13 Evolution of the Speech Code: Higher-Order Symbolism and the Grammatical Big Bang

Daniel Silverman

Our speech code may have originated as an accompaniment to a manual system consisting of iconic gestures (Tomasello 2008). In this scenario, the speech code broke away from its redundant origins, coming to replace an iconic visual-receptive system with a symbolic auditory-receptive one. This qualitative change from (hand-based) iconicity to (speech-based) symbolism may have quickly evolved to the *higher-order* symbolic status that is characteristic of language.

Herein, *first-order symbolism* refers to a one-to-one correspondence between (arbitrary) symbol and meaning. It is a consequence of single vocal symbols produced in isolation. *Second-order symbolism* evolves from first-order symbolism as two vocal symbols are juxtaposed, inevitably changing the phonetic character of both. Symbolism of the second order involves a breakdown of a one-to-one symbol-meaning correspondence, culminating in many-to-one and one-to-many correspondences between symbol and meaning. *Third-order symbolism* evolves from second-order symbolism as a consequence of string-medial phonetic content being of sporadically ambiguous affiliation between our two juxtaposed symbols, thus potentially inducing listener confusion: if both structures are sensibly interpretable, listeners may wonder, "Is the medial portion of this phonetic event part of the first symbol or the second?" As will be argued, such semantic ambiguity of structural origin triggers this phonetic string's analysis into a hierarchical constituent structure by listeners, thus paving the way for recursion.

As lower orders of symbolism naturally (and perhaps rather suddenly) evolved to higher orders, we may characterize the beginnings of the speech code as triggering a grammatical Big Bang.

13.1 Zero-Order Symbolism: The Iconic Manual Gesture

As noted, Tomasello (2008) suggests that our early communication system may have consisted of iconic hand-based gestures produced in isolation from one another, just as exists in our primate relatives today. Such iconic manual gestures were likely to have been non-symbolic—or "zero-order symbolic"—in nature. But hand-based visual signaling does not permit manual "multi-tasking," requires close, daytime contact, and possesses limited cue redundancy, likely rendering it ill-equipped to jump-start a system as complex as grammar. This is especially true of *iconic* visual symbols: regardless of the magnitude of the hand gestures or the angle from which they were viewed, if there is a direct link between action and meaning, these gestures' iconic status would likely have resisted transformation into a symbolic system.

Indeed, even if a manual iconic system has the potential to evolve into a manual symbolic one, the intervening innovation of a sound-based system quickly and irrevocably quashed that conceivable trajectory. Acoustic signaling allows for vocal-manual "multi-tasking," does not require close, daytime contact, and is particularly rife with cue redundancy (Ay, Flack, and Krakauer 2007). Any era of multi-modal communication (involving both vision and sound) was largely pruned of its visual component, settling towards a sound mode of sufficient "robust overdesign" (Krakauer and Plotkin 2004) to evolve toward higher-order symbolic status.

Perhaps most importantly, the *inherently symbolic* character of the speech code acted to unshackle its semiotic character from the invariant one-to-one relationship between action and meaning that is characteristic of an iconic, gestural system, culminating in a system possessing both *one-to-many* and *many-to-one* relationships between sound and meaning. To understand both the simple causes and the complex effects of this development, we trace the origins of first-order symbolism as sounds are produced in isolation.

13.2 First-Order Symbolism in the Speech Code: One-to-One Correspondence between Sound and Meaning

The first meaning-imbued sounds of our species (morphemes) may have quickly settled towards ones involving a sudden expulsion of air from the mouth due to an oral seal being broken (oral stops) followed by vocal fold vibration accompanying the oral opening gesture (vowels). There are articulatory, aerodynamic, acoustic, and auditory reasons for this (the four "A"s).

Regarding articulation, a complete oral closure followed by its release is quite easy to produce in comparison to other gestures that have become part of the speech code, gestures that often require extreme muscular and timing precision to achieve their characteristic aerodynamic, acoustic, and auditory traits (Ladefoged and Johnson 2011).

Aerodynamically, this simple articulatory action produces a passively energized expulsion of air from the vocal tract. As air is the medium of sound transmission, increased airflow allows for more salient and more varied sounds. Perhaps especially, upon the breaking of an oral seal and allowing air to rapidly flow from the lungs and out the mouth, the vocal folds, when properly postured, may readily engage in vibratory activity (Rothenberg 1968).

Acoustically, this sudden and forceful expulsion of air produces a speech signal of comparatively heightened energy, one in which any number of pitch/phonation (source) and resonance (filter) modifications might be encoded.

Regarding audition, the mammalian auditory nerve is especially responsive to sudden increases in acoustic energy (Delgutte 1982; Tyler et al. 1982); a quick reaction to the sudden breaking of silence provides obvious survival advantages in predation situations. The incipient speech code would likely exploit this property from the outset, as it does to this very day (Bladon 1986).

This nascent oral seal may be at the lips, but also, the flexibility of the tongue allows both its front to form a seal at the alveolar ridge, and its back to form a seal at the soft palate. The perceptual product of these distinct closure locations is three easily-distinguished speech events of exceptionally short duration. This tripartite perceptual distinction establishes the conditions for different acoustic signals to encode different meanings; we might imagine an early stage during which these three closure postures were in place, coordinated with largely undifferentiated qualities to their opening postures, perhaps resulting in three sounds, roughly, **pu**, **ti**, **ka**.

If vocal activity of this nature was indeed harnessed to encode meaning, the semiotic character of primitive speech was of a first-order state, in contrast to the zero-order state of the manual-gestural system with which it may have overlapped: each of the three sounds might encode a single meaning (maybe "Run!", "Kill!/Eat!", "Sex!"). One arbitrary event corresponds to one meaning, and one meaning is cued by one arbitrary event. Still, despite this move toward a speech-based semiotic system, this one-to-one correspondence between event and meaning is perhaps characteristic of almost all animal sound communication systems, though, to be sure, whereas early human vocalizations were probably both voluntary and situation-semantically flexible, animal vocalizations are almost surely involuntary, situation-reactive, and instinctual (Jackendoff 1999). Even sporadic deceptive and stifled animal calls are amenable to such an analysis, as such behaviors may be a consequence of a genetically inherited *probability* of use and disuse.

Nonetheless, we are far from grammar.

13.3 Second-Order Symbolism in the Speech Code: One-to-Many and Many-to-One Correspondence between Sound and Meaning

How might second-order symbolism have evolved from these modest beginnings?

13.3.1 Innovating the Juxtaposition of Two Symbols, and the Rise of Compositionality

Consider the physical consequences of producing two of our meaning-imbued sounds in quick succession. Exhaustively, these are **pu-pu**, **pu-ti**, **pu-ka**, **ti-pu**, **ti-ti**, **ti-ka**, **ka-pu**, **ka-ti**, and **ka-ka**. There is any number of ways that such complexity might develop. For example, two-sound sequences may represent an assemblage of a complex verb-like element, say "Run! Kill!/Eat!" (**pu-ti**) or "Kill!/Eat! Run!" (**ti-pu**), either of which might convey a passive predation warning ("Run if you don't want to get killed and eaten (by that animal)!") or an active predation call ("Run to kill and eat (that animal)!"). Alternatively, two sounds may be strung together to name more objects or events in a nascent form of noun-noun compounding. Both of these structure-building strategies are present in perhaps all languages, of course, but while we will return to the increased semantic complexity that results from such groupings of sounds, for now, consider their phonetic complexities, complexities that culminate in second-order symbolism.

Indeed, from the moment that a juxtaposition of two sounds is regularly produced, the speech signal is irrevocably transformed into a second-order symbolic system. Here's why: as one sound is juxtaposed to another, each of the sounds undergoes a systematic change in its phonetic character. Consider **pu-ti**. Here, the end of the first sound is systematically modified by the immediate succession of the second, and likewise, the second sound is systematically modified by the immediate precedence of the first. After all, the vocal tract posture that accompanies one sound cannot instantaneously transform into the posture that accompanies another sound. Rather, the postures affect each other, and the acoustic signal follows suit (Öhman 1966). So, whereas until this time there had been a one-to-one correspondence between sound and meaning, now—instantly and irrevocably—this correspondence is sabotaged: there is now a many-to-one correspondence between sound and meaning (*allomorphy*), a situation found in all languages (Silverman 2006). Under the plausible assumption that compositionality is maintained at these early stages, now it is *two* sounds that correspond to one meaning: **pu-** when immediately followed by **-ti** is systematically phonetically distinct—though semantically non-distinct—from **pu** in isolation; **-ti** when immediately preceded by **pu-** is systematically phonetically distinct—though semantically non-distinct—from **ti** in isolation. The juxtaposition of one sound to another thus opens the floodgates to second-order symbolism.

Consequently, as these sound complexes are repeated and repeated in their appropriate real-world contexts, *new* sounds inevitably arise. This is certainly true of oral openings when they come to immediately precede oral closures, but for now, consider the oral closures themselves. While constant repetition of juxtaposed sounds in appropriate situations may serve to reinforce their *semantic constancy*, it is this very repetition that induces their *phonetic change* (Kruszewski [1883] 1995). For example, the medial closure in our **pu-ti** example may eventually undergo a process of voicing, becoming **pu-di**; voicing a mouth closure between mouth openings is a very natural phonetic development (Rothenberg 1968), one frequently encountered in the languages of the world. At this point, both **ti-** and **-di** correspond to a single meaning (remember, we are assuming compositionality): every time **ti** (with a voiceless closure) is immediately preceded by another sound, it is replaced by (alternates with) **di** (with a voiced closure). Again, this systematic change in sound does *not* expand the inventory of meanings, but it *does* expand the inventory of motor routines put in service to encoding this meaning.

But now, with a larger garrison of sounds to deploy, a huge expansion of the semantic inventory becomes possible, one that is able to meet the needs of our species' increasingly sophisticated cognitive and social structures. Motor routines and sounds that have heretofore corresponded to a single meaning may now unhinge themselves from their predictable contexts, to be cycled and recycled in ever-increasing and unpredictable ways: -**di**, for example, may now come to be associated with an additional meaning, and thus becomes free to appear as the first element of a complex, for example, **di-bu** (as opposed to a different complex, **ti-bu**). Note that the articulatory properties of these initial **di-**s are slightly distinct from -**di** (typically involving an expanded pharynx and lowered larynx during oral closure in order to maintain trans-glottal airflow, hence voicing), but nonetheless correspond to -**di** quite well in acoustic terms.

This sort of simple and natural sound change sets in motion a massive increase in the system's complexity. For example, newly-voiced medial closures may undergo further sound changes, to be harnessed for new meanings: when the **di** of **di-bu** is placed in second position (for example, **ka-di**), it is pronounced with closure voicing, comparable to the closure voicing that had earlier been added to **-ti** in this context (for example, earlier **bu-ti**, now **bu-di**). That is, two different meanings are now cued by the same sounds in comparable or even identical contexts. We may have **bu-di** in which **-di** means one thing, but also **bu-di** in which **-di** means something else. This establishes a one-to-many relationship between sound and meaning (*derived homophony*), a development also found in all languages (Silverman 2012).

If many sounds each came to symbolize more than one meaning, listener confusion and communicative failure may result. Such a scenario will not come to pass, however (Martinet 1952; Labov 1994; Silverman 2012). Defeating the pervasiveness of this potentially function-negative development, the **di-** of **di-bu** may passively undergo another change when placed in second position: some spontaneous productions of original **-di** that possess a slight weakening of their voiced closures may evolve towards a new value, perhaps, **-zi**, so we have **bu-di** (earlier **bu-ti**), and a different form, **bu-zi** (earlier **bu-di**; still earlier, **bu-ti**). Indeed, such sound patterns are likely to take hold exactly because of their function-positive consequences: creeping phonetic patterns that inhibit undue listener confusion are likely to be replicated and conventionalized. In short, successful speech propagates; failed speech falls by the wayside.

This means we now have **di-** alternating with **-zi**, both meaning one thing, and, recall, we have **-di** alternating with **ti-**, both meaning another. The co-evolution of these many-to-one relationships between sound and meaning results in many meaningful elements of the speech signal possessing both systematic phonetic variation and semantic stability, even across varied contexts. Now, in turn, this new sound **zi** may unhinge itself from its context and be deployed to signal new meanings.

Such speech patterns are found time and again in both (diachronic) sound changes and (synchronic) sound alternations (Gurevich 2004).

It is now clear that the mere juxtaposition of two simple sounds triggers remarkable growth and complexity of both the phonetic and the semantic inventories. Both one-to-many and many-to-one correspondences between sound and meaning naturally emerge. This is second-order symbolism.

13.3.2 More Examples, More Complexity

Recall that maintaining vocal fold vibration during an oral closure in utterance-initial position is aerodynamically unnatural, often involving an actively expanded pharynx and lowered larynx. Consequently, newly-evolved **bu-**, **di-**, and **ga-** might gradually lose this voicing, thus running the risk of sounding the same as **pu-**, **ti-**, and **ka-**. If this occurs, then those spontaneous productions of original **pu-**, **ti-**, and **ka-** that possess a slight delay in voicing may emerge as new and different sounds, **p^{h}u-**, **t^{h}i-**, and **k^{h}a-**, which now, again, may unhinge themselves and acquire new meanings, thus allowing them to appear in second position: **-p^{h}u**, **-t^{h}i**, and **-k^{h}a**. Alternatively, **bu-**, **di-**, and **ga-** may come to be accompanied by velic venting during their oral closures, thus again maintaining their phonetic distinctness, **mbu-**, **ndi-**, and **ŋga-**, which, as expected now, may unhinge and recombine as **-mbu**, **-ndi**, and **-ŋga**, thus opening the gates to phonotactic complexity, say, **ti-mbu**, **ti-ndi**, and **ti-ŋga**, and of course, creating more fodder for an expanding inventory of sounds with semiotic function. Another possibility is that the pitch-lowering effect that naturally accompanies voiced closure releases may, over time, migrate to pervade the opening, coming to replace closure voicing itself, and so becoming a pitch distinction that the language may now recycle: **bu-**, **di-**, and **ga-** as distinct from **pu-**, **ti-**, and **ka-** yields to **pù-**, **tì-**, and **kà-** as distinct from **pú-**, **tí-**, and **ká-**, respectively.

All of these hypothetical developments are not merely proposed characteristics of the nascent speech-based system. Rather, they are encountered over and over again in the history of language change. This is not a coincidence. Modern-day pressures on sound patterning are not merely characteristic of the modern-day grammatical system. Rather, they may have been in place long before the grammatical system came into existence, acting as a driving and inertial pressure on the very development of grammar itself. Natural systematic phonetic changes are not merely a *result* of grammatical complexity; they are a very *cause* of this complexity.

To summarize, the juxtaposition of two simple speech sounds may evolve to convey increasingly complex meanings. Such juxtapositions necessarily change the phonetic character of both sounds in systematic ways. These sound complexes may also be harnessed to encode new meanings, thus precipitating an explosive growth in the complexity of both the phonetic and the semantic inventories. The consequent sound complexes now achieve second-order symbolic status: both many-to-one and one-to-many sound-meaning correspondences come to be present in the speech code. Still, on rare occasions, certain of these complexes may result in semantic ambiguity, hence listener confusion and communicative failure.

13.3.3 Entrenching the Juxtaposition of Two Symbols, and the Rise of Post-Compositionality

Repeated usage of these compositionally transparent two-symbol structures not only induces the sorts of phonetic changes just considered, but may actually trigger the loss of compositionality itself, resulting in even more complex sounds with semiotic value. For example, compositional **pu-ti** possesses a meaning that is transparently built from **pu** and **ti**. But through its constant use and re-use, in addition to its phonetic changes, it may lose its link to its semantic origins, and thus become stranded as a semantic primitive (Kruszewski [1883] 1995), becoming "post-compositional." The now-opaque form (perhaps **puti**, **pudi**, **p^{h}uzi**, or **pútì**) becomes a single sound that correlates with a single meaning, thus embodying a counter-pressure back towards first-order symbolism, even as the system becomes increasingly phonetically complex.

This tug-of-war between first-order and a second-order symbolic states induces a lengthening of our meaning-impregnated sounds. Whereas earlier, the juxtaposition of one sound to another involved only two mouth-opening gestures (of increasingly varied forms), now such juxtapositions may involve three or four opening gestures, for example, **puti-ka**, **puti-kati**, etc.

We are moving closer to grammar.

13.4 Third-Order Symbolism in the Speech Code: The Ambiguous Affiliation of String-Medial Content, and the Triggering of Hierarchical Constituent Structure and Recursion

Semantic ambiguity of structural origin feeds a hierarchical constituent-structural analysis, which in turn feeds recursion. Let's consider how.

13.4.1 The Tug-of-War between First-Order and Second-Order Symbolism

There are now pressures *towards*, and pressures *against* the development of third-order symbolism. We first consider a passive *resistance to* the triggering of third-order symbolism.

We have been assuming that context-induced phonetic changes to sounds inevitably trigger their "unhinging," such that they may now be assigned additional meanings, and thus come to freely combine in new ways (recall, if **pu-ti** becomes **pu-di**, the new sound involving vocal fold vibration during the oral closure, **-di**, may now be assigned an additional meaning, thus freeing itself from the shackles of its context, allowing for **di-**). Still, if more and more sounds combine into wholly unconstrained sequences, a genuinely damaging ambiguity-of-meaning may result, in the form of an excess of one-to-many correspondences between sound and meaning. For example, the string **putika** may be ambiguous between compositional **pu-tika** and **puti-ka**.

Recall that successful speech propagates and conventionalizes; failed speech falls by the wayside. Speech sounds may thus be subject to a passive curtailment in their distribution such that certain sounds are only found in certain contexts. For example, perhaps our closure-voicing development comes to be limited to sound-medial position, and never takes place sound-initially, thus **pu-tiga** and **pudi-ka**. Closure voicing now acts to cue the compositionality of the forms. Every language passively evolves such patterns, which sometimes go by the name of "boundary signals" (Trubetzkoy [1939] 1969). In our example, voiced closures indicate the *absence* of a boundary; voiceless closures the *presence* of a boundary. Heterophony and clarity of meaning is thus maintained in a decidedly passive way, simply because those speech signals that are not semantically ambiguous are likely to be the very ones that are communicated successfully, hence imitated and conventionalized. Indeed, in many languages, the phonetic properties of word-initial oral obstructions are different from these properties in word-medial position, thus often serving this boundary-signaling function.

Still, even in the absence of these particular sorts of boundary signals, most languages have extremely reliable cues to boundaries in the form of prominence or *stress*. Let's return to **putika**. Even in the absence of medial closure voicing, clarity of compositional structure may be conveyed by prominence, thus **'pu-'tika** or **'puti-'ka**; one prominence per semantic primitive. These prominence distinctions serve to structurally—and, in most cases, semantically—disambiguate strings that might otherwise sound the same.

Reflecting its proposed origins as an aid in disambiguating these early two-sound structures, prominence (linguistic stress) typically involves a binary strong-weak or weak-strong rhythmic pattern at word edges, often iteratively applied in accommodation to the inevitably increased length of meaningful elements of the speech stream, that is, words and phrases (Hayes 1995). The role of prominence as a binary *phonetic* structure that originally cued a binary *semantic* structure thus persists, in remarkably comparable function and form, up to the present day.

In sum, the juxtaposition of a very small inventory of simple meaning-imbued sounds may inevitably lead to an explosion of phonetic and semantic complexity, rife with cues to structure and meaning, of the sort possessed by all languages. This complexity now sets the stage for full-blown grammar to emerge, as second-order symbolism yields to symbolism of the third order.

13.4.2 The Ambiguous Affiliation of String-Medial Content, and the Rise of Hierarchical Constituent Structure

Boundary signals are not ubiquitous; not in grammar, and almost certainly not in these early stages of pre-grammar. In the absence of such signals, a genuine counter-functional ambiguity will, on occasion, be present in the speech code. Indeed, it is the very ambiguity of some of our increasingly complex sound strings that establishes the conditions for third-order symbolism to arise. Consider our **putika** case again (assuming the absence of any boundary-signaling content). At these early stages, recall that at least two structures and meanings may be paired with this single phonetic string: **pu-tika** and **puti-ka**.

In most cases, real-world context will serve a disambiguating function, but once in a while, genuine ambiguity prompts a deeper structural analysis by listeners ("Is it **pu-tika** or **puti-ka**?"). The very moment listeners consider competing structures and their associated meanings, they are engaging in constituent analysis: the potential for hierarchically-structured strings suddenly becomes a reality.

The semantic ambiguity exemplified by **pu-tika** versus **puti-ka** is of another, *higher* order than what we have considered thus far; it is an ambiguity rooted in *structure*. Listeners' now-conditioned expectation of binarity, coupled with the string's semantic ambiguity, triggers its deeper, higher-order analysis. Structural ambiguity, then, opens the gateway to third-order symbolism, by requiring listeners to perform a deeper structural analysis of the sounds than had been heretofore required. The ambiguous affiliation of the middle term thus opens the gates to hierarchical structure.

Of course, these multiple interpretations of particular phonetic strings should be few and far between, since most strings possess (1) sound-sequencing cues, (2) meaning-sequencing cues, and (3) pragmatic cues to the intended structure and meaning of the string. Consequently, and most interestingly, it is exactly those rarely-encountered ambiguous forms that are most important for the development of the system toward third-order symbolic status. We turn to this issue now.

13.4.3 Hierarchical Constituent Structure, and the Rise of Recursion

Consider a longer string that is ambiguous, **putikakatipu**. This string might be intended by the speaker as, say, **putika-katipu**, and yet is open to a number of interpretations by the listener. For example, imagine the ambiguous affiliation of its middle content, **kaka**. As listeners impose binarity, both **putikaka-tipu** and **puti-kakatipu** may be interpreted, assuming each of these makes sense to the listener. So far, this is exactly the scenario just considered with respect to **putika**.

Clearly though, in comparison to **putika**, this longer string is impregnable, with many more structures and meanings. Consider **[[pu-ti]-kaka]-tipu**, or **puti-[kaka-[ti-pu]]**, or **[[puti-]ka]-[[kati]-pu]**, etc., some of which might be sensibly interpretable by listeners under the appropriate real-world conditions, even if the speaker intends a "flat" non-hierarchical binary structure. Again, it is listeners' expectation of binarity, coupled with the semantic ambiguity of the string, that triggers these strings' deeper structural analyses, analyses that quickly culminate in both hierarchical and now *recursive* structures, when embedding involves elements of the same type. Indeed, recursion is considered by some to be a primary characteristic of grammar (Hauser, Chomsky, and Fitch 2002).

In sum, the phonetic product of two juxtaposed sounds of increased length may lack semantic clarity, due to the ambiguous affiliation of its middle span. These ambiguous forms prompt deeper structural analyses on the part of listeners, culminating in both hierarchical and eventually recursive configurations. In short, semantic ambiguity of structural origin drives grammatical complexity.

All the major structural components of grammar have now emerged: a lexicon, a phonology, a morphology, a syntax, a semantics. "*All the rest is commentary. Now go study.*"

13.5 Discussion

When it comes to the origins of grammar, the search for evidence typically encompasses four domains:

1. Naturally occurring "sub-language" states in child learners, pidgins, innovated signed languages, and impeded speech (due to drunkenness, semi-consciousness, or pain, for example)
2. Ape-training studies
3. Laboratory experiments
4. Computer simulations

The present proposals exploit a fifth domain of inquiry, one of "internal reconstruction" (Saussure 1879) taken to its final frontier. Internal reconstruction is a method for investigating the origins of grammar inasmuch as observing the receding of distant galaxies is a method for investigating the "real" Big Bang: we observe extant pressures on structure and change, and extrapolate them to their logical origins.

Several advantages arise from this approach to the origin of grammar.

1. These proposals properly treat language as a "complex adaptive system" (Steels 2000; Beckner et al. 2009), one that is inherently social, involving both speakers and listeners; one that is inherently dynamic, involving competing pressures, and thus allowing for adaptive change; one whose structures are wholly emergent; one that affects—and is affected by—the co-evolutionary interactions of biological, cognitive, and social structures.

2. The present approach strictly adheres to the tenets of Uniformitarianism (Hutton 1795; Lyell 1830–1833). As noted, the proposed pressures and emergent structures by which the system originated remain in place to this very day. And while Uniformitarianism does not rule out the possibility of "punctuated equilibrium" (Eldredge and Gould 1972)—indeed, the proposed grammatical Big Bang embodies this phenomenon—still, saltation itself is fully absent: *natura non facit saltum*.

3. Speaker-based approaches to the evolution of grammar and grammatical change, as compared to listener-based approaches, are not equals-and-opposites: "production" is solely relevant at the level of the speaker (not the listener), whereas "perception" crucially relies on a role for both the speaker *and* the listener. That is, perception is inherently dependent on the interlocutionary event, whereas production is not. With its emphasis on the interlocutionary event itself, the present approach properly

situates the origins of grammar in the social world, a domain necessarily involving both producer and perceiver. Thus, unlike speaker-based approaches—which sometimes propose a single mutation in a single individual as the trigger of the grammatical Big Bang (for example, Bickerton 1990; Hauser, Chomsky, and Fitch 2002)—the present approach allows for a genotypic change in a *group* of individuals that may have been in place well before its phenotypic expression.

4. There need be no debate over whether grammar has its origins in a system of cognitive *organization* rather than a system of cognitive *expression* (Bickerton 1990). Rather, as an emergent consequence of sporadic semantic ambiguity in the sound signal, grammatical structures passively come into being due to a necessary interaction between speaker and listener, and most pertinently, these emergent structures necessarily affect both organization *and* expression: speakers' structurally ambiguous productions trigger listeners' higher-order structural analyses.

5. The current proposals take a decidedly holistic or Gestalt-based view of both language structure and language evolution. Indeed, it would be incorrect to characterize the present approach as one in which "phonology precedes syntax," or "syntax precedes phonology," or anything comparable. Rather, both phonetics and semantics—the only components of language that are empirically discernible by both language users and language analysts (Kiparsky 1973)—are inherently intertwined from the outset. So-called intermediate levels of grammatical structure—phonology and syntax—passively emerge from these two components' necessary interaction (and may, in fact, have no independent structural standing).

6. Some assert that our sound communication system has achieved its *final state* in the form of spoken language (Bickerton 1990; Mithen 1996; Hauser, Chomsky, and Fitch 2002). For example, Mithen proposes that language was triggered when supposedly distinct "modules of intelligence" (Fodor 1983) eventually coalesced into one, oddly likening this supposedly fully-culminated end-state of the human mind to a Christian house of worship. The present approach imposes no such upper limit on the evolution of the system. Indeed, perhaps the very same pressures that gave rise to the system, and that continue to shape and change it, also allow its evolution towards new, as-yet-unfathomed states of complexity. For example, in coordination with vocal tract, brain, and social changes, a slow decay of linearity (in the form of increased temporal overlap of morpho-syntactic content) may result in an increase in both parallel production (Mattingly 1981) and parallel processing (Rumelhart,

McClelland, and the PDP Research Group 1986); the present-day "sentence" might shorten to present-day "word" length, and in turn, these evolved "word-sentences" may be subject to an additional level of hierarchical and recursive arrangement. The semantic content of these higher-level (fourth-order?) structures—whatever they might turn out to be—may force a re-evaluation of the present-day system as one of "infinite expressivity" (Kirby 2007). Indeed, certain present-day languages already reverberate with the stirrings of such properties: witness the "polysynthetic" languages of North America, and the "stem-modifying" languages of Meso-America and East Africa.

7. The present approach to the origins of grammar incorporates *degeneracy* as an important component in its evolution: comparable forms may have distinct functions, and single functions may be underlain by multiple, different forms. Degeneracy may be a crucial element to the introduction of hierarchical complexity in any complex adaptive system (Whitacre 2010; see also Firth [1948] for an analysis in a specifically linguistic context). First employed to characterize genetic and biological systems (Edelman and Gally 2001), degeneracy may be characteristic of any system when categories are at once sufficiently robust to fulfill and maintain their function ("stability") and also sufficiently variable to be under constant modification ("evolvability"). Clearly, the presence of second order symbolism—with its one-to-many and many-to-one relations between form (sound) and function (meaning) paving the way to third-order symbolism (hierarchical and recursive structures)—is the analog of this trait in the evolution of the speech code: a "degenerative grammar."

13.6. Conclusion: The Grammatical Big Bang

It may or may not be relevant that the acquisition of grammar by children proceeds on a trajectory that reasonably hugs the levels of complexity proposed herein for the origins of grammar itself, just as it may or may not be relevant that implicational hierarchies concerning phonotactic complexity also fit rather snugly into these proposals.

Still, there is likely no evolutionary-biological privilege bestowed upon the primordial binary configuration that is characteristic of so many grammatical structures, just as there is no evolutionary-biological privilege bestowed upon the pentadactyl configuration among our planet's tetrapods. In both cases, there was merely a sensitivity to an initial complex of conditions that culminated in these features' prominent role in the evolution of species.

Regarding these initial conditions, again, the humble origins of the speech code may have consisted of extremely short, meaning-imbued sounds uttered in isolation that first accompanied, and then replaced a manual iconic communication system. These sounds' yielding to their juxtaposition in pairs may indeed have triggered a sort of grammatical Big Bang. Phonetic and semantic pressures came to interact in a way that inexorably, and perhaps rather suddenly, led to genuine grammatical complexity: the conditioned expectation of binarity, coupled with the sporadic semantic ambiguity of these increasingly long structures, prompted listeners to perform deeper analyses in order to extract their meaning, which in turn triggered the emergence of hierarchical and recursive grammatical structures.

Again, semantic ambiguity of structural origin drives grammatical complexity.

These primordial pressures and their yielded structures, in remarkably similar function and form, continue to constrain, shape, and change the speech code, even unto to this very day, and beyond.

13.7 Acknowledgments

Many thanks to James Winters, Paul Willis, Devin Casenhiser, and to Simon Kirby and members of the Edinburgh University Language Evolution and Computation Research Unit. Thanks especially to the editors of this volume, and the reviewers of this submission. All embarrassing errors are mine and mine alone. Happy birthday Ray, and thank you for being my teacher all those years ago!

References

Ay, Nihat, Jessica C. Flack, and David C. Krakauer. 2007. Robustness and complexity co-constructed in multi-modal signaling networks. *Philosophical Transactions of the Royal Society of London / B* 362 (1479): 441–447.

Beckner, Clay, Richard Blythe, Joan Bybee, Morton H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning*, 59 (s1): 1–26.

Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bladon, Anthony. 1986. Phonetics for hearers. In *Language for Hearers*, edited by Graham McGregor, 1–24. Oxford: Pergamon Press.

Delgutte, Bertrand. 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. In *The Representation of Speech in the Peripheral Auditory System*, edited by Rolf Carlson and Björn Granström, 131–150. Amsterdam: Elsevier Biomedical.

Edelman, Gerald M., and Joseph A. Gally. 2001. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America* 98 (24): 13763–13768.

Eldredge, Niles, and Stephen J. Gould. 1972. Punctuated equilibria: An alternative to phyletic gradualism. In *Models in Paleobiology*, edited by Thomas J. M. Schopf, 82–115. San Francisco: Freeman Cooper.

Firth, John R. 1948. Sounds and prosodies. *Transactions of the Philological Society* 47: 127–152.

Fodor, Jerry A. 1983. *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Gurevich, Naomi. 2004. *Lenition and Contrast: The Functional Consequences of Certain Phonetically Conditioned Sound Changes*. New York: Routledge.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298 (5598): 1569–1579.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hutton, James. 1795. *Theory of the Earth; with Proofs and Illustrations*. Edinburgh: Creech.

Jackendoff, Ray. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3 (7): 272–279.

Kiparsky, Paul. 1973. How abstract is phonology? In *Three Dimensions of Linguistic Theory*, edited by Osamu Fujimura, 5–56. Tokyo: The TEC Corporation.

Kirby, Simon. 2007. The evolution of language. In *Oxford Handbook of Evolutionary Psychology*, edited by Robin Ian MacDonald Dunbar and Louise Barrett, 669–681. Oxford: Oxford University Press.

Krakauer, David C., and Joshua B. Plotkin. 2004. Principles and parameters of molecular robustness. In *Robust Design: A Repertoire for Biology, Ecology and Engineering*, edited by Erica Jen, 115–133. Oxford: Oxford University Press.

Kruszewski, Mikołaj. [1883] 1995. Očerk Nauki O Jazyke (An Outline of Linguistic Science). Translated by Gregory M. Eramian. In *Writings in General Linguistics,* edited by Ernst Frideryk Konrad Koerner, 43–174. Amsterdam Classics in Linguistics 11. Amsterdam: John Benjamins.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Ladefoged, Peter, and Keith Johnson. 2011. *A Course in Phonetics.* 6th ed. Independence, KY: Wadsworth, Cengage Learning.

Lyell, Charles. 1830–1833. *Principles of Geology, Being an Attempt to Explain the Former Changes of the Earth's Surface, by Reference to Causes Now in Operation*. 3 vols. London: John Murray.

Martinet, André. 1952. Function, structure, and sound change. *Word* 8 (2): 1–32.

Mattingly, Ignatius G. 1981. Phonetic representations and speech synthesis by rule. In *The Cognitive Representation of Speech*, edited by Terry Myers, John Laver, and John Anderson, 415–419. Amsterdam: North-Holland Publishing Company.

Mithen, Steven J. 1996. *The Prehistory of the Mind: The Cognitive Origins of Art and Science.* London: Thames and Hudson.

Öhman, Sven. 1966. Coarticulation into VCV utterances: Spectrographic measurements. *Journal of the Acoustic Society of America* 39: 151–168.

Rothenberg, Martin. 1968. *The Breath-Stream Dynamics of Simple Released-Plosive Production*. Basel: S. Karger.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. 1. *Foundations*. Cambridge, MA: Bradford Books/MIT Press.

Saussure, Ferdinand de. 1879. *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig: Teubner.

Silverman, Daniel. 2006. *A Critical Introduction to Phonology: Of Sound, Mind, and Body*. London/New York: Continuum.

Silverman, Daniel. 2012. *Neutralization (Rhyme and Reason in Phonology)*. New York: Cambridge University Press.

Steels, Luc. 2000. Language as a complex adaptive system. In *Parallel Problem Solving from Nature. PPSN-VI*, edited by Marc Schoenauer, Kalyanmoy Deb, Günther Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel, 17–26. Lecture notes on computer science 3242. Heidelberg: Springer-Verlag.

Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.

Trubetzkoy, Nikolai S. [1939] 1969. *Principles of Phonology*. Berkeley: University of California Press.

Tyler, Richard S., Quentin Summerfield, Elizabeth J. Wood, and Mariano A. Fernandes. 1982. Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America* 72 (3): 740–752.

Whitacre, James M. 2010. Degeneracy: A link between evolvability, robustness and complexity in biological systems. *Theoretical Biology and Medical Modelling* 7 (6): 1–17.