Time will tell. But intuition won't." (Bromberger & Halle 1995: 739)

24 A well-known difference between Saussurean and generative linguistics is that in the former language is not a mental object while in the latter it is. The important point here is that De Saussure's *form* vs. *substance* dichotomy remains valid in a completely mentalistic, internalist approach to language.

REFERENCES

Akiyama, K. et al. (2019). First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *The Astrophysical Journal Letters*, 875/1: 1–17.

Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. *Proceeding of Fourth International Conference on Spoken Language Processing*, Vol. 3.

Archangeli, D. & D. Pulleyblank. (2015). Phonology without universal grammar. *Frontiers in Psychology*, 6.

Bale, A., M. Papillon & C. Reiss. (2014). Targeting underspecified segments: A formal analysis of feature changing and feature filling rules. *Lingua*, 148: 240–253.

Bale, A., C. Reiss & D. T. Shen. (2019). Sets, rules and natural classes: {} vs. []. *Loquens.* 6/2.

Bastiaanse, R. & C. Thompson. (2003). Verb and auxiliary movement in agrammatic Broca's aphasia. *Brain and Language*, 87: 286-305.

Bastiaanse, R. & van Zonneveld, R. (1998). On the relation between verb inflection and verb position in Dutch agrammatic aphasics. *Brain and Language*, 64: 165-81.

Bergelson, E. & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796.

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of neurophysiology* 112/5: 1105–1118.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bloomfield, L. (1926). A Set of Postulates for the Study of Language. *Language*, 2/3: 153–164.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., & Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: A positron emission tomography activation study. *Brain*, 117(6), 1241–1253.

Bouchard, K. E., N. Mesgarani, K. Johnson & E. F. Chang. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature,* 495/7441: 327–332.

Brannon, E. M. (2005). The independence of language and mathematical reasoning. *Proceedings of the National Academy of Sciences*, 102(9), 3177–3178.

Bromberger, S. & M. Halle (1995). The Ontology of Phonology. *Studies in Logic and the Foundations of Mathematics* 134: 725–743.

Buchert, R., Thomasius, R., Wilke, E, Petersen, K., Nebeling, B., Obrocki, J., Schulze, O., & Schmidt, U. (2008). Sustained effects of ecstasy on the human brain: A prospective neu- roimaging study in novel users. *Brain*, 131/11: 2936–45.

Burton, M. W, Small, S., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience* 12: 679–690.

Caramazza, A. (1988). Some aspects of language processing as revealed through the analysis of acquired aphasia: The lexical system. *Annual Review of Neurosdence*, 11: 395–421.

Carr, P. (2006). Universal grammar and syntax/phonology parallelisms. *Lingua*, 116(5): 634–656.

Champagne-Lavau, M. & Joanette,Y. (2009). Pragmatics, theory of mind and executive functions after a right-hemisphere lesion: Different patterns of deficits. *Journal of Neurolinguistics*, 22: 413–26.

Charles, R. (2019). *Epistemic boundedness and poverty of the stimulus: Why phonology is not different*. Ms. Concordia University.

Chomsky, N. (1951/1979). *Morphophonemics of Modern Hebrew*. New York: Routledge.

Chomsky, N. (1955/1975). *The Logical Structure of Linguistic Theory*. New York: Plenum Press.

Chomsky, N. (1957a). Review of "Fundamentals of Language" by Roman Jakobson and Morris Halle. *International Journal of American Linguistics*, 23/3: 234–242.

Chomsky, N. (1957b) Review of "A Manual of Phonology" by Charles F. Hockett. *International Journal of American Linguistics*, 23/3: 223–234.

Chomsky, N. (1962). Explanatory models in linguistics. In Nagel, E., P. Suppes & A. Tarski (eds.), *Logic, Methodology, and Philosophy of Science*. 528–550. Stanford, CA: Stanford University Press.

Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. (1968). *Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (1976). On the biological basis of language capacities. In Rieber, R. W. (ed.), *The Neuropsychology of Language*. 1–24. Boston, MA: Springer.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Chomsky, N. (1986). *Knowledge of Language. Its Nature, Origins, and Use*. New York: Praeger.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2007). Approaching UG from below. In Sauerland, U. & H. M. Gärtner (eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*. 1–24. Berlin: Mouton de Gruyter.

Chomsky, N., M. Halle & F. Lukoff (1956). On Accent and Juncture in English. In Halle, M. (ed.), *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*. 65–80. The Hague: Mouton.

Chomsky, N. & M. Halle. (1965). Some Controversial Question in Phonological Theory. *Journal of Linguistics* 1: 97–138.

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

Chomsky, N., A. J. Gallego & D. Ott (2017) Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges. Ms., MIT, Universitat Autònoma de Barcelona, and University of Ottawa. [https://ling.auf.net/lingbuzz/003507].

Corina, D., Poizner, H., Bellugi, U, Feinberg, T, Dowd, D., and O'Grady-Batch, L. (1992). Dissociations between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language*, 43: 414-47.

Curtiss, S. (1982). Developmental dissociations of language and cognition, in L. Obler and L. Menn (eds), *Exceptional Language and Linguistics.* New York: Academic Press, 285-312.

Curtiss, S. (1988a). The special talent of grammar acquisition, in L. Obler and D. Fein (eds), *The Exceptional Brain.* New York: The Guilford Press, 364-86.

Curtiss, S. (1988b). Abnormal language acquisition and grammar: Evidence for the modularity of language, in L. Hyman and C. Li (eds), *Language, Speech, and Mind: Studies in Honor of Victoria A. Fromkin.* New York: Routledge, Kegan, & Paul, 81-102.

Curtiss, S. (1995). Language as a cognitive system: Its independence and selective vulnerability, in C. Otero (ed.), *Noam Chomsky: Critical Assessments, 4.* New York: Routledge.

Curtiss, S. (2011). Revisiting modularity, in Yukio Otsu (ed.), *Proceedings of the 11th Annual Tokyo Conference on Psycholinguistics.* Tokyo: Hituzi Syobo, 1-33.

Dahl, Veronica. (2002). On implicit meanings. In *Computational logic: Logic programming and beyond*. 506–525. New York: Springer.

Dehaene-Lambertz, G. & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport* 12/14: 3155–3158.

Dehaene-Lambertz, G., Baillet, S. & Hertz-Pennier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298: 2013–2015.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384/6605: 159–161.

Eickhoff, S. B., S. Heim, K. Zilles & K. Amunts. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367/1896: 2399–2421.

Emmorey, K. (2002). *Language, Cognition and the Brain: Insights from Sign Language Research.* Mahwah, NJ: Lawrence Erlbaum.

Emmorey, K., Kosslyn, S. M., & Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition*, *46*(2), 139-181.

Fant, G. (1969). Distinctive features and phonetic dimensions. *STL-QPSR*, 10/2–3): 1–18.

Fodor, J. A. (1980). Reply to Putnam. In M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. 325–334. Cambridge, MA: Harvard University Press.

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge: MIT Press.

Gallistel, R. C. & A. P. King (2010). *Memory and the Computational Brain. Why Cognitive Science Will Transform Neuroscience*. Malden: Wiley-Blackwell.

Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2016). Language and speech in autism. *Annual review of linguistics*, 2, 413-425.

Greenberg, J. (1966). *Universals of language*. Cambridge, MA: MIT Press.

Grinstead, J., MacSwan, J., Curtiss, S., & Gelman, R. (1997). The independence of language and number. In *Twenty-Second Boston University Conference on Language Development*.

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language*, 27: 135–59.

Grodzinsky, Y. & Finkel, L. (1998). The neurology of empty categories. *Journal of Cognitive Neuroscience*, 10/2: 281–92.

Grodzinsky, Y. & Novogrodsky, R. (2006). Syntactic movement in agrammatism and S-SLI: Two different impairments, in A. Belletti, E. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari (eds), *Language Acquisition and Development*. Newcastle: Cambridge Scholars Press, 197–210.

Grosvald, M., Gutierrez, E., Hafer, S., & Corina, D. (2012). Dissociating linguistic and non-linguistic gesture processing: Electrophysiological evidence from American Sign Language. *Brain and language*, 121(1), 12–24.

Guenther, F. H., Satrajit S. G. & J. A. Tourville. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96/3: 280–301.

Gussenhoven, C. & H. Jacobs (2017). *Understanding Phonology.* Fourth edition. London: Hodder Arnold – Hachette Livre.

Hale, M. & C. Reiss. (2003). The Subset Principle in phonology: why the tabula can't be rasa. *Journal of Linguistics*, 39/2: 219–244.

Hale, M., M. Kissock & C. Reiss. (2007) Microvariation, variation, and the features of universal grammar. Lingua, 117: 645–665.

Hale, M. & C. Reiss. (2008). *The Phonological Enterprise.* Oxford: Oxford University Press.

Halle, M. (1954). The strategy of phonemics. *Word* 10/2–3: 197–209.

Halle, M. (1959). *The Sound Pattern of Russian*. The Hague: Mouton.

Halle, M. (1961). On the Role of Simplicity in Linguistic Descriptions. *Proceedings of Symposia in Applied Mathematics* 12: 89–94.

Halle, M. (1962). Phonology in Generative Grammar. *Word*, 18: 54–72.

Halle, M. (1964). On the Bases of Phonology. In Fodor, J. A. & J. Katz (eds.), *The Structure of Language*. 324–333. New York: Prentice Hall.

Halle, M. (1975). Confessio grammatici. *Language*, 51/3: 525–535.

Halle, M. (1978/2002). Knowledge Unlearned and Untaught. In Halle (2002): 95–104.

Halle, M. (2002). *From Memory to Speech and Back*. Berlin – New York: Mouton de Gruyter.

Halle, M. & A. Marantz (1993). Distributed morphology and the pieces of inflection. In Hale, K. & S. J. Keyser (eds.), *The View from Building 20*. 111–176. Cambridge, MA: MIT Press.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, 4: 353–363.

Harley, H & R. Noyer (1999). Distributed Morphology. *GLOT International*, 4/4: 3–9.

Harris, J. W. (1984). Autosegmental phonology, lexical phonology and Spanish nasals. In M. Aronoff & R. Oehrle (eds.), *Language sound structure*. 67–82. Cambridge, MA: MIT Press.

Hart, J., Berndt, R. S., & Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction. *Nature*, 316: 439–40.

Hayes, B. (2009). *Introductory Phonology*. Oxford: Wiley-Blackwell.

Heisenberg, W. (1962). *Physics and Philosophy: The Revolution in Modern Science*. New York: Harper & Row.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13/2: 135–145.

Hickok, G. & D. Poeppel. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8/5: 393–402.

Hillis, A. E. (2010). The 'standard' for poststroke aphasia recovery. *Stroke*, 1316–1317.

Hirsh-Pasek, K., & Golinkoff, R. (1996). *The Origins of Grammar: Evidence from Early Language Comprehension*. Cambridge, MA: MIT Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (eds.), *Language, speech, and communication. Methods for assessing children's syntax*. 105–124. Cambridge, MA: The MIT Press.

Householder, F. (1965). On Some Recent Claims in Phonological Theory. *Journal of Linguistics*, 1: 13–34.

Inkelas, S. (1995). The consequences of optimization for underspecification. In *Proceedings of the North East Linguistic Society*, 25: 287–302. University of Pennsylvania: Graduate Linguistic Student Association.

Idsardi, W. & E. Raimy (2013). Three Types of Linearization and the Temporal Aspects of Speech. In Roberts, I. & M. T. Biberauer (eds.), *Challenges to Linearization*. 31–56. Berlin: De Gruyter.

Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (2015). In defense of theory. *Cognitive science*, 41: 185-212.

Jacob, P. (2010). Intentionality. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/ intentionality.

Jakobson, R., G. Fant & M. Halle. (1952). *Preliminaries to speech analysis*. Cambridge: MIT Press.

Jodzio, K., Biechowska, D., & Leszniewska-Jodzio, B. (2008). Selectivity of lexical-semantic disorders in Polish-speaking patients with aphasia: Evidence from single-word comprehension. *Archives of Clinical Neuropsychology*, 23/5: 543-51.

Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kaplan, R. M. (1987). Three seductions of computational psycholinguistics. In P. Whitelock, M. M. Wood, H. L. Somers, R. Johnson, & P. Bennett (eds.) *Linguistic theory and computer applications*. 149–188. London: Academic Press.

Karmiloff-Smith, A., Klima, E., Bellugi, U., Grant, J., & Baron-Cohen, S. (1995). Is there a social module? Language, face processing, and theory of mind in individuals with Williams syndrome. *Journal of cognitive Neuroscience*, *7*(2), 196-208.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In Fromkin, V. (ed.), *Phonetic linguistics: essays in honor of Peter Ladefoged*. 115–132. Orlando: Academic Press.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, 18: 321–334.

Lenneberg, E. 1967. *Biological Foundations of Language*. New York: Wiley.

Leonard, L. B. (2017). Phonological impairment. *The handbook of child language*, 573-602.

Lionnet, F. (2017). A theory of subfeatural representations: The case of rounding harmony in Laal. *Phonology*. 34/3: 523–564.

Lust, B. (2006). *Child Language – Acquisition and Growth*. Cambridge: Cambridge University Press.

MacGregor, L. J., Pulvermüller, F., Van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications* 3: 711.

Magrassi, L., Aromataris, G., Cabrini, A., Annovazzi-Lodi, V., & Moro, A. (2015). Sound representation in higher language areas during language generation. *Proceedings of the National Academy of Sciences*, 112/6: 1868–1873.

Marr, D. (1982/2010). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information.* Cambridge, MA: MIT Press.

Matamoros, C. & C. Reiss (2016). Symbol taxonomy in biophonology. In A. M. Di Sciullo (ed.), *Biolinguistic Investigations on the Language Faculty*. 41–54. Amsterdam: John Benjamins.

McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, J. J. (2007). Slouching toward optimality: Coda reduction in OT-CC. *Phonological Studies (Journal of the Phonological Society of Japan)*, 7: 89–104.

McCarthy, J. J. & A. Prince (1995). *Faithfulness and Reduplicative Identity*. Rutgers Optimality Archive: http://roa.rutgers.edu/files/216-0997/roa-216-mccarthy-1.pdf.

McGuire, P. K., Robertson, D., Thacker, A., David, A. S., Kitson, N., Frackowiak, R. S. J., & Frith, C. D. (1997). Neural correlates of thinking in sign language. *NeuroReport*, 8/3: 695-8.

Medawar, P. B. (1960). *The Future of Man: The BBC Reith Lectures 1959*. London: Methuen and Company.

Mesgarani, N., Cheung, C., Johnson, K, & E. F. Chang. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343(6174), 1006–1010.

Michotte, A. & L. Burke (1951). Une nouvelle enigme de la perception: le "donné amodal" dans l'expérience sensorielle. *Proceedings of the 13th International Congress of Psychology*, 179–180.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford: Oxford University Press.

Miller, G. A. (1998). Interview. In M. Gazzaniga, R. Ivry, & G. Mangun, (eds.), *Fundamentals of cognitive neuroscience.* New York: Norton.

Mountcastle, V. B. (1998). Brain science at the century's ebb. *Daedalus*, 127/2: 1–36.

Odden, D. (2013). *Introducing Phonology*. Second edition. Cambridge: Cambridge University Press.

Okada, K., Matchin, W. & Hickok, G. (2018). Phonological feature repetition suppression in the left inferior frontal gyrus. *Journal of cognitive neuroscience* 30/10: 1549–1557.

Papillon, M. (2020). *Precedence-Relation-Oriented Phonology*. PhD thesis. University of Maryland.

Phillips, C., Pellathy, T, Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, 12: 1036-55.

Pinker, S. (1994). *The Language Instinct*. New York, William Morrow.

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, 29/1–2: 34–55.

Poser, W. J. (1993). Are strict cycle effects derivable? In S. Hargus & E. M. Kaisse (eds.), *Studies in lexical phonology*, volume 4. San Diego: Academic Press.

Poser, W. J. 2004. *On the status of Chumash sibilant harmony*. Ms. University of Pennsylvania.

Prince, A. & P. Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Pulvermüller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and language* 104/3: 244–253.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge: MIT Press.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42.

Reiss, C. (2012). Towards a bottom-up approach to phonological typology. In A. M. Di Sciullo (ed.), *Towards a Biolinguistic Understanding of Grammar*. 169–191. Amsterdam: John Benjamins Publishing Company.

Reiss, C. (2021, forthcoming) Priority union and feature logic in phonology. *Linguistic Inquiry*. 1–10.

Reiss, C. & V. Volenec. (2021, forthcoming). An Internalist Perspective on Phonological Features. *Canadian Journal of Linguistics*.

Rohlf, M. (2016). Immanuel Kant. In *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/kant/.

Rondal, J. (1995). *Exceptional Language Development in Down Syndrome.* Cambridge: Cambridge University Press.

Samuels, B. (2011). *Phonological architecture: a biolinguistic perspective*. Oxford: Oxford University Press.

Sapir, E. (1933/1949). The Psychological Reality of Phonemes. In Mandelbaum, D. G. (ed.), *Selected Writings of Edward Sapir in Language, Culture and Personality.* 33–45. Los Angeles: University of California Press.

Saussure, F. de (1916/1959) *Course in General Linguistics*. New York: Columbia University Press.

Schaller, S. (1995). *A Man Without Words.* Berkeley: University of California Press.

Scharinger, M., Domahs, U., Klein, E., & Domahs, F. (2016). Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus. *Brain and language*, 163: 42–49.

Scharinger, M., P. J. Monahan & W. J. Idsardi. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research*, 55/3: 903–918.

Seikel, A. J., W. D. King & D. G. Drumright. (2009). *Anatomy & Physiology for Speech, Language, and Hearing*. 4th edition. New York: Delmar, Cengage Learning.

Skeide, M. A., Brauer, J., & Friederici, A. D. (2014). Syntax gradually segregates from semantics in the developing brain. *NeuroImage*, *100*, 106–111.

Sklar, L. (2000). *Theory and Truth: Philosophical Critique within Foundational Science*. Oxford: Oxford University Press.

Smith, N. (2003). Dissociation and modularity: reflections on language and mind. *Mind, Brain, and Language: Multidisciplinary Perspectives*, 87–111.

Smith, N. (2010). *Acquiring phonology: a cross-generational case-study*. Cambridge: Cambridge University Press.

Smith, N. V., & Tsimpli, I. M. (1995). *The mind of a savant: Language learning and modularity*. New York: Blackwell Publishing.

Smith, N., Tsimpli, I., Morgan, G., & Woll, B. (2010). *The signs of a savant: Language against the odds*. Cambridge: Cambridge University Press.

Tippett, D. C., & Hillis, A. E. (2016). Vascular Aphasia Syndromes. In *Neurobiology of Language* (pp. 913–922). Academic Press.

Versace, E., Martinho-Truswell, A., Kacelnik, A., & Vallortigara, G. (2018). Priors in animal and artificial intelligence: where does learning begin? *Trends in cognitive sciences*, 22/11: 963–965.

Volenec, V. (2019). Nasal place assimilation in Substance Free Logical Phonology. *Loquens* 3/1.

Volenec, V. & C. Reiss (2017). Cognitive Phonetics: The Transduction of Distinctive Features at the Phonology-Phonetics Interface. *Biolinguistics*, 11: 251–294.

Volenec, V. & C. Reiss (2019). The intervocalic palatal glide in Cognitive Phonetics. *Proceedings of the 49th Meeting of the North East Linguistics Society*. 255–264.

Wang, P. P., Doherty, S., Rourke, S. B., & Bellugi, U. (1995). Unique profile of visuo-perceptual skills in a genetic syndrome. *Brain and Cognition*, *29*(1), 54–65.

Wiese, R. (2000). *The phonology of German*. Oxford: Oxford University Press.

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage*, 47(4), 1992–2004.

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature communications*, 10(1), 1–7.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch processing in speech perception. *Science* 256: 846–849.

Zwicker, E. & Fastl, H. (2013). *Psychoacoustics: Facts and models*. Springer Science & Business Media.

DISCUSSION WITH IRIS BERENT

(NORTHEASTERN UNIVERSITY)

Berent, Iris. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 74-83.

COMMENTS

Volenec and Reiss's manifesto (2020) advocates both a theoretical claim and a methodological position. Their theoretical claim is that phonology is a formal system that is substance free and abstract. Methodologically, they rely heavily on logical analysis and auxiliary assumptions concerning the "epistemological precedence" of certain hypotheses over others; they present very little in the way of empirical support for their conclusions. The paucity of empirical evidence is especially striking, given that, per their *realism* assumption, Volenec and Reiss's subject of inquiry—the mind/brain of individual speakers, is closely aligned with the natural sciences. Their methods, however, decidedly are not.

Volenec and Reiss's logical argument for "substance free" phonology departs from the premise that phonology is a member of the cognitive sciences, and as such, it concerns cognition, that is, mental states, not stimuli (e.g., sounds; for summary, see (1)). Now, since per "cognitive science", cognition is a system of abstract symbols, it follows that phonology, likewise, consists of a system of abstract symbols, distinct from their sensorimotor externalizations; hence, sensorimotor substance is irrelevant to phonology. For Volenec and Reiss, the symbolic hypothesis further implies the capacity to operate over variables (I will refer to it as the *algebraic hypothesis*, see Chomsky & Schützenberger, 1963; Marcus, 2001). And since in the authors' view, abstract mental categories are not reliably marked by sensorimotor cues, phonological primitives are likely inborn as well.

(1) *The logical argument for substance-free phonology*

a) *Premise 1*: Phonology is I-Phonology—it is the study of mental states (rather than sounds).
b) *Premise 2*: The mind is a system of abstract symbols.
c) *Inferences*: Phonology is a system of abstract (substance-free) symbols
   i. Phonology is algebraic: it is endowed with the capacity to operate over variables (rules).
   ii. UG: Some phonological primitives and principles are innate.

To be clear, I happen to concur with several of Volenec and Reiss's conclusions, and I have expressed my views in dozens of publications (for reviews, see Berent, 2013l; 2017). But I believe these difficult questions ought to be decided by rigorous empirical investigation. A logical analysis alone won't do.

For starters, Volenec and Reiss's logical argument hinges on the premise that the "symbolic" view is normative in cognitive science. And by "symbolic" they mean not only that the mind encodes symbols (i.e., some representational format that carries information), but also that it operates over variables (i.e., the algebraic hypothesis). This latter position, however, is certainly not universally accepted by contemporary cognitive scientists.

The founding fathers of cognitive science (Fodor & Pylyshyn, 1988; Newell, 1980; Newell & Simon, 1961; Pinker & Prince, 1988) hypothesized that the mind operates on structured symbols, and that this structure guides mental operations. For example, Fodor and Pylyshyn's proposal (1988) implies that all geminates (e.g., *pp, dd, bb*) are assigned the same structure (XX, where X is a variable that stands for "any consonant"), distinct from singletons (X), and it is the structure of symbols (XX), not their substance (*pp*), that guides mental operations. Consequently, such operations, are by definition, sub-

stance free. And since, in this view, one's knowledge of geminates concerns their structure (*XX), then the relevant generalization should apply productively, *across the board*, even to segments that are unattested in one's language (e.g., xx, for English; Berent & Marcus, 2019). But the algebraic hypothesis has since been the subject to an ongoing controversy,

Connectionists (e.g., Rumelhart, McClelland, & Group, 1986), for example, implicitly assume symbols, but deny that the mind includes structure-sensitive operations or even discrete representations; for many connectionists, cognition is gradient all the way up (for recent reviews, cf. Berent & Marcus, 2019; Pater, 2018).

Other cognitive scientists advocate the view of cognition of "embodied". Representations, in this view, encode precisely the sensorimotor detail that Volenec and Reiss reject (for reviews, Glenberg, 2015 Pulvermüller & Fadiga, 2010 Barsalou, Kyle Simmons, Barbey, & Wilson, 2003). For example, the representation of the sound token of "ba" consists of the sensory experience evoked by this sound token and correspondingly, the articulatory motor action of producing it.

There is in fact strong empirical evidence to suggest that something of this sort is actually right. A large literature has shown that the phonetic categorization of speech sounds selectively activates motor areas of the brain (e.g., labial sounds activate the lip motor area; coronals activate the tongue motor area; (e.g., Pulvermüller et al., 2006); that the stimulation of these areas—either by transcranial magnetic stimulation (e.g., D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; D'Ausilio et al., 2009; Möttonen & Watkins, 2009; Smalle, Rogers, & Möttönen, 2014) or mechanically (e.g., Bruderer, Danielson, Kandhadai, & Werker, 2015; Masapollo & Guenther, 2019)—selectively affects the perception of the relevant categories.

In light of this, it should be clear that the algebraic hypothesis is hardly "the only game in town". Cognitive scientists *do not* agree that all cognitive processes are algebraic. In fact, most would reject the weaker claim that *some* cognitive processes are al-

gebraic. Consequently, it is plainly wrong to assert that, since all cognitive processes are patently algebraic, phonology must be likewise so as well.

To be clear, none of this proves that the algebraic hypothesis is wrong. But it does show that, the algebraic hypothesis isn't a slam dunk in cognitive science, so it cannot be so for phonology. At best, one can say that the cognitive science has demonstrated that the mind includes (*inter alia*) algebraic mechanisms (and as noted, this remains controversial); whether or not this applies to phonology is thus an open question.

This question can certainly be addressed empirically. One strategy is to examine the scope of phonological generalizations. One could evaluate whether people do in fact generalize phonological principles across the board (as predicted by the algebraic hypothesis) and, using computational simulations, one could further evaluate whether such generalizations are further *inconsistent* with non-algebraic mechanisms. To the extent both conditions are met, one could then be reasonably confident that phonology is algebraic, which, by definition, would mean it is substance free (for reviews; Berent, 2013; Berent & Marcus, 2019).

A second strategy is to evaluate whether phonology is substance free directly, independently of the question of whether it is algebraic. The strategy here is to determine whether phonological processes can be doubly dissociated from their sensorimotor correlates. First, one could show that the phonology (e.g., of spoken language) can be distinguished from the phonetic properties of the stimulus (e.g., speech)—this is the strategy that Volenec and Reiss consider in a cursory fashion. Second, one could also show the converse, namely, that stimuli of distinct sensorimotor properties (speech and sign) are assigned invariant phonological representation—this is a strategy I have followed in my own research (Berent, Bat-El, Brentari, Dupuis, & Vaknin-Nusbaum, 2016; Berent, Bat-El, Brentari, & Platt, 2020).

(2) *Open empirical questions*

a) Phonology is substance free.
b) Phonology is algebraic.
c) Some phonological principles are innate.

Finally, there is of course the innateness question. Volenec and Reiss seem to suggest that their abstraction hypothesis is closely linked to innateness, as abstract categories presumably cannot be learned from stimuli properties. The cognitive literature, however, has amply shown that humans are adept implicit learners of both rules (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and statistical regularities (Saffran, Aslin, & Newport, 1996), whereas many innate constraints are sensorimotor in nature. Accordingly, so the link between innateness and abstraction is far from evident.

Summarizing, the questions of whether phonology has (a) substance free representations, (b) algebraic mechanisms and (c) innate primitives and constraints are three independent questions that are in urgent need of empirical investigation (see (2)). While I commend Volenec and Reiss for their call to anchor phonology in broader cognitive considerations, I only wish their paper practiced more of what they preach.

Barsalou, L. W., Kyle **Simmons**, W., **Barbey**, A. K., & **Wilson**, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Sciences, 7(2), 84-91. doi:http://dx.doi.org/10.1016/S1364-6613(02)00029-3. **Berent**, I. (2013). *The phonological mind*. Cambridge: Cambridge University Press. **Berent**, I. (2017). On the origins of phonology. *Current Directions in Psychological Science, 26*, 132-139. **Berent**, I., **Bat-El**, O., **Brentari**, D., **Dupuis**, A., & **Vaknin-Nusbaum**, V. (2016). The double identity of linguistic doubling. *Proceedings of the National Academy of Sciences, 113*(48), 13702-13707. doi:10.1073/pnas.1613749113. **Berent**, I., **Bat-El**, O., **Brentari**, D., & **Platt**, M. (2020). Knowledge of language transfers from speech to sign: Evidence from doubling. *Cognitive Science*(44), 1. **Berent**, I., & **Marcus**, G. (2019). No integration without structured representations: Reply to Pater. *Language, 95*. **Bruderer**, A. G., **Danielson**, D. K., **Kandhadai**, P., & **Werker**, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1508631112. **Chomsky**, N., & **Schützenberger**, M. P. (1963). The Algebraic Theory of Context-Free Languages. In P. Braffort & D. Hirschberg (Eds.), *Stu-*

dies in Logic and the Foundations of Mathematics (Vol. 35, pp. 118-161): Elsevier. **D'Ausilio**, A., **Bufalari**, I., **Salmas**, P., & **Fadiga**, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior, 48*(7), 882-887. doi:10.1016/j.cortex.2011.05.017. **D'Ausilio**, A., **Pulvermüller**, F., **Salmas**, P., **Bufalari**, I., **Begliomini**, C., & **Fadiga**, L. (2009). The motor somatotopy of speech perception. *Current Biology: CB, 19*(5), 381-385. Retrieved from http://0-search.ebscohost.com.ilsprod.lib.neu.edu/login.aspx?direct=true&db=cmedm&AN=19217297&site=ehost-live&scope=site. **Fodor**, J., & **Pylyshyn**, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71. **Glenberg**, A. M. (2015). Few Believe the World Is Flat: How Embodiment Is Changing the Scientific Understanding of Cognition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 69*(2), 165-171. doi:10.1037/cep0000056. **Marcus**, G. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge: MIT press. **Marcus**, G. F., **Vijayan**, S., **Bandi Rao**, S., & **Vishton**, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77-80. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9872745. **Masapollo**, M., & **Guenther**, F. H. (2019). Engaging the Articulators Enhances Perception of Concordant Visible Speech Movements. *Journal of speech, language, and hearing research : JSLHR, 62*(10), 3679. doi:10.1044/2019_JSLHR-S-19-0167. **Möttonen**, R., & **Watkins**, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience, 29*(31), 9819-9825. **Newell**, A. (1980). Physical symbol systems. *Cognitive Science, 4*(2), 135-183. doi:10.1016/S0364-0213(80)80015-2. **Newell**, A., & **Simon**, H. A. (1961). Computer Simulation of Human Thinking. *Science, 134*(3495), 2011-2017. doi:10.1126/science.134.3495.2011. **Pater**, J. (2018). Generative linguistics and neural networks at 60: foundation, friction, and fusion. *Language*. **Pinker**, S., & **Prince**, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73-193. **Pulvermüller**, F., & **Fadiga**, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience, 11*(5), 351-360. Retrieved from http://0-search.ebscohost.com.ilsprod.lib.neu.edu/login.aspx?direct=true&db=cmedm&AN=20383203&site=ehost-live&scope=site. **Pulvermüller**, F., **Huss**, M., **Kherif**, F., **Moscoso del Prado Martin**, F., **Hauk**, O., & **Shtyrov**, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865-7870. doi:10.1073/pnas.0509989103. **Rumelhart**, D., E., **McClelland**, J., L, & **Group**, T. P. R. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge MA: MIT Press. **Saffran**, J., R, **Aslin**, R. N., & **Newport**, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1298. **Smalle**, E. H. M., **Rogers**, J., & **Möttönen**, R. (2014). Dissociating Contributions of the Motor Cortex to Speech Perception and Response Bias by Using Transcranial Magnetic Stimulation. *Cerebral Cortex*. doi:10.1093/cercor/bhu218. **Volenec**, V., & **Reiss**, C. (2020). Formal generative phonology. *Radical: a journal of phonology*. Retrieved from http://radical.cnrs.fr/volenec-reiss-formal-generative-phonology/.

REPLY

We agree with most of what Berent writes and are grateful for her insightful suggestions for further research. First and foremost, we agree that the questions that she outlines in her concluding remarks – about the exact nature of phonological representations, phonological computations, and phonological UG – are far from being resolved. Indeed, they are all open research questions "in urgent need of empirical investigation", as Berent says (p. 3). We also agree that many other, more specific issues that we have raised require further empirical verification. In fact, the main goal of our paper was to provide a theoretically coherent and potentially productive program for further *empirical* research. This endeavor has been motivated by a widespread practice in modern phonology which we perceive as a combination of theoretically incompatible notions, such as constructing purportedly generative theories of phonological phenomena (e.g., OT is 'constraint interaction in *generative* grammar', as the title of its founding text indicates) while failing to follow the implications of the competence/performance distinction which is the core of generative linguistics. Naturally, many of the assumptions that we have laid out in the paper need to be (further) tested empirically; we have merely tried to be explicit and clear about what these assumptions are. On the other hand, we have referred to our own previous work where we have subjected many of these assumptions to empirical scrutiny (e.g., Volenec & Reiss 2019; Volenec 2019), and, perhaps more importantly, we have built our claims on a large body of other people's previous empirical studies. For example, the section on the neural basis of Cognitive Phonetics draws from a large number of experimental studies in neurobiology of language and speech (summarized on pp. 52–53 of the paper, but see Volenec & Reiss 2017: §3.2 and §4.3 for a fuller treatment).

However, we do not agree with Berent's characterization of our reasoning on why phonology is substance-free. After reminding us that "cognitive scientists do not agree that all cognitive processes are algebraic", she concludes that "it is plainly wrong to assert that, since all cognitive processes are patently algebraic, phonology must be like-

wise so as well" (p. 2). This has not been our line of reasoning, neither in the present paper nor in any of our previous work on SFP. The substance-free conception of phonology follows from the competence/performance distinction. Following a long tradition in linguistics, we take I-language to be a form of knowledge, fully contained in the brain of a person. Substance, as it is understood at least since De Saussure (1916/1959), is a result of (what we now call) performance, of putting that knowledge into communicative action. We think that knowledge and action are not one and the same thing, which can clearly be observed, for example, when a person is in a state of deep coma: knowledge is preserved despite the absence of any kind of action. Patients who recover from coma may suffer from dysarthria or other *speech* disorders, but a comatose patient whose L1 is English will not wake up not knowing English; the absence of action does not entail the absence of knowledge.

Berent cites studies that purportedly show that cognition is "embodied" (p. 2), which she interprets as meaning that there is some phonetic substance in phonological knowledge after all. It is of interest that none of these studies are conducted within a framework that distinguishes – either theoretically or methodologically – between competence and performance; and what they study is almost always linguistic performance, i.e., speech perception and production. For example, the D'Ausilio et al. (2012) paper that Berent cites is, in their own words, about "listening to speech", not about phonological knowledge. Even papers that talk about "[b]rain language mechanisms" (p. 351), such as Pulvermüller & Fadiga (2010), are actually concerned with "perception and comprehension of stimuli" (p. 351) and with "the production of speech sounds" (p. 353), not with phonological knowledge.[25] And that is precisely the problem that motivated our paper: since the vast majority of 'sound patterns' researchers study linguistic performance, it is no wonder that they find 'substance' everywhere.

25 Failure to distinguish between competence and performance can also clearly be observed in many narrower issues, such as equating phonemes and speech sounds: "during this babbling phase (months 6–12), the sounds that babies articulate become increasingly similar to the types of speech sounds, or phonemes, that they hear frequently" (Pulvermüller & Fadiga 2010: 352).

We do not deny that the human mind *represents* phonetic substance (i.e., parses it into a symbolic form using equivalence classes); what we claim is that a mental representation of substance is not substance, at least not in the sense that this term has been used from De Saussure onward. Phonological knowledge may contain a representation of, say, tongue movements. But the mental representation itself does not contain tongue movements because there are no tongues in the brain. Likewise, a digital photograph of a 500 Euro bill – stored on a computer and encoded in a binary representational system – does not contain 500 Euros. Berent challenges us to "[r]oll up [our] sleeves and find out" (p. 1) through empirical research whether phonology is substance-free, but that seems equivalent to charging us with the task of empirically testing if we can pay the rent with photos of Euros instead of actual Euros. It should, however, be pointed out that ample empirical evidence for the substance-free nature of features already exists. The conclusion of the MEG experiment conducted by Phillips et al. (2000: 1040) was that "there is good reason to distinguish the acoustic and phonetic representations that underlie categorical perception from the discrete phonological category representations involved in lexical storage and phonological computation", and that when it comes to phonological computation, "all within-category contrasts are lost: e.g., all different tokens of /d/ are treated by phonological processes as exactly the same", irrespective of the phonetic substance that is indirectly associated with the bundle of features that we conventionally label as 'd'. Magrassi et al. (2015: 1) have shown that "the activity of language areas" is organized in terms of features even "when language is generated mentally before any utterance is produced or heard", i.e., when there is no phonetic substance whatsoever. Similarly, Okada et al. (2018) have conducted an fMRI investigation of silent word sequence production (i.e., the subjects read words in their minds) where the stimuli (different words displayed one after another on a screen) varied in the degree of feature overlap in consonant onset position. The experiment confirmed a featural organization of investigated word sequences in absence of overt speech. These neuropsychological studies suggest that phonological features cannot be equated with the pho-

netic correlates that are typically associated with them, i.e., that features and phonetic substance are two different things.

For our claim about the substance-free nature of phonological knowledge to be true, it doesn't matter if there is a high or a low degree of resemblance (homomorphism) between a representation and its correlate. A map of some terrain, no matter how detailed and hi-fi, is still not the terrain itself. The same logic holds between phonological representations and phonetic substance. We think, then, that the actual contribution of the empirical studies which Berent cites is not in showing that phonological knowledge contains substance, but rather in potentially elucidating the degree of homomorphism between them. No matter what the degree of homomorphism turns out to be (an important research question in itself), a mental representation of substance is still not substance. The main reason why we are insisting on this practically trivial but often misunderstood issue is because some phonologists (e.g., Lionnet 2017) directly encode phonetic measurements (e.g., formant values) within phonology, thereby claiming – explicitly or implicitly – that the mind directly encodes such values. If we are correct in assuming that there is no phonetic substance in phonological knowledge, then such analyses cannot be the truth about I-phonologies.

DISCUSSION WITH JOAQUIM BRANDÃO DE CARVALHO

(UNIVERSITÉ PARIS 8, CNRS UMR 7023 - STRUCTURES FORMELLES DU LANGAGE)

Carvalho (Brandão de), Joaquim. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) “Formal Generative Phonology”. *Radical: A Journal of Phonology*, 2, 84-89.

COMMENTS

Volonec & Reiss's (henceforth V&R) paper is meant to provide "a general program for the study of phonology as a branch of cognitive science [whose] main tenet is that phonological competence is to be characterized as a formal – that is, explicit, logically precise, and substance-free – manipulation of abstract symbols" (p. 59).

First and foremost, I would like to say that I am deeply committed to the formalist tradition in phonology, as opposed to "laboratory" and other "grounded" approaches. I totally agree that "it is phonetics that is grounded in phonology and not the other way around" (p. 54). Most of my own work follows a similar line of thought, which, by the way, long predates generative grammar: besides Sapir (quoted in the article), both the Prague school and glossematics, albeit differently, have regarded features as abstract entities.

Vis-à-vis V&R's Formal generative phonology (henceforth FGP), I want to confess: (i) a deep agnosticism as to whether features are or not innate – I stick strictly to internalism, and I thank the authors for recalling that this entails neither acceptance nor denial of nativism; (ii) a complete lack of knowledge about the neurolinguistic literature.

V&R's paper is clear on the whole, well structured, and well argued in epistemological terms. However, several questions arise, at least for those not familiar with FGP. I will leave aside minor issues like: (i) are binary features a necessary corollary of the proposal? (ii) same question about rules and serialism; (iii) if "there is no temporal as-

pect in phonology" (p. 15), how then should affricates, diphthongs, prenasal consonants, etc. be accounted for, knowing that such segments may have "edge effects"?

I only wish to mention one point that follows from the central tenet of FGP, which seems to take up SPE minus its last chapter. V&R claim (p. 25) that features and computation maintain an arbitrary relation. If so, I see several problems that boil down to a single question: does FGP admit the many purported cases of "crazy rules"? Let us take the classic example of Velar softening (mentioned in p. 37). Is /k/ → [s] /_i as licit as /k/ → [c] /_i? Also, is the issue of why final devoicing is preferable to final voicing really irrelevant to phonology?

I am afraid it is, according to V&R. In this case, if feature-filling rules manipulate any feature ([s] is not [palatal]!), they have to explain, for example, why the two processes affecting /k/ are treated quite differently by speakers: electri[k]ity, albeit unattested, remains a possible word in English, and should be perceived as such. Or is /k/ → [s] /_i disallowed because this rule is not based on a natural class? These crucial aspects are far from clear in the article.

A related question remains: how can feature "combinability" (p. 22) be accounted for without any reference to markedness? I cannot dwell further on this topic; let me only recall that, outside OT, and contrary to V&R's claim (p. 36-37), markedness is not a matter of frequency, but of (formal) complexity, nor is frequency primarily "determined by political, social, economic, geographical and other extra-linguistic factors". This is simply preposterous.

REPLY

Carvalho gently chides us for not mentioning the Prague school and Glossematic traditions of treating features as abstract entities. He is surely correct that such work contributes to the sense in which we understand the abstractness of features, however we maintain that in the absence of a mentalist grounding to their work, it would be a bit

anachronistic to equate our notion of the abstractness of (mental) features to their abstractness of (apparently) non-mental features. Their abstract features are abstract entities with no spatiotemporal existence, whereas ours are encoded in individual nervous systems. In another paper (Volenec & Reiss 2017), we do connect our understanding of features to their antecedents in structuralism and early generative phonology, stating the similarities and differences.

We have no problem with Carvalho's widely shared “deep agnosticism as to whether features are or [are] not innate'”. Our point is that in order to lose that agnosticism and come down against innateness, one is obliged to acknowledge and refute long standing arguments (from Kant, Fodor and others) that something very much like phonological features must be innately present in all cognitive domains. Additionally, denying innateness requires a plausible story of how features could arise *ex nihilo*. Finally, denying innate universal features requires acknowledgement that no generalizations across languages are possible of the form ‘voiced obstruents tend to get devoiced in codas’' since there is, under the anti-nativism view, no such well-defined thing as ‘voiced obstruent’ recurring across languages. At this point, it is worth bringing up the argument that Chomsky & Halle (1965: 127) presented to Householder when he proposed (Householder 1965) that it would be better for phonological theory if we could postulate features on a language-specific basis:

> "As has often been noted – cf., e.g., Chomsky (1964a: 944–73 & c:65–110) – the assumption of a universal feature structure is made (often only implicitly) in every approach to phonology that is known, and clearly cannot be avoided. What is at issue only is the choice of features, not their universality. To repeat the obvious once again, suppose that we were to approach, let us say, English with no assumption at all about universal features (i.e., suppose we were to follow Householder's proposal). Consider only monosyllables of the form CVC. We find certain phonetic elements in initial position and others (phonetically distinct from them) in final position. Phonemic analysis of any sort requires that we somehow identify initial and final phones. With no assumption about feature structure, we can do this in any way we like. No way is more 'useful' than any other, or more 'simple', in any abstract sense. For example, if we wish to show that the maximally simple system is to identify initial p with final t and initial t with final p, it is only necessary to invent a physical feature (there are innumerably many of these) which is common to initial p and final t, and excludes initial t and final p. Call this the feature A. Following Householder, there is no reason to prefer, e.g., the feature Labial to the feature A. And furthermore, having now (à la Householder) accepted A for our analysis, we can immediately demonstrate the absurdity of using the feature Labial for the analysis of English. For consider the complexity of the feature Labial. Thus a phone is Labial only if it is in initial position and has the feature A or in final position and has the feature non-A. Obviously such a disjunctive feature does not contribute to the 'simplicity' or 'usefulness' of linguistic description (in the vague sense of these terms that Householder apparently has in mind).
>
> It is for such reasons as this that one does not 'construct features from scratch for each language'."

So, innateness of features is to us clearly the null hypothesis (see Reiss & Volenec 2021 for further discussion).

Carvalho's next paragraph correctly identifies two weaknesses of our paper. (i) We *assume* binarity. We think privative features present several problems discussed elsewhere, such as the problem of referring to the unmarked member of a privative opposition; and we think that anything more than binarity is conceptually unnecessary given current states of knowledge. We also want to reiterate the underappreciated point that underspecification, which yields a kind of ternarity from binarity is a simplification of the feature system. (ii) As for the point about apparent sequencing of opposite feature values inside of segments, Carvalho is absolutely correct, as we acknowledge in footnote 14. We ignore such cases for now in order to work only with set-level formalisms. With David Ta-Chun Shen we are currently exploring ways of maintaining our results with slightly more complex representations of segments that allow for contours with edge effects.

The question about crazy rules is mostly an artifact of tradition and informality of rhetoric. Given our advocacy for breaking down the arrow of traditional rules, pretty much any well-known ‘crazy rule’ phenomenon will have to be recast as a sequence of primitive rules.

Carvalho says that according to us “the issue of why final devoicing is preferable to [i.e., more common than] final voicing” is “really irrelevant to phonology'”. He is correct, and we should acknowledge that this is not something we prove, but rather a principle we adopt because it is consistent with various other ideas we hold. For example, we should have made clearer the connection between our treatments of segments as (more or less) unstructured sets and our substance-free views. Other phonological models of segment structure try to derive the behavior of segments in context with other segments at least in part from their internal structure (e.g. head-dependent relationships among elements). Such approaches are very different from ours, and this reasonable position is what we understand Carvalho to be referring to in saying that “markedness is [...] a matter of (formal) complexity”. Such views are not incoherent or illogical, but we do not adopt them for a variety of reasons. Here we briefly mention that approaches that

massively reduce the number of primes (elements, particles or features) available for phonological representation (in a reasonable attempt at theoretical parsimony in modeling UG) are forced to enrich the overall model of human language by introducing, often without explicit acknowledgement, a tremendous amount of language-specific phonetics. In other words, if two languages are said to contain "the same segment" but each pronounces that segment differently, then a whole other module is needed in the speech chain. This is exactly what the universal transduction of our Cognitive Phonetics is meant to avoid.

Finally, let's clarify our "preposterous" claim that markedness is primarily "determined by political, social, economic, geographical and other extra-linguistic factors". We do not mean here that the kind of formal markedness Carvalho seems to work with is subject to such factors – there is a problem of the term 'markedness' meaning too many things to different people. All we mean is that widespread properties of, say, English-type grammars, like do-support, have become extremely common in the I-languages of the world for non-linguistic reasons like the British domination of the seas in the 18th and 19th centuries. A survey of currently living humans would show that lots of them have I-languages that have do-support… because lots of people speak English. No linguistic conclusions about the complexity or simplicity of do-support can be drawn from its distribution on earth.

DISCUSSION WITH ALEX CHABOT

(UNIVERSITÉ CÔTE D'AZUR)

Chabot, Alex. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 90-112.

COMMENTS

In their article, Volenec & Reiss build on work in Volenec & Reiss (2018) to lay out a complete research program for phonology, which is built on two related wings: 1) the computational/representational system of phonology itself (which they call *Logical Phonology*) and 2) a system of *universal phonetics* that interprets the output of phonology (which they call *Cognitive Phonetics*) and results in the production of speech. At the heart of this paper is a very bright line between the computational/representational system underlying the human faculty of phonology – viewed here as a property of mind instantiated in human brains – and the properties of the physical world (henceforth *the real world*). The program elaborated is one that is *substance-free*; that is, in stark contrast to models of phonology which yoke phonology to acoustic or articulatory properties, here phonology is deaf and blind to the real world outside of mind.

They begin by making an argument for treating phonology as a cognitive science. To this, I have little to add, since Volenec & Reiss do what I consider an admirable job of outlining their model and arguing extensively across the domains of linguistics and neurobiology to support their arguments. In short, I am entirely in agreement with their argument that phonology is something that happens in human minds/brains, and consequently, that I-language is the object of inquiry for phonology.

In fact, I should say now that I am generally extremely sympathetic to the larger arguments made by Volenec & Reiss. I agree with Volenec & Reiss that language is innate, and so the component of mind which generates I-language is fruitfully conceptua-

lized as a module of mind (Fodor, 1983). This module consists of a set of primitive symbolic objects, which we call representations, and a set of computations that operate over the set of representations. We thus have two aspects to any given phonological system: the representational wing, and the computational wing.

I should like to frame my discussion in the context of minimalism as argued for by Chomsky (1995, 2002). That is, following Boeckx (2013:68), I would like to consider UG by asking "How little can be attributed to UG while still accounting for the variety of I-languages attained?" With this in mind, I will point out arguments made by Volenec & Reiss with which I do not agree, and why. The points I will address in what follows represent to my mind the main topics of debate – and thus departures from Volenec & Reiss' position on features – in Substance-Free phonology.

To wit: the nature of phonological features and their exponence as neuromuscular instructions for speech. While proponents of the Substance-Free program all (seemingly) agree that phonological computation is arbitrary and not subject to any demands from the real world, there seems to be two camps as far as features and the exponence of features as physical objects are concerned. The first camp is represented by Volenec & Reiss and continues the position established by Hale & Reiss (2000, 2008) and sees features as objects with UG determined phonetic expressions. Volenec & Reiss do an admirable job of making their position clear, as on page 25 when they argue that the relation between phonological features and their real-world correlate is "not random or arbitrary," but rather, as they say on page 25, is "lawful." The other camp is argued for in Blaho (2008), Iosad (2017) and Odden (forthcoming) among others (see the volume edited by Chabot, forthcoming), and views features as objects whose expression is language specific and emerges according to ambient linguistic data. I should like to briefly highlight those two positions in this discussion, since ultimately the consequences for the Substance-Free program are far reaching.

The debate hinges on a critical question: are features innate with a biologically predetermined phonetic correlate or do they emerge through language acquisition with an

arbitrary and learned phonetic correlate. Since I agree with Volenec & Reiss that phonology is a natural object – a product of biology instantiated in human brains – the question is fundamental since it bears on the nature of Universal Grammar (UG): are features endowed by UG or do they emerge based on language use as a product of domain general cognitive capabilities? If features are endowed by UG and thus a part of the human genome then UG is correspondingly "bigger." Alternatively, if features arise through the interaction of stimulus (input to the language learner) and domain general cognitive capabilities, then UG ends up looking substantially smaller.

From the perspective of minimalism (Chomsky 1995; 2007; Samuels 2011a), this is a desirable goal. As argued by Hauser et al. (2002) a small UG is more plausible than a big one, given the relatively short period of time during which the evolutionary leap resulting in UG occurred in the human branch of hominids. Any aspect of language which is domain-specific to UG is also specific to human beings, by virtue of being an element of that human-specific evolutionary leap.

Features. The second part of Volenec & Reiss' paper establishes the nature of the symbolic representations used by the phonological module – that is, the physical units in the brain over which phonological computations[26] are carried out. Since Trubetzkoy, phonemes or segments have not been conceptualized as atomic units, but rather as unities which can be decomposed into features. In SPE, phonological features look a lot like phonetic features, the feature [+labial] is a phonological symbol which is directly translated into an articulatory cue.

In Substance-Free phonology, as articulated lucidly by Volenec & Reiss, features are symbolic objects that have neither acoustic nor articulatory correlates; since they are realized in the brains/minds of speakers, they cannot be of "like-kind" with the real-world objects they represent. This seems obvious upon reflection: "blue" may describe

26 Phonological computation is the other major wing of Volenec and Reiss' paper. See Odden (2013, to appear) for another model of a formal, substance-free account of phonological computation.

both a physical object (light of a certain wavelength) and the cognitive representation of the color blue, but there is no light in the brain, and so no blue wavelengths of light.

Following earlier work in SFP (Hale & Reiss 2000a, Hale & Reiss 2008, Reiss 2018), and developing arguments in Volenec & Reiss (2018), Volenec & Reiss argue that features must be innate. Their reasoning hinges on what they call a *logical argument* based on a philosophy of *rationalism:* in order for learning to take place, the human mind must contain innate knowledge of basic units which cannot themselves be learned. As Hale & Reiss (2003) put it, "ya gotta start with something."

This logical argument is not based on empirical evidence and is confronted with an alternative point of view which argues that features emerge through language use (Mielke 2008) and are language specific. Indeed, it seems there are many things that humans do not have innate knowledge of, but experience allows them to form categories and make differences: between the sound of a dog barking and the sound of a motorcycle engine, for example. Boersma (1998:161), Cowper & Hall (2015:160) and Odden (to appear) argue innate phonological feature values are not necessary for acquisition of language.

In the phonological and phonetic literature there are scattered accounts of unusual, typologically rare segments, as for example clicks (see Halle 1995 and Miller 2011). Some segments are exceedingly rare indeed as noted by Ladefoged & Everett (1996) who describe the case of a "voiceless apico-dental plosive [followed by a] voiceless labio-labial trill" in the Chapakuran languages Wari‘ and Oro Win, spoken in Brazil.

This sound, at least in the context of its native phonological system, is entirely conventional, with a conventional phonological identity: it is contrastive, it is not limited in the word forms it can be realized in, and it is a regularly produced and recognized sound by speakers of Wari' and Oro Win. Ladefoged & Everett's observations are interesting (1996:799):

> All the Chapakuran languages are severely endangered. It seems likely that [t͡ʁ̥] will not occur in any of the world's languages in the latter part of the next century; but that is no indication that this sound will not develop again elsewhere. Some high elitist group in England may decide that babies are [t͡ʁ̥at͡ʁ̥a], and the word may catch on and spread to other things. So we can never really tell what features will be needed for describing languages. In principle it is the complete set of human vocal sounds that can be integrated into the flow of speech, and that are sufficiently distinct from one another; but this is too cumbersome a notion to be of practical value for working linguists describing languages.

Ladefoged & Everett's observations have an interesting implication for *any* theory of universal features: if the set of features is universal, then we would expect exceedingly rare segments to appear with greater frequency in the typological record, even despite possible performance pressure in perception or articulation. There is no reason to assume that Chapakuran [t͡ʁ̥] or Xhosian [ǁʰ] should not occur here and there as the result of regular variation and language change. This is because speakers would all be drawing from a single well of possible features and all be subject to the same perceptual and articulatory pressures. Yet, some typologically rare sounds are restricted to isolated or areal phenomena, suggesting their genesis is an exceptional, language specific occurrence.

There are two possible solutions to the problem of rare sounds which do not require the jettisoning of a UG endowed set of universal features. The first is to allow "slack" in the realization of phonological features: a feature can vary in its realization to a certain extent, so that /t/ might be realized as dental, alveolar, or retroflex, for example. But admitting "slack" into the theory is not enough – there is no principled way for deciding when a phonetic form entails one phonological feature or another. In the Kabyle Berber vowel system (Kossmann & Stroomer 1997), the three vowels /i a u/ are realized variably in the vowel space, such that /a/ may be realized as [æ], or /i/ as [ɛ]. There is no

principled way outside of phonology to know when /i/ is realized so low that it is perceived as something else (/a/ for example).

Volenec & Reiss (2018, 2020) chose another tack, arguing that there may be a very large number of features indeed, which give rise to an even larger number of possible segments (on the order of hundreds of thousands), which are all innate and genetically determined. This "combinatoric explosion" accounts for the slack instantiated in the relationship between some phonological objects and their phonetic exponence. In this view, every variant of a phoneme has a distinct composition in the form of features. In turn, some languages engage relatively more features to realize a given phoneme than do others which exhibit less variation. The reason for typological rarities, in this view, is a question of performance, not of competence: there are factors outside of phonology proper (in perception, articulation, or elsewhere) which favor some sounds at the expense of others.

In fact, this leads Volenec & Reiss to reject arguments made in Keating 1985, 1988; Kingston & Diehl (1994); Cohn (1998); and Kingston (2007) among others that the phonetic realization of phonological features is *language specific and learned* and as a consequence phonological features may be variably realized; as an example, languages that exploit [+round] may vary in the degree to which lip-rounding is manifested. Instead, Volenec & Reiss argue that a system of innate features provides such rich combinatory power that all variation can be explained by specific combinations of feature variables (see Hale et al. 2007), and there is thus no need to suggest that phonetic variability is the result of variability in the realization of phonological objects. Further, they argue on page 23 that with no language specific phonetics, there is a reduction in "the sequence of conceptual steps needed to account for the externalization of language." This reduction in conceptual steps comes with a price, however: a concurrent increase in the size and remit of UG, since every feature of every language must be present in UG.

I am in agreement with Volenec & Reiss when they say on page 23, "the general research agenda is to build bio-linguistically plausible cognitive models of language." In

turn, we must ask what leads UG to develop a universal system of features which leaves any given number of features fallow in the brains/minds of speakers whose language makes no use of those features? For example, if, *contra* the language specific phonetic position, variation in the amount of lip-rounding across languages is to be accounted for by different features ([+roundA], [+roundB], [+roundAF], [+roundXA], etc) all of which result in different degrees of lip-rounding across languages, why is there such a cognitive over-abundance in phonological features when so many features are destined to go unused in the brains/mind of speakers? This seems to be contrary to the minimalist program, and not a likely evolutionary path, given that cognitive resources are metabolically expensive. Using features to account for variation also means we are faced with another conundrum: gradience is infinite but features are finite, meaning that not all gradient effects can be reduced to feature specification. One possible solution to this is by invoking performance factors, but doing so means feature-based gradience (phonology) and performance-based gradience (phonetics) become indistinguishable.

An alternative point of view proposes that features emerge (Mielke 2008) according to the needs of speakers; a mode more in line with a minimalist conception of phonology. Volenec & Reiss reject this proposal, which I will return to below.

Transduction. Unlike in the systems such as SPE where phonological features are lawfully translated into phonetic features with the same label by switching from binary values in phonology to gradient, numerical values in phonetics, in SFP features are purely cognitive objects and are not of like-kind with their real-world correlates. In keeping with a modular theory of mind (Fodor, 1983), in order for one module to produce output to another module, or to receive input from another module, there must be an interface between modules which translates the vocabulary of one into that of the other. If the basic units of phonology contain no information related to their instantiation as physical objects in the real world, there must be some device that carries out the translation of features as cognitive objects in minds/brains to neuro-muscular instructions that result

in articulatory and acoustic objects. This device is known as the phonetics/phonology interface, and various treatments regarding can be found in the literature (see for example Boersma & Hamman 2008; Hamman 2011; Scheer 2014).

Since Hale & Reiss (2000a), the interface between phonetics and phonology has been called *transduction*[27], the process that converts phonological vocabulary to phonetic vocabulary. While Volenec and Reiss maintain that there is no phonetic substance in the features themselves, they argue that the features are indirectly connected to their real-world instantiation by means of transduction, which imbues features with substance. Crucially, For Volenec & Reiss, the relationship between primes and their phonetic correlates is non-arbitrary and "lawful."

As they say on page 53, "transduction is *deterministic*, which means that it assigns the same neuromuscular schema to each feature every time that feature is transduced." Although features have no information relevant to their phonetic form, they also cannot be decoupled from phonetics; Hale & Reiss argue that this must be so, because if it were not, then language acquisition would not be possible.

Transduction proceeds according to universal schemata, features cannot cheat destiny: [+labial] must be realized with specific neuromuscular correlates even if in phonology itself [+labial] doesn't mean anything at all, being invisible to computation except as a symbol which can be operated over. In this conception of phonology primes look much like those of Chomsky & Halle (1968) in that they form a universal set.

A final point to be made regarding the deterministic transduction of features is that it is not clear how transducers are linked to features, if the features themselves are all alike and contain no labels but *must* be associated with specific transducers (in contrast to a scheme where the link between features and phonetic exponence emerges through acquisition or the ambient linguistic environment). The solution proposed by Volenec & Reiss on page 23 is interesting: "Possibly, the actual form of all features is the same—a

27 Transduction is a general term in cognitive science and refers to any process by which vocabulary is translated; from light waves to action potentials at the retina, for example. That is, transduction translates the vocabulary of one module into another.

neural spike (i.e., an action potential). But more importantly, the unique location of the spike and/or the rate of its repetition is how the transducer determines the identity of the feature and 'knows' which neuromuscular schema (e.g., labiality and not, say, nasality) to assign to it." Indeed, as pointed out, on page 21, in cognitive neuroscience it is assumed that symbols are distinguished by some combination of place coding, rate coding, or time coding of neuronal activity in brain tissue.

However, this is also what general audition would look like, and at the present time it is impossible to tell what neurophysiological patterns underlie phonological computation (Gallistel & King 2009; Gazzaniga 2010). It is clear that acoustic and articulatory speech patterns are manifest in variations in amplitude and topographic patterns in neural tissue, but as Grimaldi (2018:87) notes "finding a ubiquitous system for speech-specific processing in line with the phonological features hypothesis is difficult. When we look at the data, it is hard to disambiguate between N1m/N1 evidence suggesting pure acoustic patterns and those indicating abstract phonological features." That is, it is not clear in neurobiological terms how the auditory cortex differentiates between basic auditory activity and phonological computation; we can not tell the difference between what is acoustic – coming from the real world – and the phonological (see Gazzaniga et al. 2014:482). For the time being then, such hypotheses are purely speculative.

On the arbitrariness of the interface. The final point regarding features and the phonology/phonetics interface that I would like to address concerns the deterministic nature of the interface argued for by Volenec & Reiss. As they argue on page 25, since in their model, the instantiation of a phonological feature is not arbitrary – but rather *lawful* – phonological objects should always be realized in the same way. The example they use is [+round], saying that if the realization of phonological features were arbitrary, [+round] could be realized here "as a lowering of the velum," and there as "raising of the tongue dorsum, and so on."

However, it has been argued that the phonological identity of a segment cannot be predicted based on its phonetic properties alone (Kaye 2005; Purnell 2009). As such, phonetic properties are in a many-to-one relationship with phonological objects. This observation led Kaye to posit his *phonological epistemological principle*: it is only by phonological behavior that phonological identity can be discovered (see also Gussmann 2001). Volenec & Reiss argue that this many-to-one relationship is an effect of performance and not relevant to the nature of phonological features and their expression.

This would mean, then, that learners must have some way of separating the phonological wheat from the phonetic chaff, as it were. There are good reasons to assume this is the case: cases of "phonetics/phonology mismatches" (Hamann 2014), show that the phonological identity can be belied by its phonetic qualities: fricatives that act like glides as in Argentine Spanish (Harris & Kaisse 1999), a sibilant that acts like a vowel in Blackfoot (Goad & Shimada 2014), a sonorant that acts like a stop in Quechua (Gallagher 2019), and a fricative [ð] that acts like a sonorant in Woods Cree (Starks & Ballard 2005). In all these cases speakers derive the phonological identity of segments despite misleading phonetic garb.

If the transduction of features is lawful, performance can introduce sufficient distortion to make lawful transduction essentially opaque. Yet, learners are able to see past distortion introduced by performance, showing that lawful transduction is not necessary for the success of the discovery procedure. If this is true in some cases, as in the case of phonetics/phonology mismatches, why should transduction be deterministically lawful elsewhere? Put another way, the mapping between phonological features and their phonetic expression is essentially arbitrary in some cases, so by Occam's razor we must assume that it can be arbitrary in all cases.

This is essentially the argument made by Chabot (2019), applied to the case of rhotics. Rhotics represent a phonological class with extensive surface variation, but a stable phonological identity. In Chabot (2019), it is argued that performance factors cause the wide range of variation exhibited by phonological rhotics, but that performance factors

do not play a role in the stable phonological role of rhotics – the phonetics/phonology interface handles an arbitrary mapping in the case of rhotics. If the mapping is arbitrary for one class of phonological segment, then in order to avoid stipulating exceptions, it must be arbitrary for all classes of phonological segments. The consequence is that natural classes are open, unrestricted, and determined by phonological behavior.

Finally, Volenec & Reiss briefly touch on the nature of features in sign languages on page 31. They rightly remark that if features are innate, then the set of features may be quite large indeed, in order to account for the phonological features used in signed languages[28]. If spoken language makes use of a universal set of innate features with a deterministic exponence, then features are modality-dependent: there is one set of features for spoken languages, and one set of features for signed languages because the real-world correlate for each is different in kind. Again, from the perspective of minimalism, this seems like an unwelcome conclusion. Are signed language features laying dormant in the brains of speakers of spoken language, and is the reverse true for speakers of signed languages?

A model whereby features emerge based on ambient linguistic information does not encounter this doubling of feature sets: features emerge the same way for signed languages as they do for spoken language – speakers build categories across modalities based on usage patterns. Sandler (2014) shows that in Al-Sayyid Bedouin Sign Language, there is reason to suppose that primes are in fact emerging, and not innate. As argued by Sandler, since features are not innate to sign languages, is it not worth assuming that the same is true for spoken languages?

On the neurolinguistic front, Poeppel et al. (2012) provide several neurological correlates implicated in *both* spoken and signed languages, bolstering the position that language processing is modality independent, and that the same kinds of features underpin both modalities, suggesting that primes are both substance-free and emergent.

28 There can be little doubt that there is phonology in sign languages, and that signs, like vocal segments, can be decomposed into phonological primes – discrete, meaningless units of contrast (Stokoe 2005[1960]; Brentari 2011; Sandler 2012).

An alternative to innate features and deterministic transduction is the foundation of so called "radical" forms of SFP (Hall 2007; Blaho 2008; Samuels 2011b; Iosad 2017; Odden to appear), where the link inherent in transduction between features and substance is dissolved. Radical Substance Free Phonology (RSFP) argues that there is no reason to assume primes have phonetic content, or that they should be fated to be universally realized via specific articulatory configurations.

Both camps of Substance-Free Phonology have in common a place for explanations outside of UG itself, so-called "third-factor" explanations (Chomsky 2005) that are not a part of UG. That is, there is a "filter" on surface patterns in the real-world that is a reflection of facts about human perception, the mechanics of speech production, and the vagaries of history, but not about UG; the difference is that in emergent models UG does not have to account for the set of features. Instead, they are the product of domain-general capacities such as categorization (Harnad 2005).

On computation and natural classes. Volenec & Reiss reject emergent features and argue that an advantage of their model of computation, dubbed *Logical Phonology*, over a model in which features are emergent (see for example Odden to appear) is that rules can only be formulated over natural classes, making the formal system explicit. They argue that in an emergent system, rules may somehow refer to something *other* than a natural class, making the formal system impossible to evaluate: a phonologist cannot tell a possible rule from an impossible rule.

I do not fully understand this argument: it seems to me that in a system with emergent features, if /r u k i/ participate in some phonological process, either as a united class of triggers or targets, then by definition /r u k i/ constitute a natural class. Since the earliest work in Generative Phonology, it has been recognized that processes which share structural properties (identical environments or identical structural changes) are functionally united (Kissberth 1970). That is, a natural class is defined by its participation in a unified rule, which presumably targets a shared feature. However, I argued

above that it is not by examining surface properties of segments that phonological features can be identified.

That is, again falling back on Kaye's epistemological principle, *a natural class can only be identified by phonological behavior*. The alternative, which I understand to be Volenec & Reiss' position, would be to identify natural classes based on surface properties: one could deny the unity of /r u k i/ by saying there is no phonetic property which unambiguously unites the disparate segments, and *ipso facto* they do not constitute a natural class.

For example, they argue on page 41 that "if a language has a rule that refers to the vowels /i/ and /u/, then the rule must apply equally to the vowel /y/ = {+ROUND, +HIGH, –BACK}, if this vowel also occurs in the language, because /y/ is also a superset of the generalized intersection ∩ {i, u}." Here, the determination is that /y/ *must* contain a shared feature with /i/ and /u/ because /i/ and /u/ are high vowels, and /y/ is also a high vowel. In this case, the judgment is based on surface properties (substance), not on phonological behavior; this is an example of what Hale & Reiss (2000b, 2008) refer to as *substance-abuse*: a misuse of phonetic properties in the analysis.

The only way to determine that /y/ is a member of the set of high vowels is if it patterns with them, merely examining the phonetic qualities of the segment is not enough. If the phonological behavior of a set of segments can be fruitfully described as a single phonological process, then that set of segments constitutes a natural class, which is one of the strengths of a model with emergent features.

Conclusion. Volenec & Reiss have written an engaging article which, to my mind, makes an indispensable contribution to the epistemology of phonology. I find the main thrust of their argument logical, well argued, thoroughly supported, and a useful roadmap for a research program in phonology. The points I have highlighted above constitute what I consider to be the main problems for Volenec & Reiss. Those same issues constitute the subject of ongoing debate and research in Substance-Free Phonology, but

do not threaten the program itself. I am left with a few minor questions, such as what Logical Phonology ends up looking like if features are unary instead of binary, as argued for by Odden (to appear), but those few remaining details seem unlikely to change the program outlined by Volenec & Reiss in any substantial way.

Blaho, Sylvia. 2008. *The syntax of phonology: A radically substance-free approach*. Tromsø: University of Tromsø dissertation. **Boeckx**, Cedric. 2013. Biolinguistics: forays into human cognitive biology. *Journal of Anthropological Studies* 91. 63–89. **Boersma**, Paul. 1998. *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics. **Boersma**, Paul & Silke **Hamann**. 2008. The evolution of auditory dispersion in bidirectional constriant grammars. *Phonology* 25. 217–270. **Brentari**, Diane. 2011. Sign language phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory*, 691–721. Oxford: Blackwell. **Chabot**, Alex. 2019. What's wrong with being a rhotic? *Glossa: a journal of general linguistics* 4(1): 38. 1–24. **Chabot**, Alex (ed.) Forthcoming. Features and phonetic substance: Implications for phonological theory. *The Canadian Journal of Linguistics*. **Chomsky**, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press. **Chomsky**, Noam. 2002. Minimalism. In Juan Uriagereka (ed.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–156. Cambridge, MA & London: MIT Press. **Chomsky**, Noam. 2007. Approaching UG from below. In Uli Sauerland & Martin Gartner (eds.), *Interfaces + Recursion = Language?*, 1–29. Berlin: Mouton de Gruyter. **Chomsky**, Noam & Morris **Halle**. 1968. *The sound pattern of English*. New York: Harper & Row. **Cohn**, Abigail C. 1998. The phonetics-phonology interface revistied: Where's phonetics? In Amanda R. Doran (ed.), *Proceedings of the 1998 Texas Linguistics Society conference*, 25–40. Austin, Texas: Department of Linguistics, University of Texas at Austin. **Cowper**, Elizabeth & Daniel Currie **Hall**. 2015. *Reductiō ad discrīmen*: where features come from. *Nordlyd* 41(2). 145–164. **Fodor**, Jerry. 1983. *The modularity of mind*. Cambridge, Massachusetts: MIT Press. **Gallagher**, Gillian. 2019. Phonotactic knowledge and phonetically natural classes: the plain uvular in Cochabamba Quechua. *Phonology* 36(1). 37–60. **Gallistel**, C. R. & Adam Philip **King**. 2009. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Chichester, U.K.: Wiley-Blackwell. **Gazzaniga**, Michael. 2010. *Neuroscience and the correct level of explanation for understanding mind*. Trends in Cognitive Sciences 14. 291–292. **Gazzaniga**, Michael S., Richard B. **Ivry** & George R. **Mangun**. 2014. *Cognitive neuroscience: The biology of the mind*. New York & London: W.W. Norton & Company, Inc. **Goad**, Heather & Akiko **Shimada**. 2014. In some languages, /s/ is a vowel. *Proceedings of the 2013 Meeting on Phonology*. 1–12. **Grimaldi**, Mirko. 2018. The phonetics-phonology relationship in the neurobiology of language. In Roberto Petrosino, Pietro Cerrone & Harry van der Hulst (eds.), *From sounds to structures: Beyond the veil of Maya*, 66–104. Berlin & New York: Mouton de Gruyter. **Gussmann**, Edmund. 2001. Hidden identity, or the double life of segments. In Katarzyna Dziubalska-Kołaczyk (ed.), *Constraints and preferences*, 229–249. Berlin & New York: Mouton de Gruyter. **Hale**, Mark, Madelyn **Kissock** & Charles **Reiss**. 2007. Microvariation, variation, and the features of universal grammar. *Lingua* 117. 645– 655. **Hale**, Mark & Charles **Reiss**. 2000a. Phonology as cognition. In Noel Burton-Robers, Philip Carr &

Gerard Docherty (eds.), P*honological knowledge: Conceptual and empirical issues*, 161–184. Oxford: Oxford University Press. **Hale**, Mark & Charles **Reiss**. 2000b. “Substance abuse” and “dysfunctionalism”: Current trends in phonology. *Linguistic Inquiry* 31(1). 157–169. **Hale**, Mark & Charles **Reiss**. 2003. “The subset principle in phonology: Why the t*abula* can't be *rasa*. *Journal of Linguistics* 39(2). 219-244. **Hale**, Mark & Charles **Reiss**. 2008. *The phonological enterprise*. Oxford: Oxford University Press. **Hall**, Daniel Currie. 2007. *The role and representation of contrast in phonological theory*. Toronto: University of Toronto dissertation. **Halle**, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26(1). 1–46. **Hamann**, Silke. 2011. The phonetics-phonology interface. In Nancy Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Continuum companion to phonology*, 202–224. London: Continuum. **Hamann**, Silke. 2014. Phonetics-phonology mismatches. Paper presented at the Eleventh Old-World Conference in Phonology, Leiden, January 22-25, 2014. **Harnad**, Stevan. 2005. To cognize is to categorize: Cognition is categorization. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 19–43. Amsterdam: Elsevier. **Harris**, James W. & Ellen M. **Kaisse**. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 16. 117–190. **Hauser**, Marc D., Noam **Chomsky** & W. Tecumseh **Fitch**. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298. 1569–1579. **Iosad**, Pavel 2017. *A Substance-free Framework for Phonology. An Analysis of the Breton Dialect of Bothoa*. Edinburgh: Edinburgh University Press. **Kaye**, Jonathan. 2005. GP, I’ll have to put your flat feet on the ground. In Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster (eds.), *Organizing grammar. studies in honor of Henk van Riemsdijk*, 383–288. Berlin: Mouton de Gruyter. **Keating**, Patricia A. 1985. Universal phonetics and the organization of grammars. In Victoria Fromkin (ed.), P*honetic linguistics: Essays in honor of Peter Ladofoged*, 115–132. New York: Academic Press. **Keating**, Patricia A. 1988. The phonology-phonetics interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey volume 1*, 281–302. Cambridge: Cambridge University Press. **Kingston**, John. 2007. The phonetics-phonology interface. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 401–434. Cambridge: Cambridge University Press. **Kingston**, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70(3). 419–454. **Kissberth**, Charles W. 1970. On the Functional Unity of Phonological Rules. *Linguistic Inquiry* 1(3). 291-306. **Kossmann**, Maaren G. & Harry J. **Stroomer**. 1997. Berber phonology. In Alan S. Kaye (ed.), *Phonologies of Asia and Africa*, 461–475. Winona Lake, Indiana: Eisenbrauns. **Ladefoged**, Peter & Daniel **Everett**. 1996. The status of phonetic rarities. *Language* 72(4). 794–800. **Mielke**, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press. **Miller**, Amanda L. 2011. The representation of clicks. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Maiden, Mass.: Wiley-Blackwell. 416-443. **Odden**, David. 2013. Formal phonology. *Nordlyd* 40(1). 249–273. **Odden**, David. To appear. Radical substance free phonology and feature learning. T*he Canadian Journal of Linguistics*. **Poeppel**, David, Karen Emmorey, Gregory Hickok & Liina Pylkkänen. 2012. Towards a new neurobiology of language. *Journal of Neuroscience* 32(41). 14125–14131. **Purnell**, Thomas. 2009. Phonetic influence on phonological operations. In Eric Raimy & Charles E. Cairns (eds.), *Contemporary views on architecture and representations*, 337–354. Cambridge, Massachusetts: MIT Press. **Reiss**, Charles. 2018. Substance free phonology. In S.J. Hannahs & Anna R.K. Bosch (eds.), *Handbook of phonological theory*, 425–452. London & New York: Routledge. **Samuels**, Bridget. 2011a. A minimalist program for phonology. In Cedric Boeckx (ed.), *The Oxford handbook of linguistic mini-*

malism, 574–594. Oxford: Oxford University Press. **Samuels**, Bridget D. 2011b. *Phonological architecture: A biolinguistic approach*. Oxford: Oxford University Press. **Sandler**, Wendy. 2012. The phonological organization of sign languages. *Languge and Linguistics Compass* 6/3. 162–182. **Sandler**, Wendy. 2014. The emergence of the phonetic and phonological features in sign language. *Nordlyd* 41(2). 183–212. **Scheer**, Tobias. 2014. Spell-out, post-phonological. In Eugeniusz Cyran & Jolanta Szpyra-Kozłowska (eds.), *Crossing phonetics-phonology lines*, 255–275. Newcastle upon Tyne: Cambridge Scholars Publishing. **Starks**, Donna & Elaine Ballard. 2005. Woods Cree /ð/: An unusual type of sonorant. *International Journal of American Linguistics* 71(1). 102–115. **Stokoe**, William C., Jr. 2005[1960]. Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of Deaf Studies and Deaf Education* 10(1). 3–37. **Volenec**, Veno & Charles **Reiss**. 2018. Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics* 11. 251–294.

REPLY

In his commentary on our paper, Chabot endorses many of the philosophical and linguistic ideas on which our theories are predicated, concluding that the paper can serve as a "useful road-map for a research program in phonology" (p. 102), which is exactly the role we envisioned for it. Naturally, there are some specific claims from the paper that Chabot challenges, such as the relationship between features and their correlates, the innateness of features and its implications for typology, and the definition of a possible rule through the notion of a natural class. We think that a part of these apparent disagreements stems from slight misinterpretations of our claims, so we will use this opportunity to further clarify our perspective. We will also comment on the few areas where our positions genuinely seem to diverge, aiming to sharpen these competing proposals so that they can be investigated in future research.

Chabot attributes to us the claim that "features are symbolic objects that have neither acoustic nor articulatory correlates" (p. 92). However, as stated on p. 25 of the paper, we *do* think that features have phonetic correlates, and in an earlier paper which Chabot refers to we have stated that "[w]e understand distinctive features […] as a particular kind of substance-free units of mental representation, neither articulatory nor acoustic in themselves, but rather having articulatory and acoustic correlates" (Volenec & Reiss

2017: 253). Consistent with Hale et al. (2007), we think that while the correlates of features are variable – a phenomenon known as *lack of invariance* – they typically fall within some reasonably well-defined articulatory and acoustic space. Roughly, this means that if we are dealing with, say, the substance-free feature +BACK, then we expect that during the pronunciation of a vowel that contains that feature the tongue dorsum will typically be positioned in a certain posterior space in the oral cavity and we also expect its F2 to be in a certain low region. Every repetition of what is mentally that same vowel will lead to correlates that fall within these spaces (barring particularly strong coarticulatory effects), even though precise measurements will show that no two repetitions are physically identical. The automatic, deterministic nature of transduction is what ensures that the vast majority of feature correlates fall within those spaces, while the tremendous complexity of all other performance components gives rise to lack of invariance.

Chabot uses typological insight of the kind presented in Ladefoged & Everett (1996) to argue that we shouldn't assume that features are innate: "if the set of features is universal, then we would expect exceedingly rare segments to appear with greater frequency in the typological record, even despite possible performance pressures in perception and articulation" (p. 94). Upon finding the "exceedingly rare" speech sound [tʁ̥] in the Chapakuran languages Wari' and Oro Win, Ladefoged & Everett (1996: 799) conclude that "we can never really tell what features will be needed for describing languages. In principle it is the complete set of human vocal sounds that can be integrated into the flow of speech." We do not think that it is in the purview of UG to account for the relative frequency of particular kinds of segments in the I-languages of the world. Instead, as the biologically pre-specified component of the language faculty, UG merely determines what a *possible* phonological segment is. In other words, UG determines, for example, that many of the "human vocal sounds", to use Ladefoged & Everett's phrase, such as slurping and laughing, are linguistically irrelevant: since UG provides no features for representing such vocalizations, they will never participate in the shaping of

phoneme inventories. What's rare or common is primarily determined by the geographical, social, and political factors that influence the PLD. If, for the sake of the argument, the Wari' speakers, whose I-languages contain /t͡ʁ̥/, somehow conquered the world and enforced their language upon all subsequent generations, then what was formerly an exceedingly rare segment would now be completely common (while the segment /ɹ/ of the rapidly vanishing English would become a typological curiosity).[29] But clearly this has nothing to do with UG and everything to do with how the influence of geo-socio-political factors shapes the PLD.

Chabot thinks that our solution for "the problem of rare sounds" is to expand the innate feature set: "Volenec and Reiss […] argu[e] that there may be a very large number of features indeed, which give rise to an even larger number of possible segments (on the order of hundreds of thousands), which are all innate and genetically determined" (p. 95). Apart from the aforementioned fact that we do not claim that UG should account for the "problem of rare sounds", two further misinterpretations occur here. First, we do not claim that we need a "very large" feature set. We claim that due to combinatoric explosion, even a set of 25 features, as is usually proposed in the literature, yields 847 billion *possible* segments (assuming binary features and the possibility of underspecification); and any increase in the innate feature set leads to an exponential increase in the number of *possible* segments warranted by UG. Second, it is misleading to claim that this "even larger number of possible segments" is "innate and genetically determined". We do not claim that billions of segments are innate; we claim that what's innate are features (a few dozen of them at the most), and that the segments, i.e., the phoneme inventories of particular I-language (a phenotype), are determined through an interaction between UG (the genotype) and the PLD (the environment). This conception of the phonological UG seems in line with the research agenda of biolinguistic minimalism: how few features can be attributed to UG while still accounting for the variety of I-

29 The example may seem preposterous, but it is analogous to the spread of Spanish and English in the Americas from the late 15th century onward, a domination that made the properties of many indigenous American languages become "exceedingly rare."

phonologies attained? We think the answer cannot be 'zero', because it is a logical impossibility that the categories that humans use to induce phonemic inventories from environmental stimuli be learned themselves. According to emergent models of features, it is possible to learn language-specific features on the basis of "domain-general capacities such as categorization" (Chabot, p. 101). But what does a human categorize acoustic stimuli *into*? The sorting of experience into categories implies the prior existence of at least some categories.[30] So where do these initial categories come from? In the case of phonological features, they clearly do not result from the properties of auditory perception. In various experiments, the human auditory system was reported to be able to discriminate between at least 1300 levels on a single pitch scale (Fastl & Zwicker 2013). But this capacity seems to have nothing to do with phonological competence: we do not have 1300 different levels for vowel backness (which correlates with F2 fairly well). The resolution of our perceptual systems and the granularity of our phonological representations are of a completely different order – features can't be identified with the 'just noticeable differences' of auditory perception. For a complex sound such as a vowel, an emergent model that relies on auditory categorization predicts that there could be thousands of different categories/features, which is not a hypothesis that can seriously be entertained. Finally, note that the principle of the innateness of all *primitive* categories does not mean that *all* categories are innate and therefore that no categories are learned. It merely means that the atomic categories such as features are innate and unlearnable, while higher-level categories such as segments are constructed by combining the atomic ones.

After we deny the existence of language-specific phonetics in our paper, we say that what is commonly interpreted as language-specific phonetic phenomena should be re-

30 Imagine if we asked you to categorize all of the sounds that you hear in a day. Naturally, you would ask "how?" or "categorize the sounds into what?" Otherwise, either the process could not even start or each individual would end up with a completely different and incommensurable categorization, based on random factors such as moods and recent experiences, neither of which seems to be the case with phonological features. The bottom line is that without knowing what the categories are, one cannot categorize.

analyzed as either (a) language-specific *phonological* phenomena (for which we offered the Logical Phonology model); (b) non-phonological and therefore non-language-specific transduction of surface representations (for which we offered the Cognitive Phonetics model); or (c) systematic biomechanics of speech production and random situational factors which modulate the acoustic output from the body. Therefore, it is not the case that we think that "all variation can be explained by specific combinations of feature variables", as Chabot says on p. 95. We think that phonetic variation and cross-linguistic differences are best explained by carefully analyzing them along these three mutually exclusive dimensions.

Chabot expresses worry that eliminating language-specific phonetics from the account of the externalization of language "comes with a price", namely "a concurrent increase in the size and remit of UG, since every feature of every language must be present in UG." Also, since apparently no single I-language utilizes all of the features given by UG and certainly not all of the possible features combinations (i.e., out of millions or billions of possible segments, only a small subset of them is attested), Chabot asks "what leads UG to develop a universal system of features which leaves any given number of features fallow in the brains/minds of speakers whose language makes no use of those features?" Our position is that introducing a few more features in order for linguists to be able to re-analyze some alleged language-specific phonetic phenomena in line with option (a) above is a much smaller price (i.e., scientifically more parsimonious) than postulating an entire module of the mind, with all the necessary machinery, to account for those phenomena. As we said, the free combination of a relatively small set of features gives rise to a very large number of possible segments, which can elegantly account for a large part of what is commonly interpreted as language-specific phonetics. So the innateness of features and the banishment of language-specific phonetics not only follows from theoretical considerations (which we discussed at length in the paper), but is also warranted methodologically. Chabot's implied reasoning is a bit backwards, we think: UG does not "develop a universal system of features". Yes, lin-

guists use language data to *determine* (figure out, deduce) the features that are in UG; but it is the UG feature inventory that *determines* (causes) the (intensionally defined) set of possible languages. The first sense of 'determine' is epistemological, about how scientists come to know something, whereas the second is ontological, about what exists in the world; this parallels the dual meaning of the term 'UG' itself, both as an innate component of the language faculty and as a scientific theory of that component. Languages don't need to worry about not using all the phonological features that UG provides any more than French needs to worry about leaving the possibility of *do*-support "fallow".

Chabot points out that "not all gradient effects can be reduced to feature specification". We agree, although we think he is mixing up too many kinds of "gradience" including interspeaker variation in speech (assuming identical mental representations); context dependent realizations, such as coarticulation; cross-linguistic differences in realizations; and typical performance effects (presence of peanut butter and tongue muscle fatigue). We merely proposed, following *SPE*, as well as Hale et al. (2007), that surface differences may sometimes reflect a failure by linguists to recognize a difference in mental representations. If we are correct in doing so, then there is no gradience, just a failure by linguists to recognize a distinction. If we are *not* correct, and one favors a theory of features that 'emerge according to the needs of speakers', the question arises of why the +ROUND (or whatever) of Ludwig's Swahili grammar should be identified with the +ROUND of Ahmed's Hmong grammar. You can't talk about gradient output, across languages or within a language, for the realization of a category without having that category. As Hammarberg (1976) points out "phonology is logically and epistemologically prior to phonetics". Although this is not an argument against emergent features on its own, we should recognize that accepting emergent features entails a rejection of all innate OT constraints, typological markedness generalizations, etc., as we suggested in our response to Carvalho.

On the computational side, Logical Phonology (LP) provides *explicit* definitions of the notions 'phonological rule' and 'natural class' (see p. 41 for definitions). Chabot

doesn't see the benefit of that, arguing that "if /r, u, k, i/ participate in some phonological process, either as a united class of triggers or targets, then by definition /r, u, k, i/ constitute a natural class. (…) a natural class is defined by its participation in a unified rule" (p. 101). The benefit is that whereas Chabot uses intuition to determine what counts as a single phonological process, we offer an explicit theory that can be interpreted and evaluated unambiguously. In the absence of a formal theory of rules and natural classes, just about any descriptive statement can be regarded as a rule. In LP, a rule is a function mapping between phonological representations, whose target and trigger are necessarily stated in terms of natural classes. If the set of segments {r, u, k, i} cannot be characterized through a generalized intersection of features to the exclusion of all other segments that do not participate in the observed alternation, then LP predicts that the observed alternation does not result from a single rule. Clearly, having a precise, formalized theory of rules and natural classes is a step forward from determining them on the basis of intuition, no matter how much that intuition may be grounded in experience. Another related issue that Chabot raises is the following: if LP claims that rules must be stated in terms of natural classes, then how does an analyst working in LP determine the natural classes. He claims that "it is not by examining surface properties of segments that phonological features can be identified" (p. 102); rather, "it is only by phonological behavior that phonological identity can be discovered" (p. 99). In turn, our practice of determining features by observing substance (phonetic properties of speech sounds) amounts to the very substance-abuse that we are warning against. We agree that observing substance or measuring it instrumentally is not the only way of determining the features that comprise segments and define natural classes. Neither is such a method completely reliable and foolproof. Certainly, analysis of phonological patterning is also an important source of evidence. But because the correlates of features fall in predictable articulatory and acoustic spaces *most of the time*, as discussed above, it seems that observing substance is a useful, straightforward method of determining features and seg-

ments.[31] Using phonetic substance as evidence in reasoning about phonology is not substance-abuse. Substance-abuse is attributing phonetic substance to phonological competence. Until we can infer the properties of linguistic competence directly by observing neural activity, using performance as a window – albeit a highly distorted one – into competence is unavoidable. Therefore, while "phonological behavior" can shed a different light on the status of phonological units, one that the observation of substance does not reveal, we don't know what this "phonological behavior" is until we infer it on the basis of observing substance. In fancy terms, phonetic substance is part of the *epistemic toolkit* of the working phonologist, a source of evidence about phonological patterns, but it is not the object of study of phonology, strictly speaking (although it may be part of the object of study of phonetics). The phonetic substance is not part of the *ontology* of phonology, which includes things like features (and maybe syllables, moras and feet) and the components of rules or constraints (depending on one's preferred model). Many parallels come to mind, such as the use of grammaticality judgments in syntax, where the object of study is a system of syntactic representation and computation that syntacticians attempt to infer by using grammaticality judgments, eye-tracking devices, brain imaging technology, reaction time experiments, and whatever else promises to yield insight.

31 Suppose we approach a description of a completely unknown I-language. Surely, one cannot begin the construction of a generative grammar of that I-language by first observing the "phonological behavior" of the phonological units in the mind of a speaker. As is typical in linguistic fieldwork, one begins with eliciting and transcribing some simple utterances, revising the description as the analysis progresses. But the core, inescapable principle is that if we hear a speaker utter a voiceless bilabial stop, our default assumption is that it corresponds to the *surface phonological* feature matrix [p], until some further evidence (e.g., some interesting "phonological behavior") compels us to revise that initial assumption. In any case, the main *source of evidence* for linguistic competence is linguistic performance.

DISCUSSION WITH GUILLAUME ENGUEHARD

(UNIVERSITÉ D'ORLÉANS, CNRS/LLL)

Enguehard, Guillaume. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 113-125.

COMMENTS

The following comments reflect my understanding of the paper as a structuralist linguist. As I am neither concerned nor interested in cognitive issues, most of my remarks will be mainly focused on section 3. The proposal of the authors is that phonology should be conceived as a set of several disciplines dealing distinctly with the form and its realization. Although I support this approach, I am surprised that it is regularly presented as an innovation in recent literature. This idea has been present since the earliest days of European structuralism, notably within the Copenhagen School, and has only been faded by the growing influence of functionalism and American structuralism from which generativism originated. I regret that this older literature, although it addresses comparable ideas, is hardly mentioned in the current papers. For example, the quotation from Gallistel & King (2010) (p. 38) motivating the economy of atoms by cognitive considerations is a strong reminder of the way Martinet (1960:178-197) motivates the economy of language by taking into account its causal and teleological dimensions. This is a general observation that does not concern this paper in particular, but it is interesting to question the reasons for this selectivity.

Concerning the substance of the proposal, I will group my remarks according to two distinct themes: **i.** the notion of form in phonology, and **ii.** the logical modelling of transformation rules.

1.1. PHONOLOGY AND FORM. The purpose of phonology. In their conclusion (p. 60), the authors draw a connection between their proposal and Saussure's work concerning the dichotomy between "form" and "substance". However, in many respects this comparison appears to be limited. The "form" discussed in the article is clearly an aspect of language that belongs to the psyche and is therefore inherently individual. In Saussure's case, on the other hand, "form" or "langue" is a collective aspect that is opposed to its individual manifestation, "speech" (Saussure 1916:30). Furthermore, Saussure defines the language as a system of signs, whereas the formalism defended in this article is a system of derivation rules. Therefore, the study of form as mental computation that is proposed here is much more a study of Saussure's speech process than a study of the social fact that he calls "langue". This remark is echoed by the question raised at the end of the conclusion: « is phonology on the right track? ». The answer depends, of course, on what one considers to be the goal of phonology. The authors propose to define two disciplines: logical phonology, which studies the cognitive relationship between an input and an output (p. 37), and cognitive phonetics, which studies the realization of the output (section 4). But what about the discipline that studies the representation of the input? As it stands, the units of this representation seem to be admitted de facto but without knowing precisely by which method they are defined.

1.2. The difference between form and substance. The method by which the units of representation are défined depends on the definition of the concepts of form and substance. However, contrary to what is stated at the beginning of section 4: « From the preceding discussion it is apparent that phonology is vastly different from phonetics. » (p. 50), the use of these notions in the paper does not, in my opinion, allow for a clear distinction to be drawn between phonology and phonetics. First, the segments of the representation have inherent properties that are only defined by a vocabulary belonging to phonetic description. Secondly, no distinction is made between processes that affect only the realization of the segments and processes that also affect the functional rela-

tionships between the segments. Without this distinction, it seems that logical phonology manipulates the intrinsic properties of the segments and not their extrinsic properties. Whether or not these intrinsic properties are considered to be physical in nature, they form a substance (what an entity is in itself) and not a form (what an entity is in relation to others). The term "substance-free" thus seems to refer only to the fact that the syntax of rules is independent of the intrinsic properties of the segments they manipulate. But is there a difference of principle with SPE in this?

1.3. On what basis is the form defined? It seems (p. 52) that phonological features designate an invariant type of articulatory movements, whereas the smallest variations of these same movements belong to the sphere of realisation, strangely considered from the point of view of cognition alone, although the authors admit the intervention of non-cognitive factors (p. 52-53). It is obvious that not all aspects of phonetic realisation can attain the status of a phonological feature, but the question is how to determine that a class of articulatory movements forms an invariant? There are several possibilities, some proposed by European structuralism (commutation), others by functionalism (relevant features[32]), and still others by American structuralism (distribution). The approach used in this paper to define units of form seems to be neither that of European structuralism nor that of functionalism. An absolute relationship is assumed by the authors between features and their realization: « If such Saussurean arbitrariness were applicable to the realization of features, then it should be possible that +ROUND in one language gets realized as a lowering of the velum, and in another as a raising of the tongue dorsum, and so on. » (p. 25). Same thing concerning the order of the segments and the linearity that realizes it (p. 41). This "phonetic realism" is not universally accepted, and the authors' statement is reminiscent of a debate between Hjelmslev (1936, 1943) and Marti-

32 A distinction must be made between the notion of a relevant feature and commutation. In the Copenhagen School's structuralist approach, two sounds that cannot commutate are the realizations of one and the same entity even if they do not share any features. In the functionalist approach, it is necessary that at least one feature forms a common basis for both realizations (Troubetzkoy 1939:52).

net (1946:37). In sum, the definition of units of form seems to be based, in this paper, on a description of their realisation, not a description of their function: /u/ (p. 40) contains a +ROUND feature although this is not distinctive in a /iau/ system. The choice of a descriptive vocabulary creates a bias that is not questioned by the authors: why choose the HIGH, ROUND and BACK features? For example, the authors admit that /i/ and /u/ cannot be part of a natural class that excludes /y/ because their common features belong to /y/ as well. This phonetically motivated conclusion is relative to the vocabulary used to describe the matrix of each of these segments: one could imagine an alternative where /i/ and /u/ have a +EXTREME-F2-VALUE feature absent from /y/.

1.4. The place of phonological opposition? The preference for phonetic realism rather than phonemic opposition also appears in the treatment of rules, since the authors address the formalization of the transformation from a segment A into a segment B regardless of the opposition relationship the two may hold. For example, in the absence of any indication to the contrary, the proposed rule (p. 38) t > s /_i seems to have the same formalisation whether it leads to contextual variation or neutralisation. However, this phonetic realism is relative, because, when dealing with the representation of [t,d] in Turkish (p. 45), the authors treat as phonologically distinct units segments that have the same phonetic realisation but are not subject to the same transformation rules. This formalism does not correspond to the notion of archiphoneme of European structuralism, which is fundamentally phonological, but to that of morphoneme, whose phonological status is a little more debatable, because it is treated sometimes by phonology and sometimes by morphology.

1.5. How do we access the units of the form? This is probably the main weakness of the paper: how are the units of a phonological representation defined? The answer to this question is necessary to understand why the number and quality of the HIGH, BACK and ROUND features are admitted rather than another set of features, why oppo-

sition seems to have no role in the definition of these features, and why the morpheme is considered as a phonological entity and not the result of a set of morphological processes. As it stands, all of these aspects of the paper seem to be implicitly part of the North American tradition and would merit a somewhat more explicit exposition. In sum, this proposal seems to lack the foundation of the discipline that brings together the methodological elements that allow us to define and describe the units of the representation. What the early generativists disdainfully called "taxonomy".

2.1. LOGIC AND RULES. Definitions. In their presentation, the authors undertake to formalize some definitions commonly used in phonology. However, there are a few points that I think need to be clarified. The first example concerns the definition of the natural class (p. 50) in its intentional form using the proposal {x: x ⊇{+HIGH}}. It seems to me that this definition leads to the conclusion that a phoneme alone can form a natural class (e.g. /i/ = {x: x ⊇{+HIGH, -ROUND, -BACK}}). Was it the intention of the authors to broaden the notion of natural class to include the notion of phoneme? If not, the broad inclusion symbol should probably be replaced by the strict inclusion symbol. More importantly, the authors use symbols from the phonological tradition with symbols from the mathematical tradition, which creates confusion. For example, the symbols "+" and "-" are defined (p. 40) as elements of pairs. However, it is clear that +F and -F have values that depend on each other and that the symbols "+" and "-" therefore belong to the domain of functions, perhaps double negation and negation? A clarification would be welcome, especially since the definition of the priority union is closely linked to the definition of these symbols: it is said (p. 44) that only those elements of B whose "opposites" do not belong to A are united to the elements of A.

2.2. The need for a priority union. In addition, I can see three reasons to doubt the usefulness of the priority union. The first is epistemological, the second is empirical, and the third is theoretical. I understand that the authors blame the arrow of the SPE for-

malism for two things when they talk about "deconstructing" it: **i.** its informality, and **ii.** its empirical inadequacy. First, I do not believe that the arrow is an informal symbol. It represents the logical implication, because A → B can be read as "A becomes B" as well as "if A (is present in phonological form), then B (is realized in phonetic form)". However, I agree that the representation of the rule in the SPE framework lacks formalism since +F → -F is only true if +F is false. So it should probably be noted /+F/→ [-F] or f(+F) → g(-F). Secondly, I can see the interest of the priority union in the example from Turkish exposed page 46, but this example seems to me too abstract to be perfectly convincing, as the authors admit in a footnote. The problem comes from the morphonological and very theoretical character of /D/, from the somewhat too hasty definition of the symbols "+" and "-", and from the problematic difference between "+", "-" and "Ø" if one applies the principle of the excluded third party. Thirdly, the authors justify the introduction of the priority union by observing that the union of {+F} and {-F} results in an inconsistent set. But what prevents rules from containing such a set as a target of the transformation rule (e.g. {+F, -F} ∪→ {+G})? The impossibility of these sets goes beyond the simple context of the union operation: it concerns the whole syntax of the rules.

2.3. Some ambiguities. Finally, there are still some ambiguities regarding the formalisation of the rules. In their respective applications, the relation ∪→ implements the appearance of a new feature while the relation ∪← implements and the appearance of a new feature and/or the change of one feature by another. For this reason, there are cases where these two functions are freely interchangeable (e.g. 5a, p. 44). Isn't it a problem to have several possible formulations for the same change? Moreover, this framework does not seem to be able to implement the loss of a feature. This implies that the representation of a falling segment must necessarily involve the adoption of a binary feature system in order to represent Ø as a matrix of features [-VOC, -CON, ...]. In other words, this formalism depends on a particular theory of representation whose descriptive me-

thodology is not explained in the article. The problem is even more salient when we approach the examples of free variation realizing the result of a neutralization (e.g. the free use in French of [e] or [ɛ] to realize /E/ in open non-final syllable). In this case, it is impossible to say whether the output of the derivation contains a +ATR or -ATR feature. It would therefore be necessary to admit a rule allowing to suspend the presence of one or the other of these two features. The proposed formalism does not allow such a neutralization to be derived by a loss of the relevant feature but only by a change of the feature. What then is the place of variation? Concerning now the formalization of the transformation, it does not seem to me to be totally explicit. The arguments of the SPE rule are asymmetrical because the first represents the target of the change and the second the result of the change. In the framework of logical phonology, the arguments of the transformation rule are doubly asymmetric since not only is the priority union not commutative, but the first argument continues to designate the only target of the change regardless of its role in the union relation. Indeed, if the proposition $A \cup\rightarrow B$ was a simple union (even a priority union), it would affect both A and B within the lexical items of the language. Therefore, there is a function symbolized by the precedence relation between the arguments : $A \cup\rightarrow B$ does not have the same value as $B \cup\leftarrow A$. This invisible function does not exist in the SPE framework, since $A \rightarrow B$ and $B \leftarrow A$ have the same value there, or rather, this function is symbolized by the arrow. In other words, I don't think we can say that the union really "deconstructs" the arrow: its symbol replaces the symbol of the transformation, but the transformation is still implicitly present within the rule. A more explicit formalism would probably be to represent rule $A \cup\rightarrow B$ as follows: $A \rightarrow (A \cup\rightarrow B)$ or $(A \cup\rightarrow B) \leftarrow A$.

Hjelmslev, Louis. 1936. « On the Principles of Phonematics ». In *Essais Linguistique II*, [1973], 157-62. Copenhagen: Nordisk Sprog- og Kulturforlag. **Hjelmslev**, Louis. 1943. *Prolegomena to a Theory of Language*. [1969]. Madison: University of Wisconsin Press. **Martinet**, André. 1946. « Au sujet des fondements de la théorie linguistique de Louis Hjelmslev ». *Bulletin de la Société de Linguistique de Paris* 42: 19-43. **Martinet**, André. 1960. *Eléments de linguistique générale*. [1980]. Paris: Armand Colin. **Saussure**, Ferdinand de. 1916. *Cours de lin-*

guistique générale. 5th edition. Paris: Editions Payot & Rivages. **Troubetzkoy**, Nikolaj Sergeevich. 1939. *Principes de phonologie*. Traduit par J Cantineau. [1949]. Paris: C. Klincksieck.

REPLY

The main goal of our paper is to provide "a general program for the study of phonology as a branch of cognitive science" (p. 59) since we consider the object of phonological study to be a particular module of the human mind. In his own words, Enguehard is "neither concerned nor interested in cognitive issues". Therefore, there is little to discuss as we are not talking about the same subject even though we unfortunately sometimes use the same terminology. Because there isn't a coherent definition of language that does not reference human cognitive faculties, we embrace the mentalist perspective on language, with all of its philosophical, theoretical and methodological consequences.

So, for us, phonology is a system of rules and representations that is fully contained in the mind of an individual. In his commentary on our paper, Enguehard keeps trying to interpret our claims in the context of structuralism, where language is seen as a non-mental entity. Naturally, this leads to severe misinterpretations of our claims, arguments and examples, since for the most part the two perspectives are incommensurable. On the other hand, in the few cases where we tried to point out the *similarities* between structuralist and generative phonology, Enguehard decides to focus on the *differences* in an attempt to show that our "comparison appears to be limited".

Enguehard is of course correct in stating that the terms 'form' and 'substance' do not have identical meanings in structuralist and generative linguistics. At the end of the paper, we said that even vastly different linguistic frameworks such as structuralism and generativism agree on an important idea, namely that the distinction between language and speech must be maintained in both theory and practice. This is true despite the obvious fact that the two traditions define 'language' (and also 'form') differently. We merely argue that the 'language vs. speech' and the 'form vs. substance' distinctions can also be exploited productively in a mentalistic phonology after they've been appropri-

ately adapted. After pointing out that our notion of form is mental and individual while the structuralist notion of form is non-mental and collective, Enguehard tries to situate the generative notion of form into a structuralist notion of language, showing that the two are incompatible. Indeed they are, but that is not relevant for the point that we are making. It would be a futile exercise, for example, to demonstrate that the notion of universal grammar from the 17th century French grammarians cannot be used in the current biolinguistic framework without substantial adaptation. But the mere fact that substantially dissimilar frameworks converge on a particular point, that they share certain ideas, suggests that it might be worth our while to try to extract lessons from those ideas, which is what we tried to do.

According to Enguehard, "the main weakness of the article" is its failure to provide a satisfactory answer to the question "how are the units of a phonological representation defined?" Our "proposal seems to lack the foundation of the discipline that brings together the methodological elements that allow us to define and describe the units of the representation". This is a puzzling assessment due to how explicit and detailed our discussion of features is. We define features as a particular kind of neural symbols (p. 21); we provide both a general theoretical foundation for interpreting that definition (§1 and pp. 21–22) and we ground it in experimental work (pp. 23, 55–56); we discuss not only the properties of features *qua* neural symbols such as combinability and efficaciousness but also the implications of those properties for phonology and phonetics (p. 22); we discuss the relationship between features and their correlates (pp. 24–26); we explain why we *call* features by referring to their typical phonetic correlates even though their ontology is non-phonetic (pp. 23–24); we explore the issue of the innateness of features, relating it to work in philosophy, cognitive science and phonology (pp. 27–35); we discuss the crucial role that features have in phonological computation (§3); we provide a theory of their phonetic interpretation (§4), and so on. At the very least, our proposals regarding the nature of features are theoretically coherent and explicit enough to be evaluated empirically.

Despite our fundamental disagreements with Enguehard, we are grateful for his close reading. The technical issues are complex, and of course, still being worked out, so we are happy to attempt further clarification here on several questions raised in Section 2 of his review, "Logic and Rules".

Enguehard: *It seems to me that this definition leads to the conclusion that a phoneme alone can form a natural class (e.g. /i/ = {x: x {+HIGH, -ROUND, -BACK}}). Was it the intention of the authors to broaden the notion of natural class to include the notion of phoneme?*

It is true that we specifically want a natural class to be able to **contain** just one member. There are rules that apply to only one segment, so why not? Also, given that we define a natural class as a set of sets of valued features, it is consistent that a natural class can contain a single member—a set with one member is still a set. We would not want a rule notation that differs just when the target set has one member.

Because regular set intersection is a binary operation, requiring two arguments, we specifically make use of *generalized* intersection which is a unary operation whose argument is a set of sets. Since a segment like *i* is a set (of valued features), the generalized intersection whose argument is {i} is the natural class that contains just the segment *i*. See Bale & Reiss 2018, Unit 45 for discussion.

Note that in general the set $\{a\}$ is a set, and not the same thing as the element a. In other words $\{a\} \neq a$. So we have **not** broadened the notion of natural class to include the notion of segment/phoneme. We have provided a definition of natural class that can contain any number of elements.

Enguehard: *The authors use symbols from the phonological tradition with symbols from the mathematical tradition, which creates confusion*.

The question about + and - is valid. We could not go into all details of so-called Logical Phonology here, but, yes, we are assuming that the symbols combine like multipli-

cation by 1 and -1, so that – – = +. Enguehard is correct that such things need to be made explicit. See Unit 51.3, page 452 of Bale & Reiss 2018 for more discussion.

Enguehard: *I can see three reasons to doubt the usefulness of the priority union.*

On the "need for priority union", there are a few things to say. First, as indicated, we have used another system involving unification and set subtraction elsewhere. Our point here is not that priority union is the "right" version of what UG gives, but rather that looking at the results one gets from adopting *some* explicit formal system is what Substance Free Logical Phonology is about. Second, our assumption that segments (and the characterization of natural classes) must be consistent, and not allowing expressions like {+F, -F} is just that, an assumption. We need to adopt some system and see what it gets us. Enguehard should definitely pursue his system. The point is to be explicit and test some system. Our idea is captured in this quotation from the Preface of Chomsky's (1957) *Syntactic Structures*:

> By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. (Chomsky, 1957: preface of *Syntactic Structures*)

Enguehard: *Isn't it a problem to have several possible formulations for the same change?*

For us, a rule *is* a formulation using an explicit characterization of a formal system, not an informal statement of a posited process. Using priority union we can get feature

changing voicing assimilation using one rule; and using unification and subtraction, we need two rules. These two different formulations of what we call voicing assimilation are embedded in incommensurable models of UG. See the next point for further discussion.

Enguehard: *Moreover, this framework does not seem to be able to implement the loss of a feature.*

Exactly correct. This is one reason why we think priority union might be wrong. If *derived surface underspecification* exists then things look bad for priority union and good for the unification and subtraction model. There is a discussion of the issue in Bale, Papillon and Reiss (2014), without specific reference to priority union, and then a further note in Reiss' forthcoming *LI* paper on priority union. Enguehard's reasoning is *exactly* what we think phonologists should be doing—thinking about the implications of different formal systems.

Enguehard: *This implies that the representation of a falling segment must necessarily involve the adoption of a binary feature system in order to represent Ø as a matrix of features [-VOC, -CON, ...]. In other words, this formalism depends on a particular theory of representation whose descriptive methodology is not explained in the article. The problem is even more salient when we approach the examples of free variation realizing the result of a neutralization (e.g. the free use in French of [e] or [ɛ] to realize /E/ in open non-final syllable).*

More good questions here. We assume that "falling" segment refers to **deletion of a segment**. The discussion of operations here is only relevant to operations *inside* of segments. Metathesis, deletion and insertion of segments have not been addressed. So, segment deletion is **not** accomplished by set subtraction (or priority union). We distinguish an empty set (a segment with no features) from the absence of a segment (nothing). The empty set is not nothing—it is a set; just as an empty box is not nothing, it is a box. The

most recent discussion of these issues is a forthcoming paper in *Loquens* by Bale, Reiss and Shen called 'Sets, Rules and Natural Classes: {} vs. [ ]'.

Enguehard: *Concerning now the formalization of the transformation, it does not seem to me to be totally explicit. The arguments of the SPE rule are asymmetrical because the first represents the target of the change and the second the result of the change. In the framework of logical phonology, the arguments of the transformation rule are doubly asymmetric since not only is the priority union not commutative, but the first argument continues to designate the only target of the change regardless of its role in the union relation.*

Following Bale *et al.* (2014), we use the unification symbol in phonological rules to represent the function that maps segments that fit the structural description to the same segment unified with the structural change. In other words, we are not using it to represent the unification of the structural description with the structural change. The same logic holds for priority union. We agree with Enguehard that more work needs to be done to further formalize these operations, and Reiss (2021) is a step in that direction.

DISCUSSION WITH GILLIAN GALLAGHER

(NEW YORK UNIVERSITY)

Gallagher, Gillian. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 126-132.

COMMENTS

Overview. In 'Formal Generative Phonology', Volenec and Reiss expound on phonology as a cognitive science, with *I*-language as the object of study. The majority of the monograph is a broad ideological outline, with brief examples of concrete analyses towards the end. Volenec and Reiss view much current research in phonology as focusing on phonological performance, *E*-language, as opposed to phonological competence, *I*-language, though they acknowledge that the only access we as researchers have to *I*-language is through *E*-language.

In this short response, I first make some direct comments about the manuscript and then contextualize a piece of my own work as consistent with the study of phonology as a cognitive science, where understanding a mentalist, symbolic system is the ultimate goal. I feel this is relevant, because my work mostly falls under the umbrella of "laboratory phonology" and I believe would be characterized by Volenec and Reiss as collecting data on phonological performance.

Remarks on the manuscript. I disagree with Volenec and Reiss that there is extreme fragmentation and misunderstanding within the field of phonology about the object of study. I do believe that different researchers model different types of data, and that researchers differ as to whether they are principally concerned with primary data or with theory development. I view the changes in phonological research methods over the last 20 years or so to be centered around increased interrogation of what the data is that a

phonological grammar (a mentalist, symbolic grammar in almost all cases) must account for. Specifically, much research has moved from using descriptive grammars as sources of data to using behavioral research with native speakers. The goal of this change is to increase the cognitive plausibility of our understanding of the phonological grammar. Similarly, research into language corpora and computational modeling are thoughtfully and carefully exploring what information is available to a learner and how that squares with a mentalist grammar.

While there are certainly researchers who question the validity of a symbolic grammar or explicitly argue against it, I still find their research findings to be useful and relevant in the pursuit of a model of cognition. By investigating what aspects of language data can be modeled *without* reference to a symbolic grammar, it becomes clear what a symbolic grammar must be able to do. From another angle, detailed investigations of *E*-language are imperative for understanding what the symbols are in a putative symbolic grammar.

Turning to the manuscript itself, the exposition of the Chomskyian grammatical enterprise is clearly and thoroughly stated, and is useful reading for anyone wanting an ideological refresher, or for a beginning student in linguistics. I found the inclusion of quotes from other works largely unhelpful, since these seemed designed to foster argument. Instead of selecting broad, ideological claims to criticize, I would have found it more helpful to see a critique of the actual content – that is, the exposition of data, the argumentation and methods used for building an analysis, the rationale for conducting an experiment and interpreting the result – in a range of published work in phonology.

In a similar vein, I wish more of this manuscript had been devoted to case studies and examples of analyses that Volenec and Reiss would endorse. The few pages of the analysis at the end of the manuscript left me full of questions, but without enough concrete proposals to engage with. For example, Inkelas's (1995) proposal that Turkish has [-voice], [+voice] and unspecified stops is taken at face value and presented without comment. What does this mean for a theory of representations? How do Turkish speakers

learn to set-up an unspecified class of stops? How does an unspecified set of stops emerge if the mapping between phonological features and phonetic output is deterministic and universal, as Volenec and Reiss claim?

A laboratory research question. In a current project (partially published as Gallagher 2019), I am interested in how speakers of South Bolivian Quechua (SBQ) represent the putative phoneme /q/, a voiceless uvular stop, and its distribution. The main question is: how is this sound category represented in terms of phonological features? This is a question about mentalist symbols, and *I*-language.

SBQ is described by linguists as having a voiceless unaspirated uvular stop /q/, but this segment is produced with substantial phonetic variation. The descriptive tradition is based on historical and comparative evidence that leads a linguist to postulate /q/ as the underlying phoneme. On the surface, however, this segment is produced sometimes as [q], but most often as [ʁ] and quite frequently as [χ], among other realizations. In terms of phonological distribution, the sound category patterns as a stop with regards to cooccurrence restrictions (and syllable structure, which I leave aside here). Like other stops, putative /q/ cannot be followed later in the word by an ejective or aspirated stop, as shown in (1).

(1) a. Ejective and aspirated stops may be word medial with an initial fricative or sonorant

ɾit’i	‘snow’	**m**usq^{h}uj	‘to dream’
satʃ’a	‘tree’	**ʎ**imp^{h}i	‘color’

b. Ejective and aspirated stops may not be word medial with an initial stop

*tant’a	*pusq^{h}uj
*katʃ’a	*tʃimp^{h}i

c. Ejective and aspirated stops may not be word medial with an initial putative /q/

*qap’a	*qat^{h}a

A standard phonological analysis of this pattern would work as follows. First, we would presume that SBQ speakers map all surface realizations to the single underlying phoneme /q/, which they represent with the features [-continuant, -sonorant]. The class of [-continuant, -sonorant] sounds in SBQ is then all of the voiceless unaspirated, aspirated and ejective stops /p t tʃ k q p^h t^h tʃ^h k^h q^h p’ t’ tʃ’ k’ q’/. Given this representation, the grammar of Quechua can capture the distributional restriction by referencing [-continuant, -sonorant] sounds. The analysis is shown as a pair of constraints in (2a), which could be embedded in your favorite constraint-based grammar, or considered as morpheme structure constraints holding over the lexicon.[33] Alternatively, these restrictions could be encoded as a rule, as in (2b).

(2) a. Constraints

*[-continuant, -sonorant]...[+constricted glottis]

*[-continuant, -sonorant]...[+spread glottis]

b. Rule

[-continuant, -sonorant] → [-c.g., -s.g.] / [-continuant, -sonorant] V_0C_0___

33 The ellipsis in (2b) represents that this is a non-local constraint, and is used only as a shorthand.

My interest lies in whether this analysis is correct, in the sense of being informative of mental grammars. Do speakers really represent the highly variable uvular category as a single category, and as a stop? Do Quechua speakers have a [-sonorant, -continuant] class of sounds that includes a segment that is decidedly **not** obstruent and non-continuant in production?

To answer this question, I first conduct behavioral experiments with Quechua speakers to see how they treat /q/ in nonce words. In one task, native speakers of Quechua were asked to listen to and repeat nonce words that included either a [k]-ejective combination (e.g., *[kap'i]) or a [ʁ]-ejective combination (e.g., *[ʁap'i]). The question was whether speakers made more errors on repeating forms with these absent segment combinations than on controls, and they do. Accuracy on control forms is near 100%, while accuracy on [ʁ]-ejective forms is around 60% and [k]-ejective forms are repeated accurately only 40% of the time.

This and other results support a systematic, grammatical restriction on [ʁ]-ejective combinations, despite the fact that [ʁ] is not produced as a surface stop. I believe this step is imperative for a working phonologist, as there is little point in constructing elaborate symbolic models and computational architecture if we have not first confirmed that an impressionistic or statistical description of a language pattern has some correlate for native speakers.

Once the synchronic validity of this pattern is supported, two avenues of research emerge. First, if [ʁ] is mapped to /q/ and included in the [-continuant, -sonorant] class of sounds, how is this mapping learned? Second, if speakers' grammars instead include an ad-hoc constraint on just [ʁ]-ejective combinations, how do speakers learn such a constraint without learning constraints on accidental gaps in the lexicon?

Volenec and Reiss claim that the mapping from surface structures to features is deterministic and universal, and that there are as many unique feature combinations as there are surface productions. Under this view, an analysis with a single /q/ in the lexicon is

impossible. Instead, a Quechua speaker would have numerous (3? 5? 25?) feature bundles corresponding to surface productions of /q/, and would need to learn the distribution of each of these independently. I believe implemented computational simulations over large, realistic data sets are necessary to determine if this type of learning outcome is possible, and under what conditions, and such simulations constitute one branch of this research project.

This type of work has the following general structure. First, a pattern is identified based on transcribed, idealized data (e.g., a dictionary of "words"). Second, the validity of the pattern as a systematic pattern *for native speakers* is assessed via behavioral tasks. Hopefully, several tasks and methods are used, so that task effects and performance factors can be reasonably controlled for and thoughtfully considered. Once there is support for a pattern being grammaticalized by speakers, the question pivots to how a learner would acquire the symbolic representations necessary based on the data available to them. By taking inductive learning models seriously (i.e. by exploring what information is available to the learner in the signal to which they are exposed), we can discover what cannot be learned and must be attributable to an innate language acquisition mechanism. Thus, I see behavioral studies, descriptive phonetic studies, corpus construction and analysis, and implemented computational models as central to the pursuit of understanding *I*-language.

Gallagher, Gillian. 2019. Phonotactic knowledge and phonetically unnatural classes. *Phonology* 36(1): 37-60. **Inkelas**, Sharon. 1995. The consequences of optimization for underspecification. In *Proceedings of the North East Linguistic Society*, 25: 287-302. University of Pennsylvania: Graduate Linguistics Student Association.

REPLY

We don't have much to argue with in Gillian Gallagher's comments which conclude with the following statement: "Thus, I see behavioral studies, descriptive phonetic stud-

ies, corpus construction and analysis, and implemented computational models as central to the pursuit of understanding I-language." Like us, she is interested in mentalist, symbolic, I-language phonology, including questions of acquisition and learnability. However, she makes a common error that seems to lead her to believe that she disagrees with us more than we think she actually does. This error is the conflation of the notions linguistic *performance* and *E-language*. Taking Chomsky's (1965) *Aspects of a Theory of Syntax* and (1986) *Knowledge of Language* as the source for the terms, *performance* is used to identify aspects of speech events, utterances, that are not to be explained only in terms of *competence*, mental grammar. In Chomsky's words, *performance* is "the actual use of language in concrete situations" (1965: 4). Although it is common to make this mistake, *E-language* does not refer, in Chomsky's usage, to externalized *tokens* of language behavior, utterances, performance events. The term *E-language* is used to refer to a conception of a language as an external(ized) entity, outside of individual minds, for example "a language as a pairing of sentences and meanings" (1986: 19). Chomsky is very clear about this – "the construct is understood independently of the properties of the mind/brain" – as well as about the fact that *E-language* is not a coherent notion and is thus useless for scientific purposes. In Chapter 2 of *Knowledge of Language*, he refers to the notion of *E-language* with terms like "obscure", "mere artifact" and "playing no role in the theory of language."

In sum, we never use the term *E-language* at all, and *performance* is quite different. Thus, where Gallagher says that we "acknowledge that the only access we as researchers have to I-language is through E-language", what she is probably referring to is this sentence in our paper: "the main source of evidence for I-language is linguistic performance or the use of I-language (e.g., in speaking or understanding speech)". Gallagher's statement is thus doubly misleading: *E-language* plays no role in our paper; and we are open to (without discussing in the paper) other sources of evidence, such as brain-imaging studies, knowledge of general properties of biological systems, etc.

DISCUSSION WITH JOHN GOLDSMITH

(UNIVERSITY OF CHICAGO)

Goldsmith, John. 2020. discussion in: Volonec, Veno & Reiss, Charles (auth.) "Formal Generative Phonology". *Radical: A Journal of Phonology*, 2, 133-148.

COMMENTS

Introduction. Volenec and Reiss (henceforth V and R) are to be applauded for their decision to raise deep and broad questions about the nature of phonology. But I find myself surprised again and again as I read this manuscript how *conservative* the position is that they offer us, and how little they ask us to rethink decisions that do deserve to be questioned.

There's nothing *wrong* with being conservative, of course, especially when things are going well. Why rock the boat if all is going swimmingly?

V and R are perfectly straightforward about their belief that all is well in the house of cognitive science, which was, they suggest, the fruit of a paradigm shift of the late 1950s, which they describe as mentalism (I would prefer to say as "a style of mentalism"), the only uncertainty arising out of the fact that not everyone understands that paradigm as they do.

This seems very parochial to me. I think there are many challenges to the picture of cognitive science that developed in the 1960s. Weinreich, Labov, and Herzog 1968 was a powerful critique of any view of linguistics that prioritizes the individual to the degree that it effaces the importance of the social; it was thus a critique of cognitive approaches. Weinreich and his at-the-time students described the individualist approaches as following an intellectual arc spanning Hermann Paul, de Saussure, Bloomfield, and Chomsky (and now V and R). Contemporary machine learning, beginning in the mid-1980s, has completely changed our understanding of the ways in which well-understood

algorithms learn from data, in ways that were unimaginable in the first generation of cognitive science. Neural nets, in their several manifestations, have changed our understanding of the differences between serial and deeply parallel computation.

Many linguists have followed the call of Weinreich, Labov, and Herzog;[34] some linguists have listened to the opportunities open up by machine learning. I'll suggest below why I think V and R need to delve more deeply into machine learning if they want to defend positions that they sketch here.[35]

Philosophy: the good and the bad. The first strong indication of fusty conservatism (if I may describe it thus) is in an early footnote, in which V and R cite (and it certainly *seems* be approvingly) Piatelli-Palmarini's 2018 complaint about the excitement afoot regarding Bayesian learning models and other things that involve statistics, all of which is described as *neo-empiricism*. I've written quite a bit about why contemporary empiricism has a great deal to offer anyone who wants to think about the nature of language and language learning (see Goldsmith 2015 and Chater *et al* 2015, especially chapters 1 and 2). I'll offer some simple characterizations to specify what is involved: rationalism has always prioritized views of knowledge in which logical inference plays a key role, while empiricism prioritizes the discovery of patterns, and the ability of the mind to discovery these patterns. Both empiricism and rationalism put a high priority on simple hypotheses and close fit between theory and observation. The style of empiricism which has been influenced by modern machine learning has an additional characteristic that is of considerable interest: it puts a high value on providing a quantitative answer to the question: how great is the discrepancy between theory's predictions and current observation? This question, indeed, is the one that far more than anything else separates the

34 V and R allude to this when they write that language is "fully internal to a speaker and not something 'out in the world belonging to a 'speech community' in some vague sense.'' p. 7. Their focus on the individual is matched by their dismissiveness with regard to those who work on the social character of language. It is disappointing that they can dismiss the work of a large proportion of card-carrying linguists without even naming it or providing a single reason, let alone an argument, for doing so.

35 Some of the topics I mention are discussed in greater length in Goldsmith and Laks 2019.

work of mainstream LSA-oriented linguists and computationally-oriented ACL-oriented linguists. One cannot publish in computational linguistics without an explicit and quantitative description of exactly how great the distance is between one's predictions and one's observations. *That*, more than anything else, is the modern neo-empiricism that linguists should think seriously about.

Mind/brain. V and R write as if they have the answer to complex philosophical questions, writing that "the mind is just the functioning of the brain viewed from a sufficiently abstract point of view," and this insight allows us to conclude that "language can ultimately be regarded as a biological object—a subsystem of the brain."(7) This seems to me to be a naive point of view, and I don't understand how thoughtful people can take it seriously. The deepest questions of the relation of the brain and the mind remain mysteries for us, of which the most important for us is this: how do we understand the relationship between *causal* relations in the brain and *knowledge-based* relations in the mind, where by "knowledge-based relations" I mean both relations of deductive and inductive inference. That is the central problem of mind: when we observe the mind, we find objects and relations of sorts that go beyond what we find in the non-human part of the universe. Behaviorists like Quine rejected the observation; Quine himself saw nothing but events in the spatio-temporal world that could be understood in terms of causation. But the rest of us can't. Those who (in my view, naively) think that the brain and the mind form a comfortable bond need to understand that they are standing on the edge of a Quinean behaviorism.

V and R suggest that skeptics of mentalistic linguistics place expectations on linguistics that would never be taken seriously if they were placed on our colleagues in the physical sciences. (p. 17) They cite Pierre Jacob in defense of the view that scientists don't need to characterize what phenomena rightfully fall into their discipline, because Jacob says real scientists don't need to worry about such demands. I'm particularly unhappy whenever I encounter this defense. V and R seem to think that there really are no

serious questions that need to be answered about treating 'mental aspects of the world' in a fashion that is just like the way chemists treat chemical aspects of the world; that's one point. And they seem to think that linguists shouldn't be expected to have to answer questions about the object of their science if physicists and chemists aren't expected to. That's the other. Well. The physical sciences *did* need to answer the skeptics, and it caused great upheaval in centuries gone by. In the 19th century, there was tremendous dispute about the existence of atoms and the nature of heat, and serious questions needed to be answered with experimental results. I've seen experimental and theoretical physicists square off today over whether M-theory deserves to be treated as a serious theory of physics for similar reasons.

More importantly, though, there *is* real work that needs to be done to try to answer questions like how is it possible to develop a computational system that can be implemented in a neural structure, a computational system that allows one to compute the results of autosegmental spreading and OT tableau ranking. There is, we must remember, no paper and no whiteboard in the brain. I've tried to develop neurally implementable models of accent and sonority,[36] but I have been stumped when it comes to autosegmental representations. When I read V and R, they don't seem to be aware that there are real, deep problems here that need to be addressed; I could be wrong, but they seem to think that as long as phonologists have models that they can write down on a piece of paper, then it's a scientific model and that's good enough. You don't ask the physicist to look for her equation under a microscope or in the heavens, right? So why should a linguist have to worry about how her model is going to make the transition from paper to neurons?

We *do* have to worry about those transitions, just like other scientists do. Knowing the laws of thermodynamics in the 19th century was not enough; it was necessary to establish the laws as emerging from the behavior of atoms and molecules which were subject to measurable forces.

36 See Goldsmith 1994, and other work on dynamic computational networks.

But I would be misrepresenting my position if I left it there. It's not that I think that phonologists need to worry *now* about how to implement our current phonological theories in the brain; I actually don't. I think that any achievements we make on this front would be great, but we actually make lots of progress even while we forget about making a link to brain functions. People who study languages and grammars learn new things all the time about phonology.

It's innate? V and R write, at one point (p. 18),

> …there is the logical argument. Suppose that learning a given language involved learning that noun phrases like *dogs* come after adpositions like *with*. Such a rule can only be learned if the learner has access to the categories 'noun phrase' and 'adposition'

This seems to me quite backwards. We do indeed need to understand how words are learned, how categories are established in language-particular and language-general ways, and what sort of things we have ('rules') that expresses their distribution. But if V and R mean (as they seem to mean) that this problem is solvable only if there is an innate *and cross-linguistically identifiable* category that we can identify with what we call "noun phrase" in English linguistics, then I think they are wrong, and its probably because they are not paying attention to machine learning.

Knowledge of a system and its use in context. V and R suggest that it is of paramount importance to distinguish between our knowledge of language and how we might use that knowledge in particular contexts. This is a position that has developed out of important antecedents, though they seem to think that it is based in Chomsky's distinction between competence and performance. In the 19th century German context, the charge of *psychologism* was made against those who failed to see the difference, those who,

like John Stuart Mill, saw our knowledge as characterizable in psychologists' terms. The alternative to psychologism is a view of knowledge in which what is *known* transcends the individual (I can, after all, say that you know everything I know about number theory or how to cook frogs' legs), while the individual puts that knowledge to use in particular ways they choose themselves. Philosophers worry about this: how can your knowledge and my knowledge each be internal (or internal states) and yet we can say that they are the same? How can some state inside me come into any sort of interesting logical relationship with something inside you? The answer is that knowledge is not internal; if an internal state could not be linked in some fashion to something that is *not* internal, it would not deserve to be called knowledge. It might be a headache or hypertension or arhythmia, but it is not knowledge. We know that *my headache* is not the same thing as *your headache*, but my belief about syllabification may be the same as your belief about syllabification. There's a difference there that even common sense recognizes. Statements about knowledge cannot be described in strict internalist language.

There are many serious difficulties that arise when we try to say that our knowledge is internal, and not social or public. In the English-speaking world, the best known critique of this view is the work of the later Wittgenstein. If I were to summarize it all in one sentence, I would put it this way: whatever language you use to express what you think it is that someone knows in an internal way (yes, I know that sounds odd), you cannot avoid using language that is grounded and meaningful only in a supraindividual way, referring to external experiences shared by groups of people. I don't expect V and R to solve these problems, but they don't seem to be aware of them, and airily wave away any need for further discussion.

A bit more about empiricism. V and R emphasize the importance of distinguishing between the object of study in linguistics and the source of evidence (p. 13), and they say that the object of study is an individual's knowledge or competence, while the evidence is spoken utterances. Furthermore, they believe that they can detect confusions on

the part of their disciplinary colleagues who do not understand the difference, or who perhaps fail to apply it, when they talk about gradience and categoricality in phonology. (pp. 13-14) They give a couple of examples of comments by Mirjam Ernestus to illustrate this.

What V and R succeed in doing is showing that Ernestus does not agree with them. But have phonologists who have created and defined the field agreed with V and R? It matters, doesn't it? If it is not just Ernestus but everyone else who disagrees with V and R in their actual work, then we as readers are in a funny position: we want to take what they (V and R) say seriously, if only for the time while we read their paper, but if they are lone travelers in the desert, comforted only by the knowledge that they agree with Chomsky on the important questions, we have to scratch our heads and wait for more to come.

Phonology began, it seems to me, with the discovery that in each language there is an inventory of sounds that are contrastive, in that differences between sounds in the inventory can be used to mark lexical contrasts. This is traditionally called the phonemic inventory, of course. Contemporary theories integrate the concept as best they can and as they see fit, but it is not a notion that is going to fade away; it remains central to both synchronic and diachronic linguistics.

How do we determine whether two sounds are contrastive? There are the obvious cases (*pin/pen*), but the cases we phonologists care about are the ones where the connection between evidence and the answer to the question is complex, and may involve asking new questions and making new distinctions. The difficulty does not lie in a theory of how humans work, the theory of V and R, but rather in how we as linguists, as scientists of language, do our work. It is a non-starter to lay down by decree that those people who figure out if two sounds are contrastive for John Goldsmith aren't phonologists (they are something else, like phoneticians or something), because the question is fundamental and central to phonology, not to phonetics. But that seems to be what V and R are saying: if your attempt to determine whether a speaker has a contrast between two

sounds brings you past the concepts that they embed in their logical phonology, then what you are doing isn't phonology. But why? It surely can't be just because they say so, but if we study the history of the practice of phonologists who are trying to answer the questions that have evolved, and what we find does not map to logical phonology/ non-logical phonology, what should we do?

To put the point another way: just because you say you want to build a theory in which variation and gradience has no role, that doesn't mean that you will succeed in answering the questions that have traditionally been central to the field. Of course, you *can* change what the questions are (Chomsky has succeeded in doing that on occasion), but it's difficult so long as the phonologists know what they're doing and why they're doing it.

The *word*? Really? I'll offer one example of what V and R wrote that strikes me as a good example of what I feel they are missing. This is the case of the concept of the *word*. They think that the value of the term, for their argument, is that linguists have established that the word makes sense only from an internalist perspective, and that attempts to characterize a word in purely physical ways have been disastrously ineffective.

So far, so good. But what they fail to note (though surely they know this?) is that even internalists, even generative grammarians, do not have a finished story to give about what a word is. To understand why and how linguists talk about words, it is necessary to actually go and read linguistic analyses by linguists who have thought deeply about the problem. (I recently wrote a paper along those lines and the word—Goldsmith in press—so I have some thoughts on this which the reader can read.) The Bloomfieldian tradition felt that the word was old school, that it was a worn out concept that was protected by old-fashioned scholars of Indo-European languages, and that American linguists knew that talk about words should be replaced by talk about morphemes. That discussion is still going on, as anyone who reads morphology knows, and the argumen-

tation is a complex skein of facts and theoretical argumentation. (At this moment, I think more linguists believe in words than in morphemes, but that is the reverse of the situation in the 1950s.) Here is the point: the arguments proceed based on cross-individual observations, and rarely (if ever?) are based on anything we know about activities in someone's brain. I don't doubt that V and R would want to frame statements about English or Swahili as statements about individual speakers' brains, but there are no grounds for such a statement other than the ideological, by which I mean beyond insisting that it must be true independent of the facts (facts of what linguists actually *do*). We find arguments for and against this by looking at linguistic data and developing traditional linguistic arguments. V and R are not in a position to say that they have a theory that can stand up and say, Yes we know what a word is; linguists are still figuring that out (I have my own theory), and those arguments are not based on language being internal.

I know what science is, even if you don't. At the end of the day, what I find a bit difficult to take about V and R's essay is a tone that suggests that they know what science is, and you don't.

> "In every scientific domain, it is the theory that determines whether something counts as relevant data or not…while linguistics *is* an empirical science, it gives epistemological precedence to theoretical constructs of great explanatory depth over the observables. Again, this is a perfectly normal state of affairs in the natural sciences." (p. 19)

Well, of course that is a little bit true, but it is also only a part of the truth, and not even the largest part. And the style exudes the sort of self-confidence that Chomsky in his younger years displayed when he complained about linguists around him not really understanding what science was. It was never appreciated coming from Chomsky, and it is a tone that does not need to be emulated today. The reader may sense some asperity in

my own remarks, which I regret, but it's a response to being told that we should do things this way because that's what real scientists do.

Quite often one person's explanatory depth is another person's unsubstantiated hypothesis. V and R offer the example of black holes (p. 20) as something predicted by physics before being observed. What they don't point out is that there are many, many examples in the sciences when a theory makes a prediction, even one as awesome as the existence of black holes, and then the predictions don't pan out. You don't get credit in the sciences for believing in the existence of something heretofore unobserved *until it's been discovered*. Science is not about believing predictions; it's about being excited after predictions turn out to be true. Such a huge difference! I think that some popularizers of science have muddied the waters by romanticizing the situation of the scientists who believes in their theories but have to wait a long time for recognition. Sorry – that's how science works; you do have to wait for experiment to show you were right. (It's just like investing a lot of money in a stock that is selling for one dollar a share. If it soars to fifty dollars, then you can tell everyone how smart you were. But till then, it is just an investment, not a brilliant investment.)

Learning and innateness. V and R cite a recent argument from Ray Jackendoff.

> Exactly how does the mind/brain encode instances it has encountered, and what are the dimensions available for encoding them? A crucial constraint on such a theory is that the dimensions in terms of which experience is encoded cannot themselves be learned. They form the basis for learning; they are what enable learning to take place at all. So they have to somehow be wired into the brain in advance of learning. (p. 28)

V and R add, "it is noteworthy that even some researchers in the AI machine learning domain have recently recognized the need to build 'priors' into their system."

This is really not right, and again I chalk my concerns up to the writers not being conversant with machine learning. Most problems in machine learning can be posed along the lines like this: we have a large amount of data which can be expressed only in highly observation-laden ways, but we know that there is a systematicity to it underlyingly, and we need to discover what that is; sometimes this is described as trying to find a manifold of lower dimensionality that is hiding under the data when that data is expressed in terms that are strictly observational. That is what machine learning is about, and there are many answer that have been given that ought to be of great interest to any linguist interested in questions of cognition. I am not entirely sure what Ray (Jackendoff) was referring to, but it does sound like he was wrong; it absolutely *is* possible for the dimensions in which a system is analyzed to be learned — indeed, that's the *normal* case. A very simple example is singular value decomposition, in which a lower dimensional vector space emerges out of analyzing data, and there's no upper limit on the complexity of problems that are addressed this way. Misha Belkin, Partha Niyogi, and Vikas Sindhwani (just to mention three friends) developed very interesting work in the early 2000s that sought to discover lower dimensional manifolds that were close to large collections of data that had been original described in a much higher dimensional space.

And that V and R should say that machine learning people are finally using the notion of priors – I don't know how to respond. I think they do not understand the role that priors play in statistical analysis, and hence in machine learning. The term "prior" is most obvious in bayesian discussions, because Bayes's rule is often described as being the tool that shifts our knowledge from a prior to a posterior. Priors are always there, and have nothing to do with one's position on an empiricism–non-empiricism spectrum or any other (although there are interesting ideas to discuss about Kolmogoroff complexity; again, see Goldsmith 2015). I have the impression that V and R don't really command this material, and I suppose it is the purpose of people making comments, like I am, to make such observations. There you are. Please don't take V and R's comments about machine learning to be accurate.

A second go at it. And yet. Despite everything, I do agree with some of what V and R are excited about. What resonates is the belief that we have only *begun* to grasp the kinds of computational tools we will want to integrate into the general theory of phonology. I could give some examples of what I think are good candidates for this treatment, but hopefully so can you. The point is to be inclusive, not exclusive. If you find it boring to think about how certain phonological contrasts sit there on the fence, partly showing the behavior of a contrast and partly showing the behavior of subphonemic allophony, fine. Don't worry about that problem. I think it is a waste of time to try to show that people who are interested in things that don't interest you are not really doing phonology.

Worse yet, it can be offensive. I know a lot of linguists who have been told (or feel that they have been told) that what they do is not linguistics. Today we are considering the special case: what Mirjam Ernestus does is not phonology (we leave open whether it is linguistics – that depends on finding a different subdiscipline that will accept the research as its own). I say: Don't do that. It is not a step forward for our discipline to be able to demonstrate that what Mirjam Ernestus published is not really phonology, and that she does not understand that. And if you do, think hard about what your credentials are for engaging in the operation. And what counts as a real argument in favor of your position.

The focus in this paper on features and markedness is too bad, in my opinion. I think that features are peripheral to the big questions that phonology tries to answer, and I think that markedness is rarely cogently defined by people who try to take it seriously. The deeper questions in phonology involve the structure of representations, and the interplay between sound-based and grammar-based accounts of morphophonology and phonology, and indeed how the distinction between morphophonology and phonology can be best understood.

I'll end these comments where I began them: with the observation that V and R's paper is astonishingly conservative: to summarize, Chomsky got it right, all along. This proposal doesn't seem like a step forward (nor a step backward); it feels like a position paper describing how 1960s cognitive psychology might want to organize (some parts of) 21st century phonology. It would be more interesting if it took seriously the very smart people it disagrees with. Labov is one obvious example, though his name is not mentioned; Drs. Volenec and Reiss, why not try to understand what someone who disagrees with you would tell you about your views?

Belkin, Mikhail, Partha **Niyogi**, and Vikas **Sindhwani**. 2006. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research.* 2399–2434. **Chater**, Nick, Alexander **Clark**, John **Goldsmith**, and Amy **Perfors**. 2015. *Empiricism and Language Learnability.* Cambridge: Cambridge University Press. **Goldsmith**, John. 1994. A Dynamic Computational Theory of Accent Systems. In *Perspectives in Phonology*, edited by Jennifer Cole and Charles Kisseberth, pp. 1-28. Stanford: Center for the Study of Language and Information. **Goldsmith**, John. 2015. "Towards a new empiricism for linguistics." In Chater et al 2015. **Goldsmith**, John. In press. American descriptivist morphology in the 1950s. To appear in *Oxford Research Encyclopedia of Linguistics*, edited by Rochelle Lieber and Mark Aronoff. Oxford: Oxford University Press. **Goldsmith**, John and Bernard **Laks**. 2019. *Battle in the Mind Fields*. Chicago: University of Chicago Press. **Weinreich**, Uriel, William **Labov** and Marvin **Herzog**. 1968. In Winfred P. Lehmann and Yakov Malkiel, eds., *Directions for Historical Linguistics*, pp. 95–195. Austin: University of Texas Press.

REPLY

It is no surprise that John Goldsmith, editor of a 1993 collection of essays called *The last phonological rule: reflections on constraints and derivations* should look askance at our attempt in 2020 to push a rule-based model of phonological computation. Goldsmith seems to agree with us concerning the lack of originality of our paper. We tried to be as clear as possible that we adopt a well-established approach grounded in the tradition of linguists and other cognitive scientists including Morris Halle, Zenon Pylyshyn, Jerry Fodor, Eric Lenneberg, Ray Jackendoff, and of course Noam Chomsky. This perspective

entails acceptance of notions like the competence-performance distinction, the I-language conception of language as *internal, individual* and *intensional*, and a strong commitment to nativism. Goldsmith's mostly negative characterization is that the paper is "astonishingly conservative: to summarize, Chomsky got it right, all along," so your feelings about our paper and the particular passages that offend Goldsmith will probably be closely tied to how you view what we might call orthodox Chomskyan Linguistics of the past six decades.

In a paper published in 2015, *Towards a new empiricism for linguistics*, that Goldsmith calls "my favorite paper among those I've written" he describes his approach as both "agnostic about psychology" and "non-cognitivist". Our paper is as clear as can be in adopting the cognitivist and psychological assumptions of Chomskyan Linguistics, so there is not really much to discuss, since Goldsmith rejects, or at best doesn't care about, our grounding assumptions. We are not sure what his approach *is*, but it seems to be incommensurable with our own, to a greater extent than others with whom we disagree, but find well-grounded in the generative tradition (of internalism and so on), such as, say, Prince & Smolensky's (1993) Optimality Theory.

Our paper advocates using the term *phonology* (not historical phonology, not computational phonology) only in the generative cognitivist manner, and it takes the old-fashioned question of 'psychological reality' seriously. Thus, one can understand, if not necessarily agree with Goldsmith chiding us "to be inclusive, not exclusive," but some of his suggestions are downright bewildering. For example, in response to our paper arguing that many phonologists who *seem* to accept the tenets of "classical" generative linguistics are being inconsistent, Goldsmith refers to the work of William Labov, the founder of sociolinguistics. We have no idea why.

Similarly, we can't make heads or tails of Goldsmith's discussions of nativism. We present the orthodox (consistent with Kant, Fodor, Jackendoff etc.) view:

> A crucial constraint on [a language acquisition] theory is that the dimensions in terms of which experience is encoded cannot themselves be learned. They form the basis for learning; they are what enable learning to take place at all. So they have to somehow be wired into the brain in advance of learning. (Jackendoff 2015: 187)

Goldsmith's response is that he is "not entirely sure what Ray [Jackendoff] was referring to, but it does sound like he was wrong". Goldsmith suggests that all the supposed problems that UG is meant to solve would disappear if we (Volenec and Reiss, and apparently Ray, Jerry, Noam and Immanuel, too) knew more about machine learning. It would have been nice to have an actual critique of our premises or reasoning. Goldsmith urges the reader not to take our remarks about machine learning to be accurate. This is probably good advice, but not very relevant since the extent of our comments on the topic were a reference to a paper that says this:

> The question [we address is] summarized in the debate between the nativist Gary Marcus and the pioneer of Machine Learning Yann LeCun [...]: Shall we search for a unitary general learning principle able to flexibly adapt to all conditions, including novel ones, or structure artificial minds with driving assumptions, or priors, that orient learning and improve acquisition speed by imposing limiting biases? (Versace et al. 2018)

Goldsmith, responding to our characterization of this paper as making the point that AI and Machine Learning is starting to accept the need for priors, says "I don't know how to respond" and "Priors are always there [in the AI and Machine Learning literature], and have nothing to do with one's position on an empiricism-non-empiricism spectrum or any other." Maybe someone should tell Gary and Yann that they were wasting their time.

However, Jon Rawski [p.c.] points out that the Versace et al. (2018) paper is not an isolated case of the discussion of priors in the current AI literature, as shown by work such as Tony Zador's (2019) *A critique of pure learning and what artificial neural networks can learn from animal brains*. Zador discusses a typical machine learning image categorization task that required ten million (10^7) labeled question-answer pairs in order to mimic human behavior. He points out that:

> [a]lthough the final result of this training is an A[rtifical] N[eural] N[etwork] with a capability that, superficially at least, mimics the human ability to categorize images, the process by which the artificial system learns bears little resemblance to that by which a newborn learns. There are only 10^7 seconds in a year, so a child would need to ask one question every second of her life to receive a comparable volume of labeled data; and of course, most images encountered by a child are not labeled. There is, thus, a mismatch between the available pool of labeled data and how quickly children learn.

The point of Zador's paper is that "[e]very learning model must contain implicit or explicit restrictions on the class of functions that it can learn"; that "much of animal behavior is innate, and does not arise from learning. Animal brains are not the blank slates, equipped with a general-purpose learning algorithm ready to learn anything, as envisioned by some AI researchers"; and that an AI that can approach biological intelligence should be built "according to the same design principles" as brains. However one feels about Zador's proposal that "the genome doesn't encode representations or behaviors directly; it encodes wiring rules," it is clear that discussion of the existence, source and nature of priors is still a lively topic, and that it is clearly *not* the case that "priors are always there" in the AI and machine learning literature.