



CAROL A. FOWLER  
University of Connecticut

## MEANING IN PHONOLOGY AND OTHER DEPARTURES FROM MODULARITY IN THE LIVING LANGUAGE

I review evidence of three kinds relating to leakages in modularity within language domains and between linguistic and nonlinguistic action. One kind of evidence shows that the form-meaning “rift” in language that enables the important principle of duality of patterning and the particulate principle of self-diversifying systems is bridged in many ways. Segmental language forms have iconic meanings, and form-meaning correlations of other kinds emerge cross linguistically. A second kind of evidence occurs in parallel transmission of linguistic prosodic information with iconic and emotional information conveyed suprasegmentally. The final kind of evidence shows the integrality of linguistic and nonlinguistic action (deictic points, speech-accompanying gestures, head motions, facial expressions, etc) in conveying communicative information in public language use. I suggest that these violations of modularity within language and between linguistic and nonlinguistic action reflect the dynamic effects of sets of competing and cooperating constraints including, among others, parity and learnability of language forms that shape communicative actions in social activity.

*Key words:* modularity, form-meaning correlations, iconicity, parity, learnability of language forms

### Introduction

A central issue for theories of cognition has been modularity (e.g., Fodor, 1983). With respect to language, the issue arises in two domains: within the language itself and between language and other cognitive domains. Within language, the issue is whether a useful theory, for example, of phonology can be developed that makes no reference to other language domains, for example,

semantics and syntax. Likewise, in reference to language processing, for example, do perceivers of speech identify atomic language forms (say, phonemes) without reference to information in other language domains? The other modularity issue is whether language itself is a modular system with respect to other cognitive domains. Can realistic theories of language or of language processing be developed that make no reference to extra-language domains?

The facts seem to be that valuable theories of phonology (or syntax, etc) have been developed that do not make reference to information in other domains of language. Likewise, theories of language developed within linguistics have provided illuminating accounts of language without making reference to nonlinguistic domains of cognition. With respect to processing (perception, comprehension, production), the issue is highly controversial. However, my take is that processing is convincingly nonmodular pretty much everywhere.

Here, I address modularity and nonmodularity from a different perspective, that of public language use. I suggest that language systems, whether subsystems within language such as phonology or else language itself, are only roughly modular in the “living language,” language as it is manifest in human social action. The systems are as modular as they need to be to do their linguistic work, but constraints arising within the social communities in which language is used foster leaks in modularity. Phonological systems need to have roughly meaningless elements for languages to have duality of patterning, that is, to make infinite use of finite means at the level of the lexicon. Languages themselves have to have roughly consistent sets of conventions (ways of using words, ways of producing structured sequences of them) to make infinite use of finite means at the level of utterances. However, they do not need, and do not have, strict modularity in these respects. Thanks to phonological language forms being shaped in social use of language, segments and sequences of them have iconic meanings as well as other form-meaning correlations. In addition, language grades into nonlanguage in its real-world between-person use. Prosodic information does linguistic work; but the same acoustic variables that convey prosodic metrical structure and intonation also convey emotion and iconic meanings. Finally, a whole variety of actions that occur in between-person face-to-face exchanges, linguistic and nonlinguistic, contribute integrally to speaker meaning.

None of the departures from modularity is sufficient to prevent languages from having creative power at the levels of the lexicon or syntax. However, the departures are significant. They mark language as *primarily* a kind of social action that is deeply embedded in a larger context of social action. Language use is one component of the social actions in which participants engage to achieve joint aims. Because language use is embedded in social action, language is continually shaped by a variety of constraints relating to successful achievement of joint aims. Those constraints weaken modularity but suit language to its public use.

## Meaning in phonology

The phrase, *meaning in phonology* should be an oxymoron, because *phonology* refers to language form, and, in language, ostensibly there is a “rift” (Levelt, Meyer, & Schiller, 1999), between form and meaning. The rift is fundamental to the open, creative power of language at the level of the lexicon. That is, languages have duality of patterning (Hockett, 1960) whereby meaningless atomic language forms combine to form meaningful words of the language. In being meaningless, the forms can be combined in an indefinitely larger number of ways consistent with the phonotactics of the language to generate larger forms with meanings related arbitrarily to the component atoms. In this way (and in a different way at the level of syntactically structured sentences), languages make “infinite use of finite means” (von Humboldt, 1836/1972). Relatedly, languages are among a very small number of systems, including chemical compounding and genetic combination, characterized by “the particulate principle of self-diversifying systems” (Abler, 1989; Studdert-Kennedy, 1998) whereby atomic particles can combine without blending (compare combining the ingredients of a cake) to create new forms, with the possibilities for novel forms constrained but unlimited in number.

However, this is language in the ideal, as it is characterized by linguistic descriptions. In the “wild,” where language emerges and evolves in the course of between-person language use, meaning pervades phonology. The languages in use by community members are only as true to the linguistic ideal as they need to be to do their work in the joint activities of social groups. Duality of patterning and the particulate principle do underlie the openness and creativity of language at the lexical level, but the principles can operate without the form-meaning rift being absolute.

In the following, I will outline respects in which atomic language forms are meaningful by reviewing evidence for meaning in consonants and vowels and in prosodic forms. The review will show that meaningfulness of primitive forms does not lie at the periphery of language, but rather infuses it (see also, Perniss, Thompson, & Vigliocco, 2010). The shading of forms into meanings is one way in which the living language departs from the ideal of linguistic description. Another way, to which I will also allude in the following review, is that language shades into nonlanguage so that traditionally nonlinguistic actions that accompany spoken language use are integral to meaning conveyance in activities that include language.

## Meaningfulness of atomic segmental language forms

In one sense, the meanings of words are mostly arbitrary. In English, the word for “canine animal” is *dog*; in French it is *chien*; in Finnish *koira*, and, in Croatian, *pas*. The forms differ segmentally and in their numbers of syllables, but

their meanings are the same to a close approximation. Of course, the mappings of forms to meanings are not arbitrary in terms of usage. If English speakers wish to refer to canine animals, they must use *dog*; *koira* will not work. The arbitrariness is in the fact that there is nothing about any meanings of /d/, /ɔ/, and /g/ that, in combination, add up to “canine animal.” Any phonetic segments, combined in any way permitted by the phonotactics of a language, can refer to “canine animal” if members of the language community use it that way.

There are exceptions to a claim that word forms relate arbitrarily to meanings. One example is onomatopoeic words, that is, words whose pronunciations are somewhat like the sound to which they refer. Examples from English are *moo* and *boing*. Another class of words that are not wholly arbitrary in their meanings is the class of phonesthemes. These are non-morphemic sound sequences appearing in sets of words that share some meaning. Examples from English include such words as *gleam*, *glitter*, *glisten*, and *glow*, in which the /gl/ sequence is associated with meanings having to do with glittery-ness and *sneeze*, *sniff*, *snout*, *snooty*, *snoop*, etc. in which /s/+ nasal /n/ somehow denotes nose-related things or actions.

Onomatopoeic words and phonesthemes constitute a tiny proportion of the lexicons of languages. But meaning in form is far more extensive than this. People agree that *kiki* is a better name than *bouba* for forms like that on the left side of Figure 1, and *bouba* is better for forms like that on the right (see, e.g., Ramachandran & Hubbard, 2001; and Kohler, 1929). Even children as young as two and a half years of age make this judgment (Maurer, Pathman, & Mondloch, 2006). Somehow there is something sharp-evoking about phoneme sequences such as /kiki/ and something round-evoking about sequences such as /buba/.

There are subtler form-meaning associations as well. English speakers presented with antonym pairs in English and in another language (Chinese or Hindi) were able to decide which order of the foreign-word pair matched that in the English word pair (e.g., which word in a Chinese pair meant *fat* and which meant *thin*; Brown & Nuttall, 1959). Similar findings have been obtained for English

Figure 1. Shapes like those for which adults and children show naming preferences. Names such as /kiki/ are preferred over names such as /buba/ for the figure on the left. The figure on the right evokes the opposite preferences



speaker's judgments of antonym pairs in many other languages (Kunihira, 1971; Nygaard, 2010). In these studies, judgments do not rise far above chance, and the bases for the judgments often are not obvious. (Which Japanese word, *akarui* or *kurai*, means *bright*; which means *dark*?) However, performance is consistently, significantly above chance.

Similarly, Remez, Fellowes, Blumenthal and Nagel (2003) found small but consistent preferences in solving such analogies as the following:

Which is better: *ee* is to *ah* as *rough* is to *smooth* or as *smooth* is to *rough*?

Which is better: *ee* is to *ah* as *Las Vegas* is to *Salt Lake City* or as *Salt Lake City* is to *Las Vegas*?

Whether analogies were perceptual as in the first case or abstract as in the second, participants agreed in their choices on most comparisons to an extent that significantly exceeded chance levels. In short, /i/ and /a/ are not meaningless. Somehow they evoke meanings that listeners can access in explicit judgments.

The findings just reviewed, among others, suggest that the meanings evoked by segments have an impact on how words are coined or how their pronunciations change (or resist change; Joseph, 1987) over time. Language users “prefer” words that are somehow apt for their usage. This same conclusion is suggested by other observations as well.

One observation, noted by Perniss, et al. (2010), is that some languages make use of reduplication to refer to iterated events. An example from Japanese is *gorogoro*, which refers to “a heavy object rolling repeatedly”; one from Siwi is *biribiri*, which means “sowed at close intervals.” Another, as Perniss, et al. also note, is the use in Japanese of a voicing distinction to differentiate heavy vs light (e.g., *goro* “heavy object rolling” vs *koro* “light object rolling”) or thick vs thin (e.g., *boro* vs *poro* for thick or thin “liquid hitting a solid surface”).

Another form-meaning link occurs in the way that suffixes are appended to verb stems. Bybee (1985) sampled 50 languages selected to represent the languages of the world. She identified a variable “relevance” that is reflected in the cross linguistic ordering of suffixes with respect to the verb stem. Relevance refers to the extent to which the suffix changes the meaning of the stem verb. For example, a person marker changes the meaning less than an aspect marker. Across the languages, suffixes higher in relevance tended to be placed closer to the verb stem than suffixes lower in relevance.

In most of the foregoing examples, form-meaning relations likely have a basis in iconicity. However, not all instances of form-meaning relations are iconic. One finding is that statistically distinct phonological patternings differentiate words of different grammatical categories. Nouns and verbs are different grammatical categories with different stereotypic meaning types (i.e., thing-ness vs action).

The two grammatical classes are also distinguished statistically by their characteristic phonological properties. Perhaps because they tend to occur in different rhythmic contexts (e.g., Kelly, 1988), disyllabic nouns and verbs tend to differ in their stress pattern. Nouns tend to have trochaic stress (strong-weak), whereas verbs have iambic stress (weak-strong). English is a stress-timed language in which strong and weak syllables approximately alternate. In an examination of both written and spoken corpora, Kelly (1988) found that verbs are much more likely to be followed by unstressed syllabic inflections (e.g, *-ing*) than are nouns. This fosters a stress pattern for the verbs in which final syllables are strong so that they are both preceded and followed by weak syllables when there is a syllabic inflection. Likewise in a corpus study, disyllabic nouns Kelly and Bock (1988), found that nouns were followed by a strong syllable more frequently than were verbs; verbs were more likely than nouns to be preceded by a strong syllable.

Nouns and verbs exhibit other phonological differences as well. Farmer, Christiansen, and Monaghan (2006) compared over three thousand nouns and verbs from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) by aligning their onsets, nuclei, and codas and computing their featural differences. For each word, they computed its mean phonological distance from the set of nouns and its mean distance from the set of verbs. In a two-dimensional plot with these distances as x and y axes, nouns and verbs occupied distinct but overlapping regions of the space. This clustering has an impact on word recognition in adults. In a self-paced reading task in which noun-like or verb-like nouns appeared in sentences, latencies to read noun-like nouns were shorter than to read verb-like nouns. In a second experiment, latencies to read verb-like verbs were shorter than to read noun-like verbs (Farmer, et al., 2006). Other research shows that the clustering of phonological properties distinguishing word classes facilitates word learning (Monaghan, Chater, & Christiansen, 2005).

Farmer, et al (2006) reviewed evidence showing that lexical categories in other languages than English also cluster phonologically; but they suggest that the clusters of properties that distinguish nouns and verbs differ across languages. It is not that some phonological properties intrinsically evoke noun-ness or verb-ness naturally. Rather, any statistical differentiation makes learning the class difference easier, and language communities exploit that fact.

These are just some of the ways that forms in language can be correlated with meanings at the level of spoken words and below. None of them is potent enough to sabotage the fundamental arbitrariness of the form-meaning relation underlying duality of patterning and the particulate principle of self-diversifying systems. Rather, they underscore how language forms emerge and change over time. They do so in the course of public, between-person language use. In this domain, there is ongoing competition among multiple forces shaping language forms. These apparently include forces promoting the creative power of duality of patterning and therefore arbitrary sequencing of phonological atoms, but also

learnability of forms, aptness of a label for its usage, and probably others that promote correlations between forms and meanings. (See a speculative partial list of constraints in the final section of the paper.)

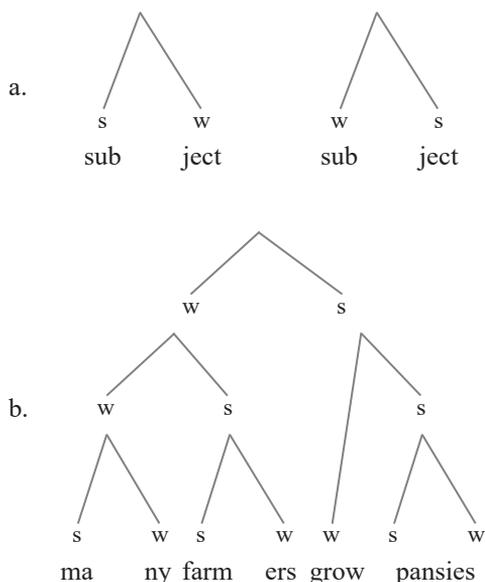
### **Meaning in suprasegmental patterning: Language grades into nonlanguage**

The previous section provided a review of some of the evidence for form-meaning relations in the segmental component of the phonology. The present section reviews evidence for form-meaning correlations in the suprasegmental component. However, the evidence in the two sections is not strictly parallel. In the segmental component, evidence suggests that atomic segments (for example, /i/ and /a/ in the research of Remez, et al. (2003)) or else sequences of them (as in /gl/-initial words or as in *bouba* and *kiki*) that, in the ideal, are meaningless are not in fact. Although this may be the case as well for some of the atomic elements of the prosodies of languages (for example, atomic tones, H and L, of intonation contours), the present review will focus on a kind of parallel transmission of information conveyed linguistically and nonlinguistically by the same set of acoustic variables. Evidence here shows leakage, not within language domains (form-semantics), but between language and nonlanguage.

As for the linguistic expression of suprasegmental information, prosody refers to a collection of properties of utterances that are suprasegmental in scope. Included are rhythmic or metrical properties such as the patterning of strong and weak syllables in utterances and their intonation contours. Acoustic manifestations of prosodic properties include intensity, duration, and spectral manifestations of segments, particularly vowel quality, for expressing syllable prominence and fundamental frequency for realizing the intonation contour.

Metrical patterning is described by theories of metrical phonology (e.g., Liberman & Prince, 1977; Hayes, 1995; see Ladd, 2008 for an overview). In these theories, pairs of adjacent syllables differ in relative prominence. For example, the noun and verb *subject* are represented in Figure 2a. Relative prominence in longer utterances can be represented by “trees” of hierarchically-organized branching relationships of stronger and weaker relations between branching pairs (or, in other approaches, by metrical grids; see Liberman & Prince, 1977; Hayes, 1995). An example of a longer utterance with its hierarchical patterning of relative prominences is shown in Figure 2b. Whether branches are left- (s w) or right- (w s) dominant can differ in different prosodic domains (feet, phonological words, phonological phrases, intonational phrases) and differs between languages. Languages can differ in other ways as well (e.g., whether “trees” are binary branching or can have any number of branches, whether assignment of s and w is sensitive or not to syllable weight; see Hayes, 1980).

Figure 2. Patterns of metrical prominence in otherwise homophonic words (a) and in a sentence (b)



Intonation contours are pitch contours aligned with respect to metrically structured utterances generally of sentence-length. In the theoretical perspective on intonation on which I focus here (cf. Ladd, 2008; Pierrehumbert and Hirschberg, 1990), primitive intonational units are high (H) and low (L) tones. Alone or in combinations of two, these compose pitch accents; individually, they serve as edge tones. Intonation contours are composed of sequences of pitch accents that are aligned with metrically prominent syllables in an utterance and two kinds of edge tones, phrase accents, and boundary tones.

This is just one theoretical account of intonation among others. (See Ladd, 2008, for a discussion of this one and others.) However, it will serve in the present context as a framework within which to discuss meaning in this domain.

Individual H and L tones do not have inherent meanings in theory just as, in the ideal, segmental phonemes are meaningless. However, pitch accents and edge tones may have meanings. Together, the pitch accents, phrase accents, and boundary tones constitute a “tune,” the meanings of which according to Pierrehumbert and Hirschberg (1990) are compositional, reflecting the meanings of the component pitch accents and edge tones.

Pierrehumbert and Hirschberg (1990) suggest that, in general, components of an intonation contour convey the speaker’s intention to change what the hearer

believes to be mutually believed in the conversation. Pitch accents are produced on selected metrically prominent words (see below); then the choice of pitch accent expresses whether the speaker intends for the referent of the accented word to be included or excluded from what the interlocutors take to be mutually believed. Phrase accents and boundary tones convey the intended relationship between, respectively, intermediate phrases and intonational phrases. In their example:

The train leaves at seven or at nine twenty five  
 H\*    H\*            H\*    H        H\*                    H\*    L L%

the H\* pitch accents indicate information meant to be added to the hearer's beliefs about the talker's and hearer's mutual beliefs. The H intermediate phrase tone indicates continuity of information in the first phrase with information in the second. The L intermediate phrase tone and L% boundary tone indicate finality or a lack of relation of the second intermediate phrase and the whole intonational phrase to what follows.

As noted, in the ideal, H and L elements are meaningless. I am not aware of any tests for the meaningfulness of H and L tones. However, given the association of higher pitches in speech with happiness (or joy) and lower ones with sadness (e.g., Banse & Scherer, 1996; Scherer, 1986; 2003), chances are that they are no more bleached of meaning than are consonant and vowel segments.

That issue aside, the same suprasegmental acoustic channels that convey linguistic prosody convey other meanings as well. Accordingly, in parallel with expression of linguistic meaning, prosodic variables convey other meanings. One respect in which suprasegmental components of utterances, like segmental components, express meanings beyond those signaled by utterances as described linguistically is in expression of emotion (e.g., Banse & Scherer, 1996). However, there are other respects as well.

As for expressions of emotion, Scherer (2003) summarizes the literature on acoustic correlates and reports acoustic correlates of such emotions as anger, fear, sadness and joy including effects on f0 mean and variability, intensity, and speaking rate (or duration), the same variables that encode prosodic variables. However, it is not just that the same acoustic variables that convey linguistic prosody also convey emotion. Information for the emotion of a speaker encoded in those variables affects how linguistic information is recognized.

In particular, listeners use acoustic information that expresses emotion as information for word meaning. Nygaard and Lundes (2002) had actors produce words in neutral, happy or sad tones of voice. Included among the words were differently spelled homophones (e.g., *die-dye*; *bridal-bridle*, the meanings of which were differently valenced so that one meaning was positively or negatively valenced and the other meaning was more neutral. Listeners heard the words among fillers and spelled what they heard. The emotional tone of the words was a between-participants variable so that participants only heard homophones

produced with one emotion. Listeners tended to spell homophones in a way consistent with the tone of voice. For example, a homophone *dye-die* was spelled *die* more frequently when it was produced in a sad tone of voice than when it was produced neutrally or especially than when it was produced in a happy tone of voice. This is an interesting finding in showing that lexical access (whether *die* or *dye* is accessed by /dai/) is affected by a nonlinguistic property of the signal. A related study by Nygaard and Queen (2008) led to a similar conclusion. Nygaard and Queen found faster naming times to words whose meanings had positive or negative emotional valence when the words were spoken in a compatibly-valenced tone of voice than when they were produced neutrally or in an oppositely valenced tone.

Acoustic variables that convey prosodic meanings also conveyed emotional content in the foregoing studies. In the following study, the same variables provide information for non-emotional dimensions of meaning. Words in the experiment by Nygaard and Lunders (2002) were found to differ in  $f_0$ ,  $f_0$  variability, RMS amplitude and duration depending on tone of voice. These same dimensions of variability also differentiated words produced with a range of intended non-emotional meanings in a study by Nygaard, Herold, & Namy (2009). In that study, talkers were instructed to use infant-directed speech in producing a carrier sentence *Can you get the \_\_\_ one?* in which nonsense disyllables such as *blicket* and *daxen* were embedded. Speakers produced them without knowing the meaning of the nonsense disyllable or else in contexts in which they were to name, for example, a big object visible in a picture rather than a small object also shown in a picture or else a small object rather than a big object. Contrasts elicited from talkers were on the six dimensions, big-small, hot-cold, happy-sad, yummy-yucky, tall-short and strong-weak.

Utterances of the same nonsense word produced to convey different meanings differed along the dimensions of  $f_0$ ,  $f_0$  variation, RMS amplitude and duration. Different patternings of acoustic structure along those dimensions differentiated words depending on the meaning dimension. For example, a word such as *blicket*, produced to mean the larger of two pictured objects had significantly greater amplitude and duration than when it was produced to refer to the smaller object. The same word produced to refer to the stronger of two pictured entities had higher amplitude than when it was produced to refer to the weaker entity. To an extent, the acoustic reflections of these semantic dimensions are iconic. For example, large, strong things, on average, produce higher amplitude signals than smaller, weaker things.

In a perceptual study, in which listeners heard sentences, e.g., *Can you get the blicket one*, with *blicket*, having been produced to refer to the stronger, or bigger or yummier pictured entity, they performed with above-chance accuracy in picking out the intended referent. Accordingly, acoustic variables that provide prosodic meanings in utterances at the same time can provide other semantic information as well.

## Intermediate discussion

The foregoing review shows that language forms convey meanings that lie outside the meanings that derive from their linguistic roles. Consonants, vowels and sequences of them evoke reliable impressions that underlie agreements across speaker-hearers on what such forms as *bouba* and *kiki* should mean, on analogies involving the vowels /i/ and /a/, and on which members of antonym pairs in unknown languages map on to which pair members in English (for English judges). Prosodic dimensions of variation in the acoustic signal also convey meanings beyond those resulting from mapping of forms to meanings in the language. Acoustic variation in  $f_0$  and other prosodic variables express emotion and can convey impressions of size and strength among others.

In the final discussion, I will return to a consideration of how linguistic utterances come to convey these meanings that fall outside the scope of linguistic analysis. First, however, I address another issue, namely the general grading of linguistic into nonlinguistic action in social use of language.

One way of looking at the foregoing findings and many others not covered in the present review concerns modularity, or, rather, the lack thereof in language. Within the realm of language as noted earlier, an issue has been whether different linguistic levels are modular with respect to one another. Clearly phonology, syntax and semantics are distinct linguistic domains, but an issue is whether they are they distinct modules, encapsulated from one another in some or all respects. For example, linguistic theories of phonology (e.g., Chomsky & Halle, 1968, Goldsmith, 1976; Prince & Smolensky, 2004) have been developed as if the phonology is modular with respect to other linguistic domains in the sense that phonological rules or constraints only make reference to phonological variables, never, say, to semantic or syntactic ones. In my view, these theories have been of considerable value in exposing the character of phonological language forms and processes or constraints that characterize different phonological systems. However, the findings reviewed in the two previous sections show that, for language users, the distinct domains, of form and of meaning, have consequential leaks. Thus, a realistic accounting of phonological systems as they emerge, develop and are used, has to include an accounting of the leaks. (For a consideration this idea from a different perspective, one of language processing, see Dell, 1984.)

As noted in the introductory section, modularity arises in a larger context as well. Language use is a kind of activity that seems quite distinct from other kinds of actions in which humans engage. It has been enormously valuable for linguists and psycholinguists to focus on linguistic activities as if they were modular, in order to develop theories of language structure. However, it is also significant that there are leaks here as well. In the wild, linguistic actions are embedded in other human actions to an extent that the work that gets done by talking is not done by talking alone, but by talking embedded in a larger context of human actions.

In the case of manual gesturing that accompanies language use, for example, it has been an issue (e.g., Kendon, 2000; McNeill, 1985) whether or not (e.g., Krauss, Dushay, Chen, & Rauscher, 1995) gestures are kinds of language forms.

In this domain, language appears to shade into nonlanguage. However valuable it may be for there to be a discipline, linguistics, that studies language in itself, understanding meaning in language use will require scientists to study embedded language use. I next briefly review just some of the ways in which actions other than phonological ones contribute to meaning in language use.

### **Further departures from modularity in language: Integrity of linguistic and nonlinguistic action**

We use language in many settings. We can talk to ourselves or to our pets. That is, we can talk when there is no one other than ourselves to appreciate what is being said. We also use language successfully in between-person contexts in which interlocutors cannot see one another, for example, in phone conversations. However, in settings in which interlocutors are co-present, and in which they can see one another, strictly linguistic activities are just a subset of actions that contribute to interpersonal communication. Moreover, the actions that accompany linguistic ones appear to be integral to the linguistic actions in serving jointly with them to communicate. In a larger context still, communicative actions jointly can be integral to social activities in serving with noncommunicative actions to achieve social aims. For example, when people work together, say, to move bulky furniture through a narrow doorway into a new apartment, their speech, facial expressions, gestures, eye gazes, etc. can all foster achievement of the joint nonlinguistic aim to get the furniture moved.

As for communicative actions, clear instances of conventionally nonlinguistic actions that are integral to linguistic actions in their communicative function are deictic points. In answer to a question, such as *Where is the hand truck?*, a co-present interlocutor may say *It's over there*, while pointing to the object's location. For the hearer to know where *there* is requires that s/he follow the point.

Pointing can be achieved in other ways as well. Imagine two furniture movers in separate rooms. One may say to the other: *Where is the hand truck?* If the other says: *It's over here* (cf. Hanks, 1992), the first person can know where the hand truck is by auditorily localizing the other speaker's voice. Deictic points can be achieved with eye gaze as well. A speaker who says: *I really like that car* may reveal to interlocutors which car s/he means by looking at it.

Other nonlinguistic communicative actions include facial expressions, head movements (e.g., Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004), and manual gestures. Manual gestures occupy a kind of grey region between language and nonlanguage that is of particular interest in the present context.

Some gestures, known as emblems (e.g., McNeill, 1985) can communicate without accompanying language. Examples include the OK sign or V for victory sign. Other gestures, however, are not as conventionalized as emblems and accompany language use. Although it has been controversial whether gestures serve to communicate or, instead, occur for the benefit of the talker (e.g., for purposes of lexical access; Krauss, Dushay, Chen & Rauscher, 1995), in our view, there is convincing evidence in favor of the former view (not necessarily against the view that they benefit the talker as well). See Goldin-Meadow and Beilock (2010) for an excellent review.

At least one theorist considers gestures linguistic accompaniments of speech (McNeill, 1985). My own view, is that there is no real point in deciding whether to call them linguistic or not. They are less conventionalized than the phonemes, words, and syntax of the spoken language, and they more easily dispensed with (as in telephone conversations and writing). However, they appear to be integral to language use in a face-to-face communicative context in their contribution to interlocutors' understanding of what is said. There is likely no clear line to draw between what is language and what nonlanguage in communicative activities.

Research by McNeill and colleagues demonstrates the integrality of speech and gesture in a variety of ways. Manual gestures tend occur synchronously or slightly in advance of the speech they relate to (e.g., McNeill, 1985). McNeill (1992) has shown that the vocal and manual gestures are temporally coupled by using delayed auditory feedback as speakers narrate a cartoon. Delayed auditory feedback is an effective way to create halting stuttered speech. The finding is that when speech is delayed, so is an accompanying gesture.

Integrality is shown in a different way by a finding that information can be provided by gesture that is not redundant with that provided by accompanying speech. One example (McNeill, 2005) is from a narration in which a cat, Sylvester, in a cartoon that participants viewed and narrated swallows a bowling ball and exits a drainpipe. The speaker says *And Sylvester [came out] of the pipe*. Coincident with the bracketed speech, the manual gesture indicates bouncing. That Sylvester bounced out of the pipe is information true to the event in the cartoon that is not expressed in the speech.

Finally, McNeill's research has shown (e.g., McNeill, 2005) that, for the listener/observer of these narrations, memory for information conveyed by speech and by gesture are integrated. Having heard the statement and seen the gesture above, a participant might recall having heard: *And the cat bounces out of the pipe*. In research, listeners sometimes cannot recall whether information was provided by speech or gesture. Meaning is provided jointly by vocal-gestural and manual-gestural forms.

### **What is the significance of the integrality of form and meaning and of language and nonlanguage in the living language?**

Chomsky (1986, p. 31) has written that public language (as contrasted with internal language) "appears to have no significance" and, slightly more moderately,

that (Chomsky, 2011) the externalization of language by the sensorimotor system is a secondary property of language. This confusion (see Millikan, 2003) has fostered some very wrong ideas about language and particularly about the modularity of language within itself and with respect to other human actions. That atomic “meaningless” language forms are meaningful, conveying meanings in parallel with their conventional meanings, and that the prosodic forms of language convey meanings in parallel with emotional expression and iconic expression fairly shout that languages are first and foremost public kinds of social actions that emerge and grow in the context of meaningful engagements that shape them.

There are other indications as well. The sound inventories of languages reflect communicative constraints. Small vowel inventories predictably include vowels that are located far apart in acoustic space (e.g., Liljencranz & Lindblom, 1972 ). However, one vowel, /ɜ/ (as in American English, *church*), that is acoustically quite distinctive is avoided, presumably, because its articulation is complex, involving three simultaneous vocal-tract constrictions. Similarly, small consonant inventories tend to include only “basic” (articulatorily simple-to-produce) consonants; larger ones include those and “elaborated” consonants. Only the largest inventories have the articulatorily most complex (“complex”) consonants (Lindblom & Maddieson, 1988). Together, suggest Lindblom and Maddieson (1988), the segment inventories of languages exhibit a dynamic balancing of perceptual distinctiveness and of articulatory ease, with both constraints shaping both vowel and consonant inventories to different extents. Why? Communicative constraints shape segment inventories because phonological language forms emerge and change in the course of between-person language use in which constraints relevant to communicative success (itself relevant to successful achievement of between-person joint aims) shape how interlocutors talk. Perceptual distinctiveness is especially important to between-person achievement of parity (A. Liberman & Whalen, 2000), that is, successful recognition on the part of hearers of the forms produced by the talker. Favoring articulatorily simple forms may foster learnability as well as efficiency in talk.

Another indication of the central role of public language use in shaping language is language change. Languages change over time; different communities of speakers of the “same” language speak it differently, and those facts reflect that constraints of a social nature affect use of language.

The departures from modularity within language and between linguistic and nonlinguistic action as reviewed here are just further reflections of the significance of public language and the primary, not secondary, import of its public use. Humans are social organisms who get together to achieve joint aims. Linguistically significant vocal tract actions constitute just one tool for fostering achievement of those aims. Other tools (coordinated action itself; facial expressions, manual gesturing, etc.) are available as well and tend to be used concurrently and integrally.

It is probably accurate to think of linguistic actions being dynamically shaped by a whole collection of constraints that can compete with one another or, alternatively, can reinforce one another in the ongoing development of language. The unifying characteristic of those constraints is that they reflect requirements relating to humans acting together. Some constraints, say, the parity constraint, will foster development of segments that are easy to say and to distinguish from others. Learnability constraints may also foster inventories of segments that are easy to say. Learnability as well as likelihood of being coined may foster invention of word forms that have iconic meanings or in which phonological properties correlate with grammatical categories. Aptness of word form and meaning may serve as a constraint for maintaining forms with nonarbitrary meanings. An opposing constraint, however, would be one for maintaining form-meaning arbitrariness that underlies the creative power of language at the level of the lexicon.

These constraints and no doubt many others both foster and weaken modularity within language and between linguistic and nonlinguistic (and somewhat linguistic) actions. Modularity within language enables its richness and creative power. But communicative constraints, or even more broadly constraints arising when humans get together to achieve joint aims will foster leaks in modularity. However, leaks that do no major harm to the power of language will be readily tolerated.

## References

- Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12 (1), 1–13.
- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database*. Philadelphia: Linguistics Data Consortium.
- Banse, R. & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70 (3), 614–636.
- Brown, R. & Nuttall, R. (1959). Method in phonetic symbolism experiments. *Journal of Abnormal and Social Psychology*, 59 (3), 441–445.
- Bybee, J. (1985). *Morphology*. Amsterdam: John Benjamins Publishing Company.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chomsky, N. (2011). Language and other cognitive systems: What is special about language? *Language Learning and Development*, 7 (4), 263–278.
- Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Dell, G. (1986). A spreading-activation theory of retrieval in speech production. *Psychological Review*, 93 (3), 283–321.

- Farmer, T.A., Christiansen, M.H., & Monaghan, P. (2006). Phonological typicality influences sentence comprehension. *Proceedings of the National Academy of Sciences*, 103 (32), 12203–12208.
- Goldin-Meadow, S. & Beilock, S. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5 (6), 664–674.
- Goldsmith, J. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell Publishers.
- Hanks, W. (1992). The indexical ground of deictic reference. In A. Duranti & C. Goodwin (Eds.), *Rethinking Context* (pp. 43–76). Cambridge, UK: Cambridge University Press.
- Hayes, B. (1980). *A Metrical Theory of Stress Rules*. Cambridge, UK: MIT Press.
- Hayes, B. (1995). *Metrical Stress Theory*. Chicago: University of Chicago Press.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 89–96.
- Joseph, B.D. (1987). On the use of iconic elements in etymological investigation: Some case studies from Greek. *Diachronica*, 4 (1–2), 1–16.
- Kelly, M.H. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, 30 (2), 107–137.
- Kelly, M.H. & Bock, J.K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14 (3), 389–403.
- Kendon, A. (2000). Language and gesture: Unity or duality. In D. McNeill (Ed.), *Language and Gesture* (pp. 47–63). Chicago: University of Chicago Press.
- Kohler, W. (1929). *Gestalt Psychology*. New York: Liveright.
- Krauss, R., Dushay, R., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31 (6), 533–552.
- Kunihira, S. (1971). Effects of expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, 10 (4), 427–429.
- Ladd, D.R. (2008). *Intonational phonology (Second ed.)*. Cambridge, UK: Cambridge University Press.
- Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge, UK: MIT Press.
- Liberman, A.M. & Whalen, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4 (5), 187–196.
- Liberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8 (2), 249–336.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48 (4), 839–862.
- Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. Hyman & C.N. Li (Eds.), *Language, Speech and Mind* (pp. 62–78). London: Routledge.

- Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9 (3), 316–322.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92 (3), 350–371.
- Millikan, R.G. (2003). In defense of public language. In L.M. Antony & N. Hornstein (Eds.), *Chomsky and His Critics* (pp. 215–237). Malden, MA: Blackwell Publishing, Ltd.
- Monaghan, P., Chater, N., & Christiansen, M.H. (2005). The differential role of phonological distributional cues in grammatical categorisation. *Cognition*, 96 (2), 143–182.
- Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15 (2), 133–136.
- Nygaard, L. (2010). Cross-linguistic sound symbolism. Presentation at the Conference on Sound Symbolism: Challenging the Arbitrariness of Language, Emory University, March 26, 2010.
- Nygaard, L.C., Herold, D.S., & Namy, L.L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33 (1), 127–146.
- Nygaard, L.C., & Lunders, E.R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, 30 (4), 583–593.
- Nygaard, L.C., & Queen, J.S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, 34 (4), 1017–1030.
- Perniss, P., Thompson, R.L., & Vigliocco, G. (2010). Iconicity as a general property of language; Evidence from spoken and signed languages. *Frontiers in Psychology*, 1 (227), 1–17.
- Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in Communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Prince, A. & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Publishers.
- Ramachandran, V.S. & Hubbard, E.M. (2001). Synaesthesia -- A window into perception, thought and language. *Journal of Consciousness Studies*, 8 (12), 3–34.
- Remez, R.E., Fellowes, J.M., Blumenthal, E.Y., & Nagel, D.S. (2003). Analysis and analogy in the perception of vowels. *Memory & Cognition*, 31 (7), 1126–1135.
- Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99 (2), 143–165.

- Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40 (1–2), 227–256.
- Studdert-Kennedy, M. (1998). The particulate origins of language generativity: From syllable to gesture. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 202–221). Cambridge: Cambridge University Press.
- Von Humboldt, W. (1836/1972). *Linguistic Variability and Intellectual Development*. Translated by G.C. Buck & F.A. Raven. Philadelphia: University of Pennsylvania Press.