Chapter 7

The origin of words: how frame-stage patterns acquired meanings

7.1 Introduction
7.2 Baby talk and the first words
7.3 Parental terms as first words
7.4 Other possibilities for first words
7.5 Summary
7.6 Coda

7.1 Introduction

How did hominids actually make their first words?

So far I have presented a theory that includes a hypothesis about what the sound patterns of the first words were like. But that is only one part of the story of speech, and, in the view of most, the less important part. The other part—the part that made vocalizations become speech, because speech is a component of language—was the pairing of particular sound patterns with particular concepts. When this happened, the sound pattern took on meaning, becoming a symbol for the concept to which it was paired. This pairing is what made words, and language proper began with words. I share the view of Jackendoff (2002) and many others (e.g., Deacon, 1997; Donald, 1991) who, in Jackendoff’s words, see “symbol use as the most fundamental factor in language evolution” (Jackendoff, 2002, p. 239). And the origin of symbol use could well have been the most important event in the evolution of humans.

The Frame/Content theory is a theory about how we evolved the sound system we have today. It is a system whereby consonants and vowels are programmed into syllable structures. This system is what allows the production of an indefinite number of words in present-day languages. I have argued that the motor basis for this system—the frame stage of output—evolved before speech, in the service of vocal grooming, and then was pressed into service for the first words. So my task will be to present a possible scenario for the origin of words by pairing of concepts with pre-existing frame-stage patterns in particular. There might have been other pairings of concepts with sounds. I will consider some possibilities later. And these may have even preceded the pairings of concepts with frame-stage patterns. But for me, speech, and therefore language, truly began with the latter pairings because only they provided the initial groundwork for the speech sound system that we have today. Remember from Chapter 3 that what was required to have an open-ended vocabulary of words was an implementation of the particulate principle whereby “elements drawn from a finite set…are repeatedly permuted and combined to form larger units” (Studdert-Kennedy, 1998, p. 203). Both syllables and segments can be considered to have this property.

What concepts are is a very complex issue, one that I must necessarily gloss over here with the excuse that this book is primarily about the sound level, not the concept.
level. But note that the concepts that initially became associated with sound patterns must have been very simple. Certainly they must have been simple in the sense that they were available prior to the explosion in conceptual complexity made possible by the ability to pair them with sound patterns. More importantly, however simple they were, there must have been compelling reasons for them to acquire a signaling component. I will try to evaluate possible candidates for first words against this criterion.

It is relatively uncontroversial that living primates most closely related to us have concepts, by which I mean mental representations that can stand for particular aspects of the world of their experience. For example, Jackendoff (2002) take[s] it as established by decades of primate research…that chimpanzees have a conceptual structure in place, adequate to deal with physical problem solving, with navigation, and above all with rich and subtle social interaction incorporating some sense of the behavior patterns and perhaps intentions of others. (p. 238)

What then was necessary for the evolution of speech was to pair pre-existing concepts with vocal patterns that symbolized them. But when we raise this issue we immediately encounter a major stumbling block, an obstacle so important that it deserves a digression into the history of thought about the origins of speech.

In the ruins of the ancient city of Babylon, in Iraq, lie the remains of what is now known as the Tower of Babel, originally an impressive seven-story, stepped-pyramid construction. According to the Bible (Genesis 1:11, or see Pinker, 1994, p. 231, depending on your affiliation), this tower was intended to reach to the sky and thus bring the residents closer to God. But, the story goes, God was incensed at the presumption behind this endeavor, and responded by assigning different languages to the different subgroups of construction workers, thus preventing them from communicating with each other. As a result, the project ground to a halt.

This is the biblical account of how humans came to have many languages. Prior to this, the story goes, there was only “ adamant language,” the language God provided for Adam and Eve. The implication is that in adamant language there was somehow a natural relation between concepts and sound patterns. The story of the Tower of Babel seemed to explain the lack of correlation across languages between the meaning of a word and its sound pattern. For example, the word for “tooth” in German is “zahn,” in Irish it is “fiacl,” in Greek it is “dhondi,” in Polish it is “zab,” in Finnish it is “hammas,” in Hungarian it is “fog,” in Basque it is “ortz,” etc. How on earth could the world’s languages have evolved so many different sound patterns to symbolize our various words?

Since 1859, the Tower of Babel story has fallen out of favor except among religious fundamentalists. But what story of the origin of words do we replace it with? Sound-meaning relationships could not have been arbitrary in the beginning. There must have been a time at which the first word was invented. And we have to assume that it came to signify an object or event which was present in the environment when the sound pattern was made. The connotation of the word “arbitrary” suggests that if the initial sound-concept pairing was arbitrary, the speaker simply consulted a table of random numbers each of which signified a particular pre-existing sound pattern, picked a number, and assigned the associated sound pattern to the concept. But there could have been no such table of random numbers. Some sound pattern must have been produced in the presence of the entity for which we had a conceptual representation leading someone, and eventually many people, to take, by convention, the sound pattern to signify the concept. The question is, What were the circumstances that led to a particular sound pattern being produced in the presence of the entity for which presumably two observers had a common conceptual structure? This occasion must, in some sense, have arisen naturally.
It must have had some natural causes. It must have in some sense come naturally. So what naturally occurring events might have led to this original sound-meaning pairing?

To see this stumbling block of apprehending the original nature of the concept-sound pairing better in the presence of current arbitrariness, we should note that there is a way of avoiding it if we assume that the first word was a word in sign language. For example, if two people were looking at a stationary bird and one flapped her arms up and down in unison, it’s not difficult to imagine that the other person might be able to realize that the flapping motion stood for, or was a symbol for, the bird. Then on the next occasion on which a bird was in common view, the flapping action could be accepted as meaning bird. Here there is a natural relation between the concept and the symbol for it. This is called “iconicity.”

But returning to the problem of linking vocal symbols to concept, you might suppose that historical linguistics would have something to say about this. Unfortunately, this field is in disarray. There is a major split within the discipline between a rather timid orthodoxy, reluctant to push their questions back much beyond about 7,000 years, or across very many languages (e.g., Campbell; Dixon; Hamp; Goddard), and a smaller group of mavericks whose more adventurous suggestions are considered by the orthodoxy to be ill-conceived (e.g., Dogolopolsky; Greenberg; Ruhlen; Shevoroshkin, ). (See discussion in “Nostratic: Sifting the evidence” [Salmons & Joseph, 1998]).

Disappointingly, neither the orthodoxy nor the mavericks have had much to say about the actual origin of the stage of pairing sounds with meanings. In particular, as far as I can ascertain, the predominant view of phonological organization is that it did not evolve from simple to more complex but has always been the way it is today. Bengtson and Ruhlen (1994), members of the maverick group, did suggest a list of 27 possible protoword roots or global etymologies—in effect, words of the first language. These words appeared to be plausible candidates from my phonetic perspective in that they primarily involved CV syllables and consonants and vowels that are relatively frequent in infants and in modern languages. Davis and I (2000) analyzed the sound patterns of these words and found that they showed a mixture of frame-stage and frame/content-stage patterns. But Boe et al. (in press), in a subsequent analysis of the methodology that Bengtson and Ruhlen used to come up with these words, showed that it involved a deeply flawed methodology, flawed to the point that the proposed sound patterns could be simply a result of chance. Boe et al. concluded that “They used too few global etymologies, too many equivalent meanings, too many languages per family, and too many phonological equivalences for a too small number of different phonological shapes” (p. 114).

7.2 Baby talk and the first words

There is one remaining domain that seems to allow us to develop a story about how frame-stage patterns in particular became linked to concepts. It is the domain of baby talk. This genre, which allows us today to trade concepts with infants, might have originally also allowed us to begin to trade concepts with each other.

What then is “baby talk,” and how can it bear the heavy theoretical burden I am putting on it? Ferguson (1964) defines baby talk as “any special form of a language which is regarded by a speech community as being primarily appropriate for talking to young children and which is generally regarded as not the normal adult use of language” (p. 103). Since baby talk is a two-way affair, we ought to expand its definition to include the
vocalizations made by infants to adults. Baby talk is known to be widespread in different language communities, and it may even be universal, although it’s known that in some cultures, communicative interaction between caregivers and infants is minimal. In a study of baby talk in 6 languages, Ferguson (1964) found that, with respect to words, the special lexical items typically number from 25 to 60. They include the following categories (accompanied here with English examples using Ferguson’s notations: kin terms and appellations (“mommy”, “daddy”); bodily functions (“wee-wee”, “poop [oo]”); certain simple qualities (“ææ” [meaning don’t], “teenie”); and vocabulary concerning animals, nursery games and related items (“doggie”, “kitty [-cat], “bye-bye”, “peek-a-boo”). Ferguson gives us 211 examples of these terms in 6 languages.

The first thing needed in order to hypothesize that baby talk might contain the key to the first words is to establish that the form of baby-talk terms is indeed analogous to the form of infants’ babbling and first words. Because we have provided a precise characterization of the infant patterns, we are in a position to make, for the first time, a quantitative estimate of how similar baby talk is to infant productions in general. As to sound inventory, Ferguson points out that, just as in babbling and early speech, “baby talk items consist of simple, more basic kinds of consonants, stops and nasals in particular, and only a very small selection of vowels” (p. 109). As to serial organization, as might be expected, the CV form, as in every other aspect of speech we have considered, is the most favored syllabic form. The most favored word type is CVCV, the 80 examples of which make up 38% of the words in Ferguson’s corpus. And intersyllabically, as in babbling and early speech, reduplication is most characteristic. For example, the CVCV utterances are totally reduplicated 53% of the time, as compared with the 50% of successive pairs of CVs we found to be reduplicated in babbling (Davis and MacNeilage, 1995).

But if these forms are to be considered truly similar to babbling and first-word forms of infants, it’s crucial that they also have the three CV and VC co-occurrence constraints that are characteristic of these two stages. Table 7.1a shows the co-occurrence patterns of front, central, and back vowels with coronal, labial, and dorsal consonants in the CV syllables that occur in the 80 CVCV words. Table 7.1b shows also that, as expected, the VC patterns are very similar to the CV patterns. The two individual patterns were both significantly different from chance at the .01 level.

There was also a tendency in these forms to favor the labial-coronal (LC) sequence over the coronal-labial (CL) sequence as we had found in infants’ first words. There were 7 instances of the LC pattern but only 1 instance of the CL pattern.

Of course it’s not surprising that baby-talk words are, in the main, similar to forms that the infant produces during the babbling and first-word periods. Baby talk is used by both infants, starting at about the time they otherwise produce the first words of their language, and by parents, and in order for this to be possible, given the infants’ vocal limitations at the time, we can assume that this special language is at least to some degree tailored to conform to these limitations. But although the parallels of baby talk with babbling and infants’ first words might have been expected, it is useful to have a specific idea of what they are. And the finding of close parallels is further evidence of the robustness of the serial-organization patterns we have identified in speech acquisition.

But while the forms of baby-talk words are, in general, similar to those of babbling, baby-talk words have meaning whereas babbling doesn’t. The question is whether one could nevertheless say that baby talk is a separate genre from language
proper, and therefore the origin of the sound-meaning pairings found here needs to be considered separately from the origin of words in language proper.

Such a separation may be insufficiently motivated. Numerous people have suggested that there seems to be a tendency for baby-talk forms to get into language proper, especially in the case of parental terms. Here is Roman Jakobson’s (1960) view of the matter:

Some of the nursery forms overstep the limits of the nurseries, enter into the general usage of the adult society, and build a specific infantile layer in standard vocabulary. In particular, adult language usually adopts the nursery forms, designating each of the mature members of the nuclear family. Very frequently these intimate, emotional, childishly tinged words coexist with more general and abstract, exclusively parental terms. (p. 125)

A possible paradigm is illustrated by the situation in English noted by Jakobson (1960). He reminds us that the reduplicated nursery forms “mama (mamma, mammy, ma, mom, mommy) and papa (par, pappy, pa, pop, or dada, dad, daddy)” (p. 125) co-exist with orthodox variegated language terms “mother” and “father.” Elsewhere (Jakobson, 1962) he suggests that in this particular case, words of true language could be modifications of nursery words:

In Indo-European the intellectualized parental designations mater and p(schwa)ter were built from the nursery forms with the help of the suffix –ter, used for various kin terms.

Then, as Jakobson sees it, a sound change occurred in p(schwa)ter making the initial stop [p] and the intervocalic stop [t] into fricatives, and another such change has the same effect on the [t] of mater. Thus we end up, in English, with the true language terms “father” and “mother.”

The strength of this tendency for baby-talk terms for parents to leak into language proper can be estimated by considering the phonetic structure of parental terms across languages. This estimate can be made from data (unfortunately only on the structure of the first syllable) from a rather remarkable study of parental terms of the language proper in 474 languages by Murdock (1959).

Note that Murdock took some pains there to rule out borrowings of words from European languages “due to recent missionary and other influences.” Also, “Forms for mama and papa were excluded unless comparative data on related languages clearly demonstrated their indigenous origin” (p. 1). CV relationships in 881 syllables from Murdock’s data are shown in Table 7.2. The vowel classification he used was different from the one we have used. But his two categories of high and mid-front vowels and high and mid-back vowels can be taken to represent front and back vowels respectively because, as we saw in Chapter 3, languages don’t tend to have many low-front and low-back vowels. In addition, his low vowels can be taken to represent central vowels, because Murdock indicated that the low-central vowel [a] was the most frequent vowel in this category, as it is in languages in general.

As the table shows, all three of the usual CV co-occurrence patterns are present in this data. No other CV category exceeds chance levels. Murdock presented data only on the first syllable of parental terms. Paine has made an independent analysis of parental terms in a broad sample of 211 languages in order to consider properties of the words in general (unpublished observations). She found that, in a corpus of 141 CVCV forms involving stops and nasals, 45% of them involved syllabic reduplication, a level
characteristic of infant babbling and first words but much higher than in languages in
general, where syllabic reduplication is quite rare.

As Tables 7.3a and 7.3b show, she also found the expected consonant-vowel co-
ocurrences between coronals and front vowels, and labials and central vowels in the first
and the second syllables of CVCV forms. (As there were only 9 dorsal consonants and 9
back vowels in the corpus, these were omitted from the overall analysis.) She also found
the preferred relationship between coronals and front vowels and between labials and
central vowels in VC sequences (Table 7.3c), indicating that these forms are like infant
forms but unlike words of language proper. A total of 333 sequences were analyzed. All
three distributions shown in Table 7.3 were statistically significant. Finally, Paine also
found significantly more LC patterns (17) than CL patterns (4).

In conclusion, parental terms in language proper are highly similar to both baby-
talk words and infants’ babbling and first words. And in the two cases in which infant
and language patterns differ, namely in reduplication versus variegation and VC co-
ocurrences versus their absence, parental terms in language unequivocally go with infant
and baby-talk patterns rather than language patterns in general. As the simple frame-stage
patterns of infants and of baby talk tend to be found in kinship terms in actual languages,
it becomes more difficult to dismiss the baby-talk patterns as irrelevant to the structure of
language proper.

7.3 Parental terms as first words

There is one further aspect of sound-meaning relationships in baby-talk terms that
also carries over into languages proper, and it may provide a crucial cue as to how the
sound-meaning pairings were made in the first place. It is that baby-talk words for the
female parent tend to favor nasal consonants while words for the male parent tend to
favor oral consonants. For example, Table 7.4 shows the parental terms in the 6 languages
of Ferguson’s baby talk corpus. Note that all the words for female parent have nasal
consonants and all the words for male parent have oral consonants.

This pattern is also present in parental terms in languages proper. We know this
from Murdock’s (1959) study. He found that 75% of the consonants in words for female
parent contained a nasal consonant while only 25% of the consonants were oral
consonants. In contrast, only 19% of the consonants in words for male parent contained a
nasal consonant while 81% of the consonants were oral consonants. Paine’s findings
were similar. She found that 77% of the consonants in maternal terms contained a nasal
while only 23% contained a nasal. In contrast, only 25% of the consonants in paternal
terms were nasals, the remainder being orals.

To my knowledge, this is the only phonetic regularity consistently distinguishing
two opposing concepts across the world’s languages. Could the basis of it have been
established in the first language and have been maintained as subsequent languages
descended from the first one? One could argue that the necessity for phonetic terms to be
simple in order for them to work in parent-infant communication preserved them against
change. For example, Ferguson (1964) presents evidence of the persistence of baby-talk
words for “food,” “drink,” and “sleep” for some two thousand years in the
Mediterranean area: “The Roman grammarian Varro (116-27 B.C.) cites Latin bua and
papa or pappa as baby talk for ‘drink’ and ‘food’ respectively, and the use of Latin
naenia ‘dirge’, ‘lament’ in the baby talk meaning of lullaby is attested” (p. 104).
Are we therefore more entitled to push the origins of modern baby-talk terms and the phonetically simple modern words of the language from which they derive back in time than we are for words in general?

Perhaps not. In standard historical linguistics, one encounters a good deal of pessimism on the question of whether any present-day word could have been an original word. It is the consensus among historical linguists of any stripe that sound change is ubiquitous, and it goes on all the time, often in front of our very noses (Labov, 1994). We know that it can be relatively rapid. The Old English of the classic 8th-century English text *Beowulf* is virtually unintelligible. For example, its opening line is “Hwaet! We Gar Dena in Geardagum” (Allman, 1990, p. 62). As mentioned earlier, orthodox historical linguists are in general dubious as to whether we can trace any word back beyond about 7,000 years (e.g., Dixon, 1997). From this standpoint it could be argued that present-day baby-talk terms and parental terms of language couldn’t possibly be very old.

Regardless of whether these words are stable across time or not, it’s common in linguistics to consider these patterns to be irrelevant to the question of actual language genesis even though they occur in words of language. For example, the historical linguist Hock (1986) concludes that “words derived from the babbling-based variety of ‘baby talk’ are generally considered unreliable evidence in attempts to establish genetic relationships” (p. 559). He thought this was true because of “the greater possibilities of chance similarities” (p. 555). However, the maverick historical linguist Ruhlen (1994) believes that attributing to chance any similarities in baby-talk forms across language, as orthodox historical linguists do, reflects bias. He thinks the conclusion was “so readily accepted” in historical linguistics because “it explained similarities which, for want of a better explanation, would have to be attributed to common origin, thus undermining the supposed independence of many language families” (p. 41). Instead, he believes that “the supposed independent development of kinship terms like mama and papa in the world’s languages has been greatly exaggerated; in most such cases we are probably dealing with historically connected forms rather than with independent creations” (p. 41).

To address this difference of opinion on the origin of parental terms we turn to a paper, available on the Internet, by a traditional historical linguist, Larry Trask, entitled “Where do mama/papa words come from?” (Trask, 2005). He begins by noting that there are a number of people (Ruhlen presumably included) who take what he calls a “Proto-World” stance according to which the characteristic nasal/oral dichotomy in parental terms originated with the first language. But Trask, like other traditional historical linguists, asserts that this is impossible because of the fact that sound change in languages is ubiquitous and continuous. To support this generally accepted contention, he first reviews evidence regarding the fate of “*gwena-,” the supposed Proto-Indo-European (PIE) word for “woman.” (The asterisk here indicates a protoform, and the hyphen indicates the option of a suffix.) He concludes that one of three things has happened to it:

1. it has disappeared completely; (2) it has changed its meaning to something quite different (as with English *queen*); or (3) it has changed its pronunciation so much that it is no longer recognizable as the same word (as with Irish *bean* or Greek *yineka* or Persian *zan*). There is, in fact, scarcely a single IE language in which the PIE word survives with a recognizable form and the same meaning. (pp. 7-8)

Trask also gives evidence that the words “boy,” “girl,” and “child” have undergone similar histories (pp. 8-9).
Thus, sound patterns of words in general seem to have an ability to cover up their phonetic tracks. But could it nevertheless be, as suggested earlier, that parental terms are different from other words in that they don’t change because, however indirectly, they are kept on a leash by the necessity that they maintain forms accessible to infants? Apparently not. Trask reviews evidence that they are simply like other words in being subject to continual change. Not only are they continually being created but they are sometimes being lost and then recreated in similar forms.

It’s worth considering some of Trask’s evidence to get a feel for the issue. In Old Japanese, for example, the word for “mother” was *papa which, with its oral consonants, was already an exception to the general trend, but its modern equivalent is haha, with no articulatory consonants at all. In the virtually extinct language Manchu, the modern forms—eme for “mother” and ama for “father”—have changed from earlier forms by participating in a recently developed generalized gender-marking system whereby female words have the vowel “e” and male words the vowel “a.”

Several further patterns come from Indo-European languages where the protoforms were *mater and *pater, which had even then acquired the kinship suffix –ter which also occurs in PIE *brater “brother” and *dhugater “daughter.” The equivalent forms in Spanish are madre and padre; in English, mother and father; in Swedish, mor and far; in French, mère and père; and in Irish they are pronounced [ma:hir] and [ahir]. In other languages, the PIE words have been lost and replaced by what Trask calls other mama/papa words—by mama and tata in Roumanian, for example, and by mam and tad in Welsh.

A further example comes from Turkic languages many of which still have the inherited words ana and ata. But in Turkish, ata is now instead an elevated term meaning “forefather” or “ancestor,” as in the surname chosen by the founder of the modern Turkish republic Kemal Ataturk, meaning “Father Turk.” But now there is an everyday mama/papa word for “father”—baba. Meanwhile, the language Uyghur has replaced ata with dada, and Turkmen has replaced it with kaka.

In many languages, inherited words that are traditional and more formal coexist with newer forms that are informal or intimate, and these new words are mama/papa words. For example, Modern Greek has mitera and mama for “mother” and pateras and babbas for “father.” And in this case babbas is definitely a new word, because in the history of Greek there was a stage where /b/ was totally replaced by /v/.

Trask provides dozens of further examples, and one must agree with his conclusion that there is an “endless re-creation and recycling of mama/papa words” (p. 15). But his evidence that instances of this form are recreated, emphasizes the deep-seated nature of this tendency which is consistent with it being present at the time when the first words are formed.

When Trask comes to give his own answer to the question of “where are these words coming from?” (p. 15), he suggests a scenario that we know to be untrue. First he asserts, correctly in my opinion, that the forms of parental terms come from babbling, and singles out the shared tendency toward reduplicated CVCV forms of babbling and parental terms as evidence for this origin. But then he asserts that the details of the process of formation of parental terms are triggered by successive members of a chronological sequence of events in modern babbling, a sequence that does not exist. Ironically, he makes his claims on the basis of a conception of ease of articulation similar
to the one I will use later in the chapter when talking about the phylogeny of speech. But his conception does not fit the facts of babbling. He asserts that the first vowel in babbling is the vowel [a] because it is the easiest vowel to produce, and the first consonants are labial because, unlike coronals, they don’t require tongue activity. In addition, among labials he asserts that nasal labials come first because they are easier in not requiring the palatal elevation that oral labials require.

As a result of this supposed sequence of events, the first babbled utterances take the form “mama” and, according to Trask, a modern mother “happily concludes that Jenny is saying ‘mother’ as well as her little speech organs will allow her” (p. 17). He goes on to say, “And, of course, one of the earliest babbling sequences to follow, usually something like papa or dada, is taken to be Jenny’s word for father” (p. 17).

But I know of no reports of such a sequence of events in the development of babbling. Instead, labial and coronal stops, nasals, glides, and vowels in the left quadrant of the vowel space seem to be present at the outset, and, as we have seen in Chapter 5, the vowel [a] and other central vowels are only favored in labial contexts, with front vowels favored in coronal contexts.

Many years ago, Jakobson (1960) suggested a naturalistic explanation for the prevalence of nasals in words for the female parent. He claimed that “the sucking activities of the child are accompanied by a slight nasal murmur, the only phonation to be produced when the lips are pressed to the mother’s breast or to the feeding bottle and the mouth is full” (p. 130). He thought that later “this phonatory reaction to nursing is reproduced as an anticipatory signal at the mere sight of food and finally as a manifestation of a desire to eat, or more generally, as an expression of discontent and impatient longing for missing food or absent nurser, and any ungranted wish” (1960, p. 130). Then, thought Jakobson, “Since the mother is, in Gregoire’s parlance la grande dispensatrice, most of the infant’s longings are addressed to her, and children, being prompted by the extant nursery words, gradually turn the nasal interjection into a parental term, and adapt its expressive make-up to their regular phonemic pattern” (pp. 130-131).

Assuming for the moment that something like this happens, how long ago might it have begun to happen? An argument that the assignment of a vocal pattern, including nasalization, to female parental terms could possibly be of considerable antiquity has recently been made by Falk as part of a more general thesis that parent-infant vocal communicative interaction may have played a key role in prelinguistic evolution (Falk, 2004).

Falk suggests that there has been an increase in selection pressures for parent-infant communication in hominids beginning as early as the australopithecine/hominid transition over 2 million years ago. The steps that set this trend in motion were two—bipedalism and brain-size increase. Bipedalism led to changes in the design of the pelvis which made it more difficult for females to deliver large offspring just at the time that brain size was increasing. The solution to this problem, apparent from present-day facts, was to deliver infants earlier in the gestation period, resulting in smaller infants that were more helpless at birth (neoteny) and took longer to mature. One consequence of natal immaturity was an inability of infants to cling to their parents, thus reversing a secure early physical link between parent and offspring ubiquitous in primates up until that time. This resulted in a tendency for females to park their babies while foraging, thus
decreasing energy use from holding and carrying them and increasing the mother’s foraging efficiency.

As a consequence of all this, a need arose for increasing use of a vocal communication channel for maintaining distal contact with babies—for remotely pacifying them and getting them to follow during foraging. Falk suggests that “these circumstances gave rise to prosodic vocalization including the universal infant-directed genre of motherese. In this context “the meanings of certain utterances (words) became conventionalized” (p. 491). Lying behind this trend, Falk suggests, was strong selection for “mothers that attended vigilantly to infants” (p. 491).

Falk suggests a scenario for a nasalized infant vocalization becoming conventionalized as a maternal parental term that is similar to Jakobson’s scenario whereby nasalized sucking noises became conventionalized. She chooses a nasalized demand vocalization, documented by Goldman (2001) in modern infants, beginning at about 2 months, and roughly translated as “mama,” though not canonical babbling. This form tends to be terminated when a parent approaches an infant or in the extreme picks it up. Falk chooses this form partly because of my argument (especially MacNeilage, 2002) based on the Frame/Content theory that the simplicity and probable antiquity of utterances such as “mama” would have been conducive to their use in early words.

A modern infant signal that seems to have properties partly overlapping the one discussed by Goldman but is better documented has been called the “grunt” by McCune and her collaborators (1996). This is a vocalization that they consider to have phylogenetic precursors. They show that it takes on a communicative function in modern infants shortly after one year of age, following two earlier stages when it is first associated with physical efforts and then with focused attention by the infants. McCune et al. conclude that, of the various non-linguistic communicative vocalizations that infants make in the first half of their second year, “only the grunt form was observed to fulfill communicative functions in the majority of children” (p. 28). They add that “These communicative grunts were directed towards the mother and might be accompanied by infant looks at mother, extension of objects, pulling at mother or reaching toward her or other objects” (p. 32).

What are the acoustic properties of these grunts? “In humans, the grunt is defined as a vocalization that results when brief glottal closure…is followed by abrupt vowel-like release (i.e., a brief egressive voiced breath) occurring with open or closed lips but no other supraglottal constriction” (p. 27). There are two possible variants of these grunts, each favoring one of the two airways above the glottis. If the lips are closed but the nasal airway remains open, the grunt will have a nasal vocal quality (“-m-”). If the lips are open, the predominant airflow will be through the mouth, and the grunt will have an oral quality (“-uh-”). Grunts occur singly or in a repetitive series, like multisyllabic babbling episodes.

Falk’s key suggestion for the association of nasality with the mother is similar to Jakobson’s, namely, that the cognitive act of pairing the nasalized vocalization with the meaning “female parent” may have been done by the infant: “After all, wouldn’t maturing prelinguistic infants, then as now, be inclined to put a name to a face that provided their initial experiences of warmth, love and reassuring melody?” (Falk, 2004, p. 503).

But if this was the origin, how would it spread, given that it would be associated by others only with the specific demand situation? My suggestion, following Jesperson
(1921/1964), is that the female parent, in the presence of the nasal demand vocalization made directly to her, decided that “This sound stands for me.” Having undergone this realization, the mother then produces from within her pre-existing repertoire the canonical babbled form /mama/ as a self-naming vocalization, perhaps first addressed to the baby, with a meaning of the form “Yes, this is your mother.” This labeling operation could then transcend the adult-infant dyad and get used by both parents as a term for the mother in a nuclear family situation, eventually becoming common currency for labeling mothers in a group.

The nasal form for female parent can be considered iconic in the somewhat specialized sense of its being a sound that is consistently associated not with the emitter but with the recipient—the mother—who I assume to be the only direct infant provider at the time this began to occur. But what could be iconic about an oral consonant for a male parent? Again following Jesperson (1933; see also Jakobson, 1960), I see operating here a step beyond iconicity—an instance, perhaps the first, of active use of the fundamental linguistic concept of contrast. Nasal and oral sounds are produced by the two possible different configurations of the upper airways. (See Fig. 7.1.) In nasals (7.1a), the nasal airway remains open, and for maximum contrastive effect the oral airway must be closed (7.1b). In oral sounds, in which the mouth is open, for maximal contrastive effect the nasal airway must be closed. It may even have been that for /mama/ both the nasal and the oral airways were originally open, giving a somewhat nasal quality, but after the oral contrast for /papa/ developed, the oral airway was closed for /mama/, making the nasality stronger and therefore more contrastive with /papa/.

What happened to the evolution of words after that? The coinage of the two parental terms must have originally occurred in a single family group. But after that the usage must have spread into the culture in general. However, two words aren’t enough to be called a language. One possibility sometimes mentioned (e.g., Thong, 1999; Trask, 2005) is the generalization of words with frame-stage properties and the nasal phonetic quality of the maternal term to other aspects of the mother. A relatively modern example close to home that supports Jakobson’s idea that /mama/ arose in a feeding context is “mammal.” Parts of Merriam-Webster’s (11th ed.) definition of the term are of interest: “mam-[insert schwa here] ... n [NL Mammalia, fr. LL, neut. pl. of mammalis of the breast, fr. L mamma breast] (1826): any of a class (Mammalia) of warm-blooded higher vertebrates that nourish their young with milk secreted by mammary glands....”

But what seems necessary at some point is a flash of what McShane (1979) called, in a developmental context, “nominal insight”—the realization that things can have names. That such a thing can indeed happen is indicated by an incident in the remarkable life of Helen Keller, the deaf blind genius (2003). She suddenly realized, when a tactile signal accompanied the pouring of water on her arm, that the signal was a label for the substance, and that this was an instance of the general case that tactile signals were labels for concepts. Once the analogous insight occurred in a single earlier hominid, then the possibility of conscious assignments of signals to concepts could have arisen, and this presumably involved developments in the frame-stage system driven by the necessity of phonetic contrast.

Let’s return to the possible initial two-word sequence of events of word invention, and consider it further in the context of what was said about the properties of the frame stage in Chapter 5. From the beginning of babbling, infants produce both nasal
and oral consonants. From a recapitulationist point of view both forms may have been initially available to hominids. Remember also that McCune et al. consider that there were phylogenetic precursors to both nasal and oral grunts. The nasal and oral forms differ in their production only in terms of whether the soft palate is left in its rest position or elevated. Now let’s go back to the model of frame-stage production in Chapter 5 in order to consider what was necessary to produce the various consonants favored in the frame stage. In that model, the frame was the constant base for all vocalizations. It was noted there, as Trask also noted, that nasalized frames require no movements in addition to the frame. The soft palate simply stays in its breathing configuration. It’s possible, then, that nasalized frames were the most available frame types in ancestral hominids, and this makes it more likely that they would have been involved in the first word.

But a complication appears to arise here. From the point of view of ease alone, nasal consonants should be the dominant consonant types in modern languages, but they are not. However, it is well known in phonetics that nasals don’t contrast very effectively with each other, either in consonants or in vowels (See Hura, Diehl & Lindblom [1992] regarding consonants, and Wright [1986] regarding vowels.) So there was presumably a trend in the evolution of speech toward more oral than nasal sounds in the progression toward modern languages.

Returning then to our model, the simplest form, the nasal form, might have been used for the female parent. And note that the nasalization would be present for the entire utterance. The most straightforward way of producing a perceptually different utterance for a paternal form would be to close off the soft palate to produce a word entirely consisting of oral sounds. Thus the initial contrast would have been a gross and therefore unmistakable one consisting of a difference that stretches across the whole word.

An additional consideration from the point of view of the model is that both the nasal and the oral sounds would have labial consonants because these could have been produced by the frame alone, without the additional tongue movements necessary for coronal or dorsal consonants. These considerations suggest that an oral-nasal contrast between labials (basically [mama] versus [papa]) would have been the most likely initial contrast from the articulatory point of view.

The choice of place of articulation of consonants in Ferguson’s sample of 6 languages is consistent with the expectation that most parental terms in baby talk would be labial consonants (see Table 7.4). There are 8 words with labials and 4 with coronals. I pointed out in Chapter 6 that coronals tend to exceed labials in babbling but that this trend is reversed in first words. However, in either case the difference is relatively small. Because of the small size of the baby-talk sample, the excess of labials is unconvincing.

But what are the relative proportions of labials and coronals in parental terms? Murdock found 430 labials and 462 coronals, while Paine recorded 69 labials and 63 coronals. The roughly 1:1 proportion of labials to coronals is much higher than would typically be found in languages in general. For example, in an analysis of the phonetic properties of conversational English, Mines, Hanson, and Shoup (1978) found a 2.5:1 ratio of coronals to labials. And in an unpublished survey of dictionary counts of Dutch, French, Romanian, and Turkish, Kern found that the coronal/labial ratio ranged from 2.5:1 (French) to 3.5:1 for Romanian. In conclusion, there are proportionately morelabials both in baby talk and in the parental terms of languages than there are in modern languages in
general, and this is consistent with the prediction regarding the forms of first words arising from the model of the frame stage.

Beyond Falk’s argument for the likelihood that terms such as “mother” might have been important in the evolution of the word for ecological reasons, what else can be said about the possibility that parental terms might have been among the earliest words? It’s interesting to note that while Trask talked about the “endless recycling of mama/papa words,” he never considered the implication that the process might stretch a long way back in time and could therefore have given rise to the first words. It would seem that the sheer robustness and productivity of the trend can be taken as evidence for the antiquity of its origin. Thus, although Trask is apparently correct that present-day mama/papa words are not themselves living fossils, they are presumably replicas of them, because the conditions that mandate their generation remain similar to what they always were.

If there has long been this strong general propensity for the mother to regard the infant’s nasal demand vocalization as a label for her, and also a tendency to then choose a simple distinctive alternative term for the male parent, then this event could, in principle, have occurred in more than one location. If it first happened in Africa before the last human migration out of Africa occurred, perhaps some 60,000 years ago, language could have been monogenetic in origin and then spread, notwithstanding the subsequent coming and going of particular forms. If it happened after that, a monogenetic basis for the spread would not have been possible because language in general, and this regularity in parental terms in particular, could not have spread by contact from a single source to all other parts of the world. Nevertheless, it’s conceivable that it happened more than once, either before or after the migratory Rubicon was crossed and each time served to ignite the language process. The latter would be consistent with the orthodox historical linguistic view of independence of different language families. Unfortunately, we don’t know when language began. My contention is simply that however many times the invention of the two parental terms happened in an infant-parent matrix, the first time it happened may have marked the first step toward the invention of words with a systematic phonology.

7.4 Other possibilities for first words

There have been a number of other candidates for first words. One well-known possibility is onomatopoetic forms. Merriam-Webster’s 11th defines onomatopoeia as “the naming of a thing or action by a vocal imitation of the sound associated with it.” Some English examples are “cockatoo,” “curlew,” “swish”, “bubble,” buzz,” and “trickle.” These have the advantage that they are iconic—they contain some semblance of the noise that the entity or action being symbolized makes. It is therefore easy to see how the sound/meaning link was made by capitalizing on properties of the physical world. However this mode of sound-meaning pairing is primarily restricted to concepts related to our experience of the physical world, in particular to entities that make noises, or events that provoke them. Furthermore, from the phonetic point of view they tend to involve complex sounds and sound patterns, patterns that might not have been as easy to stabilize in a population as more simple sound patterns of the frame stage. For example, they often involve fricatives and consonant clusters. And they do not, in themselves, demand a systematic organization based on CV syllable forms.

One additional property of onomatopoetic forms that does not sit well with the possibility that they were the first words is that forms for the same acoustic event are
different from language to language. Jackendoff (2002) points out that “After all, dogs go bow-wow in English but gnaf, gnaf in French, and roosters go cockadoodledoo in English but kikiriki in German” (p. 251).

There is a particular subclass of onomatopoeic forms that conforms to a “Size Principle” (Ohala, 1994). Terms for small objects tend to involve a vowel with a higher spectral center of gravity than terms for large objects. Consider the /i/ at the end of “tiny” and other “diminutive forms” that end in /i/ (e.g., “baby”) versus the /u/ in “huge.” Try to say each of these vowels with the same energy level and you will find that the /u/ has a lower perceived pitch.

These forms can be considered to have a deep history in communication. For simple mechanical reasons, sound frequencies vary inversely with size in animal calls and in noises that animals create when moving about the environment, and there is evidence that other animals can use these sounds as information about the size of the animal producing them—information that can be important to survival. In addition, like the parental terms, these forms lend themselves to creating a phonetic contrast, but only one, and one that is isomorphic with one basic contrast at the semantic level—the size contrast. And as they don’t involve consonants, they don’t contribute to the formation of a phonetic system centered on CV syllables. Another question that can be asked is, How important was it to vocally communicate information about the size dimension to others, relative to other semantic properties, at the time that hominids began to invent a symbolic system?

Jackendoff (2002) identifies a third class of forms that have some claim to a relatively early origin. He calls them “‘defective’” lexical items, forms that have semantics and phonology but no syntax. He lists seven subclasses of these forms, “sorted approximately by semantic class” (p. 131), and I give 6 of them (leaving out onomatopoeia) in Table 7.5 below with some of his examples, and with some added information about semantics.

Jackendoff “would like to think of such words as these as ‘fossils’ of the one-word stage of language evolution—single-word utterances that for some reason are not integrated into the larger combinatorial system” (p. 240). He notes that “Their semantic and pragmatic diversity suggests that they are island remnants of a larger system, superceded by true grammar” (p. 240). He notes too that “I am not suggesting that the actual ‘defective’ lexical items of English are historical holdovers from this stage of evolution. Rather, what is a holdover is the possibility for a language to contain such ‘defective items’; those of English are realizations of this possibility” (p. 240).

Note some properties of these various forms. For the forms in group #1, it seems more likely that we first evolved head movements, such as the head nod for “yes” and the head shake for “no”. Greeting forms in group #2 may not have been important enough to be the first words, and again, as in the assent/dissent dichotomy, gestures such as eyebrow-raising and waving may have been early available alternatives. The affective linkages of items in group #3 may make them bad candidates for generalization into true language, which is typically defined in terms of the lack of an obligatory link between symbols and affective states. Some of these at least may be members of a second affective linguistic system, perhaps with right- hemisphere involvement, that sometimes remains present in aphasics who cannot produce lexical items of language proper (see Chapter 9). Of the forms under group #4, a form such as “hey” might have had an early nonverbal
alternative that served as an early substitute for a verbal form. For example, chimpanzees are known to crinkle up leaves to get attention (Tomasello and Camaioni, 1997). The forms under group #5 are not actual words, and like many onomatopoeic words have complex phonetic properties that probably ruled out their early occurrence as part of a systematic phonology. The forms listed under group #6 can probably be ruled out as early forms because of their semantic complexity.

Considering these forms in general, one notable fact is that with the exception of onomatopoeic forms, they don’t function as nouns. And it has been argued that the first words were predominantly nouns (Calvin and Bickerton, 2000). Another fact is that unlike the use of nasal and nonnasal consonants in parental terms, none of these classes of term has consistent specific phonetic properties across languages. Notwithstanding my caveats about these vocal forms as first words, I conclude that some of them could conceivably have been around at the time that the first systematic sound contrast in words was made. They could even have contributed to the eventual and presumably necessary conscious insight that concepts could be vocally symbolized. But my argument is that they were less important than the earliest parental terms because they were not as conducive to the formation of the systematic structure of the sound level as parental terms were. As I consider that the initial form of that system, the frame stage, had to have been firmly in place in a vocal-grooming mode before anything that could be called a word evolved, I consider the initial engagement of that system to provide the underpinnings of phonological protoforms of words to be the key issue in the advent of words proper. Baby-talk parental terms qualify as candidates for that initial step. To my knowledge, there is no other systematic rationale for the sound patterns of the first words.

7.5 Summary

Having theorized how the sound system of modern speech evolved, I now sought in this chapter to explain how parts of this system might have become paired with concepts to form words. The modern phenomenon of baby talk seems relevant. It involves the pairing of frame-stage patterns with concepts. The simplicity of the frame-stage patterns that accompany particular concepts might mean that modern patterns and their links with concepts closely reflect ancient ones. From an ecological perspective, the increasing importance to survival of parent-infant communication-at-a-distance, and the intimacy, stereotypy, and attention-getting properties associated with the communicative dyad were, together, presumably conducive to the invention of sound-meaning linkages. Finally, parental terms in baby talk and languages proper are not only being continually created, which suggests a robust and therefore longstanding history, but also have one particular phonetic property redolent of systematic phonology—a tendency to contrast nasalized (female) forms with oral (male) forms. Perhaps the nasalized female form was initially derived from infant nasalized demand vocalizations and the oral male form, not being iconic in a similar way, was formed specifically for contrast, perhaps the first systematic contrast. Although there may have been some other symbolic uses of the sound patterns at the time words first evolved, the advent of parental terms in the parent-infant matrix seems most likely to have led into the origin of the phonological system of present day language.
7.6 Coda

It’s a telling point regarding the intellectual orientation of generative linguists that they almost totally ignore the question of how actual events in hominid history created linkages between the semantic and the phonetic levels in the first place (though, for a recent exception see Jackendoff, 2002). One could say this is not surprising because the origin of words is, par excellence, a performance-level issue and therefore not within the generativists’ domain of competence. Furthermore, one could hardly take the usual generative stance on causality and say that an innate tendency to pair sounds and meanings preceded the actual pairing event. The pairings must have been culturally induced.

But if indeed this event was the most important one in language evolution and therefore hominid evolution, as many people believe, it leaves generative linguistics looking a little intellectually impoverished. My attempt to explain this event, the only one specifically attempting to incorporate the origin of systematic phonology, could certainly be regarded as flimsy. But it will serve some purpose if it suggests to some that we should not really be comfortable with our understanding of ourselves until we have understood this key event in our becoming.

Table 7.1 The relation between observed and expected frequencies for consonant-vowel relationships in CVCV forms from Ferguson’s corpus of baby talk terms

<table>
<thead>
<tr>
<th></th>
<th>Coronal</th>
<th>Labial</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonant-vowel co-occurrence patterns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front</td>
<td>1.33</td>
<td>0.96</td>
<td>0.28</td>
</tr>
<tr>
<td>Central</td>
<td>0.82</td>
<td>1.33</td>
<td>0.69</td>
</tr>
<tr>
<td>Back</td>
<td>0.74</td>
<td>0.76</td>
<td>2.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Coronal</th>
<th>Labial</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowel-consonant co-occurrence patterns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.2 Observed-to-expected ratios of CV sequences in Murdock’s 1959 study of
parental terms of 474 languages

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>1.43</td>
<td>0.82</td>
<td>-</td>
</tr>
<tr>
<td>Central</td>
<td>0.75</td>
<td>1.34</td>
<td>1.00</td>
</tr>
<tr>
<td>Back</td>
<td>0.61</td>
<td>0.79</td>
<td>2.33</td>
</tr>
</tbody>
</table>

Table 7.3 Observed to expected co-occurrence patterns in CV sequences and VC
sequences in Paine’s sample of parental terms in 211 languages.

a. Consonant-vowel co-occurrence patterns in the first CV sequence

<table>
<thead>
<tr>
<th></th>
<th>Coronal</th>
<th>Labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>1.56</td>
<td>0.31</td>
</tr>
<tr>
<td>Central</td>
<td>0.64</td>
<td>1.44</td>
</tr>
</tbody>
</table>
b. Consonant-vowel co-occurrence patterns in the second CV sequence

<table>
<thead>
<tr>
<th></th>
<th>Coronal</th>
<th>Labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>1.55</td>
<td>0.27</td>
</tr>
<tr>
<td>Central</td>
<td>0.90</td>
<td>1.14</td>
</tr>
</tbody>
</table>

c. Vowel-consonant co-occurrence patterns

<table>
<thead>
<tr>
<th></th>
<th>Coronal</th>
<th>Labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>1.55</td>
<td>0.40</td>
</tr>
<tr>
<td>Central</td>
<td>0.85</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Table 7.4 Examples of “defective” word forms given by Jackendoff, 2002 (pp. 131-2) with semantic notes added

1. yes, no  (assent/dissent)

   hello,  (greeting

2. goodbye  (forms)  thanks  (gratitude)

3. ouch  (pain cry)  oops  (mishap label)  dammit  (swear word)
(attention getter)  oo-la-la (admiration marker)

(quieting signal)  psst (attention-getting signal)  tsk-tsk (disapproval signal)

hocus

6. abracadabra (incantation)  pocus (deceptive content)

Figure 7.1. Schematic views of the configuration of the soft palate for nasal sounds (a) and oral sounds (b) (Adapted from Ladefoged, 1993, Figs 1.5 and 1.6)