Conquer primal fear: Phonological features are innate and substance-free

CHARLES REISS (iD)
Concordia University, Montreal, Quebec, Canada
charles.reiss@concordia.ca

and

VENO VOLENEC (iD)
Concordia University, Montreal, Quebec, Canada
veno.volenec@concordia.ca

Abstract

We argue that the representational primes of the human phonological faculty, the so-called dis-
tinctive features, are innate and substance-free. Our arguments for the innateness of features are
built on traditional and novel logical arguments, experimental evidence accumulating over
recent decades, and somewhat detailed proposals about their neurobiological reality. As
symbols in the brain, features are substance-free, that is, they are devoid of articulatory and
acoustic content, or even any direct reference to such phenomena. This is consistent with
our substance-free conception of phonological computation, an approach that eschews func-
tionalist notions like markedness, ease of articulation, and so on. We also outline a neural
model of the phonetics-phonology interface called Cognitive Phonetics, which transduces
innate features to speech articulation and from speech acoustics. These extra-grammatical
transduction procedures are also part of the human biological endowment, which leaves no
room for language-specific phonetics in our theory of the externalization of language. We
show how Cognitive Phonetics can account for traditionally recognized intersegmental coar-
ticulation, as well as the previously unexplored intrasegmental coarticulation, strongly suggest-
ing that the basic units of speech production are transduced features.

Keywords: features, Universal Grammar, neurobiology, phonetics, phonology

CrossMark

Résumé

Nous soutenons que les objets primaires de représentation de la faculté phonologique humaine, les soi-disant traits distinctifs, sont innés et sans substance. Nos arguments en faveur de l'innéité des traits sont fondés sur des arguments logiques traditionnels et nouveaux, sur des preuves expérimentales accumulées au cours des dernières décennies et sur des propositions assez détaillées concernant leur réalité neurobiologique. En tant que symboles dans le cerveau, les traits sont sans substance, c'est-à-dire qu'ils sont dépourvus de contenu articulatoire et acoustique, ou même de toute référence directe à ces phénomènes. Ceci est cohérent avec notre conception 'sans substance' du calcul phonologique, une approche qui évite les notions fonctionnalistes telles que le caractère marqué, la facilité d'articulation, etc. Nous décrivons également un modèle neuronal de l'interface phonétique-phonologie appelé Phonétique Cognitive, qui transduit les traits innés en parole articulée, et transduit les signaux acoustiques en traits innés. Ces procédures de transduction extra-grammaticales font également partie du potentiel biologique humain, ce qui ne laisse aucune place à une phonétique spécifique à la langue dans notre théorie de l'externalisation du langage. Nous montrons comment la Phonétique Cognitive peut rendre compte de la coarticulation intersegmentale traditionnellement reconnue, ainsi que de la coarticulation intrasegmentale jusqu'alors inexplorée, suggérant fortement que les unités de base de la production de la parole sont des traits transduits.

Mots-clés: traits, la Grammaire universelle, la neurobiologie, phonétique, phonologie

1. INTRODUCTION

When a honeybee finds food, it returns to the hive to perform the famous waggle dance that conveys to hivemates the direction and distance of the food source.[1] The direction is the angle from the hive with respect to the sun at the time of dancing (not at the time the food was found), and distance is approximately reflected in the duration of each waggle interval, also taking into account the required effort due to wind speed and other factors. Somehow, the bee's observations of the location of food are encoded as information in the bee's nervous system. The angle between the bee's figure-eight loop and the down direction provided by gravity matches the angle between the direction of the food source and the direction of the sun, which is how the bees know where they have to go. The hivemates that observe the dance end up with (approximately) the same knowledge about where the food is as the reporting scout. In both cases, the information is encoded in the neurology of an individual bee-brain as a symbolic representation that constitutes that bee's knowledge of the location of the food source. What is the substance of those symbolic representations? Are they made of sun, or wind, or angles, or waggles?[2]

---

[1]This paragraph draws on Gallistel (2000) as well as the documentary *The Waggle Dance of the Honeybee* at https://youtu.be/bFDGPgXtK-U by the Georgia Tech College of Computing. See Riley et al. (2005) for discussion of how hivemates make use of the information conveyed in the waggle dance.

[2]Abbreviations: CP: Cognitive Phonetics; IFG: inferior frontal gyrus; OT: Optimality Theory; PMC: primary motor cortex; PR: phonetic representation; PTA: paradigmatic transduction algorithm; SM: sensorimotor (system); Spt: Sylvian fissure; SR: surface

These questions are misguided. The symbols by which the bees store knowledge are transduced from sensory experience (such as observations of sun, wind and flowers) into the nervous system of the scout; then they are transduced out of the scout in the form of directed waggles by what Gallistel calls one of the bee's 'behavioral read-out systems'; then they are transduced into the brains of other bees by some kind of write-in system; and then those bees can use that information to find the food. The symbols themselves do not have nectarous, visual, geometric, waggly or any other kind of property of the world outside of brain neurology.

Phonological symbols work the same way: they are neurally encoded, and they are neither acoustic nor articulatory. Our main goal in this article is to sketch a neurobiologically-grounded view of the symbolic primes of phonological representations, viz. features (section 4). Before doing so, we discuss how features are treated by the phonological computational system (section 2), and we argue that the innateness of features remains the best hypothesis for their origin in individuals (section 3). We offer these concrete arguments about various aspects of phonological *primes* in the hope of mitigating an apparent pandemic of *primal* fear in the field – a resistance to positing innate primes born of an unjustified and tenacious empiricist bias. In brief, we present a model of phonological competence in which both the computational part and the representational part are substance-free, and we argue that the primes of phonological representations, features, are innate brain-based symbols related to – but distinct from – phonetic substance.

2. PHONOLOGICAL COMPUTATION IS SUBSTANCE FREE

Using an analogy from visual perception, we attempt here to clarify what was originally meant by Hale and Reiss (2000, 2008) in using the term "substance-free" in the context of phonological computation.

Consider the two sides of Figure 1. It seems fair to describe the left side as representing a dark gray rectangular object partially occluding a light gray one. The right side appears to depict a light gray rectangular object partially occluding a dark gray one. These descriptions and our ability to see these scenes obviously rely on differences of 'visual substance'. For each side of Figure 1, your mind 'sees' that there are three regions, and that the two non-contiguous regions have the same hue/texture/tone. Your mind also 'infers' that the same-looking regions are parts of the same object, occluded by the other object. This process of inference is substance-free, in that the merging of two regions of appearance x into a single object partially occluded by an object of appearance y is not dependent on the properties of x and y, as long as they are distinct to some degree.[3] As the two sides in Figure 1 show, x and y can be switched without affecting the inference. 'Light gray occludes dark gray' involves the same

_____

(phonological) representation; SMA: supplementary motor area; STA: syntagmatic transduction algorithm; STG: superior temporal gyrus; STS : superior temporal sulcus; UG: Universal Grammar; UR: underlying representation.

[3]There are complications, of course, such as ones involving apparent light sources, as in the checker shadow illusion: <https://en.wikipedia.org/wiki/Checker_shadow_illusion> or in Delwiche (2012).

Figure 1: Occlusion with dark and light gray regions.

abstract inference as 'Dark gray occludes light gray.' The interpretation of object constituency is substance-free in the same sense that we intend the notion to apply in phonology. Deletion of +Round before − High is no different *computationally* from the deletion of −High before +Round.

We are not claiming that absolute and relative properties of *x* and *y* can never bias certain interpretations of foreground vs. background. However, it is clear that there are infinitely many substance-y ways in which *x* and *y* can differ (e.g., in terms of colour, texture and patterns), so the operation used by the mind to decide that one object is occluding another must be substance-free. There are no necessary and sufficient conditions on the 'look' of the regions. Furthermore, it is not necessary that each region has a constant 'look' – the two regions of an occluded object can be coloured with a gradient surface, for example. There are a couple of random examples in Figure 2. Obviously, such examples can be multiplied without bounds.

The Eval function of 'classic' Optimality Theory (OT) (e.g., Prince and Smolensky 1993, McCarthy and Prince 1993) is substance-free, because this function counts markedness violations the same way whether they occur in a constraint referring to +Nasal or −Labial. However, OT, in the classic version and in many commonly adopted later versions, is not completely substance-free in that it encodes putative markedness in Con, the universal constraint set. There are constraints against front rounded vowels and against voiced obstruents in codas, but there are no constraints against back rounded vowels or voiceless obstruents in codas. Subtance-free phonology posits that a neuroscientist of the future could program a human to have a grammar with coda *voicing*, despite the phonetic unlikelihood of such a system arising naturally. This would not involve changing the built-in Universal Grammar (UG). In contrast, a classic OT phonology with universal innate constraints does not allow the possibility of coda voicing unless the neuroscientist is able to rewrite Con, a posited component of UG itself. So, classic OT is substance-abusing, overall.

Let's suppose for the moment that features have a 'semantics' via transduction[4] to and from acoustics and articulation (see section 4 for concrete discussion). Even

---

[4]This is Pylyshyn's (1984) term for an 'intermodular bridge' between the physical and the symbolic. He imports the term into cognitive science from engineering, where, for example, a microphone transduces sound vibrations into electrical signals.
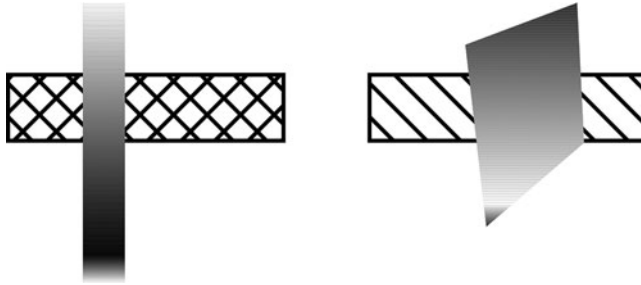
Figure 2: Occlusion with more complex surfaces

so, phonological computation remains substance-free in the same sense that formal symbol manipulation is typically understood:

(1) Brian Cantwell Smith (Smith (1996)), *On the Origin of Objects*
Because formal symbol manipulation is usually defined as 'manipulation of symbols independent of their interpretation,' some people believe that the formal symbol manipulation construal of computation does not rest on a theory of semantics. But that is simply an elementary, though apparently very common, conceptual mistake.

Symbols must have a semantics – i.e., have an actual interpretation, be interpretable, whatever – in order for there to be something substantive for their formal manipulation to proceed independently of.

The diagram in (3) of section 4 should clarify this point – the horizontal phonology mapping is independent of the vertical phonetics mapping. In other words, we claim that phonology is substance free in the same sense that a rule of inference like *modus ponens* is substance free:

(2) If P then Q
$$\frac{P}{\therefore Q}$$

This rule of inference applies regardless of whether the contents of P and Q refer to mortals, dogs, quarks, coronals or deadly sins.

Now, a given language may have a rule that involves deleting the valued feature +VOICED in syllable codas; and another language may have a rule deleting +LATERAL from a liquid after a front vowel; but the mechanism of deletion will presumably be the same in each case, for example via set subtraction (Bale et al. 2014, Bale and Reiss 2018). The process of phonological acquisition involves determining what valued features (from an innately determined set; see section 3) are present in the rules of a language, but the kinds of operations available for rule construction rely in no way upon which particular features occur in a rule. UG does not determine that some operations are for ±VOICED and others for ±LATERAL. That's what we mean when we say that phonological computation is substance free.

3. FEATURES ARE INNATE

Jerry Fodor (1980) famously demonstrated that there are no serious theories of learning from scratch. Learning is always a form of hypothesis confirmation that depends on some pre-existing, built-in primes, which cannot themselves be learned. Echoing Fodor, Jackendoff (1990: 40) says that "in any computational theory, learning can consist only of creating novel combinations of primitives already innately available." Fodor's argument is often ignored, but has never been convincingly refuted, despite attempts like the symposium on "Solutions to Fodor's Puzzle of Concept Acquisition" at the Annual Cognitive Science Society Meeting held in 2005 in Stresa, Italy, where Fodor himself framed the issue thus: "primitive concepts can't be learned, simply because they're needed to formulate the hypotheses in question". We consider features to be a kind of primitive concept. Even among scholars familiar with nativist ideas, in discussions of the source of cognitive structures, there appears to be a bias favoring learning from a blank starting state over nativist explanation. In fact, Berent (2020) shows in a series of experiments that humans have an innate bias toward empiricism and dualism, which explains the origin of the 'primal fear' we seek to overcome. See our article, Volenec and Reiss (2020), for examples of denial of the relevance of the Argument from the Poverty of the Stimulus by phonologists. For concreteness, and in recognition of the prominence to which anti-nativism has risen in recent phonological work, consider, in addition to articles in this issue, the following representative titles: 'The arch not the stones: Universal feature theory without universal features' (Dresher 2014) and 'Phonology without universal grammar' (Archangeli and Pulleyblank 2015).

    Our position, in contrast to the apparently dominant view, is that some version of nativism should generally be taken as the null hypothesis in light of Fodor's logical demonstration. Given this perspective, we will not offer a full rebuttal to every claim put forth *against* the innateness of a set of specific (but still to be determined) universal phonological features, but rather poke, in a "brief, cavalier and dogmatic" (Tolman 1948: 207) fashion, at some of the ideas circulating in the anti-nativist camps. We feel compelled to repeat here ideas that we have expressed elsewhere because we think the arguments we present (often derived from the work of others) are both correct and widely ignored. For example, we do not find them addressed in either of the two works just cited (Dresher 2014, Archangeli and Pulleyblank 2015).

    It is gratifying to read, in David Odden's contribution to this issue (Odden 2022), a critique of the Card Grammar analogy argument for innate features (Hale and Reiss 2003, 2008, Isac and Reiss 2013). Odden's presentation is clear and subtle, but we think his approach fails to address the logical problem discussed by Fodor and Jackendoff because his 'collapse' of physically different 'Jack of clubs' cards starts out with various basic discrete 'J' symbols and a basic '♣' symbol. These are the primes that you need in order to start representing input. Like us, he acknowledges physical distinctions that may be ignored, such as the size of a card, and he appears to accept, like us, an 'innate' equivalence class of 'Jack' in the context of

the analogy. Odden's mechanism for 'discarding' information that is irrelevant to a particular grammar is very much like our process of lexicon optimization, which erases features from earlier stored representations, but we maintain that his version, which applies at the level of sublinguistic transducers, is only coherent when applied to representations that already contain linguistic primes, that is, discrete built-in features.

As noted, we share with Odden the approach of viewing acquisition as the collapsing or *loss* of pre-existing contrasts at some level. For Odden, the level is sublinguistic, whereas for us, the contrasts reflect innate featural distinctions (see section 3.2 below). It is useful to briefly compare our version of this position to Elan Dresher's work, represented not only in the article cited above, but in many other works. He claims that the "concept of a contrastive hierarchy is an innate part of Universal Grammar," (Dresher 2014: 165) but the contrasts for a specific language, encoded via (non-innate) features, are *invented* by phonological learners: "features are not innate. Rather, they are created by the learner as part of the acquisition of phonology" (p. 166). A few points should be made. First, loss of contrasts is the only choice available to us to explain cross-linguistic differences, since we assume that all possible phonological contrasts are present at the initial state – early input is parsed (modulo performance errors) into maximally specified featural representations. Second, we reject altogether the relevance of contrast to the phonological grammar (Reiss 2017). Contrast is used by *phonologists* who look for minimal pairs to gain insight into grammar, but that does not mean that anything in grammar refers to contrast. The learner also needs to appeal to contrast in order to decide if an apparent featural distinction should be maintained in the lexicon – but this is only possible if the learner has access to the featural contrast. In fancy terms, examination of contrast is part of the epistemic toolkit of phonological theory, and perhaps part of the learner's algorithm, but it is a mistake to attribute the notion of contrast to the ontology of phonological grammars. Obviously, our position is at odds with the copious work, such as Nevins (2010), that allows phonological computation to make reference to whether or not a feature is contrastive in a particular context. Finally, we view appeals to contrast as a symptom of a pernicious legacy of functionalist reasoning. Contrast, as manifested in minimal pairs, is related to the phonological encoding of contrasting *meanings*, and thus to the use of language to communicate. Since, in our view, grammars are dumb computational systems with no goals or desires to communicate unambiguously (and in fact, the existence of structural ambiguity in syntax and neutralization in phonology show that grammars can actually impede communication by masking underlying contrasts), we see no reason to posit a role for contrast in grammatical computation.

3.1 A little UG goes a long way

One recurrent objection to positing innate features is that we'd need too many of them – it is implausible that UG should contain so much information. Such claims tend to be vague about how much is too much, and typically don't even discuss

basic notions like orders of magnitude in relating the genome to information structures.[5] One of the contributions in this issue (Samuels et al. 2022) is refreshing in its engagement with our combinatoric arguments. In this section, we reiterate and clarify our views and hope to allay the concerns raised in their section 2.3 discussion of the vast superficial variation in the realization of the high, front unrounded vowels of the German dialect continuum. Our goal is to maintain the plausibility of a 'reasonably' sized innate feature set along with a deterministic transduction system – without language (dialect) specific phonetic implementation.

Phonologists tend to work with 15 to 30 features, and in this context positing a thousand features seems radical, but merely raising the number to, say, 50 is trivial in terms of the amount of encoded information involved (since adding one bit of information doubles the number of distinctions that can be encoded). As we'll see when we do the basic math, the consequences of such a minor change are tremendous.

If Universal Grammar (UG) provides a set of 20 binary features as the building blocks of phonological segments, then UG defines *intensionally* a very large set of possible segments.[6] If each feature must be present in each segment with some value (e.g., +ROUND or −ROUND), the number of segments is about a million: $2^{20} = 1,048,576$. If underspecification is allowed, that is, if segments can be 'incomplete', lacking a value for some of the 20 features, then the number of segments rises to the billions: $3^{20} = 3,486,784,401$. In light of these basic calculations (Reiss 2012, Matamoros and Reiss 2016, Bale and Reiss 2018, Reiss 2022), the null hypothesis for phonological acquisition should be that a small amount of universal, innate knowledge is sufficient to account for the potential diversity of phonological inventories. This conclusion is at odds with the contrary suggestion that to learn any possible language to which they were exposed, learners would "need a (very large) *a priori* set of possible features to choose from" (Cowper and Hall 2014). If fact, with 20 features, the number of segment inventories that UG intensionally defines is $2^{1,048,576}$ (which Google calculator returns as "infinity" – this is the conservative number you get assuming no underspecification; with underspecification you get $2^{3,486,784,401}$). You get this number by considering how many ways there are of answering "yes" or "no" about whether each segment is found in a particular language – that's just the size of the power set (the set of all subsets) of the set of all possible segments. The number of particles in the universe is, by some estimates, only around the

---

[5]We are outside of our intellectual comfort zone here, but it is worth pointing out that the human genome (the totality of our DNA) consists of around six billion individual nucleotide bases <https://www.genome.gov/human-genome-project/Completion-FAQ>. So having a few base pairs which would code for the proteins that build the brain tissue whose activity corresponds to features doesn't seem unreasonable (especially if that same brain tissue is involved in other functions). The burden of proof for the claim that that is too much innate information is on those suffering from primal fear.

[6]This, of course, does not mean that children are born with lots of segments encoded in their brains. It means that they have the capacity to build a lexicon containing morphemes built from segments, each of which is one of the many, many possible combinations of features. UG intensionally defines this (finite, but large) set of segments in the same sense that an English-type grammar intensionally defines an infinite set of sentences.

order of 2^{285} (around 10^{85}). So, our simple calculations show that claims like that of Cowper and Hall (2014) are factually incorrect. The burden of proof is on those who deny universal, innate features to show that a fairly small set is *not* sufficient to account for linguistic diversity.

This perspective on the combinatorics of innate primitives of phonology is consistent with Gallistel and King's (2009: 74) discussion of how combinatoric explosion solves the problem of the power of the representational capacity of biological systems:

> [T]he number of symbols that might have to be realized in a representational system [like the brain–*our addition*] with any real power is for all practical purposes infinite; it vastly exceeds the number of elementary particles in the universe, which is roughly 10^{85} (or 2^{285}), give or take a few orders of magnitude. This is an example of the difference between the infinitude of the possible and the finitude of the actual, a distinction of enormous importance in the design of practical computing machines. A computing machine can only have a finite number of actual symbols in it, but it must be so constructed that the set of possible symbols from which those actual symbols come is *essentially infinite*. (By 'essentially infinite' we will always mean greater than the number of elementary particles in the universe; in other words, not physically realizable.) This means that the machine cannot come with all of the symbols it will ever need already formed. It must be able to construct them as it needs them – as it encounters new referents.

So on the one hand, we must have some innate primes on the basis of which more complex data structures like segments are constructed during the growth of language in the brain, while on the other hand the number of these primes can be fairly small due to combinatoric explosion. Our perspective on combinatorics is thus consistent with Chomsky's (2007) point that "the less attributed to genetic information (in our case, the topic of UG) for determining the development of an organism, the more feasible the study of its evolution". We don't need a lot of UG to get a lot of empirical coverage.

We have made this combinatoric argument repeatedly, yet, apparently not convincingly, so let's bring the discussion back down to something more concrete. With just the five most common features used to discuss vowel systems – HIGH, LOW, BACK, ROUND, ATR – we can define $3^5 = 243$ different vowels (well, 216 if you insist on ruling out any superset of {+HIGH, +LOW}).[7] If you imagine a five-vowel system as containing one member from each of five regions of the vowel space, then with a single partitioning of the space, there are over 40 choices for each region, each of the five vowels. Let's say we cut that in half under the assumption that the vowels in a five-vowel system are not going to be too close to each other – no two vowels will lie near a common region boundary. This leaves us with 20 vowels per region. What this means is that what we phonologists tend to reify as 'the common five-vowel system' is really a set of $20^5 = 3.2$ million

---

[7]There are nine logical possibilities for these two features, given underspecification: all combinations of '+', '−' and 'absence' for the two features. Excluding co-occurrence of +HIGH and +LOW leaves eight possibilities. Multiply 8 by $3^3 = 27$ possibilities for BACK, ROUND and ATR to get 216.

vowel systems. We've simplified for the sake of calculation, but this argument applies in principle to the German vowel systems discussed by Samuels et al. (2022) – vowel systems that look alike can differ in many ways.

If the set of 'basic' vowel features turns out to be a bit larger than five, say, ten, then the number of definable vowels in each region rises to 3^{10}, which is about sixty thousand. Making the same cuts as above (ruling out co-occurrence of +HIGH and +LOW, and then cutting the remainder in half), we still get about twenty-six thousand vowels, so over five thousand per region. So then we get an even larger typology of five vowel systems, namely 5000^5 which is more than 3×10^{18} distinct vowel systems.

This discussion has been concerned with the variety of vowels attributable to featural differences, but there are other factors relevant to observed differences in outputs between languages or even closely related dialects. Such factors again do not require that each variety have a language-specific phonetic module. There can be interactions among the transduction of featural representations and the transduction of other properties such as pitch or stress. If, descriptively, you find languages with two consistently different [i]'s, it may be that they are featurally different, but it can also be that they are the output of I-languages with, say, different pitch accent systems. The interaction of different pitch accent systems with the exact same feature combination will give rise to different "pronunciations of [i]". If we have these several kinds of explanation for variation in, say, vowel pronunciation across languages, why would we want to posit language-specific *phonetic* learning as well (see Hale et al. 2007)?

3.2   Claims about 'emergence'

We will briefly address two claims in the literature concerning putative 'emergence' [8] of features. Mielke (2008) argues that features are not innate and thus must be learned in the course of language growth. The argument is based on the claim that there are many rules that do not refer to (phonetically) natural classes of segments. The discussion has several problems. First, this claim cannot be evaluated since Mielke does not provide a theory of what counts as a phonological rule. As we explain in detail in Volenec and Reiss (2020), positing/determining rules on an intuitive basis is scientifically inadequate. Relatedly, without natural classes of environments, defined by features, it is unclear how a learner could even detect the patterns that supposedly instantiate a putative rule. Finally, as we discuss in Volenec and Reiss (2020), even in a model in which rules are *defined* as referring to natural classes intensionally, the extensionally defined set of forms to which a rule applies may not constitute a natural class, because of bleeding rule interactions.

---

[8]'Emergence' is a poor choice of terms for the assumption that features are not innate, but rather learned on the basis of exposure to external stimuli. It stems from a tacit assumption that 'innate' means 'present at birth', which is clearly incorrect. In humans, teeth are innate, yet they are not present at birth, they 'emerge' several months after birth; puberty is innate, but not present at birth, it 'emerges' later in life, and so on. So something that is innate can also 'emerge' during the growth of an organism, which is clearly not what the so-called 'emergent' theories of features are supposed to convey.

Another kind of argument against nativist feature theory is that the substantive correlates of (some) features reflect perceptual categories shared with other species. What such discussions fail to point out is that it is only a subset of such shared biologically determined perceptual categories that get used in human phonology.[9] We have a name for that subset: *phonological features*. Thus, features perhaps did emerge from acoustic or perceptual categories in the course of evolution, but that is actually an argument *against* 'emergence' in each individual human's development. Evolution involves changes to the genome.

Our view is not original: Chomsky and Halle (1965) concluded in their response to Householder (1965), who claimed that features ought to be posited on a language-specific basis, that "all phonology breaks down if we do not assume analysis […] in terms of universal […] features" (p. 119) and that "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" (p. 127).

3.3   Be careful what you wish for

Before presenting our view of the neurobiological basis of features, we want to point out some consequences that follow if we turn out to be wrong. If you are dead set on rejecting our nativist null hypothesis, you'd better be ready to face the consequences.

3.3.1   The scholar's complaint

First, if there are no innate features, then a tremendous portion of the phonology literature is completely worthless, because there can be no such thing as, say, ATR harmony or ROUND harmony. This is because without innate features there will be 'what we'll call ATR in Bongo', and there will be 'what we'll call ATR in Dagik' and there will be 'what we'll call ATR in Yoruba', and there will be 'what we'll call ATR in Québec French', but there will not be a coherent general notion of 'ATR harmony' because there will be no universal ATR feature as part of the innate human language faculty. All those papers on ATR harmony will be about completely different things, by definition.[10]

3.3.2   The learner's complaint

Second, the idea that phonological feature space is just a learned structure imposed on the space of auditory perception potentially raises more problems than it solves.

---

[9]And a tiny subset at that. For example, the human auditory system can discriminate between more than 1000 levels on a single pitch scale (Fastl and Zwicker 2006), but we do not have 1000 different levels for vowel height (which correlates with F1).

[10]Aside from the fact that we *do* believe in innate features, as non-substance abusers we are not bothered by the fact that ATR harmony has no *formal* properties that set it aside from, say, ROUND harmony. We are *not* claiming that any particular feature system in the literature is correct and that the literature is therefore *not* largely worthless – we are just saying that, if there are no pre-determined features at all, this problem becomes one of principle.

Under the innate features view, the child of a bilingual mother is constrained (i.e., guided by a hand of nature that cannot be disobeyed) to parse all input in terms of combinations of a small set of predetermined categories. However, as the existence of language-independent voiceprint technology demonstrates, there are purely acoustic properties of specific human voices that are detectable regardless of which language is being used. How does a child learn to ignore those properties that tell him that mommy is mommy, or that mommy is happy, and use the ones that allow him to discover French vs. Italian patterns? Pure acoustics provides lots of ways (maybe an infinite number) to categorize signals, most of them linguistically irrelevant, so without innate features, we need another way to explain how kids find the particular structure that works.

Given the astronomical (or maybe infinite) number of ways in which a raw acoustic signal can be categorized, it is miraculous that we get any cross-linguistic generalizations at all, even more so that they are so robust, and that there has never been a need to posit more than about thirty features to account for the attested patterns. All this seems much less miraculous if we assume that humans are innately equipped with a relatively small set of features.

3.3.3   The syntactician's complaint

Third, *Why not?* Unless you have an alternate account, maybe involving oscillators and energy levels" (Tilsen 2019), of why languages have things like negative polarity items and *wh*-movement, you probably believe in innate aspects of syntax and semantics. Isn't it odd to accept that evolution could have managed to 'nativize' those really abstract categories (NPIs don't all have the same formant transitions cross-linguistically), but reject the idea that the abstract (but less so) phonological features could also be innate? Or maybe you're going to have to tell your syntactician friends that all *their* features, like *wh-*, don't exist either as innate categories, and that cross-linguistic recurrent patterning is just a mirage.

3.3.4   More on our null hypothesis

Fourth, a further *Why not?* point relates to the pernicious empiricist prejudice noted above, the notion that positing that a cognitive capacity is learned from scratch is somehow 'better' than positing that it reflects some innate endowment; nativism is rejected as the null hypothesis. However, it behooves us to recall that there is very strong evidence for innate knowledge in many domains and many species. For example, Jon Rawski (p.c.) directs our attention to Zador (2019: 4):

> A striking example of a complex innate behavior in mammals is burrowing: Closely related species of deer mice differ dramatically in the burrows they build with respect to the length and complexity of the tunnels. These innate tendencies are independent of parenting: Mice of one species reared by foster mothers of the other species build burrows like those of their biological parents. Thus, it appears that a large component of an animal's behavioral repertoire is not the result of clever learning algorithms – supervised or unsupervised – but rather of behavior programs already present at birth.

Lorenz and Tinbergen's classic example of innate knowledge is the reaction of newly hatched chicks of various bird species to a dummy figure moving over their cage. Moving in one direction, the figure appears to have a short neck and a long tail, like a hawk or some other bird of prey; moving in the other direction the figure appears to have a long neck and a short tail, more like a harmless goose or other non-prey species. The 'hawk' elicits a panicked reaction in the lab-hatched chicks, but the 'goose' does not (see Schleidt et al. 2011 for discussion and references). Anti-nativists have to explain why they are ready to attribute highly specific innate structures to bird brains, but resistant to attributing innate structures like a phonological feature set to human brains (without ever offering a convincing and comprehensive explanation of *how* exactly features are learned from the environment).

There are countless more examples. Turtles orient themselves toward water immediately after hatching even though they've never seen water and water is not in their line of vision when they hatch; spiders raised in isolation weave a web just like spiders who had a chance to see a web being woven; ants display the path integration ability (dead reckoning) in absence of relevant experience; honeybees raised in isolation use the waggle dance to signal the direction of a food source like other honeybees, and so on. Even when some learning is involved, it is necessary to posit an innate apparatus. There are species of ants and bees that navigate on the basis of the position of the sun in the sky, and part of this capacity appears to be learned to account for variations in latitude. If you ask a typical linguistics undergraduate (at least one who isn't studying in Tromsø and who hasn't been a member of the Nome (*sic*) Chomsky Fan Club) to point in the direction of the sun at midnight, they will typically stare at you in a slack-jawed muddle. However, those sun-navigating ants and bees, when presented at midnight with a bright lamp, navigate as if the light is due north, which of course is the direction of the sun at midnight (in the Northern Hemisphere). The insects' innate endowment 'completes the circle' of the sun's movement from east in the morning to south at noon to west in the evening, allowing the bug to 'know' that the sun is to the north at midnight (Jander 1957). Do we really want to deny innate endowment to humans, who have about 86 billion neurons (Lent et al. 2012), but grant it to ants with a mere two hundred and fifty thousand?

3.3.5   *The breather's complaint*

Fifth, Lenneberg et al. (1967: Chapter 3) points out that "it is quite clear that breathing undergoes peculiar changes during speech" due to "respiratory peculiarities that have evolved" which are specific to speech. Speech is speech, and not grunting or whistling, by virtue of being an externalization of a phonological (featural) representation. Feature nativists (like us) view the evolution of speech-specific respiratory mechanisms as dependent on the existence of speech, and therefore, on the existence of features. If one denies the existence of innate features, one needs to imagine that what ends up as a speech-specific respiratory mechanism in each individual could have evolved without the existence of a set of features present in each individual. It seems a bit odd to think that children are not innately endowed with features,

upon which speech is predicated, but have an innate evolved system of respiration for speech. Implausibly, one has to posit an innate respiratory system that has evolved to support a feature system that itself is not innate. It would be like positing an innate evolved mechanism that is only manifested for the learned process of moving chess pieces.

3.4   Reiteration

Never in the history of philosophy or science has it been demonstrated that a true blank slate can learn. Even when a prominent article like Silver et al. (2017) makes explicit reference to a "tabula rasa", the 'tabula' is anything but 'rasa' – it includes loads of built-in assumptions and preconfigured learning machinery. Where does that machinery come from? It has not been learned, but has been built into the system by human engineers. True blank slates, like rocks, don't learn; that which can learn has some kind of a learning mechanism which itself hasn't been learned; therefore the mechanism is bestowed by biology (or by an intelligent programmer) – it's innate.

Innateness is omnipresent in the biological world – unless you believe that genetics doesn't exist. Since brain development is guided by genetics and since the growth of the language faculty displays all the relevant traits of other genetically-driven developmental processes like puberty and menopause, we should expect that phonology has an innate core just as much as we think that puberty has an innate core. Clearly, the superficial aspects of both puberty and phonological development can be determined by external factors (in phonology, this amounts to 'learning'), but the core general properties of both of these processes are given by biology, not learned by observation. In the case of phonology, they cannot be learned purely by observation because the sorting of experience (which, objectively, is a welter of undifferentiated stimuli) into categories implies the prior existence of at least some categories. Otherwise, either the process of categorization 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; this does not happen with phonological features.[11] Therefore, the null hypothesis should be that all *elementary* units of language – those irreducible units from which all other, more complex representations are built – are given *a priori* (i.e., provided by genetics) and that I-languages are constructed through the interaction of those innate primes with environmental factors.[12]

Phonologists should keep in mind that the notion of innate special purpose mechanisms is not an outlandish idea that is entertained only in those branches of cognitive science with a "guilt-by-association" (Mielke 2008: 38) relationship to

---

[11]For an accessible presentation of this point see Isac and Reiss (2013: Ch. 11), which includes Chomsky's (1980) explanation of how a rich innate endowment prevents us from being "mental amoeboids".

[12]Again, it is conceivable that innate features are not necessarily present in a fully mature form at birth, but the results of speech perception studies on infants like those discussed by Werker (1995) are not particularly supportive of that possibility.

Chomskyan syntax. For example, Gallistel (2000: 1179) points out that innate specialized mechanisms are the rule in biology:

> One cannot use a hemoglobin molecule [for] light transduction and one cannot use a rhodopsin molecule as an oxygen carrier, any more than one can see with an ear or hear with an eye. Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms.

He goes on to point out that there is no justification for rejecting this approach in studying learning and cognition in the biological world:

> In this light, it is odd but true that most past and contemporary theorizing about learning does not assume that learning mechanisms are adaptively specialized for the solution of particular kinds of problems. Most theorizing assumes that there is a general purpose learning process in the brain, a process adapted only to solving the problem of learning. Needless to say, there is never an attempt to formalize what exactly that problem is. From a biological perspective, this is equivalent to assuming that there is a general purpose sensory organ, which solves the problem of sensing.

Let's allow Gallistel to assuage our primal fear – it is okay to posit innate features as part of a specialized biological system that parses and represents linguistic input. This is the species-specific specialized capacity called phonology.

4.   THE NATURE OF PHONOLOGICAL FEATURES AND THEIR RELATION TO PHONETIC SUBSTANCE

Phonological computations operate over atomic units out of which phonological representations are built. At the lowest level of granularity, these units are phonological features. In this section we clarify how features are conceptualized in Substance Free Phonology, focusing primarily on their ontology and their relationship to phonetic substance.

The most basic question that can be asked about features is "What are they?" Following a long tradition in generative phonology, we take features to be abstract mental units that have a lawful but highly indirect relation to phonetic substance.

> Considerations of this nature [that languages do not make free use of acoustic values or articulatory properties] were much in our minds […] 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: 94)

> The relation between a phonemic system [which is built out of bundles of features – *our addition*] and the phonetic record […] is remote and complex.   (Chomsky 1964)

As phonology in the 20th century progressed from the taxonomic and mostly antimentalist structuralism to the cognitively and neurobiologically oriented generative

perspective, attempts have been made to connect feature theory to human neural structures.

> In articulatory terms each feature might be viewed as information the brain sends to the vocal apparatus to perform whatever operations are involved in the production of the sound, while acoustically a feature may be viewed as the information the brain looks for in the sound wave to identify a particular segment as an instance of a particular sound.   (Kenstowicz and Kisseberth 1979: 239)

> [F]eatures correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems. In speech perception detectors sensitive to the properties […] are activated, and appropriate information is provided to centers corresponding to the distinctive feature[s] […]. This information is forwarded to higher centers in the nervous system where identification of the utterance takes place. In producing speech, instructions are sent from higher centers in the nervous system to the different feature[s] […] about the utterance to be produced. The features then activate muscles that produce the states and configurations of different articulators […].   (Halle 1983: 109–10)

Continuing this tradition and sharpening its main claim about the neural reality of features, we adopt the cognitive-neuroscience framework proposed by Gallistel and King (2009). In this framework, the atomic elements of mental representations are called *symbols*. Symbols are "physical entities in a physically realized representational system" (Gallistel and King 2009: 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 neural 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 precisely how features qua neural symbols are realized in the brain, but experimental studies are consistently emphasizing the importance of neural activity in the superior-most part of the superior temporal gyrus (STG), superior temporal sulcus (STS), and Brodmann (BA) areas 44 and 6 (Figure 3).

The representations of articulatory correlates of features are coded in the posterior inferior frontal gyrus of the left hemisphere, traditionally known as Broca's area (Okada et al. 2018). More specifically, Hickok (2012: 138) reports that *pars opercularis* (BA 44) and the ventral-most part of BA 6 store articulatory programs needed to reach the auditory targets imposed by features (roughly corresponding to the blue (left) circle in Figure 3). These auditory targets (i.e., the representations of auditory correlates of features) are coded in the STG and the STS (roughly corresponding to the red (right) circle in Figure 3). Mesgarani et al. (2014) showed that acoustic phonetic information is represented in the STS and is distributed along five distinct areas, each corresponding to a general 'manner of articulation' parameter. By measuring the responses in implanted electrical cortical grids placed along the superior-most part of the temporal gyrus, they found that one electrode responded selectively to stops, one to sibilant fricatives, one to low back vowels, one to high front vowels and a palatal
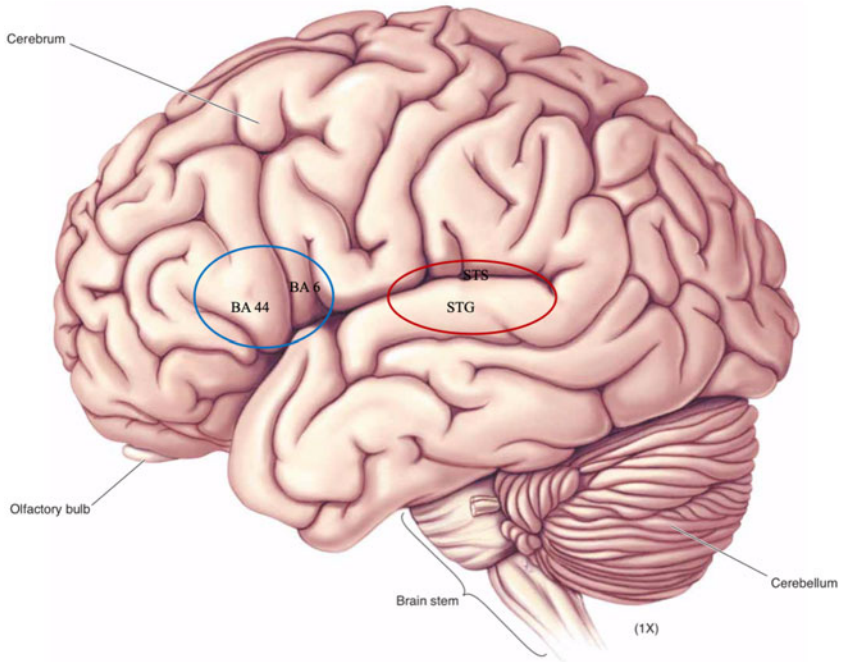
Figure 3: Regions of the human brain (left hemisphere) assumed relevant for the coding of phonological features.

glide, and one to nasals. Similarly, Bouchard et al. (2013) constructed an auditory-based 'place of articulation' cortical map in the STG, confirming labial, coronal and dorsal place features with different electrodes, and cutting across various manner classifications. Using magnetoencephalography, Scharinger et al. (2012) localized three vowel features – height, frontness and roundness – in different parts of the STG.

Features qua neural symbols also meet the criterion of *combinability*. A hallmark of modern phonology is the notion that features can be grouped into sets to construct higher-level, non-atomic data structures. An unordered, unstructured set of features constitutes a phonological *segment* (see Volenec and Reiss 2020, section 4) while a particular organization of segments constitutes a data structure of the next higher taxonomic level, namely 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 behaviour* – different structures can behave alike because they contain identical substructures.

As we saw above, features are also an *efficacious* way of coding information, since their combining leads to combinatoric explosion. For example, if we assume that the brain stores and uses only 30 binary features with the possibility of underspecification, then from this small set of primitive symbols we can construct 3^{30} or about

206 trillion different segments. 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). 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 (as in Keating 1984, 1990), which simplifies the sequence of conceptual steps needed to account for the externalization of language (see Volenec and Reiss 2020 for elaboration).

Recent neuropsychological studies have shed light on some other aspects of features that are significant for phonological theory, namely their discreteness, binarity and potential underspecification. By eliciting magnetic mismatch fields in an oddball paradigm, Phillips et al. (2000: 1038) have shown that the left hemisphere auditory cortex has access to representations of discrete and binary phonological categories. In other words, their study has "demonstrate[d] the all-or-nothing property of phonological category membership," where this category membership is determined on the basis of phonological features and not on the basis of general categorical auditory perception. This finding contradicts the claim that gradient articulatory gestures serve as basic units in phonological computation (as in Browman and Goldstein 1989). Furthermore, Scharinger et al. (2016) found that a less specified vowel compared to its more specified counterpart resulted in stronger activation in the left STS, thus providing some insight into the neural underpinnings of phonological underspecification.

Even though some progress has been made in discovering the neural reality of features, we are still far from being able to refer to particular features by stating their neurobiological substrate, and therefore have to resort to using symbols (labels, names) to refer to 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. In other words, 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 below) interpret the identity of a feature by the place of the neural activity in the brain and its temporal properties (Khalighinejad et al. 2017). 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 and King 2009: 73). Possibly, the actual form of all features is the same – a neural spike (i.e., an action potential). But more important, the unique location of the spike, and 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 such as *alpha*.

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*, or *correlate*, or *the represented*. Phonological features are symbols that 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 and King 2009: 56) because "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 below), and features do not encode speech-related information in any straightforward way. 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.[13]

It should also be noted that there is ample experimental evidence for the substance-free nature of features. Phillips et al. (2000: 1040) have concluded 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, that is, 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

---

[13]In some earlier work, we referred to features as 'substantive' to indicate the regularity of the transduction processes – the feature we call +HIGH cannot be transduced in into, say, spreading of the vocal folds, which is correlated with a feature we typically call −VOICE. This older usage of the term 'substantive' led to confusion and thus turns out to have been a bad choice of terminology.
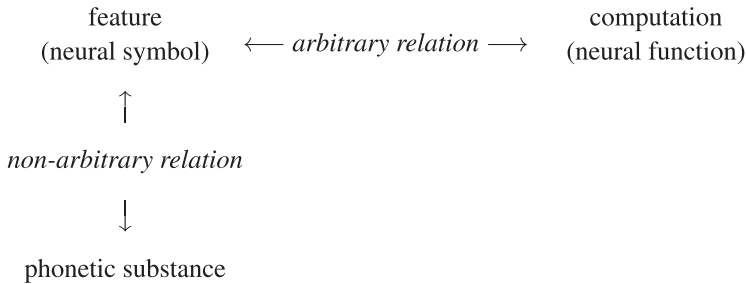
investigated word sequences in absence of overt speech. These neuropsychological studies suggest that phonological features cannot be equated with the phonetic correlates that are typically associated with them, which is to say, that features and phonetic substance are two different things.

The results of these studies should not surprise us, as they merely reflect a more general principle in cognitive neuroscience. In cognitive domains that are unrelated to language, a decoupling of a mental/brain representation from the stimuli used to elicit it has frequently been demonstrated. In a series of experiments (Quiroga et al. 2005, Cerf et al. 2010) Itzhak Fried and colleagues have shown that an individual neuron in the medial temporal lobe is consistently selectively activated when – to take one specific example – a mental representation of the actress Halle Berry is invoked. Crucially, the same 'Halle Berry neuron' was activated in the following disparate situations: when Berry was shown in a photograph, when a video of her was played, when she was seen dressed as Catwoman, when she was seen without the costume in various instances (with various outfits, hairstyles and accessories), when only her name was written on a screen, when her name was said out loud, and when the participants imagined her. Such a neuron corresponds to a mental representation that is clearly not tied to any particular sensory modality, and whose relation to substance/stimulus is highly indirect. We assume that the same principle holds for phonological primes.

We do not think that the relation between any given feature and its correlate is random or arbitrary. If it were, then any feature could in principle be realized by any possible human articulation, just as 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 would be possible that +Round sometimes gets realized as a lowering of the velum, sometimes as a raising of the tongue dorsum, and so on.[14] But there seems to be no convincing evidence for anything like that, so the simplest assumption is that it is not the case. Instead, we assume that there is a non-arbitrary, lawful relation between features and their correlates. The nature of this relation is described in more detail below, but at this point it is important to emphasize that the lawful relation between features and their correlates is *phonologically* irrelevant, as we stated in section 2. 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. The relation between features and phonological computation is thus completely independent from the relation between features and phonetic substance, as can be seen in the diagram in (3).

---

[14]This is not relevant to the difficult question of the relationship between the features of signed languages and spoken languages. In the present discussion, we are talking about the treatment of a given feature by a single transduction system.

(3) Phonology (horizontal) and the phonology-phonetics interface (vertical)

feature
(neural symbol)   ⟵ *arbitrary relation* ⟶   computation
(neural function)

↑

non-arbitrary relation

↓

phonetic substance

The diagram in (3) also reflects the distinction between computation and transduction, which we will make use of below. Computation – corresponding to the horizontal relationship in (3) – is the formal manipulation (reordering, regrouping, deletion, addition, etc.) of representational elements within a single cognitive system, and without a change in the 'representational alphabet' characteristic for that system. Transduction – corresponding to the vertical relationship in (3) – is a process of transmuting an element in one form into a distinct form, that is, a mapping between dissimilar formats.[15]

As substance-free symbols, features do not contain information on the temporal coordination of muscle contractions, on the spectral configuration of the acoustic target to be reached, and so on. Yet without this information, the respiratory, phonatory and articulatory systems cannot produce speech. The sensorimotor (SM) system which is in charge of speech production requires information about substance and time in order to arrange the articulatory score (Guenther 2016), therefore this information has to be integrated into a representation before being fed to the SM system.[16] We thus posit a transduction component that connects phonological competence with the vastly different SM system. Our theory of that component and the component itself are called *Cognitive Phonetics* or CP (Volenec and Reiss 2017).

CP proposes that the phonology-phonetics interface consists of at least two transduction procedures that convert the substance-free output of phonology into a representational format that contains substantive information required by the SM system in order to externalize language through speech. The inputs to CP are the outputs of phonology, that is, surface phonological representations (SRs). SRs are strings of segments, each of
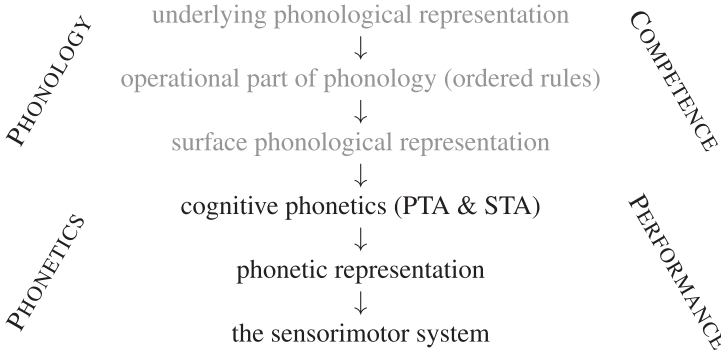
---

[15]At the neural level of description, we tentatively assume that transduction of features corresponds to the activity in the anterior part of the left insula and in parts of the basal ganglia and the cerebellum. See Figure 4 and the accompanying text for elaboration. So, what can, at an abstract cognitive level, be described as a mapping between dissimilar formats (i.e., transduction) corresponds to a difference in the localization of neural activity at a more concrete neural level.

[16]Since the term 'sensorimotor' system is used ambiguously in the linguistic literature, 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.

which is a set of features. Each feature of SRs is transduced and subsequently receives interpretation by the SM system (see Lenneberg et al. 1967: §3). This transduction is carried out by two algorithms (see Marr 1982: 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, and STA distributes that activity temporally. These transduction algorithms yield an output representation of CP, which then feeds the SM system. The output of CP is called the *phonetic representation* (PR),[17] 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 can now be expanded to accommodate the transduction performed by CP, transforming it into a more complete 'speech chain' set out in (4):

(4) Phonology and Phonetics

PHONOLOGY
underlying phonological representation
↓
operational part of phonology (ordered rules)
↓
surface phonological representation
↓
COMPETENCE

PHONETICS
cognitive phonetics (PTA & STA)
↓
phonetic representation
↓
the sensorimotor system
PERFORMANCE

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 and Reiss 2019 for further examples). We will see that PTA and STA have considerable implications: they open the possibility of elegantly accounting for subtle yet systematic interactions of two kinds of coarticulatory effects, which is only possible if we assume that the basic units of speech production are indeed transduced phonological features. Suppose that a hypothetical I-language contains SRs [lok] and [luk]. Each segment is a set of features, and

---

[17]In generative linguistics literature, the output of phonological grammar, what we call a surface phonological representation, is sometimes called a phonetic representation (see Hall 2001: 29). Note that we use the term very differently – our phonetic representations are not part of grammar and do not contain the phonological features that are in URs and SRs.

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. 2019) – which lead 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 *intrasegmental 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 *intersegmental coarticulation*, where transduced features from *different* segments interact. Returning to SRs [lok] and [luk], it is now apparent (a) that PR_{+Round} is different for [o] than for [u] due to its intrasegmental coarticulation with PR_{High}; and (b) that [l]'s inherent PR_{-Round} is now temporally overlapping with the PR_{+Round} from the adjacent vowels because 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 (see Volenec and Reiss 2017, section 5 for further implications of CP).

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 language externalization by a great number of factors. Transduction is accompanied
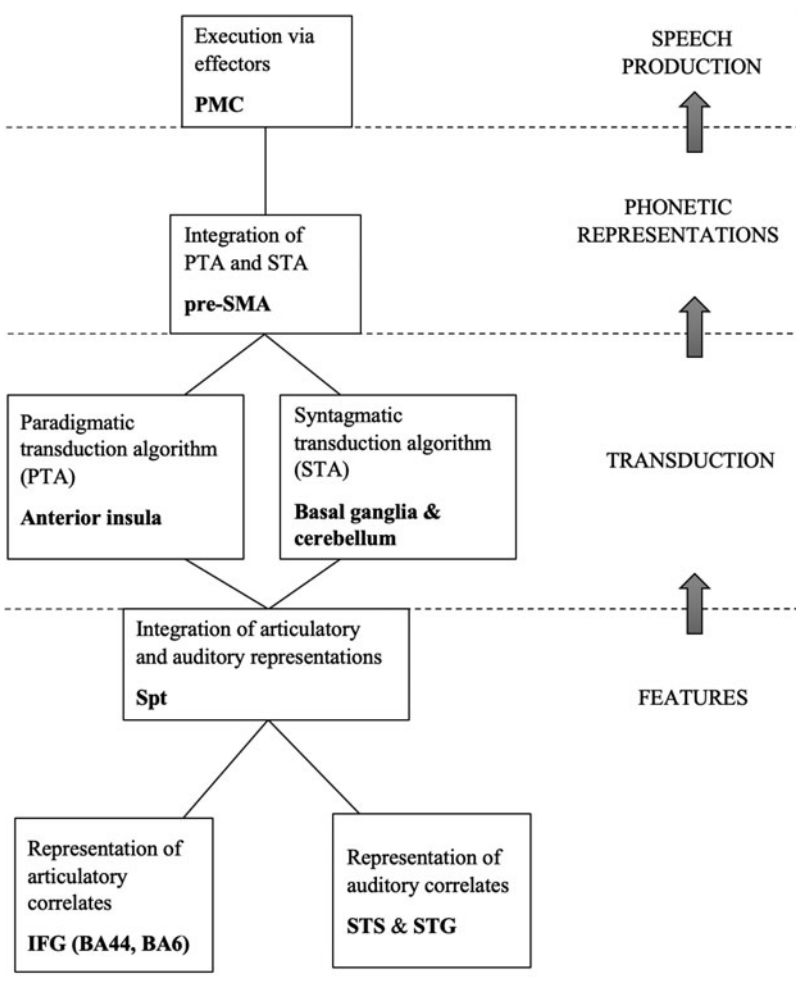
Figure 4: Neural substrate of phonological features and of the phonology-phonetics interface.

by other performance factors that have no bearing on either phonology or transduction, factors like muscle fatigue, degree of enunciation, interruptions due to sneezing, trying to achieve a certain intensity level, 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 even 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. However, this apparent lack of invariance in the realization of a cognitively invariant category is not a matter of transduction, but rather is a result of accidental performance factors.
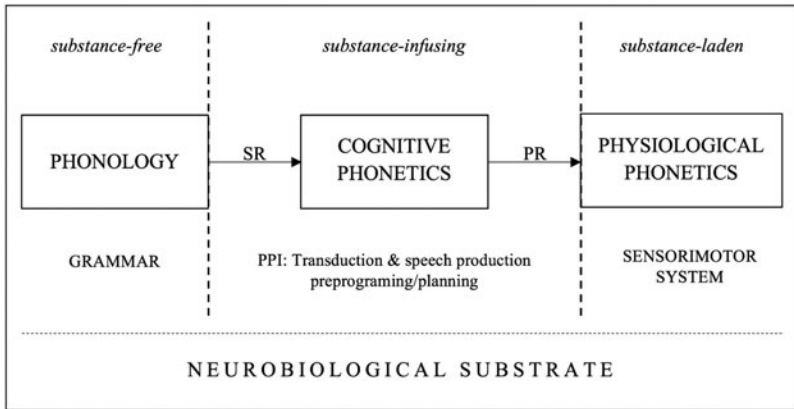
Figure 5: The architecture of the phonology-phonetics interface in the theory of Cognitive Phonetics.

Recent neuroscience evidence is consistent with the idea that CP transduces features, which are to be understood as symbols in the brain, into temporally distributed neuromuscular activities (elements of PRs), thus relating phonological competence to the vastly different SM system (Dronkers 1996, Hickok and Poeppel 2007, Eickhoff et al. 2009, Hickok 2012, Guenther 2016, Okada et al. 2018). 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 transmitted 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 integrated 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 4.

Figure 5 summarizes the general architecture of the phonology-phonetics interface 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.

5. FINAL REMARKS

In lieu of a conclusion and still in Tolman's "cavalier and dogmatic" style, let us summarize our main points. The human language faculty exists in the human brain, and

only there (this is *internalism*). Phonology is one part, or one *module*, in technical terms, of that cognitive, brain-based faculty. This module consists of computational and representational aspects. The computations are ordered logical operations, rules, and they apply in a manner that is blind to phonetic substance. The representations are made from elementary units, features, which are symbols in the brain, in the sense of Gallistel and King (2009). These symbols are innate (this is *nativism*) – they are "knowledge unlearned and untaught" (Halle 1978), which is bestowed by human biology. They are also devoid of phonetic substance. Thus, neither the computational nor the representational aspect of phonology contains phonetic substance (this is *formalism*), which is why this approach is called *Substance Free Phonology*.

By drawing on traditional and novel logical arguments, on experimental evidence accumulating over recent decades, and on somewhat detailed proposals grounded in neurobiology, we hope to have mitigated the widespread primal fear manifested as bias against the innateness of phonological primes. Beyond phonology, we propose that our approach applies to the necessary innateness of all elementary (indivisible, irreducible) units of cognition.

Lila Gleitman intended irony when she quipped that "empiricism is innate" (quoted from Gallistel 2018: 276), but the experiments reported by Berent (2020) suggest that resistance to nativism really is pervasive and deeply ingrained in humans; so as scholars, we need to work to overcome 'primal fear' in the same way we need to overcome our naive belief that heavier objects fall faster than lighter objects. With our discussion of the relation between phonological competence and phonetic substance, we hope to have provided a further "illustration of the necessity of getting behind the sense data of any type of expression in order to grasp the intuitive […] forms which alone give significance to such expression" (Sapir 1925: 45).

REFERENCES

Archangeli, Diana, and Douglas Pulleyblank. 2015. Phonology without universal grammar. *Frontiers in Psychology* 6: 12–29.

Bale, Alan, Maxime Papillon, and Charles Reiss. 2014. Targeting underspecified segments: A formal analysis of feature changing and feature filling rules. *Lingua* 148: 240–253.

Bale, Alan, and Charles Reiss. 2018. *Phonology: A formal introduction*. Cambridge, MA: MIT Press.

Berent, Iris. 2020. *The blind storyteller: How we reason about human nature*. Oxford University Press.

Bouchard, Kristofer E., Nima Mesgarani, Keith Johnson, and Edward F. Chang. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495(7441): 327.

Browman, Catherine P., and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2): 201–251.

Cerf, Moran, Nikhil Thiruvengadam, Florian Mormann, Alexander Kraskov, Rodrigo Quian Quiroga, Christof Koch, and Itzhak Fried. 2010. On-line, voluntary control of human temporal lobe neurons. *Nature* 467(7319): 1104–1108.

Chomsky, Noam. 1964. Formal discussion in response to W. Miller and S. Ervin. In *The acquisition of language,* ed. Ursula Bellugi and Roger Brown, 35–39. Chicago: University of Chicago Press.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + recursion = language?: Chomsky's minimalism and the view from syntax-semantics,* ed. Uli Sauerland and Hans-Martin Gärtner, 1–24. Berlin: Mouton de Gruyter.

Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1: 97–138.

Cowper, Elizabeth, and Daniel Currie Hall. 2014. Reductio ad discrimen: Where features come from. *Nordlyd* 41(2): 145–164.

Delwiche, Jeannine F. 2012. You eat with your eyes first. *Physiology and behavior* 107(4): 502–504.

Dresher, B. Elan. 2014. The arch not the stones: Universal feature theory without universal features. *Nordlyd* 41(2): 165–181.

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

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

Fastl, Hugo, and Eberhard Zwicker. 2006. *Psychoacoustics: Facts and models*, vol. 22. Heidelberg: Springer Science and Business Media.

Fodor, Jerry. 1980. On the impossibility of acquiring "more powerful" structures. In *Language learning: The debate between Jean Piaget and Noam Chomsky*, ed. M. Piatelli-Palmarini, 142–162. Harvard University Press.

Gallistel, C. Randy. 2000. The replacement of general-purpose learning models with adaptively specialized learning modules. In *The new cognitive neurosciences,* ed. Michael S. Gazzaniga, 1179–1191. Cambridge, MA: MIT Press, 2nd ed.

Gallistel, C Randy. 2018. Finding numbers in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1740): 20170119.

Gallistel, C. Randy, and Adam Philip King. 2009. *Memory and the computational brain: Why cognitive science will transform neuroscience.* Chichester, West Sussex, UK: Wiley-Blackwell.

Guenther, Frank H. 2016. *Neural control of speech*. Cambridge, MA: MIT Press.

Hale, Mark, Madelyn Kissock, and Charles Reiss. 2007. Microvariation, variation, and the features of universal grammar. *Lingua* 117(4): 645–665.

Hale, Mark, and Charles Reiss. 2000. "substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31(1): 157–169.

Hale, Mark, and Charles Reiss. 2003. The subset principle in phonology: Why the tabula can't be rasa. *Journal of Linguistics* 39(02): 219–244.

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

Hall, Tracy Alan. 2001. Introduction: Phonological representations and phonetic implementation of distinctive features. In *Distinctive feature theory*, ed. Tracy Alan Hall, 1–40. Berlin: Mouton de Gruyter.

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

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In *Linguistic theory and psychological reality*, ed. Morris Halle, Joan Bresnan, and George Miller, 294–303. MIT Press.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1(1): 91–105.

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

Hickok, Gregory, and David Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5): 393–402.

Householder, Fred W. 1965. On some recent claims in phonological theory. *Journal of linguistics* 1(1): 13–34.

Isac, Daniela, and Charles Reiss. 2013. *I-language: An introduction to linguistics as cognitive science*. Oxford: Oxford University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. MIT Press.

Jander, Rudolf. 1957. Die optische richtungsorientierung der roten waldameise (formica ruea l.) [The optical directional orientation of the red wood ant]. *Zeitschrift für vergleichende Physiologie* 40(2): 162–238.

Keating, Patricia. 1984. Universal phonetics and the organization of grammars. *UCLA Working Papers in Phonetics* 59: 35–49.

Keating, Patricia A. 1990. Phonetic representations in a generative grammar. *Journal of phonetics* 18(3): 321–334.

Kenstowicz, Michael, and Charles Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.

Khalighinejad, Bahar, Guilherme Cruzatto da Silva, and Nima Mesgarani. 2017. Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience* 37(8): 2176–2185.

Lenneberg, Eric H., Noam Chomsky, and Otto Marx. 1967. *Biological foundations of language*. New York: Wiley.

Lent, Roberto, Frederico A.C. Azevedo, Carlos H. Andrade-Moraes, and Ana V.O. Pinto. 2012. How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience* 35(1): 1–9.

Magrassi, Lorenzo, Giuseppe Aromataris, Alessandro Cabrini, Valerio Annovazzi-Lodi, and Andrea Moro. 2015. Sound representation in higher language areas during language generation. *Proceedings of the National Academy of Sciences* 112(6): 1868–1873.

Marr, David. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Matamoros, Camila, and Charles Reiss. 2016. Symbol taxonomy in biophonology. In *Biolinguistic investigations on the language faculty,* ed. Anna Maria Di Sciullo, 41–54. Philadelphia: John Benjamins.

McCarthy, John, and Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. *Technical Report# 3, Rutgers University Center for Cognitive Science*.

Mesgarani, Nima, Connie Cheung, Keith Johnson, and Edward F Chang. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343(6174): 1006–1010.

Mielke, Jeffrey. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

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

Odden, David. 2022. Radical substance-free phonology and feature learning. *Canadian Journal of Linguistics* 67(4): 500–551.

Okada, Kayoko, William Matchin, and Gregory Hickok. 2018. Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin and Review* 25(1): 423–430.

Phillips, Colin, Thomas Pellathy, Alec Marantz, Elron Yellin, Kenneth Wexler, David Poeppel, Martha McGinnis, and Timothy Roberts. 2000. Auditory cortex accesses

phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience* 12 (6): 1038–1055.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar (technical report)*. New Brunswick, NJ: Rutgers Center for Cognitive Science.

Pylyshyn, Zenon. 1984. *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Quiroga, R. Quian, Leila Reddy, Gabriel Kreiman, Christof Koch, and Itzhak Fried. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435 (7045): 1102–1107.

Reiss, Charles. 2012. Towards a bottom-up approach to phonological typology. In *Towards a biolinguistic understanding of grammar,* ed. A.M. di Sciullo, 169–191. Philadelphia: John Benjamins.

Reiss, Charles. 2017. Contrast is irrelevant in phonology: A simple account of Russian /v/ as /V/. In *Arguments of substance: Beyond markedness,* ed. Bridget D. Samuels, 23–45. Amsterdam: John Benjamins.

Reiss, Charles. 2022. Plastics. In *Ha! Linguistic Studies in Honor of Mark R. Hale,* ed. Laura Grestenberger, Charles Reiss, Hannes A. Fellner, and Gabriel Z. Pantillon, 327–330. Vienna: Dr. Ludwig Reichert Verlag.

Riley, Joe R, Uwe Greggers, Alan D Smith, Don R Reynolds, and Randolf Menzel. 2005. The flight paths of honeybees recruited by the waggle dance. *Nature* 435(7039): 205–207.

Samuels, Bridget D., Samuel Andersson, Ollie Sayeed, and Bert Vaux. 2022. Getting ready for primetime: Paths to acquiring substance-free phonology. *Canadian Journal of Linguistics* 67(4): 552–580.

Sapir, Edward. 1925. Sound patterns in language. *Language* 1: 37–51.

Scharinger, Mathias, Ulrike Domahs, Elise Klein, and Frank Domahs. 2016. Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus. *Brain and Language* 163: 42–49.

Scharinger, Mathias, Philip J. Monahan, and William J. Idsardi. 2012. Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research* 55(3): 903–918.

Schleidt, Wolfgang, Michael D. Shalter, and Humberto Moura-Neto. 2011. The hawk/goose story: The classical ethological experiments of Lorenz and Tinbergen, revisited. *Journal of Comparative Psychology* 125(2): 121–133.

Seikel, J Anthony, David G Drumright, and Daniel J Hudock. 2019. *Anatomy and physiology for speech, language, and hearing*. San Diego: Plural Publishing.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. 2017. Mastering the game of Go without human knowledge. *Nature* 550 (7676): 354–359.

Smith, Brian Cantwell. 1996. *On the origin of objects*. Cambridge, MA: MIT Press.

Tilsen, Sam. 2019. *Syntax with oscillators and energy levels*, vol. 7. Berlin: Language Science Press.

Tolman, Edward C. 1948. Cognitive maps in rats and men. *Psychological Review* 55(4): 189–208.

Volenec, Veno, and Charles Reiss. 2017. Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics* 11: 251–294.

Volenec, Veno, and Charles Reiss. 2019. The intervocalic palatal glide in cognitive phonetics. In *Proceedings of the 49th Meeting of the North East Linguistics Society*, 255–264.

Volenec, Veno, and Charles Reiss. 2020. Formal generative phonology. *Radical: A Journal of Phonology* 2: 1–148.

Werker, Janet. 1995. Exploring developmental changes in cross-language speech perception. In *An invitation to Cognitive Science,* ed. Daniel N. Osherson and Lila R. Gleitman, vol. 1, 87–106. Cambridge, Mass.: MIT Press, 2nd ed.

Zador, Anthony M. 2019. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications* 10(1): 3770.