X
Interpreting "phonetic interpretation" over the lexicon

MARY E. BECKMAN AND JANET PIERREHUMBERT

X.1. The metaphors invoked

The phrase "phonetic interpretation" (included in both special themes of the 6th Conference on Laboratory Phonology) invokes a particular view of language sound structure in which phonetics is defined as the semantics of phonology, which implies, conversely, that phonology is the syntax of phonetics. (See Pierrehumbert, 1990, *inter alia*, for earlier explicit articulations of this view.) The paired metaphors are characteristic of a class of phonological theories built on symbolic formalism, just as the theory of "ordinary" syntax is. They also delineate a class of psycholinguistic theories about how our memory of past experiences constrains our own phonological behaviour and our understanding of others' phonological behaviour. The utterances of other speakers are not apprehended and remembered as a collection of unanalyzed audio-visual patterns. Nor are our own utterances produced by dredging up from memory the unanalyzed articulatory records of some collection of previously produced utterances. Rather, utterances are decomposed into (or composed from) smaller parts that can be recognised as instances of particular syntactic categories standing in particular syntactic relationships to other categories in the context. Thus, phonological behaviour (like "ordinary" syntactic behaviour) can be seen as a particularly elaborate type of categorization, which allows for very complex and flexible stimulus generalization. In this paper, we explore the cognitive function of categorization, to try to understand its ecological advantage.

Shepard (1987) defines stimulus generalization as the cognitive act of determining the probability that a newly encountered stimulus will have a particular consequence, given that a similar stimulus previously encountered had that consequence. For example, a bear becomes ill after eating nightshade berries. Later encountering similarly small, scarlet berries on another similar-looking plant, it might infer that eating these will have a similar effect. Shepard proposes a theory of stimulus generalization in which a particular consequence is represented as a connected region in the relevant sensory dimensions in the organism's mind. The organism evaluates the probability that the new stimulus falls into this "consequential region" by estimating the region's location and size. Shepard applied his theory to explain results of multidimensional scaling analyses of identification responses to stimuli that are systematically varied in dimensions that are known to be relevant for identifying some target category. Myung & Shepard (1996) show that, given certain simplifying assumptions about the shape of the consequential region of the category, the estimations of its location and size can be based on observed distances between “consequential” stimuli, ignoring regions occupied by “non-consequential” stimuli.

In nature, however, these simplifying assumptions often do not hold, and there is no experimenter to specify the targeted response category. Imagine the difficulty for the bear if one variety of nightshade produced fruit larger than the smallest wild tomatoes. Does the bear forgo eating ripe edible fruit of intermediate size? Does it set up two independent consequential regions for the smaller and larger varieties associated with the common consequence? Also, how does “illness” become identified as a common consequence to identify these berries as “illness-causing” in the first place? Unless it is innate, this category must be learned from the sensory dimensions representing experiences of illness. That is, the bear must generalise a class of encounters in the sensory space for nausea and nausea-related events in order for “illness-causing” to become an index for equating different berry stimuli.

Categorization thus seems typically to be a more elaborate cognitive act than simple stimulus generalization. It involves a space of intervening consequences, which allows the organism to learn robust, efficient, and potentially quite arbitrary associations between otherwise incommensurable domains of experience. This is the ecological advantage of phonological categorization, which is the theme of this paper. The organism recognises the immediate stimulus as a member of one of several contrasting conceptual classes. These classes then stand as an independent conceptual space between the sensory representation of the proximal stimulus and the sensory representation of its associated distal consequence. Because the dimensions of this intermediate conceptual space are not isomorphic to those of either stimulus space, categorization promotes more robust stimulus generalization in two ways.

First, even limited experience with categorization in a particular domain promotes selective attention to regions of the stimulus space where there are salient consequences for some classificatory dimension of the conceptual space. This differential attentional weighting sharpens perceptual acuity at boundaries between categories and blurs distinctions around prototypical values well within a category. For example, the bear quickly learns to pay attention to size when a berry is a bit small (for a tomato). Second, accumulated experience with categorising stimuli in their natural environments promotes a flexible, context-dependent partitioning of the stimulus space. Sensory dimensions of the context can be associated with the competing categories in the conceptual space, so that different attentional weights can readily alternate for each other in different contexts. The bear attends to size for red berries growing on small plants with large dark-green lobed leaves, but not for blue berries growing on woody shrubs with pale smooth round leaves. The noxious berry is a bit small, *for a tomato.*

Syntactic structure allows a further elaboration on this type of intervening conceptual space, in which the category of a stimulus itself constitutes a relevant dimension of the context of another stimulus at a comparable level of description. Syntactic classes have a contextual (syntagmatic) categorization as well as an “ordinary” (paradigmatic) categorization. A syntactic object is classified as being the same as or different from other stimuli that have been previously encountered in similar contexts, and also as belonging to a particular context class for the recognition of other syntactic objects. For example, the element /æ/ is classified in an utterance of the wordform *cat* by being different from /e/, /A/, etc., and also by the way that members of its syntagmatic class provide a recurring environment for differentiating /k/ from /g/, /p/, etc.

As stated above, categorization molds attention to the most relevant available stimulus dimensions. Categorization also allows the listener to make inferences about hidden stimulus dimensions. If a listener recognises a region of the signal as an example of /æ/, he can infer from the spectral peaks in the signal that the speaker lowered the tongue body. This inference can be made even if he is listening to the speech on a tape recording and has no direct information about the articulation. Equally, however, a speaker can articulate an example of an /æ/ while having the auditory percept of her own speech blocked by white noise played through headphones. In this case, she can infer the existence of spectral peaks that she cannot observe directly. In the terms of formal semantics, complexes of events in the physical world which provide examples of the category /æ/ are elements of the “extension” of /æ/. Because /æ/ is a syntactic object, moreover, part of its extension involves the relationships in the physical world which can be inferred from its syntagmatic class. When the listener recognises a region of the signal as an example of /æ/, he also recognises it as providing a potential following context for a /k/. He then can infer from changes in the spectral peaks over time that, just before the speaker lowered the tongue body, she first raised it to momentarily contact the palate.

A robust linguistic generalization that falls out of adopting this phonetics-as-semantics metaphor is that phonological objects and relationships should be “natural” ones. The conceptual structures that acquiring a human language imposes on the child’s mind reflect often very robust correlations with properties and events in nature. This is why a child can acquire the phonetic and phonological concepts of the ambient culture. This is true also of the correlations between semantic and syntactic properties that define the “lemma”, the conceptual category that a wordform names. (See, e.g., Levin & Rappaport Hovav, 1996; also, Malt, 1996, for a review of the anthropological linguistics literature on this point.) It is unlikely that language could have evolved otherwise.

We can contrast the naturalness of these phonology- or syntax-specific associations to the typically arbitrary nature of the cross-domain association between wordform and lemma. Although speech communities can (and do) exploit recurring associations, in both sound symbolism and word play, these correlations between wordform and lemma are not predictive in the way that domain-internal associations typically are. A new member of the community (such as the field linguist) can use his native-language categories to guess the approximate relationship between phonological categories and their observed phonetic referents. Similarly, after encountering enough verbs of motion, he might be able to guess, for example, that the language has distinct words for running in place and running toward a goal. However, unless the language is closely related to his own, he is not likely to be able to predict what the two wordforms are for these two lemmas. The lack of strong correlations between the phonological and phonetic properties of wordforms and the homologous properties of their associated lemmas is a very robust generalization, the principle of *l'arbitraire du signe*. To acquire the language of the ambient culture, a child has to make many such associations, most of which are entirely conventional—i.e., arbitrary and culturally specific.

L'arbitraire du signe seems to be related to another important generalization about the organization of the human memory for words, the property of "duality of patterning" (Hockett, 1960): The principles of compositionality on the wordform side of the "primary" semantic association are different from those on the lemma side. Even the smallest "meaningful" forms in a language—the wordforms and grammatical morphemes that associate to lemmas and recurring syntactic relationships among lemmas—can be decomposed further into smaller elements which have no "meaning" on their own in the sense of associating to some category on the other side of the *signe-signifiée* divide.

Because humans who share a common culture share the categories on both sides of this divide, existing categories of various orders of complexity can be combined to make new categories. These new composite categories will have predictable syntactic and semantic properties, derived from the syntactic and semantic associations of the smaller objects that were combined to make the larger ones. This is true not just of "ordinary" syntactic objects such as lemmas and combinations of lemmas, but also of the syntactic objects internal to wordforms such as phonemes and gestures. Hearers and talkers can productively apply their knowledge of the internal composition of known wordforms to recognise and reproduce novel wordforms such as the name of a person met at a party, a newly encountered technical term, or the name of a novel object borrowed from another culture. Also, the syntax and semantics of phonological categories such as /æ/, /k/, and /t/, and of their composition into /kæt/, are independent of the "ordinary" syntax and semantics of lemmas and lemma combinations. Therefore, hearers can "understand" a new wordform in terms of categories and properties internal to the phonology. Moreover, they can do so even if the new wordform names an object or event that does not fit naturally into any existing lemma category—indeed, even if the new wordform does not name any conceptual category at all, as in the nonce "words" presented in many speech experiments.

Another fundamental linguistic principle—that of "positional contrast"—falls out from the fact that wordforms are structured in time. Recognising the syntagmatic context of a syntactic object such as the /k/ in *cat* involves learning such temporally defined properties as "the beginning of the wordform" and "before /æ/" as well as the relationships between this dimension of classification and various different sensory dimensions. Such syntactic properties must be abstracted away from the semantics of the /k/ itself. Abstracting away the temporal syntax is important, because the mapping between our experience of articulatory maneuvers and our audio-visual experience of the consequences of these maneuvers is not always the same for different positions. For example, the audible consequences of lowering the tongue dorsum after forming a /k/ closure typically will differ between "before /æ/" as in *cat* and "before /s/" as in *axe*.

To recapitulate, then, we view a phonological category, such as the phoneme /æ/ or /k/, as a formal syntactic object. It exists in memory not simply by virtue of its semantics. It is not stored or accessed just via the indexical associations to sensory representations of real-world events, such as the gesture of making and releasing a closure that blocked airflow and the burst that results. Rather, it is remembered and invoked from memory by virtue of two other kinds of indexical relationship to other categories in memory. These relationships are the syntagmatic associations to other syntactic objects which can occur in the object's context, and the paradigmatic associations to other syntactic objects which can occur in the object's place in the given context. Moreover, in adopting the metaphor, we attribute to the human mind—to the language user and not simply to Language—the four principles characterising the phonological organization of all human languages: (1) the "naturalness" of phonological categories, (2) *l'arbitraire du signe*, (3) duality of patterning, and (4) positional contrast. We propose that a formal model of phonological competence will accurately mimic human performance only if it provides an account of how these four characteristics can emerge in acquisition as attributes of the memory store for wordforms of the ambient language. While no one has yet offered a formal model that can account for acquisition and control of a reasonably large lexicon, it is possible to find evidence that these principles govern individual phonologies using very simple experiments, such as the following priming test.

X.2 The word-association experiment

Priming refers to experimental observations suggesting that retrieving a word or other item from memory also activates related items that might be associated with it in memory. We devised a word-association task to see whether accessing two words together could differentiate between associative links among similar wordforms and associative links among related lemmas, as predicted by the principle of *l'arbitraire du signe*. We also examined the priming effects that we found to see whether they suggested a simple all-or-none association between similar wordforms, or a more complex network of associations that is structured by the principles of duality of patterning and positional contrast.

We presented 60 Ohio State University undergraduates with 24 pairs of priming words (see Table X.1), each followed by a blank in which to write down the first word that came to mind. In each pair, the second word (the one immediately preceding the blank) had 5 phoneme-sized segments, as transcribed in the *Hoosier Mental Lexicon* (*HML*, Pisoni *et al.*, 1985) and a familiarity rating of 7 on the *HML* scale of 1-7. We will refer to this word as the "immediate prime." The paired non-immediate primes also were all highly familiar words, and each was either a wordform associate (a word with 2 to 4 segments in common with the immediate prime) or a lemma associate (a word with a related meaning). We chose wordform associates from different semantic domains (to minimise semantic relatedness) and lemma associates with no shared segments in identical positions. A split-plot design insured that each participant saw an equal number of lemma associates and wordform associates, without any individual participant seeing both pairings for the same immediate prime. Pairs containing lemma associates and pairs containing wordform associates were interleaved in pseudo-random order.

Response forms for participants who were not self-identified as native speakers of English were discarded, as were forms in which any blanks were unfilled or filled with long proper nouns (e.g. "Emancipation Proclamation (document)" in response to "fifty, treaty, _____"). The remaining responses were scored for their similarity in form and meaning to the immediate prime.

For similarity in form, three different objective scores were computed, based on the *HML* transcriptions (see below). Similarity in meaning was determined by obtaining judgments from a second pool of Ohio State University undergraduates. New response sheets were created, each containing all 24 immediate primes, but paired this time with the corresponding response words written down by one of the participants in the association task rather than with the other (non-immediate) prime. Participants in the scoring task rated each pair on "how closely related their meanings are" using a scale from 1 (for "NOT related") to 7 (for "VERY related"). Again, a response form was discarded if there were missing ratings or if the rater was not self-identified as a native speaker of English. This left 39 sets of responses, about evenly divided between the two lists on the association task. The mean semantic relatedness rating was then calculated by averaging over the 18 or 21 similarity judgments obtained for each immediate prime in each pairing condition in the association task.

To compute wordform similarity, we looked up the *HML* transcription for each response word from the association task, and tabulated the number of responses that shared at least one segment with the immediate prime. We also calculated the number of segments shared, in two different ways—once without regard for position in the word and once counting only identical segments in identical position relative to either the beginning of the word, the end of the word, or the stressed vowel in the word. Since the *HML* transcription system encodes some allophonic and positional features directly in the symbols used, we first recoded some symbols. For example, we recoded syllabic "R" and its unstressed counterpart "X" as "r". This uncovered, for example, the shared "r" in "lover" [l^vX]=[l^vr] as a response to "hug, friend", as well as the "r" segment shared by the wordform associates "dart, desert" (hence the count of 3 rather than 2 segments in common for this pair in Table X.1).

Table X.1. Stimulus words for the association test. Transcriptions of the target and wordform associate from the *HML*, with segments in common underlined.

target word	HML	no shared segments	wordform associate	HML	lemma associate
splash	spl@S	3	split	splIt	fountain
juggler	J^gLX	2	jumbo	J^mbo	circus
sermon	sRmxn	2	syrup	sRrxp	bible
package	p@kIJ	2	panda	p@ndx	mail
pleasure	plEZX	2	plant	pl@nt	vacation
strong	strcG	3	stripe	strYp	muscle
usual	yuZuL	2	youth	yuT	never
balloon	bxlun	2	tycoon	tYkun	party
crank	cr@Gk	4	drank	dr@Gk	pull
friend	frEnd	3	attend	xtEnd	hug
snicker	snIkX	3	wicker	wIkX	tease
excel	IksEl	2	lapel	lxpEl	grade
treaty	triti	2	fifty	fIfti	war
motive	motIv	3	octave	aktIv	excuse
wrinkle	rIGkL	3	ritual	rICuL	forehead
staple	stepL	3	simple	sImpL	tack
Chinese	CYniz	3	cheese	Ciz	Greek
cancel	k@nsL	3	camel	k@mxl	renew
famous	femxs	3	focus	fokxs	star
clamp	kl@mp	3	cap	k@p	wood
filling	fIlIG	3	footing	fUtIG	cavity
desert	dEzXt*	3	dart	dart	sand
kitchen	kIC|n	2	keen	kin	stove
machine	mxSin	2	moon	mun	gear

*Before tabulating forms in common, we changed this from the homograph verb [d|zRt]. However, responses such as "pie", "pastry", and "chocolate" to "dart, desert" suggest that many of our subjects read *desert* as the verb's homophone *dessert*.

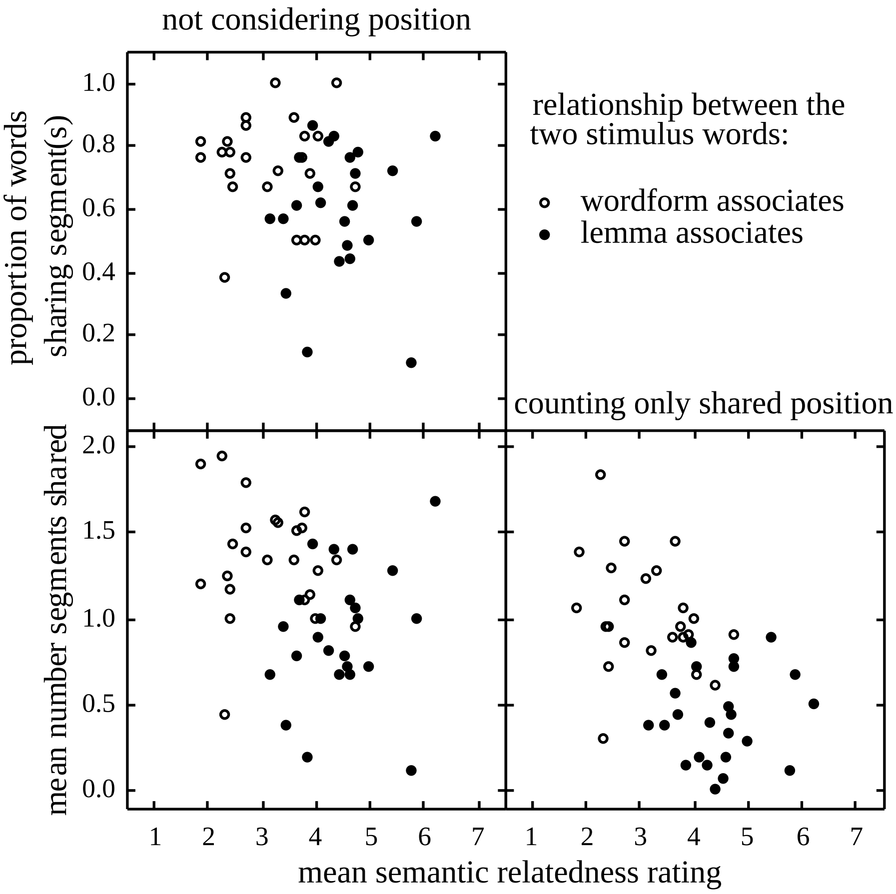

Figure X.1. Three measures of mean wordform relatedness plotted against mean lemma relatedness between each immediate prime and all participants' responses to wordform priming pairs (open circles) and to lemma priming pairs (filled circles). Measures: (a) proportion of responses sharing a segment with the immediate prime, (b) mean number of segments shared between responses and immediate prime, and (c) mean number of shared segments in identical position with respect to the wordform's beginning, end, or stressed vowel.

Figure X.1 shows the results. In each panel, the x-axis plots the mean semantic relatedness rating between the response and the immediate prime, and the y-axis plots some measure of wordform relatedness. For example, the top left panel shows the proportion of responses that shared a phoneme with the immediate prime. Note that this count disregards position in the wordform, so that the non-initial "k" symbols in "book", "subscribe", "check", and "smoking" and the initial ones in "credit", "keep", "caramel", "cancer", etc., all count as equivalent to the initial "k" of the immediate prime in the "renew, cancel, _____" and "camel, cancel, _____" priming pairs. The distribution of filled and open circles suggests the robustness of *l'arbitraire du signe* as a principle organising the remembered properties of these words. The proportion of responses sharing a segment with the target word is generally somewhat higher for the wordform associate pairs, whereas the mean semantic relatedness rating is generally higher for the same words in lemma associate pairs. When the clear and filled datapoints for each immediate prime are compared to each other, the proportion of responses containing shared segments is shown to be higher in the wordform associates condition for all but four items, and the mean semantic relatedness rating is higher for the responses in the lemma associates condition for all 24 items. (Both of these differences, as well as the effects described below for the other panels of Figure X.1, are significant at $p<0.0001$, by a Wilcoxon test.)

The y-axis in the lower left panel shows a somewhat finer measure of phonological relatedness—the mean number of segments in common between the response and the immediate prime, again without regard to position in the wordform. The responses to the two types of associates become better differentiated. When a priming pair prompted a response with any phoneme in common with the immediate prime, that was typically the only phonological similarity for lemma associates, whereas there were typically more segments in common when they were wordform associates. This suggests that duality of patterning also plays a role. Not only are wordforms remembered as separate entities from their lemmas, they are also parsed into smaller phonological elements such that degree of phonological association or activation is not all or none.

The last panel in the figure refines the measure of wordform similarity further by counting a segment in the response as identical to a segment in the immediate prime only if it occurred in the same position relative to the stressed vowel or to either of the wordform's edges (except when counting blindly from the edge would equate prevocalic and syllabic consonants as in "breakfast" as a response to "syrup, sermon, _____"). When phonemes in common are tabulated using this positionally-specific criterion, there is even less overlap between filled and open circles, suggesting that positional contrast constrains associations between phonologically-related items in memory. Links among wordforms are sensitive to the forms' internal syntactic structure, so that the most accessible paradigmatic associations are between elements that share not just gestural content but also syntagmatic position.

X.3. Lexical neighbourhoods and the emergence of structure

A useful metaphor for understanding these three results combines two ideas. One is the idea of spreading activation as a way to model retrieval of information from long-term memory (Morton, 1969). The other is the idea of the lexical neighbourhood as a way to account for the roles of semantic and phonological similarity in the spread of activation through the relevant part of the long-term memory store (Goldinger *et al.*, 1989). The composite "neighbourhood activation" metaphor developed in the 1980s as psycholinguists working on lexical access explored ways around the impasse of trying to reconcile the older metaphor of dictionary lookup with experiments suggesting multiple entry points to the lexicon. If words are stored in memory as a one-dimensional list organised by order of relative lexical frequency, one can explain the advantage that more frequent words have in naming and lexical decision tasks by the relative amount of time it takes for the lookup mechanism to reach a word that is entered earlier or later in the list. But then there is no easy way to explain semantic priming. By contrast, if the lexicon is organised along the principles of a thesaurus, with ample and easily traversed cross-reference links, then each word is a cell defined by the intersection of all of the links in this multi-dimensional space. And one can explain semantic priming as the result of increased activity in one cell activating all the connected neighbouring cells. Moreover, if phonologically similar words are also connected together to make a phonological neighbourhood for each wordform, then phonological priming is no more miraculous than semantic priming. As Dell (1999) puts it, "processing a word affects its neighbours, and vice versa. It is impossible to sneak into the lexicon and take out the word you want without the neighbours getting all excited." Rousting a word out of its bed requires a ruckus that wakes up the rest of the neighbourhood.

The development of the neighbourhood activation metaphor was a critical step in the application of connectionist models to the understanding of lexical access in several different linguistic domains. Some of these applications, such as Dell (1988), build on Morton's original conception (and the description above) in identifying words as actual cells or "nodes" in the connectionist network. Others, such as Bybee (1999), instead equate the nodes of the network only with the shared properties of words (i.e. conceptual features such as "past tense" or phonological properties such as "having the phoneme /k/ in initial position"). Words then are represented only as distributed patterns of activation across the associations between these primary feature category nodes. The first conception seems more consistent with our results, and there are a functional argument and a formal (architectural) argument in favor of it.

Interpreting the first result of our experiment in terms of Dell's (1999) description of the neighbourhood activation metaphor suggests the following elaboration. The two sets of neighbours—wordform and lemma—are connected via the word. Picture the word as a rowhouse with a kitchen (the wordform) and a parlor (the lemma). The parlor side of the word sits on a street with one set of neighbours, and the kitchen side of the word sits on an alley with a different set of neighbours. Recognizing a wordform in the stream of speech means that the revellers (the activity invoked by recognizing phonetic cues and phonological categories) have passed through the kitchen and out through the parlor door to the word's lemma neighbourhood. Producing a word, conversely, means that the revellers in the lemma neighbourhood have passed through the parlor and out the kithen door into the word's phonological neighbourhood. The only way that the neighbours on one side of the house can hear the ruckus in the other neighbourhood is if the word opens the front door and the service door to let the mob of revelers pass through. Each word is its own little bottleneck for the spread of activation. A synonym for "bottleneck" is "channel"—which does better justice to the advantages of this structural design. By localising structure in this way, the spread of activation is effectively channeled. If no close neighbour is rousted out of bed by mistake to open her parlor and kitchen doors, the reveling crowd will pass through the correct rowhouse into a smaller neighbourhood on the other side. The intended lemma will be recognised or the intended wordform produced without requiring extensive soothing of the word's closest wordform and lemma neighbours. The fact that slips of the ear or tongue are more common for words that are both lemma and wordform neighbours thus suggests the ecological advantage afforded by *l'arbitraire du signe*.

(a)	[FEMALE]	[SIBLING]	[LABIAL CLOSURE]	(b)	input		XOR output
mother	1	0	1		1	0	1
father	0	0	0		0	0	0
sister	1	1	0		1	1	0
brother	0	1	1		0	1	1

Figure X.2. (a) Fragment of semantic-to-phonetic mappings for words in a semantic neighbourhood. (b) Input-output mappings for the possible inputs to the logical function exclusive-OR;

As Dell et al. (1997) point out, the formal characterization of this bottleneck is very similar to the characterization of the XOR ("exclusive or") function, as illustrated in Figure X.2. The two wordforms that start with a labial closing gesture correspond to the lemmas for the female parent and the male sibling. If semantic properties [FEMALE] and [SIBLING] are the input, the labial-closure output must be activated for the input that is either a female or a sibling, but not both. As Elman (1990) has pointed out, the XOR function requires that a layer of hidden nodes intervene between the input and the output. To model *l'arbitraire du signe* using a connectionist architecture, similarly, there must be hidden nodes—i.e. words as discrete, local cells—in between the nodes that encode semantic and syntactic properties (lemma features), and the nodes that encode phonological and phonetic properties (wordform features).

These word nodes are needed so that the patterns of co-activation on the word's lemma side can be represented by changing the weights of the lemma associations without affecting the weights of the wordform properties that trigger co-activations of phonologically similar wordforms. Examination of the hidden nodes after training should show the emergence of dedicated wordform and/or lemma nodes, representing linguistic structure at this level. After sufficient exposure, co-activation patterns for words should be localised enough to support robust processing on just the relevant side of the bottleneck. In this way, a new (or less frequent) word can be produced or perceived by activating the subnetwork of intersecting properties that define its wordform, without activating the lemma neighbourhoods of the closest attested (or more frequent) words. The child can recognise the wordform as novel, and learn its associated semantic and syntactic properties without undue interference from the lemmas of similar wordforms.

X.4. The bottleneck between articulatory and acoustic representations

As we have just seen, duality of patterning is ecologically functional. It results in channeling of activation at the word level between wordform properties and the lemma properties. However, this is not the only channeling in the system. In this section, we argue that channeling between articulatory and acoustic representations is one source of sublexical units such as the phoneme. Figure X.3 illustrates the formal similarity between this bottleneck and the one discussed in the previous section. We posit channeling between acoustic and articulatory representations because research efforts to establish an isomorphism between them have generally failed. Several types of problems have arisen.

One is the clear differences between the most accessible semantic properties of some categories. The phonetic correlates of some features, such as [±cont], [±voice], or [±nasal], can be defined articulatorily in terms of a constriction degree "gesture" at some oral aperture, the glottal aperture, or the velic aperture. The phonetic correlates of the feature [±sonorant], on the other hand, cannot be unified in terms of any single articulatory constriction. Rather, the feature refers to the "effective constriction degree" (Browman & Goldstein, 1989), the acoustic consequence of how the resulting airflow channels are coupled in the gestural ensemble as a whole. Similarly, because many vowels can be implemented using more than one articulatory strategy, traditional vowel features of height and frontness are highly problematic as articulatory features. Vowel categorization in the acoustic domain is more perspicuous.

	[cont]	[voice]	[nasal]	[son]
/b/	0	1	0	0
/m/	0	1	1	1
/w/	1	1	0	1
/,/	1	0	0	0
	lip aperture	glottal aperture	velic aperture	effective const. deg.

Figure X.3. Logical ststructure for sample mappins from articulatorily-motivated features describing close or critical constriction degree values to an acoustically-motivated feature describing effective constriction degree for 4 gestural ensembles.

More generally, the relation of acoustic to articulatory levels of description is, as Nearey (1995, this volume) points out, "doubly weak". This means that each entity in the acoustic description corresponds to more than one entity in the articulatory description, and vice versa. Because of this double weakness, it is impossible to establish the kind of invariant, or one-to-one relationship between acoustics and articulation which many speech researchers previously hoped to find. To model this relationship in a connectionist architecture, it is necessary to posit a layer of hidden nodes. This layer of hidden nodes is, we would argue, the segmental level. If the relationship of acoustics to articulation were simple, there would be no impetus for anything as abstract as the phoneme.

Stops provide an especially clear example. In making or watching someone else make a /b/, the somatosensory or visual representations of oral and glottal constriction degree and of oral constriction location are continuously available, but during the closure, the spectral dynamics do not tract the articulatory dynamics. Building a layer of hidden nodes to stand between these disparate stimulus spaces allows the auditory cues to associate to the temporally localised gestural ensemble, and to do so even when the preceding or following environment is not conducive for carrying place or voicing cues. In other words, one of the reasons why an infant might acquire phonemes is that the articulatory-to-acoustic mapping is not invertible. (See Plaut & Kello, 1999, for a pilot implementation of this idea.) Thus, the articulatory-to-acoustic bottleneck is relevant to the bootstrapping question: How can localised representations of sublexical structure emerge for the infant to begin acquiring words? It is by now well-established that perceptual responses begin to be attuned to the phonological categories of the ambient language well before the infant begins to make associations between recurring wordforms and their lemmas (e.g., Kuhl et al., 1992). How can this attunement occur before the infant has a large enough lexicon for the wordform-lemma bottleneck to enforce an analysis of similarities finer than identity across the whole utterance?

X.5. The third hidden layer

A third bottleneck also is relevant to the bootstrapping question. Sublexical structure emerges not just because of the double-weak relationship between the articulatory and acoustic properties of consonants and vowels, but also because speech unfolds in time. This means that many of the contextual (syntagmatic) dimensions of the conceptual space for speech have as their semantic extensions such inherently temporal relationships as "occurs after" or "is contained within". We have already outlined above how categorization in intelligent animals such as the bear is more robust than Shepard's simple stimulus generalization model, because the intermediate conceptual space allows the organism to assess the likelihood of a category relative to apprehended properties of its context. For the syntactically-structured categories of speech, then, the relevant contextual properties include the category types that are identified for preceding time intervals and for larger time intervals containing the target stimulus. Thus, a good way to characterise the acquisition of the syntagmatic categories relevant for speech is that the child is "Finding structure in time" (Elman, 1990).

Devising some way to model structure in time is one of the most challenging problems in applying connectionist models to speech. The formal similarity to the XOR problem is illustrated in the two mappings in Figure X.4. The first shows how the context vowel category constrains the interpretation of the spectrum during a sibilant fricative. The center of gravity is in the ambiguous middle region of the speaker's range (where the listener must attend carefully to its precise value) when the consonant is likely to be /s/ or the vowel is likely to be /u/, but not both. The second example shows how contextual categories can not only lead to more efficient attentional strategies for processing the semantics of the target stimulus, but also can constrain expectations about the identity of the target category more directly. Given the transitional probabilities for English obstruents and liquids, an alveolar cluster is a likely onset when the first consonant is a stop or the second is a lateral, but not both. While Figure X.4(b) illustrates the point with an extreme case (in English, /sr/ and /tl/ have a near zero probability of occurring without an intervening morpheme boundary), there is a growing body of research showing that such "absolute" phonotactics are qualitatively similar to the effects of more gradient differences in transitional probability. (See Pierrehumbert *et al.*, this volume, for a review.)

(a)	/s/	/i/	mid-frequency mean spectrum
see	1	1	0 (high)
sue	1	0	1
she	0	1	1
shoe	0	0	0 (low)

(b)	C1 is /t/	C2 is /l/	likely onset
/tl/	1	1	0
/tr/	1	0	1
/sl/	0	1	1
/sr/	0	0	0
/kl/	0	1	1

Figure X.4. Logical structure for sample mappings (a) from context classifications to expected acoustic property, and (b) from context classifications to expected following paradigmatic category.

Elman (1990) shows that a recurrent neural network (RNN) connectionist model can learn some complex syntactic systems by exploiting such transitional probabilities. The RNN architecture stores a temporary buffer of "context units" to build associations between immediately preceding elements and the input element currently being processed. In the English syntactic subsystem illustrated in Figure X.4(b), for example, the low probability of the /tl/ sequence can over-ride the acoustic semantics of the alveolar stop burst, to cue a /kl/ instead of the intended /tl/. The RNN architecture can model such structure in time, as long as the transitional probabilities that are exploited to predict the next element in a sequence are stated over input and output units that are of the same structural type—e.g. using the previous one or two words to predict the next word in a phrase or the previous one or two segments to predict the next segment (Elman, 1990). If transitional probabilities for phoneme sequences contained within wordforms differ systematically from those that cross word boundaries, these probabilities can be used to generalise the syntax of the composite phonological structure. Thus, phonotactics can be exploited to establish the probable word count and the most probable points for the edges of words in continuous speech.

Many recent experimental findings demonstrate the importance of transitional probabilities in language acquisition as well as in adult speech processing. For example, Jusczyk *et al.* (1994) show that by 9 months, infants are sensitised to word-internal transitional probabilities for the ambient language. That is, already at the point when they begin to learn arbitrary wordform-lemma associations, they have established conceptual structures that are relevant for phonological parsing. These structures allow them to interpret novel wordforms by using phonotactic generalizations over remembered prior experiences of particular wordforms. Thus, another reason why there must be phonemes is that many of the transitional probabilities that establish the phonotactics of possible words in the language involve syntactic objects that are sequenced at this grain of time. The existence of robustly localised memory structures at this temporal grain can explain why the effectiveness of phonological priming is systematically gradient and why the size of the effect can be measured by counting the number of phonemes in common between the prime and the response. This is in keeping with the results of our association test, and of word identification experiments in Pisoni *et al.* (1985).

Of course, the infant typically is presented with other systematic variation that correlates with positionally specified transitional probabilities, variation that helps him to pick out likely places for a wordform to begin. For example, in many languages the acoustic semantics of stop-vowel sequences are systematically differentiated between initial and other positions in the wordform (Keating *et al*. 1983). Some of this semantic variation picks out structural positions defined on elements that can have a larger or smaller temporal grain than the wordform—e.g., elements such as the accentual phrase or syllable in Korean (Jun, 1998) or the stress foot in English and German (Keating, 1984). In connectionist models of lexical access, we might expect to see more or less localised representations of these other structures in the hidden nodes as well. (The separate nodes for prosodic “frames” and the positionally-specific nodes for consonant “phonemes” in Dell’s, 1988, model are an example.)

Infant perception research supports the idea that such hidden nodes will emerge early in speech acquisition. Saffran *et al*. (1996) show that a 2-minute exposure to an extended synthetic passage containing trisyllabic CVCVCV nonce wordforms is all that is required to sensitise 8-month-old infants to different transitional probabilities for CV syllables within versus between wordforms. However, infants are sensitised to syllable count considerably earlier than this (Bertoncini *et al.*, 1995), no doubt because of the typically very salient alternation in effective constriction degree that defines the canonical CV syllable. English-acquiring infants also show a well-established sensitivity to the alternation between strong and weak syllables and to the differing transitional probabilities for these two syllable types between and within wordforms (Jusczyk & Aslin, 1995). They do so several months before they begin to be sensitive to the transitional probabilities of segment sequences for the language. Thus, well before the infant has acquired any words qua words, structures are emerging in memory for the typical prosodic word "frames" of the ambient language. These structures will channel the child's attention so as to efficiently exploit positional allophony in remembering and accessing the phoneme strings of the thousands of wordforms that the child will learn in the first few years of life. The results of our association test suggest that positionally-specific semantic properties of wordform-internal elements remain an important component of the definition of a wordform's phonological neighbourhood even in adulthood. The phonological priming effect of the wordform associate pairs emerged most clearly when we counted only identical segments that also stood in identical positions relative to such salient points as the word's stressed vowel.

X.6. Granularity effects and reductionism

As the above discussion suggests, we believe that an important component of language sound structure is the channeling of information flow between different dimensions of the sensory and classificatory spaces. This channelling accounts for the appearance of "autonomy" or "modularity" that both the Structuralists and the later Generative Phonologists have noted. That is, there are granularity effects at each of the three bottlenecks described above, which structure the representations available for different phonological tasks. In looking for adequate formal models of these effects, it is important to avoid false reductionism. In particular, it is important to resist the temptation to try to explain all of the phonological generalizations that are available to the child in terms of the observed or expected granularity effects at just one of the bottlenecks.

For example, the exigencies of lexical contrast interact with the discontinuities in the mapping from articulation to acoustics to constrain probability distributions in the articulatory space (Stevens, 1989). This provides a natural, universal basis for bootstrapping into the phonemic structure of the lexicon. It is mistake, however, to try to reduce generalizations about segmental structure to the universal aspects of these constraints. Excessive reductionism in this direction results in the common misapprehension that the sequential elements that must be conceptualised to acquire a particular lexicon are properties of the signal per se—i.e., that discrete segments and their feature specifications pre-exist in nature. This makes it difficult to appreciate the role of the lexicon in structuring the phonological categories that the child sets up in order to reliably apprehend and reproduce the wordforms of the ambient language. Put simply, different languages use different phonological inventories to make up wordforms. For almost any phonetically robust contrast, it is possible to find a language in which that contrast is not exploited. But many very marginal contrasts are exploited in at least a few languages.

To see the false reductionism here it is important to appreciate why the child should abstract away discrete segments. In the preceding section, we outlined the advantages for the infant who is first learning to pick out recurring acoustic patterns in the ambient stream of speech. This first impetus toward abstraction will be reinforced when the child later on begins to associate the perceived acoustic patterns with his emerging word-motor schema. Consider, for example, the gestural dynamics that differentiate English *cap* from *cab* and *seat* from *seed*. The labial versus alveolar closing gestures harness different articulatory subsystems. But there are useful generalizations to be abstracted away from that difference in place of oral constriction. Some of these generalizations involve the timing of the oral gesture relative to the laryngeal and pharyngeal postures that will either inhibit or promote the continuation of voicing into closure. For English, however, an even more important set of generalizations involves the stiffness of the oral gesture and its timing relative to the oral target for the preceding vowel. The English-acquiring child learns the appropriate motor dynamics and the mapping to the resulting acoustics, and in the space of a few years is able to parse, remember, and reproduce novel wordforms such as *seep* and *cad* even after just one encounter with them. Making the right generalizations for this "fast mapping" (Carey, 1978) means that the properties and events in the articulatory and acoustic stimulus spaces are partitioned into those that belong to the vowel and those that belong to the following stop. This partitioning is specific to the language. The syntax and semantics of the voiced versus voiceless final stop contrast in Hindi, for example, differ considerably. Unlike in English, voicing is maintained during closure for Hindi voiced stops even in utterance-final position. Also, vowel length is phonologically contrastive, and is not co-opted to be a phonetic cue to the voicing category of the follow stop.

Another advantage of partitioning the signal into the properties that come from a vowel and those that come from an adjacent consonant is that the child can class together rather different acoustic properties or events in terms of the similar gestural ensembles that produced them. For example, the young child can parse the spectral distribution of energy in an /s/ before /u/ into the contribution of the consonant constriction (coupled with the velic closure and glottal opening) and the contribution of the rounding from the contextual vowel. Moreover, the older child can do so even for a synthesised fricative (Nittrouer, 1992).

The advantage of this syntagmatic partitioning becomes even more apparent when a familiar wordform is encountered in a novel sentence context. For example, the English-acquiring child can recognise and fluently mimic the different instantiations of /t/ at the end of *put*—with a lateral release in *Put lettuce on the list*, no release in *Put juice in the cup*, a strident release in *Put your feet on the floor*, and a hyperarticulated alveolar burst in *I said PUT the forks on the table, not THROW them.* The conceptual structures that arise from well-rehearsed parallel associations across the articulation-to-acoustic bottleneck, as the child learns to produce and perceive *put* in all these contexts, supports the common phonological classification of the word as invariably ending with a /t/. Moreover, abstracting the /t/ away from the preceding vowel allows the generalizations about variable acoustic patterns to be applied to correctly perceive other wordforms that end in /t/ in analogous contexts, and thus to infer the articulatory structure of their rhymes. The nonlinearities in the mapping between the articulatory and acoustic dimensions of the phonetic space constrain the partitioning of these spaces, but they do not fully determine it. The listening child cannot apprehend articulatory structure that is never supported by the acoustics, but he can adjust attentional weights to glean all relevant available information from the signal. The speaking child, conversely, can organise her articulations to insure that others apprehend the intended wordform.

Excessive reductionism here stems from focusing too narrowly on the structure that nature provides in this interchange of information between the speaker and listener. Because of the discontinuities in the mapping between the articulatory and the acoustic spaces, the language-user typically has a finer-grained phonetic representation of her own utterances than of another's utterances. This difference in granularity across that bottleneck is sometimes mistaken for evidence that one or the other set of phonetic dimensions is somehow more primary in the representation of the wordform—for example, to argue that lexical representations need refer only to acoustic properties, because the probability distributions along articulatory dimensions are determined entirely by the nonlinearities (e.g. Stevens, 1989).

This kind of reductionism often is associated with a particular type of modular processing model, whereby the categorical specifications for the "primary" (i.e., acoustic) categories are retrieved from lexical memory and transformed into independent control parameters for the secondary articulatory space in the course of producing or apprehending an utterance. When two "natural" categories in the acoustic space are not contrastive in a language, however, the discretised probability distributions in the articulatory space are assumed to still be available for the production module to use in the independent categorical control of "enhancing" features (Stevens *et al*., 1986). This type of model locates the discretisation into segments and features entirely at the articulatory-to-acoustic bottleneck that nature provides, and attributes to the lexicon only the function of determining whether a particular (naturally discrete) acoustic feature is distinctive in the language. Because it effectively ignores the role that the third bottleneck plays in discretising the phonetic space, the model fails to predict the different syntagmatic organizations (the different patterns of "coarticulation") that acquiring the lexicon promotes for children acquiring different languages (see Manuel, in press, for a review). Trying to explain all of phonetic structure in terms of the natural granularity of the acoustic space makes it difficult to appreciate that articulatory-to-acoustic nonlinearities do not deterministically partition the articulatory space.

Reductionism at this level, however, is not limited to models that take the acoustic space as primary in lexical representation. It also characterises Fowler's Direct Realism (Fowler, 1990). In this framework, phonological elements and their associated phonetic properties are equated with macroscopic versus microscopic levels of control for skilled movement. Phonemes in relationship to their phonetic extensions are homologous to different reaching tasks in relationship to the specification of degrees of joint rotation at the shoulder, elbow, and wrist. There is a similar claim about lexical memory in Articulatory

Phonology. But Direct Realism goes further in also making strong claims about perception and word recognition. Gestures can be perceived directly. One can perceive the gestures of the /t/ just as one perceives the movement of a door, when one hears the sound of it being slammed shut. The articulatory gestures that compose a wordform are transparently available in the acoustic signal, and no language-specific model of the other's intentions is required to apprehend their structuring of the acoustics.

This framework cannot account for the way in which the acoustic robustness of lexical contrast shapes articulation. Where the developers of Articulatory Phonology acknowledge that acoustic events and properties that are distinctive for the ambient lexicon play some role in organising the articulators into gestures and gestural scores (Goldstein, 1989; Browman & Goldstein, 1989: 226), Fowler must assume a drive toward granularity that is not ecologically situated in lexical memory. Because she does not recognise gestures as syntactic objects, her model cannot account for the effects on the conceptual motor space of learning a particular set of wordform-lemma associations. It cannot account for the fact that gestures and gestural scores reflect the particular set of syntagmatic and paradigmatic categories that come from generalising acoustic and articulatory patterns across the wordforms of a particular language. Like Chomsky (1986), Fowler must assume a predisposition for language that is purely structural, and not built on the more general cognitive functions of symbolic categorization.

X7. Granularity effects at the first bottleneck

In the preceding section, we described how well-rehearsed parallelism in the associations across the articulation-acoustics bottleneck interacts with differential transitional probabilities to give rise to (often highly language-specific) patterns of coarticulation and positional allophony. There are homologous granularity effects at the wordform-lemma bottleneck. Well-rehearsed parallelism across the *signe-signifiée* divide interacts with lexical frequencies to give rise to more or less productive morpho-phonological patterns.

A particularly well-studied case is the regular past-tense affix in English. Work by Marchman & Bates (1994) and others suggests that the child acquires the regular, productive pattern by generalising from parallels in wordform and lemma neighbourhoods across many different present-tense and past-tense wordforms. In a connectionist model, this generalization can be made if the wordform-lemma associations are encoded as activation patterns over separate sets of hidden nodes for lemmas and for their wordforms. In this way each regular verb contributes to the type frequency that establishes the coronal stop affix as the more frequent pattern despite the higher lexical token frequencies of many irregular verbs in English. That is, the parallel association between a past-tense and a present-tense verb lemma on one side of the divide, and between the shorter and longer wordforms on the other side of the divide, promotes the development of an independent but related morpho-phonological dimension of categorization. The child learns to recognise the ensemble of gestures at the ends of utterances of the wordforms *played, planned, cried, laughed, skipped,* and so on, not only as tokens of the phoneme types /d/ and /t/, but also as tokens of the regular past tense affix.

The granularity effects here stem from the relationship between allomorphy and sound change. The phonetic reductions in Old English that led eventually to the loss of the vowel in many post-tonic syllables juxtaposed the voiced coda of what was a fully syllabic affix against a root-final consonant. The synchronic pronunciations of *laughed* and *skipped* treats these wordforms in conformity with the otherwise nearly exceptionless phonological generalization that within words, obstruent sequences agree in voicing. To make the most useful paradigmatic and syntagmatic partitions of the articulatory and acoustic spaces on the wordform side of the divide, the child should categorise the endings of *laughed* and *skipped* together with the endings of *left*, *apt*, *plant*, and *plate* and distinguish them from the endings of *played* and *planned.* To make the most useful morphosyntactic partitions on the lemma side, on the other hand, the child should categorise *laughed* and *skipped* together with *played* and *planned.* Because of the way that the first type of categorization molds attention and discretises the motor control space, the dissimilarities that obtain between different positional allophones of a phoneme are typically more fine-grained than those between the allomorphs of a morpheme. Where English word-medial /d/ and word-initial /d/ differ merely in "subphonemic" detail—the one being "redundantly voiced" (Jakobson, Fant & Halle, 1952) relative to the other—the shapes of the past tense affix in *played* versus *laughed* differ "phonemically".

Classical Generative Grammar tries to explain both types of categorization in terms of the same device—transformational rules operating on minimal phonological representations of wordforms in the lexicon. The child stores the different allomorphs of the past tense affix as a single lexical entry, which is productively added to (or parsed away from) the verb root each time the past tense form is produced (or perceived). In thus attempting to reduce allophony and allomorphy to the same phenomenon, accounts of specific languages often are forced to impute to the native speaker, extremely abstract and potentially unlearnable underlying representations. (For elaborations of this point, see Broe, 1993; Odden, 1992; Steriade, 1995.) Equating allophony with allomorphy is an example of false reductionism. The false reductionism here makes it difficult to account for performance in lexical recognition tasks that differentiate the native speaker's knowledge of the phonological categories from his or her knowledge of the morpho-phonological categories.

The difficulty comes out especially clearly in the treatment of "incomplete neutralization"—i.e. cases where a differentiation between two phonological classes for one set of wordforms is not robustly supported by the phonetics, but is nevertheless maintained by salient associations on the lemma side to another set of wordforms in which the differentiation is robustly supported by the associated phonetic properties. For example, the contrast between /d/ and /t/ in word-medial position in German wordforms such as *Bünde* versus *bünte* supports the maintenance of very subtle "subphonemic" differences between the associated forms *Bund* and *bunt*. Native speakers' sensitivity to these differences can be uncovered in simple identification tasks (Port & O'Dell, 1985). The differences apparently can be exaggerated in some discourse contexts (Charles-Luce, 1985). Dressler (1985: 93) also reports a "clear" differentiation by many speakers of syllable-final obstruents in abbreviations such as *Log* 'logarithm' vs. *Lok* 'locomotive', suggesting that the parallel associations across phonological and semantic neighbourhoods need not involve obligatory morpho-syntactic categories to support morpho-phonological generalization.

In a model of lexical representation that reduces allophony and allomorphy to a single transformational process, the phonologist is forced to choose between the "underlying" and the "surface" consonant to encode the speaker's lexical representation of each wordform. The fact that morpho-phonological generalizations can sometimes result in finer-grained rather than coarser-grained categorization is an embarrassment (e.g. Manaster Ramer, 1996). A model of lexical representation that places the grammar square within the lexicon, by contrast, can easily explain the different granularity effects. It does so by distinguishing between categories that are made by generalising over patterns of phonetic similarity among all wordforms in a phonological neighbourhood and categories that are made by generalising over patterns of similarity among sets of wordforms that are also associated on the lemma side. Given a frequent enough rehearsal of the relevant associations, very detailed phonetic representations of the "underlying" category and of the "surface" category can co-exist in a single speaker's mind.

This understanding of how lexical memory is structured might also help to explain the ways in which lexical frequency affects both regular sound change and analogical leveling. Research on sound changes in progress has shown that a regular sound change first yields competing variant pronunciations for a segment in a particular phonological context in one set of words, and then spreads through the communal lexicon in ways that are influenced by the relative lexical frequencies of the individual wordforms. This influence exists because a more frequent target word presents speakers with more opportunities to exercise an analogical extension of the pattern of variation from the originally affected words. If the conditioning context is internal to the wordform, a more frequent word is affected sooner because it provides both the target and the context each time it is produced. If the conditioning context is external to the word, a more frequent word is again affected sooner, because its chances of being produced in the relevant external context are also higher (Bybee, 1999).

When the candidate set of wordforms is related also on the lemma side of the bottleneck, by contrast, the pattern of frequency effects is rather different. The spread of the change through the lexicon is conditioned by the frequency of the parallel association—i.e. the frequency of the pattern of phonological relationships between wordforms that are also linked by another kinds of relationship. This conditioning is most obvious in sets of inflectionally related wordforms that express obligatory morpho-syntactic categories. The conditioning of a sound change by such robustly localised parallelism leaves traces that can be interpreted as "analogical leveling" even millennia after the sound change has spread through the community. A less frequent pairing of associations across the wordform-lemma divide will leave less clear traces, in keeping with the less robustly localised representation of the joint associations.

An example is the analogical change resulting in the American pronunciations of adjectives such as *hostile* and *ductile*, with a weak /ˈl/ or syllabic /l/ in place of the strong /ajl/ of standard southern British. This change must have been conditioned by the association to another wordform ending in /IlIti/—i.e., *hostile* is associated with *hostility*, *ductile* with *ductility*, and so on. Phonologically related wordforms such as *profile* and *textile*, which have no parallel association to a noun ending in *-ility*, are unaffected. To understand this change, we note the low frequency of the analogical link between an originally tense vowel and a lax vowel in these *-ile/-ility* form pairs relative to the higher frequency of the link between an originally reduced vowel in the adjective and the lax vowel in the related noun in *civil-civility*, *national-nationality*, *noble-nobility*, and many other pairs involving the suffixes *-ible* and *-able*, as in *navigable-navigability*. A search of such adjective-noun pairs in the *Hoosier Mental Lexicon* found more than six times as many pairs in which there is an originally reduced vowel in the adjective. The very high frequency of the linking of a reduced vowel in the adjective with a lax vowel in the noun apparently overwhelmed the less frequent correspondence, so that the reduced vowel replaced the original diphthong in most of the adjectives ending in *-ile*. This change affected many very infrequent words such as *ductile*, *contractile*, and *motile* (each with only 1 occurrence per million words in the Kucera-Francis corpus), and was not limited to the higher-frequency adjectives such as *hostile* (at 19 ppm) and *mobile* (at 44 ppm). Conversely, the three adjectives in these pairs for which the *HML* gives the diphthong as the preferred American reading for the vowel in *-ile* also span the range of frequencies for the set. They are *puerile* (1ppm), *senile* (2ppm), and *juvenile* (18ppm). However, the change did not affect other phonological neighbours such as *profile* and *textile*, which are at least as high in frequency as *hostile*, but are not adjectives paired to a noun form ending in *-ility*.

Thus, phonological pattern frequency plays a role in the spread of a regular sound change through the phonological neighbourhood. It also plays a role in analogical leveling—although the frequencies at play are not the same. In the first case, the associations among words that contain the target pattern for the change are all on the wordform side of the *signe-signifiée* divide. Here it is the frequency of the word itself that affects its susceptibility to change. In the case of analogical leveling, by contrast, there are crucially relevant associations across the wordform-lemma divide. Here it is the frequency of the parallel correspondences in the two different "semantic" domains that shapes the outcome. The fact that frequency affects both types of changes highlights the common cognitive organization. By understanding what it means to say that phonological categories are syntactic objects, we can recognize the "purely phonological analogy" that makes regular sound change regular. The associations in lexical memory that are involved here are not qualitatively different from the associations that drive "irregular" morpho-phonologically conditioned changes. There is "syntactic structure" involved in both types of change; the only difference lies in whether the relevant syntagmatic associations lie only in the wordform neighourhood or also in the lemma neighourhood.

This affinity between the two types of change was impossible to capture in older Structuralist models, where a completely autonomous phonology was encapsulated away from the morphology. However, the exact nature of the affinity is just as difficult to capture in many, more recent phonological frameworks. The different frequencies that are at play cannot be predicted by any framework which reduces the granularity effects at the wordform-lemma bottleneck to a distinction between underlying and surface forms (King, 1969) or between lexical and postlexical rules (Kiparsky, 1995).

In summary, the multiple dimensions along which subparts of wordforms can be classified together or separately cannot be reduced to a neat hierarchy of derivational strata, each operating at a different grain of specification. Phonetic interpretation is a set of indexical associations between parts of wordforms and their semantic properties. These associations provide real-world extensions for all syntactic objects and functions on the wordform side of the wordform-lemma bottleneck, not just for the last set of phonological categories spewed out by a grammar of rules or ordered constraints. Structure is not learned in isolation from content. Children build on the semantics on both sides of the wordform-lemma divide in constructing the grammar and other conceptual structures appropriate for the culture that they come to share.

References

Bertoncini, J., C. Floccia, T. Nazzi & J. Mehler. 1995. Morae and syllables: rhythmical basis of speech representation in neonates. *Language and Speech* 38: 311-329.

Broe, M. 1993. *Specification Theory: The Treatment of Redundancy in Generative Phonology*. Ph.D. Dissertation, University of Edinburgh.

Browman, C. P. & L. Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6: 201-251.

Bybee. J. 1999. Lexicalization of sound change and alternating environments. In M. Broe & J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, pp. 249-267. Cambridge: Cambridge University Press.

Carey, S. 1978. The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (eds.) *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Preager.

Charles-Luce, J. 1985. Word-final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13: 309-324.

Dell, G. S. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27: 124-142.

Dell, G. S. 1999. Commentary: Counting, connectionism, and lexical representation. In M. Broe & J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, pp. 334-348. Cambridge: Cambridge University Press.

Dell, G. S., M. F. Schwartz, N. Martine, E. M. Saffran and D. A. Gagnon. 1997. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104: 801-838.

Dressler, W. U. 1985. *Morphonology: the Dynamics of Derivation*. Ann Arbor: Karoma.

Elman, J. L. 1990. Finding structure in time. *Cognitive Science* 14: 179-211.

Fowler, C. A. 1990. Comments on the contributions by Pierrehumbert and Nearey. *Jounal of* Phonetics 18: 425-234.

Goldinger, S. D., P. A. Luce & D. B. Pisoni. 1989. Priming lexical neighbors of spoken words: effects of competition and inhibition. *Journal of Memory and Language*, 28: 501-518.

Goldstein, L. 1989. On the domain of quantal theory. *Journal of Phonetics* 17: 91-97.

Hockett, C. D. 1960. Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.) *Animal Sounds and Communication*, pp. 392-430. Washington, DC: American Institute of Biological Sciences.

Jakobson, R., G. Fant and M. Halle. 1952. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

Jun, S.-A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15: 189-226.

Jusczyk, P. W. & R. N. Aslin. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29: 1-23.

Jusczyk, P. W., P. A. Luce & J. Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Verbal Learning* 33: 630-645.

Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286-319.

Keating, P. A., W. Linker & M. Huffman. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11: 277-290.

King, R. D. 1969. *Historical Linguistics and Generative Grammar*. Englewood Cliffs, NJ: Prentice-Hall.

Kiparsky, P. 1995. The phonological basis of sound change. In J. A. Goldsmith (ed.) The Handbook of Phonological Theory, pp. 640-670. Oxford: Blackwell.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. Stevens & B. Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 225: 606- 608.

Levin, B. & M. Rapaport Hovav. 1996. Lexical semantics and syntactic structure. In S. Lappin, ed., *The Handbook of Contemporary Semantic Theory*, pp. 487-507. Oxford: Blackwell.

Malt, B. (1996). Category coherence in cross-cultural perspective. *Cognitive Psychology* 29: 85-148.

Manaster Ramer, A. 1996. A letter from an incompletely neutral phonologist. *Journal of Phonetics* 24: 477-489.

Manuel, S. In press. Cross-language studies of coarticulation. To appear in W. Hardcastle and N. Hewlett (eds.) *Coarticulation*. Cambridge: Cambridge University Press.

Marchman, V. A. & E. Bates. 1994. Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language*, 21: 339-366.

Morton, J. 1969. The interaction of information in word recognition. *Psychological Review* 76: 165-178.

Myung, I. J. & R. N. Shepard. 1996. Maximum entropy inference and stimulus generalization. *Journal of Mathematical Psychology* 40: 342-347.

Nearey, T. M. 1995. A double-weak view of trading relations: comments on Kingston and Diehl. In B. Connell & A. Arvaniti, eds., *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, pp. 28-40. Cambridge: Cambridge University Press.

Nearey, T. M. this volume. EDITORS please fill in TITLE.

Nittrouer, S. 1992. Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 20: 351-382.

Odden, D. 1992. Simplicity of underlying representation as motivation for underspecification. *Ohio State University Working Papers in Linguistics* 41: 83-100.

Pisoni, D. B., H. C. Nusbaum, P. A. Luce & L. M. Slowiaczek. 1985. Speech perception, word recognition and the structure of the lexicon. *Speech Communication* 4: 75-95.

Pierrehumbert, J. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18: 375-394.

Pierrehumbert, J., J. Hay & M. Beckman. this volume. Speech perception, well-formedness, and lexical frequency.

Plaut, D. C. & C. T. Kello. 1999. The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (ed.) *The Emergence of Language*, pp. 381-415. Malwah, NJ: Erlbaum.

Port, R. F. & M. O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455-471.

Saffron, J. R., R. N. Aslin & E. L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926-1928.

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237: 1317-1323.

Steriade, D. 1995. Underspecification and markedness. In J. A. Goldsmith, ed., *The Handbook of Phonological Theory*, pp. 114-174. Oxford: Blackwell.

Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.

Stevens, K. N., S. J. Keyser & H. Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (eds.) *Invariance and Variability in Speech Processes*, pp. 426-449. Hillsdale, NJ: Lawrence Erlbaum.