Variables in Optimality Theory

Chris Golston
California State University Fresno

Optimality Theory (Prince & Smolensky 1993) is a theory of constraint interaction. At its core it is quite simple:

> Universal Grammar provides a set of highly general constraints. These often conflicting constraints are all operative in individual languages. Languages differ primarily in how they resolve the conflicts: in the way they rank these universal constraints in strict domination hierarchies that determine the circumstances under which constraints are violated. A language-particular grammar ***is*** a means of resolving the conflicts among universal constraints. (Prince & Smolensky 1993:3)

All sorts of other ideas can be associated with OT, but we must distinguish those that *follow from* the theory and those that can be *added to* the theory. The theory proper consists of "EVAL-mediated comparisons of candidates by a hierarchy of violable constraints. No matter how the details are executed or in what overall context it is embedded, any model with these indispensable characteristics will express the central claim and insight of OT" (McCarthy 2002:11).

An important add-on to the theory since its inception has been the claim that the set of candidates evaluated by the hierarchy of violable constraints is both universal and infinite. This paper argues that neither property follows from any principle of OT and that theory-neutral considerations put a different set of restrictions on what GEN most likely generates. (§1). I'll also show that the universality and infinity of GEN are needed mostly to prop up an untenable theory of underlying representation carelessly carried over from work in the 1960s that eschewed prosodic structure enitrely (Chomsky & Halle 1968). Once we have a better theory of underlying representation that includes a great deal of underlying prosody, the work that GEN must do changes drastically (§2) and the candidate set of surface representations turns out to be finite, indeed quite small (§3).

1. GEN

GEN is what the grammar adds to stored or underlying representations (URs) to make phonetically detailed programs for speech production or surface representations (SRs). Roughly, GEN is what is stripped from SR to make it storable or, what amounts to the same thing, what is added to UR to make it sayable. GEN is SR minus UR. To understand GEN, we therefore need a good understanding of both underyling and surface levels of representation. The more alike they are, the less GEN has to do. The less alike they are, the more it has to do.

Because GEN has no ability to output just the correct form of a given input, the ranked and violable constraints of OT require a set of candidates for EVAL to select from:

(1) An OT grammar

/kætz/	→	GEN	→	[kætz]	→	EVAL	→	[kætz]
				[kæts]			☞	[kæts]
				[kætəz]				[kætəz]
				etc.				*etc.*

That much follows from the basic tenets of OT. Two additional claims about GEN, however, are theoretical add-ons that require additional support because they do not follow from having ranked and violable constraints. These add-ons are that the set of candidates is *universal* and that the set of candidates is *infinite.*

We'll take these issues one at a time, beginning with the universality of GEN:

> GEN is universal, meaning that the candidate forms emitted by GEN for a given input are the same in every language. These candidates are also very diverse. This property of GEN has been called *inclusivity* or *freedom of analysis*. Precisely because GEN is universal, it must at a minimum supply candidates varied enough to fit all of the ways in which languages can differ. (McCarthy 2002:8)

The universality of GEN is clearly stipulative. No attempt is made to derive the universality from the theory of OT, presumably because there is no logical connection between the claim that constraints are ranked and violable (OT) and the claim that the candidate set is the same for all languages (universality). Nor is any empirical support offered for it, presumably because there isn't any. I'll argue below that abandoning universality for GEN has no negative consequences for the theory of grammar.

The *infinity* of GEN, on the other hand, is not stipulated, but the arguments for it are fatally flawed. Consider McCarthy's discussion in full:

> If GEN incorporates any recursive or iterative operations, as it surely

> must, then there is no bound on the size of a candidate and every candidate set, from every input, is infinite. This is perhaps not too surprising in syntax, where the infinity of sentences has long been accepted, but it is also true in phonology. Epenthesis is an iterative procedure of candidate-generation, so the set of candidates derived from input /ba/ must include /bati, batiti, batititi.../. No GEN-imposed bound on the number of epenthesis operations is appropriate. Rather, the economy of epenthesis should and does follow from constraint interaction. (McCarthy 2002:9)

Note first that the case for iterative epenthesis is purely hypothetical: no language epenthesizes more than a foot in any given spot. Thus the parallel with syntax, where recursive structure really is commonplace, is not compelling. Some types of OT already do without an infinite GEN, including what McCarthy calls *persistent OT* (this volume).

So let us narrow down what GEN really has to do. To make things concrete (and thus psychologically plausible) I'll assert what I'm sure *my* GEN does not generate; the reader is invited to join along with her own GEN to compare, but the conclusions will be pretty similar.

First consider actual physical objects like motorcycles and whales. Since I cannot produce these things with my vocal tract, there is no reason to think that GEN would ever generate them. One could of course devise a grammar to generate them and then filter them out with constraints, but this would serve little purpose, as Reiss has pointed out with respect to a hypothetical constraint NOBANANA:

> Note that the claim intended by the constraint NOBANANA, that no representation of a sentence contains bananas, is probably true for all human languages. However, there are an infinite number of true claims of this type. No language requires speakers to dance a jig to express iterativity, no language has pizza as an element of syntactic trees; *etc*... We do not want our model of grammar to express every true statement about what structures do not occur, since there are an infinite number of such statements and the grammar must be statable in finite terms if it is to be instantiated in human brains. (Reiss 2002:2)

As Bruce Morén points out (p.c.) one can readily dismiss such options if one just assumes that the grammar manipulates *linguistic elements*. That will get rid of a lot of garbage in the grammar and has no bad effects on OT as an explanatory theory of how language works. And so we may postulate a reasonable restriction on GEN, which nobody should object to:

(2) Restriction 1: My GEN only generates linguistic elements.

My own GEN has a lot of additional restrictions that have been there since I reached the age of 10 months or so and lost the ability to discriminate speech sounds and structures that are not part of my native language, English, or the one other language I speak with some degree of fluency, German. I cannot reliably perceive or produce things like [ɧœɭɒʊɘ], [ɘʓɵçɤʎ], or [ʘɰǁɥʁɴʄ]. My grammar no longer has the capacity to deal with rising-falling tones, linguolabials, voiceless implosives, ejective fricatives, voiceless nasals, or pulmonic ingressives, though it once did. Since I'm not (presently) modeling my grammar at the age of 10 months, I can reasonably add

(3) Restriction 2: My GEN only generates English and German.

We commonly do construct English OT grammars with constraints against things that no English (or German) speaker would produce or perceive, but these constraints are as pointless as the NOBANANA constraint given the actual abilities of a given adult speaker.

This is not to say that my present grammar isn't plastic to some degree or that I couldn't learn to produce and perceive speech sounds outside of the languages I currently know. But at any given point in time my grammar is pretty much blind to things that aren't found in the languages I speak. Generally, when listeners are presented with sounds or sequences of sounds that are illegal in their native language(s), they tend to assimilate them to form sounds or sequences that *are* legal (Massaro & Cohen 1983; Hallé et al. 1998; Dupoux et al. 1999; Dupoux et al. 2001).

Recall that the universality of GEN is not a necessary part of OT and that no good arguments have been given for it; what serious reason is there to think that my grammar generates and evaluates things I can neither say nor perceive? A useful parallel is the DNA in cells. The grammar of a neonate is like a stem cell that has the ability to become a hair or bone or brain cell but isn't one yet. But an *adult* grammar like the one in your head or mine is like a hair or bone or brain cell and no longer has the ability to become any other type of cell. It no longer has the plasticity of a stem cell and shouldn't be described like one that does. (Whether an infinite GEN is necessary for language acquisition I leave to those who know more about it.)

Returning to the specific limitations of my own GEN, I am quite sure that it does not normally produce things that I don't intend to talk about either: if I'm trying to say *cats* it just doesn't worry about [dɔgz], [ˌsɛ.vn.ti.ˈfɑɪv], or [ˌæv.ɹɪl.lə.ˈvin]. In the event that I do say something along these lines it is a speech error involving lexical access, not the product of my phonology.

(4) Restriction 3: My GEN only generates my message.

If I end up saying *white Anglo-Saxon prostitute* instead of *white Anglo-Saxon Protestant*, this is not because GEN included it as a candidate, but because GEN was misfed something by earlier parts of speech production. The restriction I'm proposing is related, of course, to the classical OT notion of Containment (Prince & Smolensky 1993), which requires that the output include the input. I'm proposing slightly more, that the output *only* include the input. This will be interesting when it comes to things like epenthesis, where I'll need to introduce the notion of variable, familiar enough from basic mathematics.

Finally, there's no reason to think that GEN ever considers morphological monsters like [kæ-ts] or [k-æts] when I'm trying to produce *cat-s*. GEN doesn't change basic sound-meaning relationships in this way. This is enshrined already in OT as Consistency of Exponence. See Van Oostendorp (this volume) for more reasons to think that this is a good restriction to keep:

(5) Restriction 4: My GEN doesn't change morphological affiliation.

We've seen that there are no theoretical or empirical reasons to think that GEN is universal or infinite and that there are perfectly reasonable, theory-neutral restrictions we may put on GEN, including Restrictions 1-4 above. Before we can calculate what GEN actually produces, however, we have to have a clearer understanding of what GEN operates on. And so we turn to underlying representations, the input to GEN.

2. Underlying representations

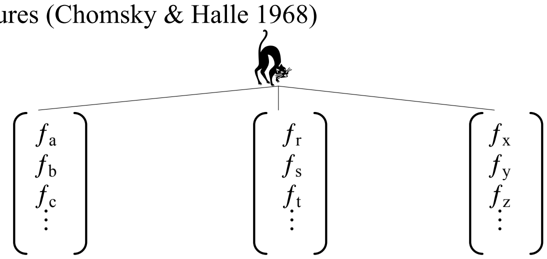
OT offers no guidelines or insights on either underlying or surface representations: "The core assumptions of OT are pretty general, and so they are compatible with a wide range of representational assumptions." (McCarthy 2002: 243). So we must look elsewhere if we want to know how people actually store things and how those things look once the phonology has processed them.

Traditional representation is segmental, so that representations consist of a string of phonemes. For Saussure the phonème "designates what we would today call phonetic segments, considered as (*ultimately unreducible*) units in acts of speaking" (Anderson 1985:38, my emphasis).

(6) Phonemes (Saussure 1916)

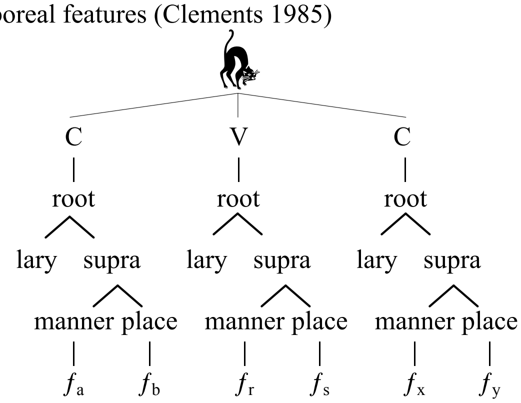
Trubetzkoy (1939) brings out the importance of distinctive features in phonological representation, but it is Jakobson (1939) who first makes them central to phonological description, so that features become the atoms of representation in subsequent theory (Jakobson, Fant & Halle 1952; Chomsky & Halle 1968):

(7) Bundled features (Chomsky & Halle 1968)



Clements argues that “if we find that certain sets of features consistently behave as a unit with respect to certain types of rules of assimilation or resequencing, we have good reason to suppose that they constitute a unit in phonological representation” and proposes that distinctive features are organized in trees (1985: 226):

(8) Arboreal features (Clements 1985)



(Note: *lary* = laryngeal node, *supra* = supralaryngeal node.)

Browman & Goldstein (1986, 1990) represent things as gestures organized like musical scores that encode timing relations:

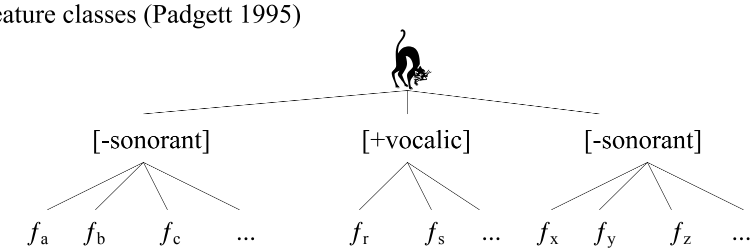
(9) Gestural scores (Browman & Goldstein 1986, 1990)

Velic		.	.	.	.	.	.
Oral	TongueBody	□	.	a	.	.	.
	TongueTip	.	.	.	.	□	.
	Lips	.	.	.	.	.	.
Glottal		.	□	.	.	.	□

There is no hierarchical structure here at all, but later versions of Articulatory Phonology add it to varying, if conservative, degrees (Browman & Goldstein 1989, 1990, 2000; Byrd 1996; Byrd & Saltzmann 2003).

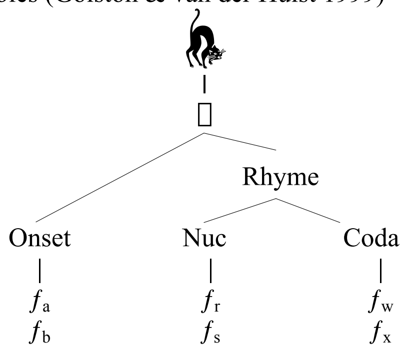
Padgett (1995) brings us back to a Chomsky & Halle type of representation in which features are just bundled, though the bundles are now rows instead of columns:

(10) Feature classes (Padgett 1995)



Golston & van der Hulst (1999) propose that underlying representations are organized into syllables rather than segments, based in part on psycholinguistic evidence reviewed below:

(11) Syllables (Golston & van der Hulst 1999)



Standard OT is of two minds with respect to representation. Surface representations are usually highly prosodified, with onsets, (no) codas, syllables, feet, and so on playing a very prominent role. Underlying representations, on the other hand, are almost always just strings of segments. No reason is given for this in the literature, and it is generally simply assumed that non-contrastive prosody is added by the grammar and not part of underlying representation. In practice, then, URs are strings of phonemes and the grammar adds structure like syllables and feet. But there are good reasons to think that this is not the case, including facts about long term and working memory.

2.1 Working memory and UR

Underlying representations must be stored in long term memory somehow and they must get into long term memory through working memory, because there's no other way in. Memory limitations put *real* restrictions on what kind of symbolic information can be memorized and stored and accessed; theories of grammar that do not respect these restrictions are therefore poor models of the human language faculty.

Evidence for how words get into long term memory comes from limits on short term (working) memory, which holds surprisingly few chunks of information (Miller 1956). Recent research suggests that working memory is different for spatial information than it is for verbal/numerical information (Baddeley 1986; Oberauer et al. 2003). Limiting ourselves to the verbal/numerical domain, the number of chunks that can be held in working memory is only 4±1 in young adults, somewhat less in older adults and children (Cowan 2001, 2005; see also Cowan 1995, 1999). Focusing on experimental conditions where chunks can be identified, capacity limits can be observed, and subjects are kept from rehearsing stimuli, Cowan finds a single, central capacity limit averaging about four chunks. His full model of working memory has two embedded levels. The first consists of long-term memory representations for which there is no limit on activation. The second level ('the focus of attention') is capacity limited and holds up to four of the activated representations. (See also Oberauer et al. 2000; Oberauer 2002)

Since morphemes have to get through working memory in order to be stored in long term memory, they must be recognized as no more than four or so chunks that have already been stored. For *cat* these chunks could be one syllable (Golston & van der Hulst), three segments (Saussure), feature-bundles (Chomsky & Halle, Padgett), or feature-trees (Clements), or five gestures (Broman & Goldstein). Happily, all of these fall into the 4±1 range and thus all of these types of representation seem to be plausible *prima facie*. But longer morphemes provide a *faciem secundam*.

Consider a monomorphemic word like *Apalachicola,* a town in Florida's western panhandle. If I say this word to a stranger who hasn't heard it before, they can usually repeat it back to me with no difficulty. But with 12 segments, /æpəlætʃɪkoʊlə/ is too long to learn if I have to get it through working memory one segment at a time, rather like my mother's eleven digit phone number (15053455480), but one unit worse. Like most people I chunk long phone numbers into prefixes, area codes, and so on, each of which is fairly short: 1-505-345-5480. Working memory limitations like this show us that *Apalachicola* does not make it into LTM one segment at a time, so we can reject Saussure's representation outright.

Breaking segments down into feature bundles doesn't help and actually makes things worse. The Chomsky and Halle model still requires us to memorize the linear order of the twelve feature bundles, something which we cannot do. Worse yet, *none* of the feature bundles is learnable, because each feature bundle consists of far more than 4±1 *unordered* features. Real-world limitations on working memory make learning even a single Chomsky & Halle segment impossible, much less an eleven unit long string of such segments. The same applies to more recent unstructured representations like Padgett's.

Clements' feature geometry makes segments learnable because it bundles the features a few at a time into chunks that can be learned. These chunks can themselves be chunked four or so at a time to make bigger chunks, and so on, till segments are learned. So we can learn sounds like [æ], [p], [ə], etc. in chunks. But there's still no way to get long strings of these sounds through working memory without chunking *them* into bigger units.

Trying to get all of this in as a bunch of coordinated gestures is out of the question unless those gestures are bundled into bigger units; so unadorned Articulatory Phonology is not a viable model of phonological representation. Models that incorporate more prosodic organization are another matter, but no explicit models yet exist for this within Articulatory Phonology as far as I am aware.

Syllable-based representation (Golston & van der Hulst) takes arboreal chunking one level higher, making it possible to store even larger chunks of sound. But if one can only store 4±1 chunks and the largest chunks we have are syllables, it is still impossible to squeeze *Apalachicola* through working memory. So it looks like none of the theories of representation phonologists have proposed so far can be the ones humans actually make use of.

But there's a moral here: more structure is better because it makes more things chunkable and thus more things learnable. *Apalachicola* in gestures is unthinkable, in segments impossible, in syllables unlikely, but *better*. The obvious solution is to assume that *feet* (groups of syllables) are the chunking mechanism used for working memory. Feet make *Apalachicola* learnable. Assuming that English feet are moraic trochees, the word can enter working memory in four chunks [æ.pə], [læ.tʃɪ], [koʊ], [lə]. If English feet aren't quantity sensitive, the word can enter in three chunks: [æ.pə], [læ.tʃɪ], [koʊ.lə]. Foot-based representation actually makes a testable prediction: no language should contain a morpheme longer than 5 phonological feet. Longer *words* should be commonplace, as long as they can be chunked roots or stems of 4±1 morphemes, each of which can be 4±1 feet, syllables, sounds, or combinations thereof. (Following Selkirk 1995 I assume that prosodic categories need not be strictly layered, so that a word like English *bereft* might be represented in LTM as a syllable [bə], a foot [rɛf] and a stray sound [t].)

A few studies of the prosodic shapes of roots suggest that natural languages fall well within this upper limit of two feet. Kager (1995), for instance, reports that canonical roots contain at most two feet in five western Australian languages, Yidiɲ (Dixon 1977), Wargamay (Dixon 1981), Mbabaɽam (Dixon 1991), Gumbayŋgir (Eades 1979) and Uradhi (Crowley 1983). The 836 roots of Dixon's (1977) vocabulary of Yidiɲ are typical, falling into the following classes prosodically (L = light syllable, H = heavy):

(12) Roots in Yidiɲ

LL	590	gala 'spear', gugaɽ 'large guanna'
LLL	219	/gindanu/ [gindaan] 'moon', /gudaga/ [gudaaga] 'dog'
LLLL	12	ɟulugunu 'black myrtle tree', yiŋgilibiy 'bee'
LH	11	durguu 'mopoke owl', giŋaa 'vine species'
LLLLL	2	/ɟilibugabi/ [ɟilibugaabi] 'next day'
LHLL	1	waɽaabuga 'white apple tree'
LLLH	1	galambaɽaa 'march fly'

Most of these root are a *single* foot (LL), or a foot plus a syllable (LLL, LH); a few are two feet (LLLL), or two plus a syllable (LLLLL, LHLL, LLLH); none is longer than this. Golston & Wiese (1995) report essentially the same facts for roots in German; Lewis & Golston (2005) report essentially the same facts for White Hmong (Heimbach 1966) and Nukuoro (Carrol & Soulik 1973); and I assume this will be found to be the case generally. There is a maximal size for roots across languages, it is comfortably smaller than four feet, and it is due to extragrammatical constraints imposed by working memory.

2.2 Long term memory and UR

Long term memory of words is best probed through studies of lexical access. The two most relevant areas for how the phonological forms of words are stored in the brain are tip-of-the-tongue (TOT) states and malapropisms. The two types of evidence bear directly on the mental representations speakers store and it is significant that they agree with one another point for point in showing that prosodic information is not only available but more prominent than segmental material (Cutler 1986:173; Levelt 1989:355).

An influential study by Brown & McNeill (1966) shows that speakers who cannot think of a word tend to know three things about it: the initial segment or onset, the number of syllables, and the stress pattern. When a speaker is primed for *sextant*, for instance, but cannot remember it, two-syllable (xx) words like *secant* and *sextet* come to mind rather than one-syllable words like *sect* (x) or three-syllable words like *sacrament* (x..). Much subsequent research has confirmed these results (e.g., Koriat & Lieblich 1974; Brown 1991; Meyer and Bock 1992; Miozzo & Caramazza 1997; Vigliocco, Antonini, & Garrett 1997; James & Burke 2000; Abrams, White & Eitel 2003). This strongly suggests that speakers store words as syllables or other prosodic groups and not (just) as strings of segments.

Classifications of speech errors include a category called sound-related substitutions (Fromkin 1973) or malapropisms (Fay & Cutler 1977). These involve mis-selection of a word that is phonologically similar to the intended word but semantically different. Typical cases include ('F' from Fromkin; 'FC' from Fay & Cutler):

(13) Malapropisms

Intended	*Spoken*	
white Anglo-Saxon Protestant	white Anglo-Saxon prostitute	(F)
a routine proposal	a routine promotion	(F)
the conquest of Peru	the conquest of Purdue	(F)
prohibition against incest	prohibition against insects	(F)
week	work	(FC)
open	over	(FC)
constructed	corrected	(FC)

As these cases illustrate, the overall prosody of the target is matched by the overall prosody of the error, both in terms of overall stress pattern and in terms of syllable count:

(14) Stress pattern and syllable count of malapropisms

Intended	*Spoken*	*Stress Pattern*	*Syllable Count*
Protestant	prostitute	(x..)	3
proposal	promotion	(.x.)	3
Peru	Perdue	(.x)	2
incest	insects	(xx)	2
week	work	(x)	1
open	over	(x.)	2
constructed	corrected	(.x.)	3

What we do not generally find in sound-related substitutions are cases like *protester* (xx.) for *Protestant* (x..); *propinquity* (.x..) for *proposal* (.x.); *perdition* (.x.) for *Peru* (.x); or *insecticide* (.x.x) for *insects* (xx)—all forms we would expect if words were stored as segment strings.

The criteria for phonological similarity here are identical to those found in TOT states: same onset, same stress pattern, same number of syllables. Data like this has led researchers like Crompton (1982), Fromkin (1985) and Butterworth (1989) to posit a phonological sub-lexicon within the mental lexicon, i.e., a way of finding and storing morphemes using phonological addresses. *White Anglo-Saxon prostitute* is produced when *prostitute* is mis-selected because of its proximity to *Protestant* in the phonological sub-lexicon, that is, by a similar phonological address in LTM. Recent work by Brown (2004) suggests that sound substitutions in speech errors are governed *solely* by prosody, so that errors involve features in authentic units of prosody (onset, syllable, etc.) but never segments. None of this makes any sense if predictable prosodic information like syllable count and stress pattern is not somehow stored.

It is significant that two quite different sources of evidence converge on the same criteria: word-onsets, stress pattern and number of syllables. Any psychologically real model of grammar must come to terms with this and admitting prosody into underlying representation seems like the necessary first step. Phonologists raised on the 1960s notion that everything predictable is left out of UR should recall that OT presents a new paradigm in which this claim is completely irrelevant.

The proposal is less novel than it first appears. It amounts to saying that URs are simply what SRs have been for decades in phonology: *trees* full of information, not structureless strings of segments. There has never been good reason to ban prosodic structure from underlying representation, and there are no reasons at all to ban it within OT.

All of this suggests an answer to a longstanding question: what is the prosodic organization in natural language for? In a few languages the prosodic organization of an utterance is actually contrastive, as has been shown for Scottish Gaelic, where monosyllabic words contrast minimally with bisyllabic words (Clements 1986b; Bosch 1998; Ladefoged et al. 1998):

(15) Scottish Gaelic

tu.an	'hook'
tuan	'song'
pal$^{\gamma}$.ak	'skull'
pal$^{\gamma}$ak	'belly'

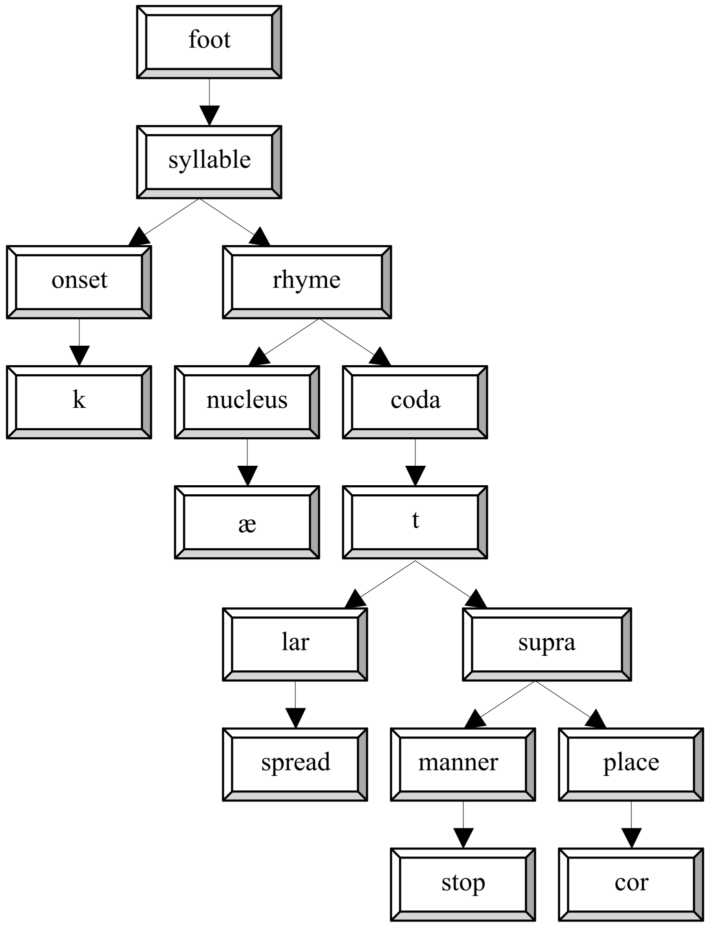
Brazilian Portuguese has similar contrasts, with minimal pairs like monosyllabic [pais] 'parents' and bisyllabic [pa.is] 'country' (Alice Meyer, p.c.).

But in most languages the prosodic organization of an utterance is not contrastive. In many languages, prosody helps position where stress goes (see Hayes 1995) but since predictable stress of this sort has no clear function either, locating stresses can hardly be the general goal of prosodic organization. Languages with prosodic organization but no stress make this point especially clear: Japanese, for instance, organizes things into moraic trochees for purposes of morphology and poetic meter but has no stress at all (Poser 1990). All languages have prosodic organization but they don’t all use it to the same ends and often don’t seem to use it for any ends at all.

So what does it do and why is it there? I propose that *prosody is a data compression device* which allows lots of distinctive features to be compressed into bigger and bigger units that can be memorized; these memorized chunks can then be used to chunk still larger units like *Apalachicola* which would otherwise be unstorable. If this is correct, prosodic structure is used in all languages to shepherd distinctive features through working memory into long-term memory.

To tie this proposal into something that has been proposed on independent grounds, I adopt from Marcus (2001) a neural model of representation involving linked and nested register sets that he and Jackendoff (2002) use to model stored semantic, syntactic, and morphological representations. The underlying representation for *cat* would then be an arboreal graph of a foot (with some abbreviation to make it fit on a page):

(16) *cat*

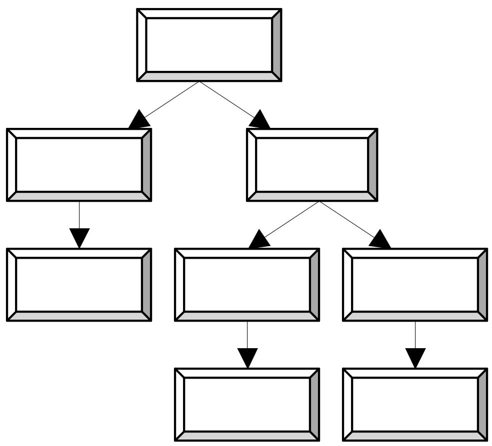


Representations like this are memorizable because you can chunk things from the bottom and store the resulting constituents for later. As successively larger chunks make it into long term memory, the 4±1 limit on working memory allows larger and larger units to get in a few at a time: once I've got *kiss* and *fat* I can parse *cat* with an onset and rhyme I already have stored in LTM. In this way, even the six syllables of *Apalachicola* can make it into LTM by coming in as feet. It would seem that long morphemes like this *must* be stored with foot and syllable structure because there is no way to get all of those segments through working memory.

A reviewer asks 'If storage is limited to four or five items in the horizontal dimension of a representation, why does this not hold for the vertical dimension?' The answer is essentially that the vertical dimension represents what has already been memorized (for other words). A speaker who knows the UR for ‘cat’ in English has stored its sounds, organized into a syllabe that constitutes a foot. The (vertical) prosodic organization is the means of storage for the (horizontal) distinctive features/gestures.

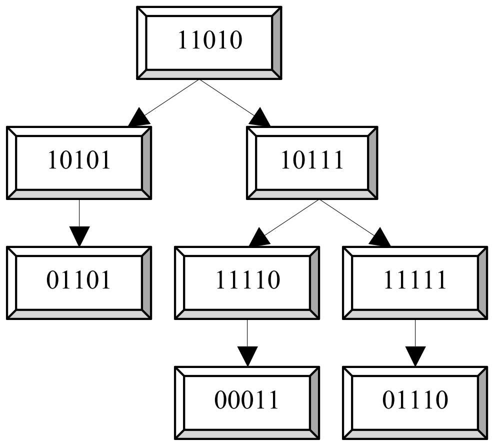
Returning to Marcus’s model, he proposes a computationally tractable model of treelets stored and used in the brain. An empty treelet is something like a stem cell, ready to take on whatever representational configuration is necessary:

(17) Empty treelet



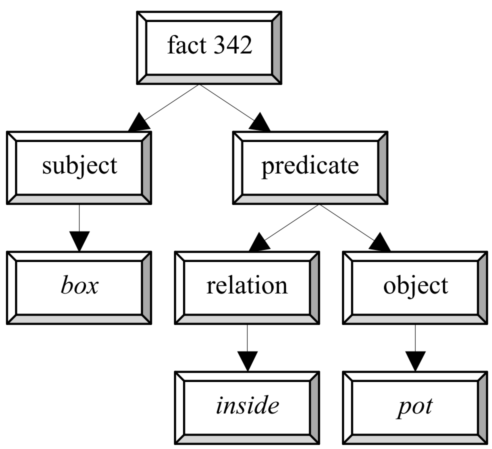
A filled treelet like the following stands for propositions held in LTM. The five digit numbers encode atomic elements already held in LTM:

(18) Filled treelet



When the numbers are cashed in for their LTM equivalents you get a representation like the one for *cat* in (16) or the following for a clause:

(19) What the numbers encode



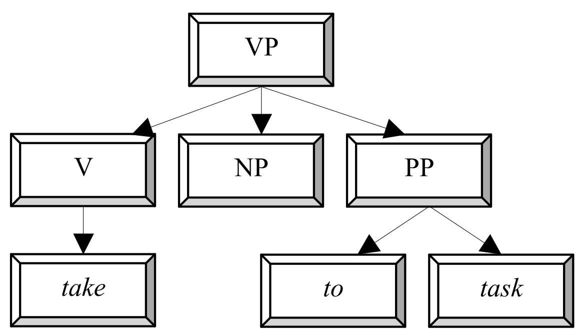
Learning a new fact amounts to setting values of the register sets within a treelet:

> Fundamental to my proposal are the assumptions that the mind has a large stock of empty treelets on hand and that new knowledge can be represented by filling in an empty treelet (that is, by storing values in the register set) or by adjusting the values contained in an existing treelet. (Marcus 2001, 108)

As we have seen, this solves the chunking problem for getting long morphemes though working memory into LTM. If LTM stores treelets that correspond to sounds, syllables, and feet, these stored chunks can be used to pass information through working memory and into LTM in successively larger chunks.

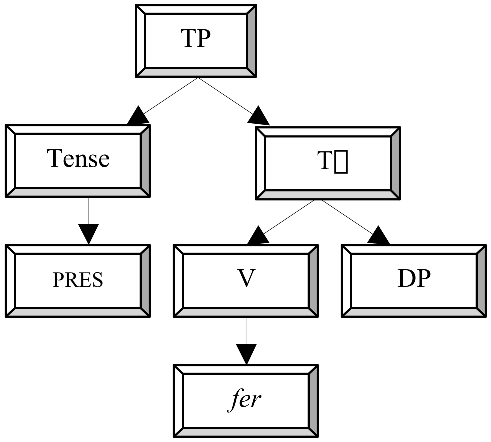
But the treelets are as yet too fixed to handle phonological alternations. For this we need *variables*. A major part of Marcus' proposal is that not all parts of a treelet are filled in—and the empty register sets in a tree represent variables. Jackendoff (2002) proposes a number of linguistic structures that make use of such variables along the following lines. An idiom like *take (x) to task* has a variable position that must be filled by material with material which matches the category type of the variable (NP in this case). The idiomatic reading shows that the whole thing must be stored and the fact that all sorts of things can go in the middle slot shows that the representation must contain a variable of some sort:

(20) *take (x) to task* (Jackendoff 2002: 172)

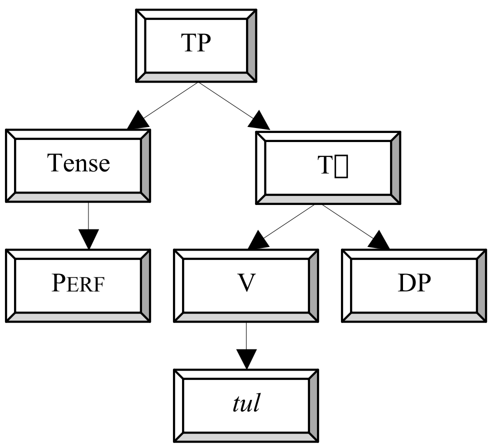


Moving a step closer to phonology, we can treat suppletion along similar lines but with information prespecified. The root for *carry* in Latin is *fer*- (*fer-o* 'I carry') in the present system, *tul*- (*tul-i* 'I have carried') in the perfect, and *lat*- (*laat-us* 'be carried') in the supine (see Aronoff 1994, 31-59). Treelets with different roots and different aspects handle this otherwise very problematic type of case straightforwardly:

(21) *fer- 'carry (present)'*



(22) *tul- 'carry (perf)'*



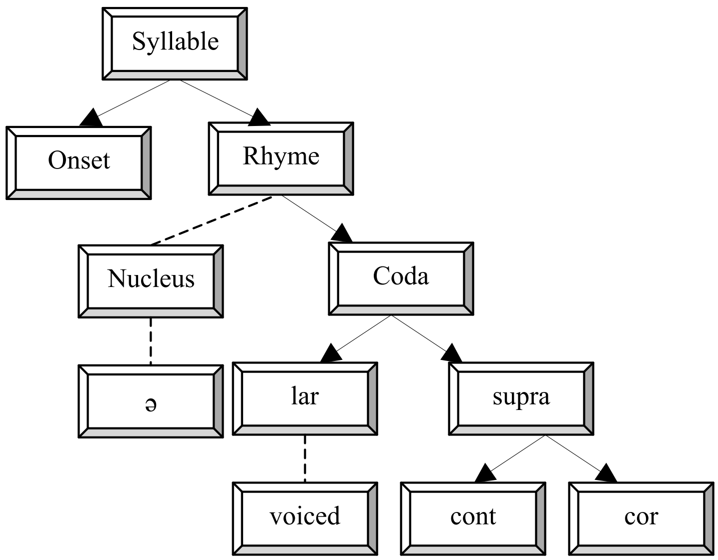
This is compatible with extant treatments of allomorphy in OT (see, for instance, McCarthy 2002:152-156).

3. Surface representations

If underlying representations are trees with variables, surface representations are trees with those variables solved. This means that GEN has merely to supply the values that the variables stand for. As long as the variables have a finite number of values, the number of surface candidates EVAL has to consider remains finite. Consider how epenthesis and voicing assimilation might work for English plural *–s*.

There are a lot of ways to pluralize nouns in English, including many that must be listed in the lexicon (*deer:deer, datum:data, locus:loci*) and thus, I assume, stored with the roots they go with in trees like those given above for Latin *fer- tul- laat-* 'carry'. But the common way of pluralizing is to add *–s*. As is familiar, plural *–s* has three shapes depending on the final sound of the stem to which it attaches, [z, əz, s]. The UR for *–s* is then just a tree with variables, along the following lines:

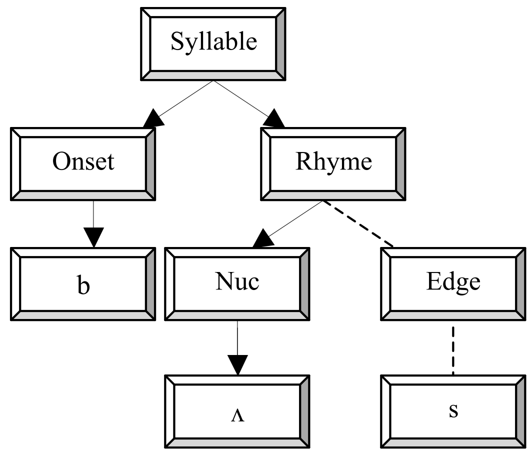
(23) [z, əz, s] 'PLURAL'



The dotted lines connecting Nucleus to Rhyme and [voiced] to the laryngeal node of the coda are realized or not depending on EVAL. This gives us four possible ways of realizing the plural [əs, əz, s, z], one of which [əs] is never used for reasons we'll see below.

A root like *bus* can surface in one of two ways: [bʌs] as in *bus* (with everything in the same syllable) and [bʌ.s] as in *bussing* with the [s] in the onset of the next syllable. The UR for *bus* thus has everything nailed down except the final consonant, which is potentially the coda for the root syllable or the onset for the following suffixal syllable. I represent this with the variable Edge (= Ons/Nuc).

(24) UR for *bus*



Given two realizations of the root and four of the suffix there are eight candidates that need to be generated. A few constraints will suffice to pick out the winner as is common in OT:

(25) AGRLAR Consonant clusters agree in laryngeal features.
(26) OCP No sibilants are adjacent.
(27) ONSET Syllables begin with consonants.
(28) MAXVOI Underlying voiced is licensed prosodically.
(29) MAXə Underlying schwa is licensed prosodically.

A tableau lays out the candidates and how they are evaluated, with most of the surface prosody left out to save space:

(30) *bus-s* [bʌsəz]

	OCP	AGRLAR	ONSET	MAXVOI	MAXə
[bʌss]	*!			*	*
[bʌsz]	*!	*			*
[bʌss]	*!			*	*
[bʌsz]	*!	*			*
[bʌs.əz]			*!		
[bʌs.əs]			*!	*	
[bʌ.səs]				*!	
☞ [bʌ.səz]					

The last candidate violates none of the constraints and bests the rest. The first four candidates fail on the OCP because of the illicit [ss] and [sz] clusters; the fifth and sixth fail on ONSET; and the seventh loses because the underlying voicing feature of the suffix isn't realized. Tableaux for *butts* and *buds* should be self-explanatory. The only point of this exercise is to illustrate how few candidates are actually needed to guarantee having the winner show up. *A candidate set that consists of the union of the set of root allomorphs and the set of suffix allomorphs must include the winner.* Other candidates are superfluous.

An anonymous reviewer points out that a generalization is lost here, since it is now coincidental that the epenthetic vowel for 3sg *-s*, possessive *-s*, plural *-s*, and past tense *-t* is always schwa. This can be addressed by replacing schwa with a variable that covers mid vowels, or vowels, or sonorants, or whatever; EVAL will then select the best of these. Here as elsewhere the key to keeping GEN small is to use variables, since a given node can be filled by only a given number of features.

Thus when Jessen and Ringen (2002) analyze German final fortition in words like *bunt* 'colorful' and *Bund* 'club' as the addition of [spread glottis], we need not consider all possible changes to the final consonant, only those that fill in permissible values of the laryngeal node. If we specify the laryngeal node of German /t/ as [spread glottis] and of /d/ as nothing, GEN will fill in [spread glottis], [constricted glottis] or [voice] for each and CON will choose the best one. Fortition processes don't require an infinite GEN if we circumscribe the range of fortition with variables. Similarly for reduction, lenition, assimilation, and so on; everything can be short and finite as long as variables are allowed in underlying representations and GEN just spells out those variables.

Other types of phonological alternation can be handled analogously, as allomorph selection (already necessary in standard OT), as morphemes with floating elements, or as morphemes with elements unspecified. The latter case will cover consonant- and vowel-harmony, where certain nodes in the tree (e.g., rounding or backness) will be empty. As long as a given node in a feature tree is limited to a small number of possibilities (e.g., Labial can dominate [round] or not, but cannot dominate [nasal]) this manner of encoding variables should always result in a finite number of candidates.

The standard method of treating variation in OT has been to have GEN produce an infinite set of candidates that includes the faithful candidate and enough nearly faithful candidates to include the winner. Variable notation, familiar from simple algebra, allows a more focused and parsimonious way of handling variation in phonology. The specific treatment of variables borrowed from Marcus and Jackendoff has the additional advantage of applicability across different domains of cognition. As Jackendoff points out, "all combinatorial rules of language—formation rules, derivational rules, and constraints—require typed variables" (2002: 65).

4. Conclusion

A better understanding of the role of prosody in lexical representation undermines a long-held assumption in phonology, that underlying representations are stored without prosodic structure (syllables or feet). This assumption has necessitated a large gulf between underlying and surface forms and has thus required a lot of work for GEN in Optimality Theory.

Working memory restrictions, TOT states, and malapropisms strongly suggest that the stored representations we call underlying representations are not stored as beads on a string but as hierarchically organized prosodic structures that may be graphed as trees. Prosodic structure most likely functions as a data compression device that gets large amounts of information (features or gestures) through working memory and into long term memory. Assuming that prosodic trees can include variables, as has been suggested by Marcus and Jackendoff for morphology, syntax, semantics, and long term memory, the difference between underlying and surface representations in phonology reduces to trees with variables and trees with those variables spelled out. This reduces the function of GEN to solving for variables, a finite and fairly well-understood task.

Acknowledgments

Thanks to audiences at California State University Fresno and the University of Tromsø for important questions and suggestions for improvement. Special thanks to Jason Brown, Tomas Riad, and two anonymous reviewers for their careful and valuable comments on an earlier draft of this paper. Any remaining shortcomings are of course my responsibility.

References

Abrams, Lisa, Kathryn K. White, and Stacey L. Eitel.
2003 Isolating phonological components that increase tip-of-the-tongue resolution. *Memory & Cognition* 31.8: 1153-1162.

Anderson, Stephen R.
1985 *Phonology in the twentieth century: theories of rules and representations*. Chicago: The University of Chicago Press.

Aronoff, Mark
1994 *Morphology by Itself: Stems and Inflectional Classes*. Linguistic Inquiry Monograph Series 22. Cambridge: MIT Press.

Baddeley, Alan D.
1986 *Working Memory*. Oxford: Clarendon Press.

Bosch, Anna
1998 The syllable in Scottish Gaelic dialect studies. *Scottish Gaelic Studies XVIII*: 1-22.

Browman, Catherine P. and Louis Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.

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

Browman, Catherine P. and Louis Goldstein
1990 Tiers in articulatoryphonology, with some implications for casual speech. In J. Kingston, & M. E. Beckman (eds.), *Papers in Laboratory Phonology: I. Between the Grammar and the Physics of Speech*. 341–376. Cambridge: Cambridge UniversityPress.

Browman, Catherine P. and Louis Goldstein
2000 Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5: 25–34.

Brown, Alan S.
1991 The tip of the tongue experience: a review and evaluation. *Psychological Bulletin* 10: 204-223.

Brown, Jason C.
2004 Eliminating the segment tier: evidence from speech errors. *Journal of Psycholinguistic Research* 33.2: 97-101.

Brown, Roger and David McNeill
1966 The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5: 325-337.

Butterworth, Brian
1989 Lexical access in speech production. In William Marslen-Wilson (ed.), *Lexical Representation and Process*. 108-135.

Byrd, Dani
1996 A phase window framework for articulatory timing. *Phonology* 13.2: 139-169.

Byrd, Dani and Elliot Saltzmann
2003 The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31: 149-180.

Carrol, Vern and Topias Soulik
1973 *Nukuoro Lexicon*. Honolulu: University Press of Hawaii.

Chomsky, Noam and Morris Halle
1968 *The Sound Pattern of English.* New York: Harper & Row.

Clements, G. N.
1985 The geometry of phonological features. *Phonology Yearbook* 2: 223-250.

Clements, G. N.
1986 Syllabification and epenthesis in the Barra dialect of Gaelic. In Koen Bogers, Harry van der Hulst, and Maarten Mous (eds.), *The phonological representation of suprasegmentals.* Dordrecht: Foris. 317-336.

Cowan, Nelson
1995 *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, Nelson
1999 An embedded-process model of working memory. In A. Miyake & P. Shah (eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, England: Cambridge University Press.

Cowan, Nelson
2001 The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24.1: 87-114.

Cowan, Nelson
2005 *Working memory capacity.* New York, NY: Psychology Press.

Crompton, Andrew
1982 Syllables and segments in speech production. In Ann Cutler (ed.), *Slips of the tongue and language production.* Berlin: Mouton.

Cutler, Ann
1986 Phonological structure in speech recognition. *Phonology Yearbook* 3: 161-178.

Crowley, Terence
1983 Uradhi. In R.M.W. Dixon and B.J. Blake, eds., *Handbook of Australian Languages, Volume III*, 307-428, John Benjamins, Amsterdam.

Dixon, R.M.W.
1977 *A Grammar of Yidiɲ*. Cambridge University Press, Cambridge.

Dixon, R.M.W.
1981 Wargamay. In R.M.W. Dixon and B.J. Blake, eds., *Handbook of Australian Languages, Volume II,* 1-144, John Benjamins, Amsterdam.

Dixon, R.M.W.
1991 Mbabaɽam. In R.M.W. Dixon and B.J. Blake, eds., *The Handbook of Australian Languages, Volume 4*, 349-402, Oxford University Press, Oxford.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier and Jacques Mehler
1999 Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*. 25(6), 1568-1578.

Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi and Jacques Mehler
2001 New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16.5-6, 491 - 505

Eades, Diana
1979 Gumbaynggir. In R.M.W. Dixon and B.J. Blake, eds., *Handbook of Australian Languages, Volume I*, 245-361, John Benjamins, Amsterdam.

Fay, David and Ann Cutler
1977 Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8: 505-520.

Fromkin, Victoria A. (ed.)

1973 *Speech errors as linguistic evidence.* The Hague: Mouton & Co.

Fromkin, Victoria A.
1985 Evidence in Linguistics. *Linguistics and Linguistic Evidence. The LAGB Silver Jubilee Lectures 1984*. Newcastle: Grevatt & Grevatt. 18-38.

Golston, Chris and Harry van der Hulst
1999 Stricture is structure. In Ben Hermans and Marc van Oostendorp (eds.), *The Derivational Residue in Phonological Optimality Theory.* Amsterdam: Benjamins.

Golston, Chris and Richard Wiese
1998 The structure of the German root. In Wolfgang Kehrein and Richard Wiese (eds.), *Phonology and Morphology of the Germanic Languages*. Tübingen: Niemeyer. 165-185.

Hallé, Pierre A., Juan Segui, Uli Frauenfelder and Christine Meunier
1998 Processing of illegal consonant clusters: a case of perceptual assimilation*? Journal of Experimental Psychology: Human Perception and Performance,* 24(2), 592-608.

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

Heimbach , Ernest E.
1966 White Hmong-English Dictionary (Linguistics Series IV). Ithaca, NY: Southeast Asia Program, Data paper: Number 75.

Jackendoff, Ray
2002 *Foundations of Language.* Oxford: Oxford University Press.

Jakobson, Roman
1939 Observations sur le classment phonologique des consonnes. *Proceedings of the 3rd International Congress of Phonetic Sciences*. 34-41.

Jakobson, Roman, Gunnar Fant and Morris Halle
1952 *Preliminaries to Speech Analysis.* Cambridge: MIT Press.

James, Lori E. & Deborah M. Burke
2000 Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26.6: 1378-1391.

Jessen, Michael and Catherine Ringen
2002 Laryngeal features in German. *Phonology* 19, 189-218.

Kager, René
1995 On foot templates and root template. In Marcel den Dikken and Kees Hengeveld (eds.), *Linguistics in the Netherlands 1995*: 125-138. Amsterdam: John Benjamins.

Koriat, Asher and I. Lieblich
1974 What does a person in a "TOT" state know that a person in a "don't know" state doesn't know? *Memory & Cognition* 2: 647-655.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind, St John Skilton
1998 Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28: 1-42.

Levelt, Willem J. M.
1989 *Speaking: From Intention to Articulation.* Cambridge, Mass.: MIT Press.

Lewis, Will and Chris Golston
2005 The randomness of the signifier. Paper presented at the Annual Meeting of the Linguistic Society of America. Oakland, CA.

Marcus, Gary F.
2001 *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge, Mass.: MIT Press.

Massaro, Dominic W. and Michael M. Cohen
1983 Phonological constraints in speech perception*. Perception & Psychophysics* 34, 338-348.

McCarthy, John J.
2002 *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

Meyer, Antje S. and Kathryn Bock
1992 The tip-of-the-tongue phenomenon: blocking or partial activation? *Memory & Cognition* 20: 715-726.

Miller, George A.
1956 The magical number seven, plur or minus two: Some limits on our capacity for processing information. *Psychological Review* 3: 81-97.

Miozzo, Michelle and Alfonso Caramazza
1997 Retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 1410-1423.

Oberauer, Klaus
2002 Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28.3: 411–421.

Oberauer, Klaus, Heinz-Martin Süss, Ralf Schulze, Oliver Wilhelm and Werner W. Wittmann
2000 Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, *29*: 1017–1045.

Oberauer, Klaus, Heinz-Martin Süss, Oliver Wilhelm and Werner W. Wittmann
2003 The multiple faces of working memory: storage, processing, supervision, and coordination. *Intelligence* 31.2: 167-193.

Padgett, Jaye
1995 Feature classes. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.)*, University of Massachusetts Occasional Papers in Linguistics* 18: 385-420. Amherst, MA: GLSA.

Poser, William
1990 Evidence for foot structure in Japanese. *Language* 66: 78-105.

Reiss, Charles
2002 The OCP and NOBANANA. Carleton University Cognitive Science Technical Report 2002-03. URL http: //www.carleton.ca/iis/Tech Reports.

Prince, Alan and Paul Smolensky
1993 Optimality Theory: constraint interation in generative grammar. Piscataway, NJ: Rutgers University Center for Cognitive Science.

Saussure, Ferdinand de
1959 *Course in General Linguistics*, translated by Wade Baskins. New York: The Philosophical Library.

Selkirk, Elizabeth
1995 The prosodic structure of function words. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*. Amherst, MA: GLSA Publications. 439-70. Also in James L. Morgan

and Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. 187-214: Lawrence Erlbaum Associates.

Trubetzkoy, Nikolai S.
1939 *Grundzüge der Phonologie. Travaux du cercle linguistique de Prague 7*.

Vigliocco, Gabriella, Tiziani Antonini and Merrill F. Garrett
1997 Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8: 314-317.