Published in final edited form as:
*New Ideas Psychol*. 2014 January 1; 32: . doi:10.1016/j.newideapsych.2013.03.007.

# Talking as doing: Language forms and public language

**Carol A. Fowler**
University of Connecticut, Storrs, CT 06269, and Haskins Laboratories, New Haven, CT 06511

## Abstract

I discuss language forms as the primary means that language communities provide to enable public language use. As such, they are adapted to public use most notably in being linguistically significant vocal tract actions, not the categories in the mind as proposed in phonological theories. Their primary function is to serve as vehicles for production of syntactically structured sequences of words. However, more than that, phonological actions themselves do work in public language use. In particular, they foster interpersonal coordination in social activities. An intriguing property of language forms that likely reflects their emergence in social communicative activities is that phonological forms that should be meaningless (in order to serve their role in the openness of language at the level of the lexicon) are not wholly meaningless. In fact, the form-meaning "rift" is bridged bidirectionally: The smallest language forms are meaningful, and the meanings of lexical language forms generally inhere, in part, in their embodiment by understanders.

### Keywords

language forms; embodiment; public language; form-meaning rift

## 1.0 Introduction

The study of language by linguists has focused on linguistic competence, that is internalized knowledge of language, rather than on performance. In psycholinguistics where performance has been addressed, language use in the mind (that is, mental processes involved in perception, comprehension, and production planning) have been studied more than language use in the world. This focus on private rather than public language has permitted development of a perspective on language forms that I suggest is unrealistic. Although language forms are the means within language for making linguistic communications public, they are assumed to have properties that prevent their making transparent or veridical public appearances.

In the present paper, I carve out a very small part of the study of public language use, invoking an insight from the study of linguistics that is mostly accurate: language structure partitions into distinct levels of description. I focus on language forms as distinct from the meaningful utterances they compose. A major aim of the discussion is to promote the idea that language forms are public actions, not private categories in the mind. As such, they are adapted to public use. Indeed, they are primarily actions of the vocal tract, not the abstract mental categories of linguistic and psycholinguistic theories. They are adapted also because

they emerge from and so are shaped by the requirements of social activities in which language use constitutes a part. The language forms that emerge in that way are perceptually easy to distinguish and are readily articulated. A second aim is to note that, although language forms are vehicles for conveying language meanings, they are also more than that. In addition, they do work themselves in serving (along with manual and other bodily actions) as coordination devices; they serve to foster successful achievement of the joint aims of groups of individuals (Clark, 1996) participating in social groups. A final aim is to acknowledge that the partitioning of language into forms and meanings on which this discussion depends is only approximately accurate, and that matters too. Precisely because languages emerge from and are shaped by social interaction, they are not the tidy formal systems of linguistic analysis. One interesting index of the untidy shaping of language by its use is that the partitioning of language into distinct levels of form vs meaning (the form-meaning "rift" (Levelt, Roelofs, & Meyer, 1999)) is only approximate.

## 2.0 Language forms

### 2.1 Some linguistic characterizations of language forms and critiques of them

In classical descriptions provided by theories of phonology, the smallest language forms are consonants and vowels. They are abstract categories in the mind that have featural attributes such as (for consonants) voicing, and manner and place of articulation. They are the meaningless building blocks that compose word forms. Across the lexicons of language, they exhibit regularities (for example, aspiration of voiceless stops in stressed-syllable initial position in English) that are expressible by rules (e.g., Chomsky & Halle, 1968) or constraints (Prince & Smolensky, 2004).

Meaningless language forms combine in systematic ways (that is, respecting phonotactic constraints) to compose meaningful word forms. This combinability, known as "duality of patterning" (Hockett, 1960), "the phonological principle"(Pierrehumbert, 2006), or more broadly, "the particulate principle of self-diversifying systems" (Abler, 1989; Studdert-Kennedy, 1998), constitutes one of the ways in which language exhibits productivity. There is no limit to the size of the lexicon of word forms for any language community. This is a critical characteristic of language underlying its openness to growth and change at the level of lexical forms.

There is something right about the foregoing characterizations of language forms provided by many theories of phonology. The characterizations at least roughly capture what language users know about their language at the level of phonological form. Featural descriptions offer insight into the nature of systematic phonological processes (e.g., vowel harmony, final devoicing) that characterize languages. They also do a good job of capturing characteristics of sublexical speech errors (e.g., Dell, 1986; Shattuck-Hufnagel, 1979), providing evidence that something like discrete segments with featural attributes are relevant to language performance. Finally, the idea that primitive forms are discrete and recombinable explains one kind of linguistic productivity as noted.

However, there is much wrong with the characterizations as well if language forms such as consonants and vowels are viewed as providing an interface between language in the mind and language in the world. In that central role, if at all possible the forms should be adapted to their public use. Yet, language forms as described in linguistic theory do not make public appearances intact. In conventional theoretical accounts, there are incompatibilities between language forms as known and the activities of the vocal tract that make linguistic utterances available to listeners.

Following are three major aspects of the presumed mismatch between language forms of linguistic competence as they are conventionally characterized and public utterance of language forms:

1. In linguistic theory, consonants and vowels are discrete. For example, the word *bus* consists of three discrete segments /b/, /ʌ/, and /s/. Each segment is characterized by a set of static featural attributes. For example, /b/ is a bilabial, voiced obstruent. However, there are no temporally discrete, static, segments either in the corresponding articulation of the word *bus* or in the resulting acoustic speech signal. (This is known as the "segmentation" problem.) The mismatch derives from coarticulation, the temporal overlap of vocal-tract gestures for sequences of consonants and vowels.

2. Another consequence of coarticulation is pervasive context-sensitivity of acoustic information for the same consonant or vowel produced in different coarticulatory contexts. Although acoustic invariants for phonetic segments have been sought (e.g., Stevens and Blumstein, 1981), they have not been found (the "invariance" problem). Accordingly, it is not just that there are no discrete, static segments in articulation or speech acoustics; there is apparently *nothing* the same in either domain when the same consonant or vowel is produced in different contexts.

3. Even as transcribed phonetically, a given ostensible phonological segment (e.g., /t/) shows endless variation, both within a speaker across contexts and speaking styles and across speakers of different idiolects and dialects.

The foregoing are the major ways in which language forms as their implementation in public action is characterized are incompatible with language forms as they are presumed to be known in the mind. Yet a conclusion that the mismatch is real is surprising. If language forms are the means within language for making linguistic communications public, should they not be adapted to their public use?

My own view is that the mismatch is not real. It has arisen, because theories of phonology have been developed without attention to the role of language forms in public use of language. Therefore, an aim of my research and that of collaborators has been to show that there is no mismatch between language forms as known (e.g., Browman & Goldstein, 1986; Goldstein & Fowler, 2003), produced (e.g., Saltzman & Munhall, 1989; Fowler & Saltzman, 1993), specified acoustically (Fowler, 1994; Iskarous, 2010; Iskarous, Fowler, & Whalen, 2010) or perceived (e.g., Fowler, 1986; 1996; Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, 2010).

One way in which some investigators have proposed to eliminate the incompatibility has been to reject the idea that abstract phonological segments are components of linguistic competence and to propose instead a close similarity between token utterances as produced in the world and memory supporting perception and production. This approach is taken, for example, by exemplar theorists (e.g., Johnson, 1997, 2005; Pierrehumbert, Hopper, & Bybee, 2001). In this approach, as language forms are preserved in memory, they are not abstracted from acoustic patterning provided by the contexts in which listeners heard or speakers produced them. For example, listeners not only store that a speaker said: *It's a new day*, they hear the utterance produced by a speaker with characterizable properties (produced by Uncle Charlie, or by an adult male speaker, a speaker from the US Midwest….), speaking at some rate, happily, gloomily, ironically, sarcastically, etc.

The proposals derive from evidence suggesting that extralinguistic information about an event in which speech occurred is not stripped from linguistic information in memory (e.g., Goldinger, 1998; Johnson, 1997, 2005; Pierrehumbert, 2006). In at least one account, the

proposed memory traces are barely processed perceptually. According to Port (2010a, p. 43): “The evidence is strong that people actually employ high-dimensional, spectro-temporal, auditory patterns to support speech production, speech perception and linguistic memory in real time.” The high-dimensional, spectro-temporal patterns are those produced jointly by the linguistic and nonlinguistic forces at work in speaking, the very forces that create the mismatch between conventional characterizations of elements of phonological competence and of speech itself.

This approach has the desirable effect of eliminating any incompatibility between language forms as known and as implemented. However, eliminating the mismatch by proposing memory traces that copy acoustic patterns is not realistic. Acoustic speech signals are “proximal stimuli” for perception but they cannot be perceptual objects. Like reflected light for seeing, skin deformations for haptic perception, etc., they have two critical properties that underlie the essential role they serve in perception, that of providers of information *for* what is perceived. First, they are lawfully structured by objects or events (“distal” objects or events) in the environment so that their changing structure over time provides information about the distal causes. Second, like reflected light, skin deformations, etc., acoustic signals stimulate activity in sensory receptors, and, in that way, impart their structure to perceptual systems. However, perception is necessarily of distal objects and events, not of proximal stimuli (e.g., Gibson, 1966; 1979). We see, feel and hear events in our econiche; we do not perceive the proximal stimuli that provide information for those distal events. This has to be the case for the very survival of organisms including humans.

A proposal that listeners store spectro-temporal auditory patterns means that they perceive proximal acoustic signals, not the distal events of speaking about which those signals inform. In turn this means that listeners do not, in fact, perceive the distal event of Uncle Charlie saying *It’s a new day* gloomily and at a slow rate, etc. Instead, they intercept and store the causal effects of that event in an acoustic medium, unparsed, its informational content untapped.

A better approach for recognizing the fundamental compatibility between language forms as known, produced and perceived derives from a realist philosophy (e.g., Kelley, 1986). In respect to language, talkers produce utterances composed of real language forms. Listeners perceive those forms, not their acoustic effects. For that to be possible, the forms must be the actions that produce perceptible acoustic effects (Fowler, 1986; 1996).

### 2.2 Embodied language forms

A useful starting point for development of compatible accounts of language forms as known and as used in public events of talking and listening is to suppose that language forms are adapted to their public use. There is, in fact, considerable evidence that language forms are so adapted at least in some respects.

For example, in languages with small vowel inventories, it is clear that members of language communities converge on vowel categories that differ acoustically as much as possible. If there are just three vowels in the inventory, they tend to be variants of /i/, /a/ and /u/, that is, a high front, a low back and a high back rounded vowel (e.g., Maddieson, 1984). Liljencrants and Lindblom (1972) have shown that it is possible to predict with good accuracy the vowels that most commonly compose vowel inventories of sizes up to about 12 by assuming that the inventories respect a principle of perceptual distinctiveness or contrast.

It is not as obvious that consonant inventories are guided by that principle. However, Lindblom and Maddieson (1988) showed that languages with the smallest consonant inventories tend to have only the articulatorily simplest (“basic”) consonants. Languages

with larger inventories have both basic and articulatorily more elaborated consonants. Only languages with the largest inventories have the most complex consonants along with simple and elaborated consonants.

In short, language communities have inventories of language forms that are easy to distinguish and easy to say.[1] These observations show that segment inventories of languages are shaped by requirements of interpersonal communication. However, they do not address the segmentation and invariance problems caused by coarticulation considered earlier. These problems for perceivers can be addressed by other characteristics that consonants and vowels must have if they are adapted to their public use:

1. They must *be* public actions of the vocal tract. Only if language forms are the immediate causes of acoustic speech signals can the signals possibly specify them.
2. The actions for successive consonants and vowels must be discrete in the sense of being separate one from the other (even if, owing to coarticulation, they cannot be temporally discrete), and they must be re-combinable as needed for duality of patterning
3. They must have properties that are not distorted (Ohala, 1981) or destroyed by coarticulation (Hockett, 1955), and, for recoverability by listeners, they must permit direct perception (Fowler, 1986; 1996; Fowler & Smith, 1986; Goldstein & Fowler, 2003) in causing acoustic signals that specify them (e.g., Fowler, 1994; Iskarous, 2010; Iskarous, et al., 2010).
4. They must be (sufficiently) shared by members of a language community. (Relatedly, as Millikan, 2003 suggests, they must be multiply "reproduced" by language community members.) Sharing or reproducing language forms has two aspects that are relevant here: a) the forms are shared among members of a language community to an extent that their reproductions by some community members count as reproductions to others; b) the forms are sufficiently the same for language users in their roles both as speakers and as listeners (cf. Garrod & Pickering (2004) and Liberman & Whalen (2000) on "parity").[2]
5. They *emerge* from interpersonal public language use. Research by de Boer (2001), Goldstein and colleagues (e.g., Browman & Goldstein, 2000; Goldstein & Fowler, 2003) and Oudeyer (2005) for example, have reported simulations in which language forms adapted to public use (e.g., vowels that are well-separated in acoustic space) emerge in the course of interpersonal communication.

Language forms having the required properties have been described many times elsewhere (e.g., Browman & Goldstein, 1986; 1992; Goldstein & Fowler, 2003). Readers may consult the cited sources for more complete descriptions. In brief, fundamental or primitive language forms are actions of the vocal tract, known in Articulatory Phonology (e.g., Browman & Goldstein, 1986, 1992, 1995) as "phonetic gestures." Canonically, the gestures are coordinated actions of more than one articulator (e.g., the jaw and two lips for /b/, the jaw and tongue body for /a/) that create and release constrictions, which vary in location (e.g., lips for /b/, palate for /a/) and in degree (closed for /b/, wide for /a/). In grain size, gestures span classical segments (e.g., /b/) and features (e.g., the velum lowering gesture

[1]It is unlikely that vowels are adapted to public use only in being easy to distinguish; whereas consonants are adapted only in being easy to say. Lindblom and Maddieson (1988) suggest that both principles affect both inventories. However, for example, vowels generally are easy to produce; accordingly, perceptual distinctiveness is more salient in vowel inventories.

[2]Obviously, they do not have to be *wholly* shared in the sense of (a). We can understand speakers of different dialects of our language, different idiolects, speakers speaking carefully or casually, speakers with foreign accents. We can understand speakers whose lexicons do not wholly overlap with ours.

for /m/). Like features, the gestures or their parametric values (constriction locations, degrees) serve a phonologically contrastive function; that is, they minimally distinguish different words (e.g., Browman & Goldstein, 1992). Likewise, the gestures of Articulatory Phonology can be appropriate referents in statements of phonological regularities or constraints, and sometimes their unique properties, for example, their dynamical character, can provide superior characterizations of the regularities than more conventional rule- or constraint-based accounts that invoke featural characterizations (e.g., Gafos, 2002).

Gestures are achieved by dynamical systems (e.g., Saltzman & Munhall, 1989) or "synergies" that emerge transiently in the vocal tract during speech. As synergies, they have the property of equifinality. That is, their essential character (constriction locations, degrees) can be achieved flexibly in a variety of ways as when, for example, lip closure for /b/ is achieved less by jaw raising and more by lower lip raising/upper lip lowering in /ba/ (where /b/ coarticulates with the open vowel /a/ that pulls the jaw down) than in /bi/. In this way (see also, the role of coarticulation resistance; e.g., Fowler, 2005), despite coarticulation, gestures are achieved without distortion or destruction of their essential properties in public vocal tract action. There is, thus, invariance in public achievement of these gestures, and, because the phonetic gestures directly cause acoustic speech signals, there is the possibilility that the signals can specify them (e.g., Iskarous, 2010).

Because gestures are themselves actions that necessarily play out over time, and because gestures overlap temporally, we should not be surprised that the acoustic signals that gestures cause are not composed of temporally discrete segments, and the absence of discrete acoustic segments should not be judged evidence against the discrete nature of language forms (as e.g., Port, 2010a,b has done). The segmental structure of speech is specified not by acoustic *segments* but by acoustic patterning that specifies discrete, but temporally overlapping gestures. Research over many decades (beginning with Liberman, Delattre, & Cooper, 1952; more recently, Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, 2010) provides converging evidence that listeners to speech are exquisite trackers of gestural information in coarticulated speech signals. They perceive the gestures that talkers produce (Liberman & Mattingly, 1985; Fowler, 1986; 1996).

Returning to the list above of required characteristics of language forms adapted to public use, gestures are produceable and perceivable by language users. Because they are produced without distortion of destruction, they can be shared between language users in events of speaking and listening. Moreover, they are fundamental language forms for the language user both as speaker and as listener. That gestures are only *sufficiently* shared, not wholly shared, among speakers of a common language underlies language change and serves a valuable, function of expressing social identity (e.g., Giles, et al., 1991). Language forms that are adapted to public use in these ways are shown to emerge from communicative interactions in simulations by a number of investigators (e.g., de Boer, 2001; Browman & Goldstein, 2000; Oudeyer, 2005).

## 3.0 Affordances of phonological speech actions

The argument of the previous section was that language forms are fundamentally embodied in being public actions of the vocal tract that constitute components of human linguistic behavior. Among the whole set of actions that occur when interlocutors speak (including manual gestures, pointing, facial expressions, head movements, etc), they are somewhat special in having a dual character (cf. Cowley, 2011): They are actions in themselves, but they are also vehicles for reproducing conventions (cf. Millikan, 2003) that endow the actions with transcendent linguistic significance. That is, appropriately produced sequences

of vocal tract actions also count as sequences of words that stand in grammatical relation to one another. Grammatically structured sequences of words, in turn, have meaning that transcends that of the vocal tract actions themselves. Those meanings do the main work in public speaking and listening.

Even so, the vocal tract actions are not *only* vehicles for conveying those meanings, and they do some work in events involving public speaking and listening. They are meaningful actions themselves in having "affordances" (e.g., Gibson, 1979) for interlocutors. That is, they invite and guide actions of interlocutors (Fowler, 1986). The purpose of the present section is to show that an important affordance of phonological speech activity is to invite and guide interpersonal coordination.

Prototypically, language use is embedded in social action where Clark (1996) suggests that it often serves as a "coordination device." For example, interlocutors can agree verbally to meet at some location at some time. Or, in another example, they can talk back and forth about how to proceed as they work together to jockey a bulky piece of furniture through a narrow doorway.

Evidence suggests that language use also serves as a coordination device at the level of phonological gestural forms. There is a large literature (see, e.g., Giles, Coupland & Coupland, 1991, for a review to that date) showing that interlocutors typically converge along a variety of linguistic dimensions as they talk. They may converge in dialect (e.g., Babel, 2010), speaking rate (Street, 1983), vocal intensity (Natale, 1975) and in the rate and duration of pausing (Jaffe & Feldstein, 1970). Some researchers (e.g., Wilson, & Wilson, 2005) report that turn taking is exquisitely timed, with speakers characteristically initiating a turn with little or no delay or and little overlapped speech with the preceding turn. Wilson and Wilson suggest that this reflects interpersonal entrainment of syllable-rate neural oscillations. Couper-Cuhlen (1993) suggests that it reflects something like crossperson entrainment of overt speech rhythms. A new speaker initiates a turn on what would have been a "beat" of the preceding speaker's utterance had it continued.

Although convergence does occur in situations that are not particularly social in nature (e.g., Goldinger, 1998; Shockley, Sabadini, & Fowler, 2004), when it occurs in social settings, it seems to reflect social affiliation or the lack thereof. For example, Babel (2010) found more convergence to an Australian by New Zealanders who showed a pro-Australia bias on an implicit association task than among New Zealanders with a pro-New Zealand bias. Likewise, Bourhis and Giles (1977) found dialect *divergence* on the part of Welsh speakers of English from a British English speaker who had insulted them. Labov (1963) found either divergence or convergence of diphthongal vowels among teenagers on Martha's Vineyard, MA with the vowels of off-islanders depending on whether the teenagers intended to stay on the island (where summering off-islanders were sources of financial distress and therefore hostility to some residents, e.g., farmers) or to leave it after graduation from high school. In short, speakers are disposed to sound like others with whom they affiliate and unlike those to whom they do not.

Public production of speech gestures serving as a coordination device manifests itself in other ways than in guiding the speech gestures of interlocutors. Condon (1976) reported that speakers move in time, not only with their own speech rhythms but also with the rhythms of interlocutors. Bernieri (1988) obtained a similar outcome with judgments of movement synchrony.

In addition, language use can foster cross person entrainment of movements that are invisible: namely, postural sway. Shockley, Santana & Fowler (2003) had participant pairs

engage in a cooperative game in which they spoke back and forth to solve a puzzle. There were four conditions reflecting two levels each of two independent variables. Either the participants could see one another, or else they were back to back. In addition they either cooperated in puzzle solving with one another, or else they individually cooperated with a confederate. Findings were that entrainment of postural sway occurred, but only among participants who worked together to solve the problem. It did not matter whether or not participants could see one another, however. Evidently the information for coordination was being carried in some way over the airways. A follow-up investigation (Shockley, Baker, Richardson, & Fowler, 2007) suggested that one critical property of speech underlying postural entrainment is speech rhythm. Under more constrained conditions than those of Shockley, et al. (2003), pair members showed more entrainment of postural sway when they produced disyllabic words synchronously that had the same stress pattern compared to synchronous speech in which words were not similar in that way.

In short, although phonological gestures of the vocal tract are not the prime movers in social events involving talking and listening, they are integral parts. They are themselves public actions and they serve a role in fostering interpersonal coordination.

## 4.0 The sound-meaning “rift”: Language as a semi-formal system

The foregoing discussion of language forms and their use in public can be justified as a proper topic of discussion by linguistic analysis in which, in the words of Levelt, et al. (1999), there exists a “rift” between meaningless language forms and meaningful aspects of language. Logically, this rift is important in particular to the “particulate principle” or “duality of patterning” discussed earlier. That is, logically, it is undesirable for language particles that compose words each to contribute meaning particles; we do not want our choice of new language forms to be restricted to those combinations of particles that somehow cohere semantically in the required way. As Studdert-Kennedy (1998) remarks, we want the phonological forms to be “bleached” of meaning. However, in nature, the form-meaning separation is not as stark as the descriptor “rift” suggests. In particular, the “meaningless” language forms are not wholly bleached of meaning, and so, even in this small way, meanings themselves are embodied in our production of “meaningless” phonetic gestures.

That the form-meaning rift is only approximate is important in highlighting how language forms arise in communities of speaker-listeners. Different communities of humans speak different languages, because the sets of conventions that constitute the syntax, morphology, lexicon and phonology of a language emerge in the course of interpersonal language use. The syntactic conventions and phonological conventions that underlie on the one hand how words combine in utterances and on the other how vocal-tract gestures combine to form words have to be sufficiently systematic and sufficiently sharable among interlocutors to enable mostly successful interpersonal use of language. The conventions that arise in different communities are different. However, at neither level do the systematicities need to be wholly formal to enable successful language use. Likewise, the particulate principle and duality of patterning require an approximate form-meaning separation. But if it is natural for interlocutors to coin and then reproduce forms that sound appropriate for what they mean, then the separation may only be approximate. An approximate form-meaning rift may be enough for language to do its work.

### 4.1 Meaningful language forms

Given a choice of a rounded shape and a pointed shape, children as young as 2.5 years old will choose the rounded shape if its name is *bouba* but the pointed shape if its name is *kaykee* (Maurer, Pathman, & Mondloch, 2006; cf. Kohler, 1947). There is something

evocative of roundness about *bouba* and of sharpness about *kaykee*. The naturalness of these pairings may reflect something about the vocal-tract shapes or actions corresponding to each word, or, for those of the view that listeners hear acoustic signals, something about the acoustic patterns that evoke roundness and sharpness respectively.

The meanings evoked by language forms are not always as obvious as those of *bouba* and *kaykee*, but they are pervasive enough that form-meaning links should not be considered exceptional cases at the periphery of language. For example, Brown and Nuttall (1959) presented English speakers who were unfamiliar with both Chinese and Hindi with antonym pairs in those languages. For each pair, they were asked which of the English translations of the pair members went with which foreign-language word. Performance was significantly above the chance level of .5 for choices in both languages. Kunihira (1971) found the same result for English speakers judging antonym pairs in Japanese. Nygaard and colleagues (Namy, Nygaard, Clepper & Rasmussen, in preparation, as described in Nygaard (2010) replicated these findings, again, presenting native English speakers with antonym pairs in languages they did not know. They found above-chance selection when antonym pairs were presented in Danish, Japanese, Russian and Shona. The far-from perfect performance in these experiments suggests that the form-meaning links in question can be quite subtle. (E.g., a pair from Kunihira (1971) is *akarui-kurai* in Japanese corresponding to *bright-dark* in English.)

A striking finding by Remez, and colleagues (Remez, Fellowes, Blumenthal, & Nagel, 2003) reinforces the finding that “meaningless” language forms evoke meaning in listeners. They presented their participants with analogies to solve. All were of the form: EE is to AH as X is to Y, or as Y is to X, where X and Y were English words. The words were sensory in nature (e.g.., EE is to AH as ROUGH is to SMOOTH, or as SMOOTH is to ROUGH?) or were abstract (e.g., COFFEE-TEA) or were arbitrary (e.g., DICTIONARY-BOWL). Strikingly, participants were in significant agreement in their choices on 22 of 32 sensory and abstract analogies and 20 of 32 arbitrary analogies. The results could not be explained based on observations that the vowel /i/ is judged higher in pitch and brighter than /a/. No obvious simple characterization of the acoustic differences between /i/ and /a/ explained the data, but, of course, *something* does. It is no-doubt related that, stereotypically, on seeing a mouse, people exclaim *eek*, whereas in sinking into a relaxing bath, they say *aaah*.

There are other markers of leakiness across the form-meaning “rift.” In a study of 50 languages, meant to represent the languages of the world, Bybee (1985) found that the spatial arrangement of affixes with respect to the verb stem reflected the extent to which the stem-affix combination changed the meaning of the composite form relative to the meaning of the bare stem. Affixes that had a greater impact on verb meaning tended to be attached closer to the verb stem than those that changed the meaning less. Bybee referred to this variable as the semantic “relevance” of the affix to the stem. (See also Givón, 1979, 1985, and Haiman, 1985a, b, for kinds of iconicity in syntax. For an overview of iconicity in signed and spoken language, see Perniss, Thompson, & Vigliocco, 2010.)

Finally, Nygaard, Herold, and Namy (2009) showed that prosodic variation can carry word meaning. They comment that, traditionally, prosody has been supposed to serve two roles in language understanding. One is structural. Intonational patterning and patterns of durational variation may mark the phrasal structure of an utterance, and, in stress languages such as English, strong syllable onsets may mark content word onsets (e.g., Cutler & Norris, 1988). A second is that prosody may provide information about speakers’ states, including attitudinal or emotional, or their conversational focus. However, Nygaard et al. show that variation in fundamental frequency, duration, and amplitude can also provide information about word meaning. Speakers’ manners of producing nonwords such as *blicket* (in infant-

directed speech) intending to indicate a big as contrasted with a small object, or a hot entity as contrasted with a cold one, provide prosodic information that permits listeners to distinguish which adjectival meaning the speaker intended.

As noted, a likely reason why these form-meaning correlations exist in languages must have to do with the fact that words are coined and used in the public domain. Word forms that are somehow apt may be more likely to be coined in the first place and then may be more likely to be reproduced than those that are somehow less apt. In addition, Nygaard, et al. (2009) have shown that English speakers *learn* antonyms better in an unfamiliar language (Japanese) when the new words map onto their correct translations (e.g., Japanese *hayai* paired with English *fast*) than when they are paired at random to English antonym words. (e.g., *hayai* paired with *blunt.*) (There is also a weaker tendency for learners to acquire mappings between an English word (e.g., *slow*) and its anonym in Japanese (e.g., *hayai* meaning *fast*) than to learn random pairings of English and Japanese antonyms.)

### 4.2 Embodied meanings

I have argued that language forms are actions of the body, specifically of the vocal tract (see also, e.g., Browman & Goldstein, 1992). As such, they are public actions that are immediately, that is, directly (Gibson, 1966; Fowler, 1986, 1996; Kelley, 1986) available to interlocutors. Although the smallest language forms are logically meaningless in being the permutable particles that can be combined in indefinitely many ways to form words that can mean anything, they are not wholly meaningless. In this final section, I briefly point out that this is not the only sense in which language meanings inhere in bodily action more broadly. This can be seen in two ways.

First, Millikan (2003) remarks that, among other aspects of language, words are *conventions* of language communities in the sense that they persist as parts of the language because community members reproduce them. Once coined, words come to have “lineages” in a language community, and language users’ reproductions of a word reflect that history of use. Accordingly, what words *mean* for individual members of a language community corresponds to that part of a word’s lineage that each has intercepted and has contributed to. In this sense, a word’s meaning *is* its public usage as experienced by individual community members, and “usages” are bodily actions produced or perceived.

Second, much research in recent decades has shown that our understanding of language is embodied quite generally. For example, participants make faster judgments that *You delivered the pizza to Andy* makes sense if the response movements are toward the responders’ body (mirroring the ostensible movement of the pizza) than away from the body; judgments that *Andy delivered the pizza to you* show the opposite patterning of relative response times (Glenberg & Kaschak, 2002). Not surprisingly, then, action words (such as *grasp* or *kick*, for example) activate parts of the motor system of the brain for carrying out those actions. (See Pulvermuller & Fadiga, 2010 for a review of this literature.) Complementarily, actions can facilitate understanding of language consistent with those actions. For example, when participants in an experiment adopt a hand posture such as “clench” or “pinch” prior to making a judgment of the sensibility of a phrase such as “aim a dart,” their response times are faster when the hand postures they adopt match those implied in the verbal phrase (Klatzky, Pellegrino, & McCloskey, 1989). In short, as language users, we embody the language forms that we produce and understand, and embodiment constitutes some of what those forms mean. As community members, we also intercept and contribute to the lineages of language forms, such as words, which constitute the meanings of those forms for the community.

## 5.0 Concluding summary

The cognitive revolution in psychology and linguistics led to important advances in our understanding of language itself and its covert use (inside the head) by language users. However, it has fostered misunderstandings stemming from the idea that language's most important function is primarily "internal, for thought" (Chomsky, 2006). A particularly ironic misunderstanding is that of consonantal and vocalic language forms as abstractions in the mind that are distorted in the activity of making them public.

Recognizing that language's most important functions are public and social motivates a different understanding of language forms and of language itself. As for language forms, they are public actions of the vocal tract that can be directly, immediately perceived by interlocutors and that do work in interpersonal language use in serving as a "coordination device" (Clark, 1996). As for language itself, its interlocking sets of conventions are always only mostly shared by language community members and are always changing. Therefore, they are only approximately formal systems. For these reasons, it is not surprising that, in addition to being only *almost* characterizable by syntactic and morphological rules, languages exhibit leakage across the ostensible "rift" between forms and meanings. Meaningless language forms are only mostly meaningless, and their meanings emerge from and inhere in bodily action.

## Acknowledgments

Preparation of the manuscript was supported by NICHD grant HD-001994 to Haskins Laboratories.

## References

Abler W. On the particulate principle of self-diversifying systems. Journal of Social and Biological Structures. 1989; 12:1–13.

Babel M. Dialect divergence and convergence in New Zealand English. Language in Society. 2010; 39:437–456.

Bernieri FJ. Coordinated movement and rapport in teacher-student interactions. Journal of Nonverbal Behavior. 1988; 12:120–138.

Bourhis, R.; Giles, H. The language of intergroup distinctiveness. In: Giles, H., editor. Language, ethnicity and intergroup relations. London: Academic Press; 1977. p. 119-135.

Browman C, Goldstein L. Towards an articulatory phonology, Phonology Yearbook. 1986; 3:219–252.

Browman C, Goldstein L. Articulatory phonology: An overview. Phonetica. 1992; 49:155–180. [PubMed: 1488456]

Browman, C.; Goldstein, L. Dynamics and articulatory phonology. In: Port, R.; van Gelder, T., editors. Mind as motion: Explorations in the dynamics of cognition. Cambridge, MA: MIT Press; 1995. p. 175-193.

Browman C, Goldstein L. Competing constraints on intergestural coordiantion and self-organization of phonological structures. Bulletin de la Communication Parlée. 2000; 5:25–34.

Brown R, Nuttall R. Method in phonetic symbolism experiments. The Journal of Abnormal and Social Psychology. 1959; 59:441–445.

Bybee, J. Morphology. Amsterdam: John Benjamins Publishing Company; 1985.

Chomsky, N. Why are there so many languages? Diversification from an underlying unity. New Haven, CT: Talk to benefit the Endangered Language Fund; 2006. November 15, 2006. (http://www.endangeredlanguagefund.org/chomsky.php)

Chomsky, N.; Halle, M. The sound pattern of English. New York: Harper and Row Publishers; 1968.

Clark, H. Using language. Cambridge: Cambridge University Press; 1996.

Condon W. An analysis of behavioral organization. Sign Language Studies. 1976; 13:285–318.

Cowley S. Taking a language stance. Ecological Psychology. 2011; 23:185–209.

Couper-Cuhlen, E. English speech rhythm. Amsterdam: John Benjamins; 1993.

Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance. 1988; 14:113–121.

de Boer, B. The origin of vowel systems. Oxford: Oxford University Press; 2001.

Dell G. A spreading-activation theory of retrieval in speech production. Psychological Review. 1986; 93:283–321. [PubMed: 3749399]

Fowler C. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics. 1986; 14:3–28.

Fowler CA. Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. Perception & Psychophysics. 1994; 55:597–610. [PubMed: 8058448]

Fowler CA. Listeners do hear sounds, not tongues. Journal of the Acoustical Society of America. 1996; 99:1730–1741. [PubMed: 8819862]

Fowler CA. Parsing coarticulated speech in perception: Effects of coarticulation resistance. Journal of Phonetics. 2005; 33:195–213.

Fowler CA, Saltzman E. Coordination and coarticulation in speech production. Language and Speech. 1993; 36:171–195. [PubMed: 8277807]

Fowler, CA.; Smith, MR. Speech perception as "vector analysis:" An approach to the problems of segmentation and invariance. In: Perkell, J.; Klatt, D., editors. Invariance and variability of speech processes. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.

Gafos A. A grammar of gestural coordination. Natural Language and Linguistic Theory. 2002; 29:269–337.

Garrod MJ, Pickering S. Why is conversation so easy? Trends in Cognitive Sciences. 2004; 8:8–11. [PubMed: 14697397]

Gibson, JJ. The senses considered as perceptual systems. Boston: Houghton-Mifflin; 1966.

Gibson, JJ. The ecological approach to visual perception. Boston: Houghton-Mifflin; 1979.

Giles, H.; Coupland, N.; Coupland, J. Accommodation theory: communication, context and consequence. In: Giles, H.; Coupland, N.; Coupland, J., editors. Contexts of accommodation: Developments in applied sociolinguistics. Cambridge: Cambridge University Press; 1991. p. 1-68.

Givón, T. On understanding grammar. New York: Academic Press; 1979.

Givón, T. Iconicity, isomorphism and non-arbitrary coding in syntax. In: Haimon, J., editor. Iconicity in syntax. Amsterdam: John Benjamins; 1985. p. 187-219.

Glenberg A, Kaschak M. Grounding language in action. Psychonomic Bulletin and Review. 2002; 9:558–565. [PubMed: 12412897]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. [PubMed: 9577239]

Goldstein, L.; Fowler, CA. Articulatory phonology: A phonology for public language use. In: Schiller, N.; Meyer, A., editors. Phonetics and phonology in language comprehension and production: Differences and similarities. Berlin: Mouton de Gruyter; 2003. p. 159-207.

Haiman, J. Natural syntax. Cambridge: Cambridge University Press; 1985a.

Haiman, J., editor. Iconicity in syntax. Amsterdam: John Benjamins Publishing Company; 1985b.

Hockett, C. A manual of phonetics. Bloominton: Indiana University Press; 1955.

Hockett C. The origin of speech. Scientific American. 1960; 203:89–96. [PubMed: 14402211]

Iskarous K. Vowel constrictions are recoverable from formants. Journal of Phonetics. 2010; 38:375–387. [PubMed: 20871808]

Iskarous K, Fowler CA, Whalen DH. Locus equations are an acoustic expression of articulator synergy. Journal of the Acoustical Society of America. 2010; 128:2021–2032. [PubMed: 20968373]

Jaffe, J.; Feldstein, S. Rhythms of Dialogue. New York: Academic Press; 1970.

Johnson K. The auditory-perceptual basis for speech segmentation. Ohio State University Working Papers in Linguistics. 1997; 50:101–113.

Johnson, K. Speaker normalization in speech perception. In: Pisoni, DB.; Remez, RE., editors. The handbook of speech perception. Malden: Blackwell Publishing; 2005. p. 363-389.

Kelley, D. The evidence of the senses. Baton Rouge: Louisiana State University Press; 1986.

Klatzky RL, Pellegrino JW, McCloskey BP. Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. Journal of Memory and Language. 1989; 28:56–77.

Köhler, W. Gestalt psychology. 2nd Edition. New York: Liveright Publishing; 1947.

Kunihira S. Effects of the expressive voice on phonetic symbolism. Journal of Verbal Learning and Verbal Behavior. 1971; 10:427–429.

Labov W. The social motivation of a sound change. Word. 1963; 19:273–309.

Levelt WJM, Roelofs A, Meyer A. A theory of lexical access in speech production. Behavioral and Brain Sciences. 1999; 22:1–38. [PubMed: 11301520]

Liberman A, Delattre P, Cooper F. The role of selected stimulus variables in the perception of the unvoiced-stop consonants. American Journal of Psychology. 1952; 65:497–516. [PubMed: 12996688]

Liberman A, Mattingly I. The motor theory revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

Liberman A, Whalen D. On the relation of speech to language. Trends in Cognitive Sciences. 2000; 4:187–196. [PubMed: 10782105]

Liljencrants J, Lindblom B. Numerical simulation of vowel quality systems: The role of perceptual contrast. Language. 1972; 48:839–862.

Lindblom, B.; Maddieson, I. Phonetic universals in consonant systems. In: Hyman, L.; Li, CN., editors. Language, speech and mind. London: Routledge; 1988. p. 62-78.

Maddieson, I. Patterns of sounds. Cambridge: Cambridge University Press; 1984.

Maurer D, Pathman T, Mondloch C. The shape of boubas: Sound-shape correspondences in toddlers and adults. Developmental Science. 2006; 9:316–322. [PubMed: 16669803]

Millikan, RG. In defense of public language. In: Antony, LM.; Hornstein, N., editors. Chomsky and his critics. Oxford, UK: Blackwell Publishing Ltd; 2003. p. 215-237.

Natale M. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. Journal of Personality and Social Psychology. 1975; 32:790–804.

Nygaard, LC. Cross-linguistic sound symbolism; Paper presented at the conference on Sound symbolism: Challenging the arbitrariness of language; March 26, 2010; Emory University, Atlanta Georgia. 2010. http://psychology.emory.edu

Nygaard LC, Herold DS, Namy LL. The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. Cognitive Science. 2009; 33:127–146. [PubMed: 21585466]

Ohala, J. The listener as a source of sound change. In: Masek, C.; Hendrick, R.; Miller, R.; Miller, M., editors. Papers from the parasession on language and behavior. Chicago: Chicago Linguistics Society; 1981. p. 178-203.

Oudeyer PY. The self-organization of speech sounds. Journal of Theoretical Biology. 2005; 233:435–449. [PubMed: 15652151]

Perniss P, Thompson RL, Vigliocco G. Iconicity as a general property of language; Evidence from spoken and signed languages. Frontiers in Psychology. 2010; 1:1–17. [PubMed: 21833184]

Pierrehumbert J. The next toolkit. Journal of Phonetics. 2006; 34:516–530.

Pierrehumbert, JB.; Hopper, P.; Bybee, J. Frequency and the emergence of linguistic structure. Amsterdam Netherlands: John Benjamins Publishing Company; 2001. Exemplar dynamics: Word frequency, lenition and contrast; p. 137-157.

Port RF. Rich memory and distributed phonology. Language Sciences. 2010a; 32:43–55.

Port RF. The reality of phonological forms: A rejoinder. Language Sciences. 2010b; 32:60–62.

Prince, A.; Smolensky, P. Optimality Theory: Constraint interaction in generative grammar. Malden, MA: Blackwell Publishing; 2004.

Pulvermüller F, Fadiga L. Active perception: Sensorimotor circuits as a cortical basis for language, Nature Reviews Neuroscience. 2010; 11:351–360.

Remez RE, Fellowes JM, Blumenthal EY, Nagels DS. Analysis and analogy in the perception of vowels. Memory & Cognition. 2003; 31:1126–1135. [PubMed: 14704027]

Saltzman E, Munhall K. A dynamical approach to gestural patterning in speech production. Ecological Psychology. 1989; 1:333–382.

Shattuck-Hufnagel, S. Speech errors as evidence for a serial-ordering mechanism in sentence production. In: Cooper, W.; Walker, E., editors. Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, NJ: Lawrence Erlbaum; 1979. p. 295-342.

Shockley K, Baker AA, RIchardson MJ, Fowler CA. Articulatory constraints on interpersonal coordination. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:201–208. [PubMed: 17311488]

Shockley K, Sabadini L, Fowler CA. Imitation in shadowing speech. Perception & Psychophysics. 2004; 66:422–429. [PubMed: 15283067]

Shockley K, Santana M-V, Fowler CA. Mutual interpersonal postural constraints are involved in the supra-postural task of cooperative conversation. Journal of Experimental Psychology; Human Perception and Performance. 2003; 29:326–332.

Stevens, K.; Blumstein, S. The search for invariant correlates of phonetic features. In: Eimas, P.; Miller, J., editors. Perspectives on the study of speech. Hillsdale, NJ: Lawrence Erlbaum Associates; 1981. p. 1-38.

Street, R. Noncontent speech convergence in adult-child interactions. In: Bostrom, R., editor. Communication Yearbook. Vol. 7. Beverly Hills, CA: Sage; 1983. p. 369-395.

Studdert-Kennedy, M. The particulate origins of language generativity: From syllable to gesture. In: Hurford, J.; Studdert-Kennedy, M.; Knight, C., editors. Approaches to the evolution of language. Cambridge: Cambridge University Press; 1998. p. 202-221.

Viswanathan N, Fowler CA, Magnuson JS. A critical examination of the spectral contrast account of compensation for coarticulation. Psychonomic Bulletin & Review. 2009; 16:74–79. [PubMed: 19145013]

Viswanathan N, Magnuson JS, Fowler CA. Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. Journal of Experimental Psychology: Human Perception and Performance. 2010; 36:1005–1015. [PubMed: 20695714]

Wilson M, Wilson T. An oscillator model of the timing of turn-taking. Psychonomic Bulletin & Review. 2005; 12:957–996. [PubMed: 16615316]

## Highlights

- Fundamental language forms are public vocal tract actions that are adapted to their public use
- Although they are largely vehicles for effective public language use, they have their own roles, for example, as coordination devices
- The form-meaning rift implied by the foregoing discussion is only approximate
- “Meaningless” language forms have some meaning
- In turn meanings of meaningful language forms emerge from their “lineages” of public use

## Highlights

- Primitive language forms are vocal tract action adapted to their public use
- They are vehicles for public language use that have roles, e.g., in coordination
- The form-meaning rift implied by the foregoing discussion is only approximate
- “Meaningless” language forms have some meaning
- In turn, meanings of forms emerge from their “lineages” of public use