The Evolution of Hierarchical Structure in Language[1]

J.C. BROWN & CHRIS GOLSTON
University of British Columbia & California State University, Fresno

1. Introduction

Pattee (1973) has argued that all problems of biology are ultimately problems of hierarchical organization. Much the same claim can be made for problems of language, where hierarchical organization is central to grammar. We propose that the scaffolding for hierarchical structure in human language is physiologically based and exapted from an internal mapping of the vocal tract. Following Cruse (2003), we assume that the reorganization of a strictly reactive system into a cognitive system (which can characterize language evolution) often requires an internal mapping of the system body. Thus, an internal map of the vocal tract was created to fine-tune motor control of articulators like the lips, tongue and larynx; the hierarchical structures in that map were then exapted elsewhere in grammar.

It has been argued that much of syntax and higher order grammatical structure was exapted from the structure of the syllable (Carstairs-McCarthy 1999). This is a desirable approach since it relates various parts of human language through a shared structure, but it leaves unanswered *where the syllable itself evolved from*. We propose that there are two crucial parts to the syllable, the embedding and the headedness, and that each had a different evolutionary source.

2. Embedding

In this section we will try to show that embedded structures (treelets) arise naturally from internal maps of the vocal tract and what one can profitably do with it. Not all parts of the vocal tract are well modeled with a treelet, but enough of them are to make treelets a good way of representing much of the speech apparatus and its output.

Embedded trees are ubiquitous in grammar and give it its hierarchical structure. We suggest that such treelets were exapted from articulation into more purely grammatical spheres to lend coherence to the messages the sound system was being used to communicate. We're interested here in showing just how

similar many of these trees are, specifically with how distinctions tend to embed in a similar way, with two binary branchings defining a three-way split. We begin here with a map of the vocal tract and how it is used in speech and note that it often involves bifurcations into two categories (eg, [lip tongue]) with a secondary distinction involving only one of the first two categories (eg, tongue = [crown dorsum]). Such dichotomies in phonetic and phonological distinctions are much more common than ternary distinctions with no sub-grouping, or quaternary distinctions with elaborations on both sides of the initial split.

Most of the distinctions we'll encounter here are paradigmatic, different optionals (like labial—coronal—dorsal) that one can take for a given parameter (like place of articulation). The little trees we'll now look at do not generally define syntagmatic, linear relations in language. These will first be countenanced when we look at how sounds are arranged into syllables. Thus we will propose that both the paradigmatic and the syntagmatic aspects of language (Saussure 1916) have phonetic and phonological precursors, specifically consonants and vowels (paradigms) and syllables and feet (syntagms). For now, let us see how more basic phonetic and phonological distinctions break down.

We propose that embedding emerged from an internal mapping of the vocal tract as follows. Long before humans split from other mammals, we would have produced sound with a laryngeal source and a supralaryngeal filter (Fant 1960), just as birds produce sound with their syrinx and a suprasyringeal filter:
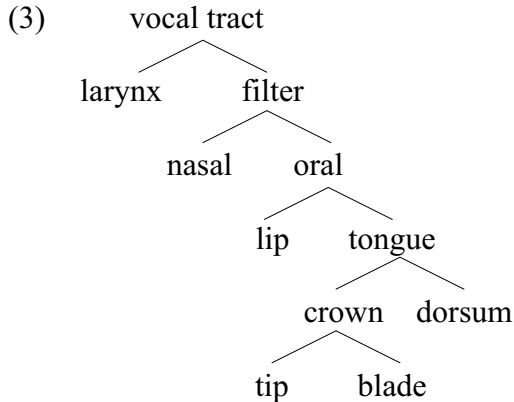
(1)　　　vocal tract
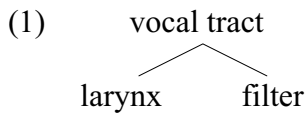
　　　larynx　　filter

As the larynx descended during human evolution, the supralaryngeal filter bifurcated into the nasal and oral cavities. As humans gained control over the nasopharyngeal port, the filter could produce both nasal *and* oral sounds, the latter being much more readily perceived because of their clearer acoustic signatures (Lieberman 1984):

(2)　　　vocal tract

　　　larynx　　filter

　　　　　nasal　　oral
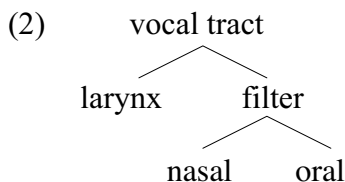
Not a lot could be done with the nasal cavity, but the oral cavity could be molded by means of two fairly mobile articulators, the bottom lip and the tongue. The tongue's crown and dorsum may be moved independently of one another, so that the tongue is itself treated as two relatively independent articulators, the crown (as much as you can grab comfortably) and the dorsum (the rest). The crown is further divided into the tip and blade, which can be used to close off the vocal
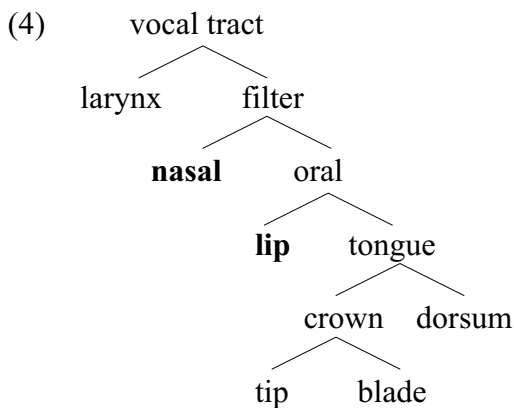
tract with a relatively narrow (tip) or relatively broad (blade) constriction against the teeth or palate. If the internal map of the vocal tract was ramified further to reflect these developments, the map would consist of a large number of embedded treelets, as follows:

(3) vocal tract
```
        vocal tract
         /      \
     larynx    filter
               /     \
           nasal     oral
                    /    \
                  lip    tongue
                        /      \
                    crown     dorsum
                    /    \
                  tip    blade
```

This internal map of the vocal tract strikes us as the most likely source for the notion of embedding in natural language.

Much of our vocal tract is similar to that of other primates, but the ability to produce and perceive the place distinctions above is limited to humans. Without the two resonating cavities a lowered larynx provides, there is no way of identifying the changes in the first and second formants that signal place of articulation acoustically. At some point in the evolution of our species, this basic physiological configuration was co-opted into the service of meaningful place distinctions in words like (labial) *pea*, (coronal) *tea*, (dorsal) *key*. Such distinctions are purely paradimatic and map directly onto the articulators used to produce them, creating a close link between meaning and the vocal tract. Thus we can characterize a sound like [m] as follows, with **nasal** and **lip** bolded

(4) vocal tract
```
        vocal tract
         /      \
     larynx    filter
               /     \
          nasal     oral
                    /    \
                  lip    tongue
                        /      \
                    crown     dorsum
                    /    \
                  tip    blade
```

because it is made with nasal airflow and constriction involving the bottom lip.

The tree in (4) is both a map of the vocal tract and a simple model of the

articulators involved in producing speech. The tree actually defines a paradigmatic space in which a number of distinct sounds (m, n, N, p, t, k) are differentiated and that paradigmatic space has a one-to-one relation to the actual vocal tract. This, we think, is how embedding crept into language. The vocal tract must be changed simultaneously along several dimensions to effectively produce a sound like [m] or [k]. And the dimensions along which the sound varies (nasal, labial, etc.) are actually linked to meaningful distinctions in the message that is conveyed, so that *mat, bat, kat* mean different things. The internal map of the vocal tract becomes a model of articulation and a source of meaningful distinctions.

Once embedded structures were used to model which articulators are involved in a speech sound, the road should have been opened to using such structures for different purposes. We look at two such areas here, involving the larynx and degrees of vocal tract constriction.

Over the course of time control over the larynx grew to allow for six-way stop contrasts: plain, voiced, aspirated, glottalized, implosive, and voiced aspirate. Feature-geometric views of laryngeals (Lombardi 1991; Iverson & Salmons 1995; Kehrein 2001) represent what it can do as follows:

(5)   larynx

       voice   glottis

   spread   constricted
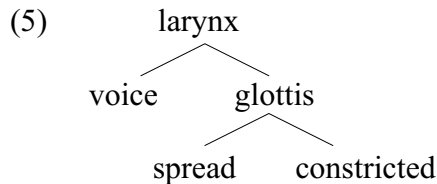
This treelet is *not* a map of the larynx and is purely paradigmatic; indeed, it shows types of laryngeal features that cannot all be distinctively ordered within the same speech sound. We propose that it is a *functional* map of the larynx that shares the same double-branching structure found in the *physiological* map of the vocal tract. Whereas (3) is both a map of the vocal tract and a model of what you can do with it, (4) is *just* a model. Its structure, we suggest, came from exapting the structure of (3) into a new domain, structuring laryngeal contrasts in terms of nested distinctions.
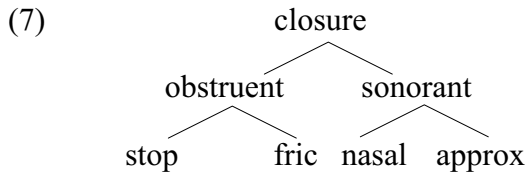
Similarly for the degree of closure in a given sound. Articulatorily there are three useful degrees of closure, which we'll call stop, fricative and resonant articulation, following Laver (1994). These notions cannot be mapped onto the vocal tract in the same way as nasal and labial can because they encode an entirely different dimension[2]. But they can still be usefully mapped with a branching tree, where the major division is between obstruents and sonorants:

---

[2] We thank Bruce Hayes for pointing out that this was a major problem with the feature-geometry of the 1980s, e.g., Clements (1985), Sagey (1986), McCarthy (1988): it was never able to satisfactorily deal with stricture issues.
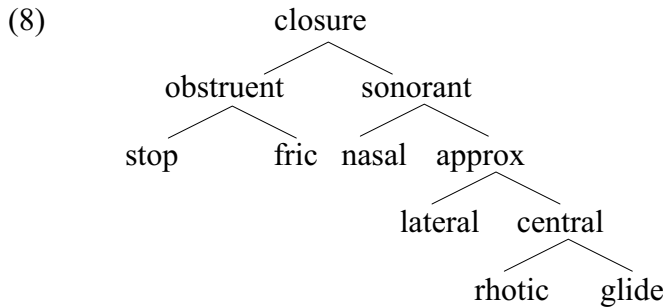
(6)
```
                   closure
                  /       \
            obstruent    sonorant
            /      \
         stop    fricative
```
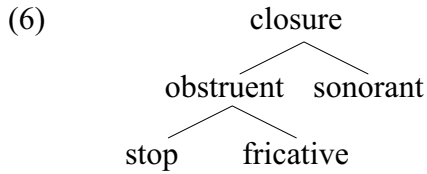
This articulatory difference also maps onto an acoustic difference, since stops and fricatives have less amplitude than sonorants. The latter group breaks down acoustically into nasals and approximants:

(7)
```
                      closure
                    /         \
            obstruent         sonorant
            /      \          /      \
         stop     fric    nasal   approx
```
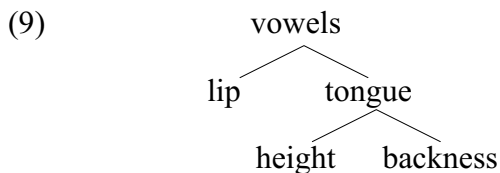
Approximants come in three types, the major distinction involving lateral articulation (along one side of the mouth) vs. central articulation (along the midline). Central approximants can be divided into r-sounds (rhotics) and semi-vowels (glides):

(8)
```
                       closure
                     /         \
             obstruent         sonorant
             /      \          /      \
          stop     fric    nasal   approx
                                  /      \
                            lateral    central
                                       /      \
                                   rhotic     glide
```

This tree does not map the vocal tract, but does provide a paradigmatic representation of stricture and sonority. Our proposal that the classes of sound that this tree maps out (sonorants, approximants, etc.) has the same hierarchical structure as the internal map of the vocal tract we began with. In this way, we think, humans exapted hierarchical structure from a physical to a cognitive domain, from articulation to abstract acoustics.

   Vowels provide another case of an acoustic mapping based on hierarchical structure. Articulatorily vowels are shaped by the lips and tongue:

(9)
```
                   vowels
                  /      \
               lip      tongue
                        /     \
                   height    backness
```

Height and backness are themselves broken down three ways into hi/mid/low (mapping onto the first formant) and front/central/back (mapping onto the second).

So much for embedding. We've seen that a number of important linguistic distinctions seem to be organized in terms of a little tree. All that is required is a basic distinction, one part of which is elaborated. This we take to have been the first stage in the development of the structural part of modern grammar. It involves nothing more than building on mental maps of the physiology and acoustic output of the vocal tract. From these phonetic and phonological relationships, the general treelet is exapted out into higher-order modules of language such as syntax and semantics.

3. Headedness

We propose here that headedness emerged from the acoustic effects of mandibular oscillation (MacNeilage 1998), which makes for periodic fluctuation in amplitude. The peak of this loudness (maximal mandibular aperture) and its less salient edges (minimal mandibular aperture) are distinct types of entity such that the identification of the edges is dependent on the perception of the peak, but not vice versa ('consonants are just ways of beginning and ending vowels'). In any case, the most important part of a syllable is the loudest part, the sonority peak:

(10)   *syllable*

     onset   *rhyme*

          peak   *coda*
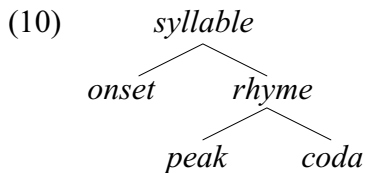
This central place for the notion peak, we propose, gave rise to the notion of head. The head of a syllable is the sonority peak it contains. This gave rise to the first *headed* embedded structures, exapted later to morphology and syntax.

Not all linguistic structures are as simple as the ones we've looked at so far. Specifically, none of the little trees we've examined treats any of its daughters differently: lip, crown and dorsum are equally parts of the mouth and none is the head part of the mouth in any sense. Similarly, stops, fricatives and sonorants are all manners of articulation, none of which is the head manner of articulation. But this is not the case with higher order grammatical constituents like syllables and sentences, which do have special (usually endocentric) units called 'heads'.

There must have been a stage in the evolution of language at which headed structures arose, and this will be the topic of the present section. We want to probe what was involved in treating one of the daughters as the *head* of a given category. The most basic structure we can find in language that has a head is the syllable (cf. Carstairs-McCarthy 1999). In developing an evolutionary scenario for language we are left with four possibilities: (i) syntactic heads predate syllabic
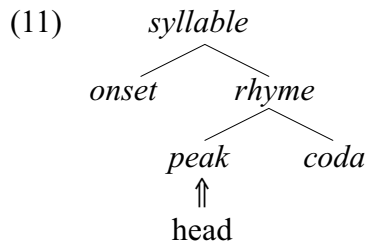
heads, (ii) syllabic heads predate syntactic heads, (iii) headedness came from elsewhere, or (iv) headedness evolved more than once. Since sentences necessarily imply syllables (in spoken language) but syllables do not necessarily imply sentences, we take it that (i) is unlikely. (iii) is a possibility, but we cannot think of any plausible analogue to grammatical headedness outside of language. (iv) is unparsimonious and unlikely, leaving us with (ii): syllables came before sentences and gave them their headedness.

So let us start with the syllable and see how it may have arisen and how the notion of head would have arisen with it. Our discussion owes much to the work of Carstairs-McCarthy, but we will try and move beyond where he has taken us.

All mammals are obligate nose-breathers except for (adult) humans, who are preferential nose-breathers. You are right now most probably breathing through your nose. This is crucial in understanding syllables for the following reason: if you now have your mouth closed (as you most likely do), you will need to open it in order to produce a noise loud enough to be heard clearly by others. The sequence is then almost always CLOSE → OPEN if you start by breathing through your nose with your mouth closed. A closed mouth speech sound is a stop (once it gets released) and a fully open mouth sound is a vowel. Hence we can go directly from preferential nose-breathing to CV syllables. We could only get VC syllables as basic if humans preferentially breathed through their mouths or preferentially breathed through their noses with their mouths wide open. But we don't.
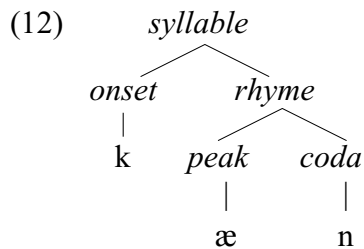
We might also link the rise of CV syllables to mandibular oscillation, following MacNeilage (1998), though we don't see how oscillation in itself can account for the CV nature of syllables. In any case, it makes good evolutionary sense to try and derive syllables from something non-linguistic (chewing, preferential nose-breathing) and then to graft other grammatical structures onto syllables (Carstairs-McCarthy 1998, 1999, 2000). An exaptationist model like this tries to derive grammatical constructs from physiology and acoustics.

All languages have CV syllables and all children begin babbling with CV syllables; it is most reasonable to think that the earliest speaking hominids had CV syllables as well (especially as they breathed through their noses most of the time). So what about headedness? There is no articulatory head to a syllable, but there is an *acoustic* head: the acoustically most salient (loudest) sound in a syllable, or *sonority peak* (Fudge 1969; Selkirk 1982; Clements 1990). The sonority peak is thus the head of the syllable because it is more prominent than the beginning. Bringing in closed syllables (CVC) does not change the picture: C articulations always involve constriction of the vocal tract relative to the V, so CVC syllables will be headed by the V just as CV syllables will be. Most languages don't allow more than CVC syllables, words like *plant* and *tramp* being quite rare cross-linguistically. So we have a basic division of C and what follows (the onset and the rhyme); and in some languages what follows the onset can be complex, either VV or VC. Syllables thus have the following endocentric structure:

(11)　　　　*syllable*

```
                 syllable
                /        \
            onset        rhyme
                        /     \
                     peak     coda
                      ⇑
                     head
```

This structure is straightforwardly syntagmatic, defining the linear order of the pieces of a syllable, not (merely) a typology of positions: onsets always precede nuclei which always precede codas within a syllable. In terms of Gazdar et al. (1985), the tree specifies *linear precedence* relations (left to right) and *immediate dominance* relations (top to bottom). This seems to be the best phonological precursor for things like word and phrase structure, both of which are inherently syntagmatic. The edges of syllables contain consonants, while the head is usually a vowel (though other sonority peaks will often do as well).

The beauty of syllables is of course that they hierarchically organize sounds, which themselves involve hierarchical organization, as we have seen:

(12)　　　　*syllable*

```
                 syllable
                /        \
            onset        rhyme
              |          /     \
              k       peak     coda
                       |         |
                       æ         n
```

As Carstairs-McCarthy and others have demonstrated the head (usually a vowel) is flanked by elements that are more like one another than they are like the head: the nucleus is typically a vowel, but the onset and coda are consonants. In short, we attribute the notion of *headedness* to syllables, which have acoustically motivated heads and which most likely predate complex words. But headedness is all that syllables need to add: the little tree predates it, at least on the assumption that sounds predate syllables. (For the idea that syllables came first, see Studdert-Kennedy 1998.)

We hypothesize that the *headed* treelet, although it originates within the phonology, was exapted from there into higher linguistic structures in much the same way that embedded trees were exapted into semantics (Brown & Golston 2002). Morphology would have been the first area to be affected, assuming that words predate sentences in phylogeny as they do in ontogeny. Roots are the heads and form stems with derivational affixes; stems form words with inflectional affixes. In the syntactic realm we have heads and complements forming X' constituents and their specifiers finishing off the phrases as XPs.

4. Conclusion

The evolutionary scenario we propose is thus: embedding comes from an internal map of the vocal tract and is later exapted to less physiological areas like sonority; headedness is added from acoustics and is later exapted to less physiological areas like morphology and syntax. Our proposal is essentially an internal reconstruction of grammar starting at morphology and syntax and working back to syllables and the sounds they organize, and ultimately to the vocal tract that produces these sounds.

References

Brown, J.C. and Chris Golston. 2002. Generalized X-bar theory and the evolution of grammar. Paper presented at the 4th Evolution of Language Conference, Harvard University.

Carstairs-McCarthy, Andrew. 1998. Synonymy avoidance, phonology, and the origin of syntax. In Hurford et al. (1998), 279-96.

Carstairs-McCarthy, Andrew. 1999. *The origins of complex language: An inquiry into the evolutionary beginnings of sentences, syllables, and truth*. New York: Oxford.

Carstairs-McCarthy, Andrew. 2000. The distinction between sentences and noun phrases: An impediment to language evolution? In C. Knight, M. Studdert-Kennedy and J.R. Hurford (eds.) *The evolutionary emergence of language: Social function and the origins of linguistic form*, 248-63. Cambridge: Cambridge University Press.

Clements, G.N. 1985. The geometry of phonological features. *Phonology Yearbook* 2:225-252.

Clements, G.N. 1990. The role of the sonority cycle in core syllabification. In M.E. Beckman & J. Kingston (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283-333. Cambridge: Cambridge University Press.

Cruse, Holk. 2003. The evolution of cognition – a hypothesis. *Cognitive Science* 27:135-155.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Fudge, E.C. 1969. Syllables. *Journal of Linguistics* 5:253-286.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag. 1985. *Generalized phrase structure grammar*. Cambridge (MA): Harvard University Press.

Goldsmith, John. 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Hurford, James R., Michael Studdert-Kennedy and Chris Knight (eds.). 1998. *Approaches to the evolution of language*. Cambridge: Cambridge University Press.

Iverson, Gregory K. and Joseph C. Salmons. 1995. Aspiration and laryngeal

J.C. Brown & Chris Golston

representation in Germanic. *Phonology* 12:369-396.

Kehrein, Wolfgang. 2001. Phonological representation and phonetic phasing: affricates and laryngeals. Doctoral dissertation, Philipps Universität Marburg.

Laver, John. 1994. *Principles of phonetics.* Cambridge: Cambridge University Press.

Lieberman, Philip. 1984. *The biology and evolution of language*. Cambridge: Harvard University Press.

Lombardi, Linda. 1995. Laryngeal neutralization and syllable wellformedness. *Natural Language and Linguistic Theory* 13:39-74.

MacNeilage, Peter F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21:499-546.

McCarthy, John J. 1998. Feature geometry and dependency: a review. *Phonetica* 43:84-108.

Pattee, Howard H. 1973. *Hierarchy theory: The challenge of complex systems*. New York: George Braziller.

Sagey, Elizabeth.1986. *The representation of features and relations in non-linear phonology*. Doctoral dissertation, MIT. Cambridge, MA.

Saussure, Ferdinand de. 1916/1960. *Cours de linguistique générale,* ed. C. Bally, A. Sechehaye and A. Riedlinger. Paris: Payot. (*Course in general linguistics,* ed. C. Bally, A Sechehaye and A. Riedlinger, trans. W. Baskin. New York: McGraw-Hill.)

Selkirk, Elisabeth O. 1982. The syllable. In H. v.d. Hulst & N. Smith (eds.), *The structure of phonological representations*, vol. 2. Dordrecht: Foris Publications.

Studdert-Kennedy, Michael. 1998. The particulate origins of language generativity: from syllable to gesture. In Hurford et al. (1998), 202-221.

Department of Linguistics
1866 Main Mall, Buchanan E270
University of British Columbia
Vancouver, British Columbia V6T 1Z1
Canada

Department of Linguistics
California State University, Fresno
Peters Business Building, Room 383
5245 North Backer Avenue
Fresno, CA 93740-8001

jcb@interchange.ubc.ca (Brown)
chrisg@csufresno.edu (Golston)