CHAPTER 6 Three Easier Comparisons

The large number of parts in the cellular system and the language systems present us with a difficult expository task. One approach is to treat each system in turn, and then present a point by point comparison. While this leads to a more unified discussion of each system, it also results in long gaps between the treatment of corresponding features. Alternatively, each subtopic of the analogical comparison can be used as a pretext to discuss evidence for that feature from the two systems. Such an interleaved approach reduces the memory load, especially for molecular or linguistic features that may be unfamiliar to the reader. The second strategy was followed because it has the added advantage of keeping the focus more firmly on the analogy, rather than on features of the individual systems--interesting as they might be.

There is another tension, introduced, in our discussion of "missing/extra" features versus "hidden" features that arises from the conflicting demands of *demonstrating* the analogy and *using* the analogy. To demonstrate the analogy, within-domain research is used to draw out both similarities and differences. This involves a certain amount of negotiation as the mapping between the two system is developed--e.g., imagining what one system would look like if it actually *was* like the other. Using the analogy to make predictions, by contrast, requires a bolder posture where the source system and a fixed mapping are used to make explicit predictions about hidden phenomena in the target system. These predictions may at times conflict with received views in the target domain. Mapping and prediction therefore need to be distinguished, but they will be interleaved as well--and for the same reason: to avoid losing sight of the structure of the analogical map.

Starting with the next chapter, we will take up each feature of the diagram presented previously (Figure 12) and consider the evidence for this structure or relation in each system. The overall order--progressing from symbols to symbol-representations to thing-representations--is shown in Figure 16. As we get further into the brain, the discussion will turn increasingly from mapping to prediction. Some relational features (e.g., that symbol chains are more flexible than symbol-representation chains) only make sense after a within-system comparison. This unavoidably requires a certain amount of prefiguring as we move along the linear path. Before beginning this long march, however, the present chapter picks out three particularly obvious similarities or exposed comparisons mentioned in the introduction--the invention of strong polymer chains, backbone and sidechain structure of symbol segments, and word recognition to help motivate the journey ahead.

Invention of strong ordered polymer chains

Rocks and minerals--one of the major components of the prebiotic world--certainly seem strong. When they are kept relatively dry, the bonds between the atoms in a mineral are in fact very strong. However, once a mineral is dissolved in water, it paradoxically becomes much less

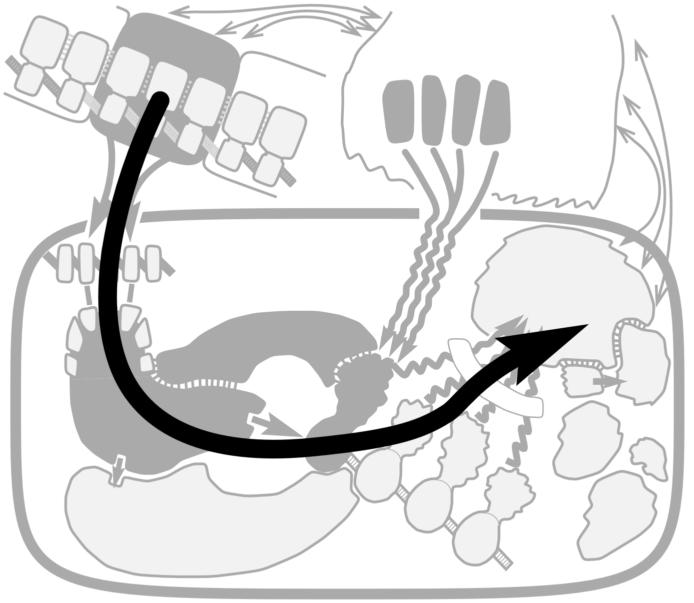
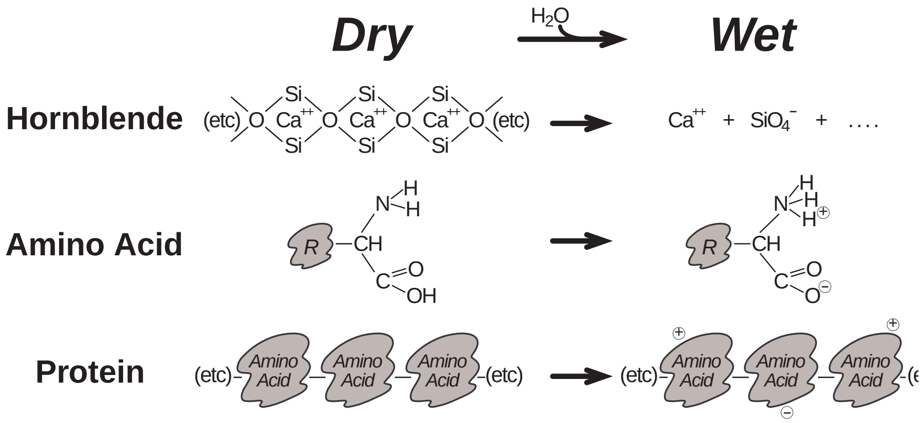


FIGURE 16: Overall order of the argument (see Fig. 12)

’strong’ than a biological molecule. At the molecular level, the biotic world is easily distinguished from the prebiotic world because of the unique ability of biological organisms to create and manipulate strong covalent bonds capable of holding together long ordered chains of units in an aqueous solution.

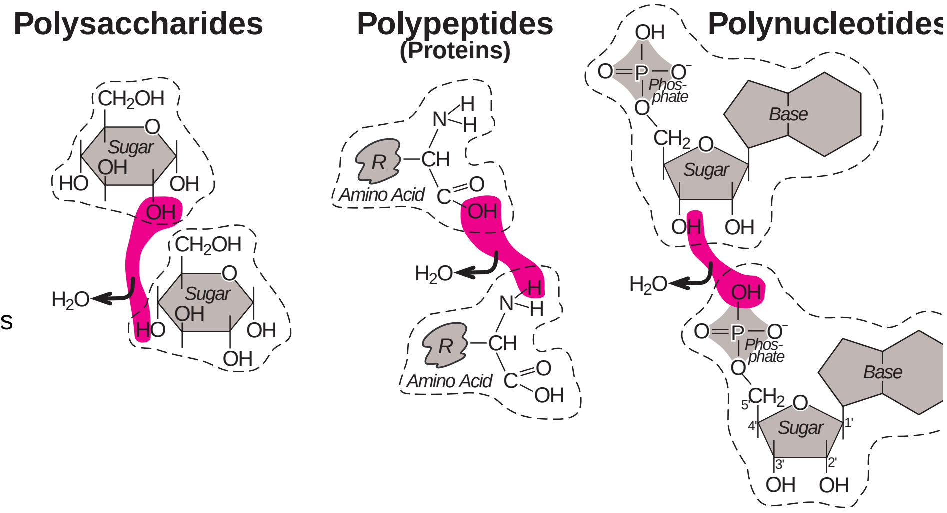
The enormous covalently bonded DNA and RNA molecules that characterize even the simplest organisms are utterly different than anything found in prebiotic solutions. Polypeptide (protein) and polysaccharide (e.g., starch, chitin) chains are equally as distinct as DNA and RNA. The isolatable DNA molecule in each *E. coli* is nearly 4 million nucleotide bases long. There are, of course, isolatable prebiotic molecules, too--amino acids, for instance (see Chapter 9). And there are, amongst the large variety of substances in the prebiological natural world, a variety of rocks and minerals consisting of regular, bonded networks of atoms (crystals), as well as many complex and less regular colloidal, gel-like, or cryptocrystalline substances (for a summary, see Cairns-Smith, 1982). The key point is that when the bonded network of a mineral--for example, the layered silicate strands of an igneous mineral like hornblende, the layered silicate sheets of a sedimentary clay mineral like kaolinite, or the 3-D framework structure typical of a metal oxide--is exposed to water, it typically breaks up into small individually-solvated pieces, each containing only a small number of atoms. Biotic molecules, by contrast, maintain their macromolecular structure in solution; the main effect of dissolving biotic polymers in water is to cause some parts of the molecule to become more negatively or positively charged (Fig. 17). Nonbiotic water-soluble macromolecular gels held together by moderately strong bonds exist (e.g., Al^{3+}gel/solutions), but are hard to study because, in contrast to biotic polymers, their covalent structures in solution are labile, forming and disassembling in response to slight changes in ionic concentrations and pH. Because prebiotic minerals do not exist as long, determinate chains in solution, their molecular structure is much less suitable for constructing ordered code chains that could serve as templates for constructing other specifically ordered, self-folding



RE **17**: Many minerals
ot maintain chains of
lent bonds in solution

structural chains. It is certainly possible that proto-living systems may have initially been built upon a mineral scaffold that was then stripped away when modern organic life achieved its present form (Cairns-Smith, 1982). At the current time, however, there appears to be a sharp divide between the prebiotic and the biotic--life essentially invented soluble but stable, long-chain molecules.

Our definition of ’stable’ needs to be deepened slightly. The crystals of a mineral often grow in sedimentary environments under nearly the same conditions that result in their degradation--dilute solutions of ions and small charged molecules. Crystallization of many minerals requires the exclusion of water--e.g., some of the water of hydration that coats a bare metal ion--but in general, this water is weakly bound and rather easy to remove. Long biological polymers, by contrast, don’t form ’naturally’ in aqueous solutions because forming the strong covalent bonds between each unit requires the removal of strongly bound water-forming groups (see Fig. 18). This dehydration process requires energy, which in a cell is typically provided by nucleoside triphosphate hydrolysis (i.e., breaking even higher energy bonds holding the second and third phosphate groups onto a nucleotide like ATP). This implies that the reverse reaction--water attacking a biotic polymer and breaking it into monomers (hydrolysis as opposed to

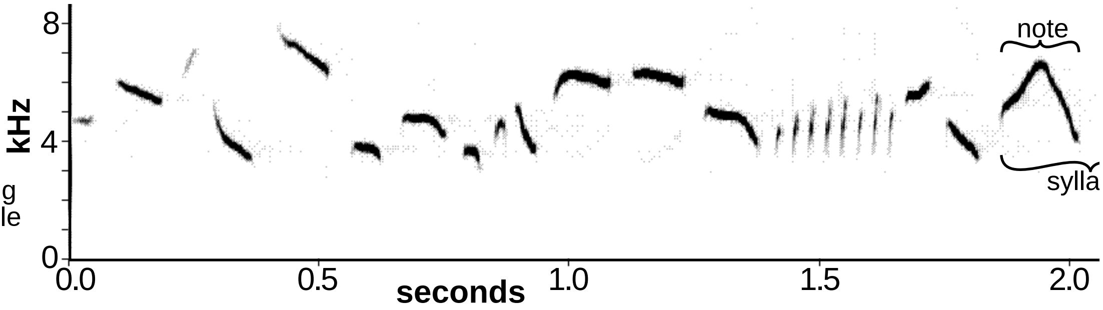


IGURE **18**: Biotic polymers
re made by dehydration

dehydration)--is actually energetically favored in aqueous biotic polymer solutions and will eventually occur given enough time (see next chapter). Nevertheless, the important point is that once the strong covalent bonds between biotic monomers have been made, they can persist in solution long enough to be used by cells for information storage and for controlling reactions--in sharp contrast to the initially strong bonds that hold dry minerals together, but which fail immediately in water (Fig. 17). It should be obvious that one main conundrum in the origin of life is figuring out how the hard covalent bonds between the monomers in long biotic chains came to be routinely generated by *dehydration* in an *aqueous* environment.

In turning to prelinguistic animals compared to humans, there is similar evidence for the invention of long ’polymer’ chains--speech sound strings and the means to produce and interpret them--with determinate, stable ordering. As a consequence of the ability to produce code as well as perceive it, the external speech sounds themselves are transient by comparison with stable DNA (and protein) strings; truly DNA-like ’speech sounds’ would be permanent, encyclopedia-length internal code-like activity patterns constituting an enormous verbatim memory that could be selectively ’listened to’. Despite this difference, there is a prelinguistic/linguistic contrast very similar to the prebiotic/biotic contrast described above--there is no sign of the immense ordered linguistic symbol chains in any other animal. There are no small languages. An otherwise unremarkable half hour conversation can consist of tens of thousands of speech symbol segments uttered and perceived in specific orders (the conversations I am hearing now over a main street in a small Italian town probably run closer to a a couple of hundred thousand segments per hour). Yet a single out-of-place word or even a single segment mispronunciation or misperception can easily be a cause for explanation, joking, or embarrassment.

Rather than bringing in syntactic theories just now, let’s instead try to examine language from an animal’s point of view--the way we viewed biotic polymers from a prebiotic point of view. Syntactic theories tend to emphasize the recursively characterizable structure of language strings; what we are after here is merely their brute lengths. There are few other examples of serial vocal behavior in animals that even remotely approach human language in scale. The closest things are the vocalizations of songbirds and whales; the vocal learning abilities of non-human primates including apes are, by comparison, quite unexceptional. Songbirds utter a series of one or more "songs" (sometimes more than 500), each of which contains a series of "syllables", where each syllable contains one or a few "notes" (Konishi, 1985; Doupe and Kuhl, 1999) (see Fig. 19). A few kinds of whales, particularly, the humpback whale, vocalize in a remarkably similar fashion, except that the pitch is lower, and the pace slower--the high notes in a whale song are several octaves below a birdsong, and a typical song takes about 10 minutes instead of several seconds (Tyack, 1993).



: **19**: Birdsong
gram (a single

Putting aside birds and whales for a moment, someone might object that there are many animal behaviors characterized by reasonably specific sequences--nest building by birds, for example. But this, and many other examples, don’t come close to the lengthy, exact sequences emitted and understood by a fluent speaker of a human language. In building a nest, it is not necessary to execute each movement in a determinate sequence; gathering and adding a little string can come before or after adding a piece of dried grass. Similarly, primate social interactions often involve sequences of actions--if Bill the baboon grooms me, then I may groom him back; but the short and long range order is tremendously looser than with language. Language is bizarrely fussy about the order of things; even in languages that have somewhat freer word order because of extensive inflectional marking capable of indicating doer, done-to, and so on, the order virtually always means something anyway; for example, it is used to establish focus or topic (Comrie, 1989). It is true that the exact order of sentences in a conversation can be varied a lot--e.g., while explaining what happened yesterday. But if something is left out, specific structures are involved in re-inserting it that don't appear in something like nest building, and that cannot be easily inferred from the sensory situation (as would be the case for something like a hole remaining in one side of the nest).

Many other things that people currently do--especially in groups--involve long strings of actions. Millions of actions must be executed in specific order to build an office building (clear the site, dig the foundation, pour concrete, put up and rivet together the frame, install heating and wiring, finish walls and doors, and so on). And many of these things depend on language. The striking thing, however, is that language has hardly changed compared to things like building buildings. Back when humans were on "the camping trip that never ends" (Pinker, 1997)--which covers most of the history of modern humans, and whose nature we have mainly inferred from fragments of Stone Age cultures that persisted until the first half of this century--they still surely used speech symbol sequences much longer than anything seen in the animal world.

As with biotic polymers, the main conundrum with human linguistic ’polymers’ is how they came to be routinely generated, understood, and reproduced, starting from a prelinguistic world lacking such long sequences. Like the prebiotic world, the prelinguistic world has a powerful tendency to select against long determinate sequences in animal behavior. Though we cannot phrase this tendency in our current theories of behavioral ’chemistry’ and brain-activity-pattern ’chemistry’ as precisely as we can in molecular chemistry, an inspection of a wide range of animal behaviors

(Hauser, 1996) suggests that despite their amazing versatility, long, determinate language-like chains are not involved.

Probably the most remarkable exception to the lack of the use of long meaningful precisely ordered chains in animals is the comprehension of basic spoken English sentences by the bonobo, Kanzi (Savage-Rumbaugh et al., 1993), which we will take up in detail later. However, Kanzi's abilities, along with similar, though not quite as prodigious feats accomplished by parrots, dolphins, and other animals, do not occur naturally without a a lot of human help, as far as we can currently tell. Songbird and whale songs, by contrast, most definitely are naturally occurring, culturally transmitted strings. Although it appears that these 'songs' (quite unlike human song, as we shall see) have no order-dependent *meanings* like human language, they nevertheless will provide us with key evidence for how one important component of human language might have gotten a prelinguistic start.

Backbone and sidechain

In this second part of our survey, we turn to consider a major parallel characteristic of the individual units or *segments* of the symbol strings in the two system (bold chain in Fig. 20)--their *constant backbone* and *variable sidechain* structure. The terminology comes from structural chemistry, where it was originally used to describe a number of different humanly synthesized chain molecules--e.g., chains based on modified 2-carbon ethylene units. When these units are polymerized, a uniform single-bonded carbon chain backbone results, with various possible sidechains on every other carbon (C-C-). The polymers with uniform sidechains are most familiar--e.g., no sidechain (polyethylene, first made in the 1860's), phenyl sidechain (polystyrene), cyano sidechain (Orlon), two-piece methyl and methyl ester sidechain (Plexiglas).

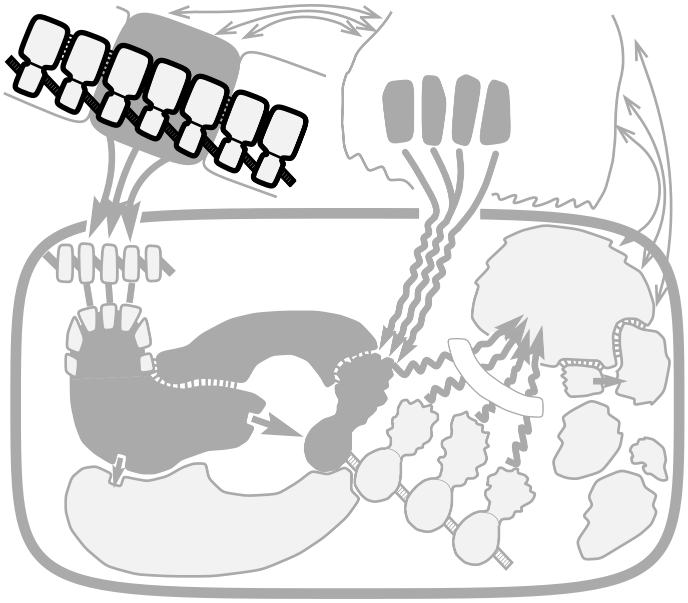


FIGURE 20: Backbone and sidechain structure of symbol segments in the symbol chain

Other nitrogen-containing backbones are also possible (e.g., Nylon, Polyurethane), which have some similarities to the N-C-C- repeating backbone unit of proteins. All of these examples have constant sidechains. Copolymers, that is molecules with variable sidechains, are also possible but not as commercially important, in part because it is difficult to generate known sequences. Turning to the linguistic level, there are a large variety of other segmented symbol strings, especially in computers. In none of these other cases, however, is a string of equally strongly-bonded-together symbol segments recognized in small groups, and then each group used to code for an arbitrarily related unit of a different kind in a parallel chain that folds--as is the case with DNA and speech streams.

When comparing molecules like DNA to the seemingly variable and slippery sounds of real spoken language (as seen by acoustic phoneticians and speech recognition software, not fluent understanders!), there is often an initial tendency to underestimate the abstractness of molecular structures--especially as drawn in an organic chemistry textbook. The schematic stick-and-letter structures (e.g., Fig. 18) are abstract since they leave out the complex 3-D electronic surface of the molecule--what is actually ’visible’ to other molecules it interacts with. But even a static 3-D representation is an abstraction--a long-time average of the real-life vibrating, twisting, stretching, and translating and rotating molecule. The recognition apparatus in the cell doesn’t have direct access to the chemist’s abstract structure (neither does the chemist, for that matter); the cell must guess at the abstract ’phonemic structure’ of each molecule that it deals with on the basis of the molecule’s surface shape, charge, flexibility, and so on during a collision. In real cells, there is also is a tremendous amount of noise in the form of other chemicals colliding with everything. Finally, it is likely that individual recognition events vary as to which part of the target molecule is most important, conditioned by the exact path and speed by which the incoming molecule ended up colliding with the recognizer. From this perspective, molecular recognition looks messier and more difficult, like speech stream recognition--where a listener is able to accept a family of different but related sounds as the same, and even use different cues on different occasions, all in the presence of noise. We now turn to examine a few of the features of the two different chains of symbol *segments*, putting aside for now the difficult problem of just how the cell and the auditory cortex manage to recognize the *word-sized* groupings in them.

Phosphate-sugar backbone

The well known subunits of the DNA (deoxyribonucleic acid) symbol chain are nucleotides, each of which consists of the sugar, 2’-deoxyribose, with a phosphate group esterified (i.e., attached by means of an oxygen) at the 5’ sugar position and one of four aromatic bases attached at the 1’ sugar carbon (see next section for the bases). The segments are connected into phosphate diesters (i.e., there is a second oxygen in the chain) at the 3’ sugar carbon (see Fig. 18--the sugar ring in the lower nucleotide is numbered). The "acid" in nucleic acid comes from the negatively charged non-chain oxygens on each phosphate group. That nucleotides were the proper subunits or segments of DNA had been clear before Watson and Crick (1953) discovered the stereochemical nature of DNA. It was known, in fact, even before DNA was thought to be the information carrying component of the genetic substance; the polymeric structure of nucleic acids had earlier been taken as evidence that nucleic acids served as an inert scaffold with a simple repeating sequence (the "tetra-nucleotide" hypothesis), upon which hypothetical information-carrying proteinaceous elements were arrayed (Olby, 1974).

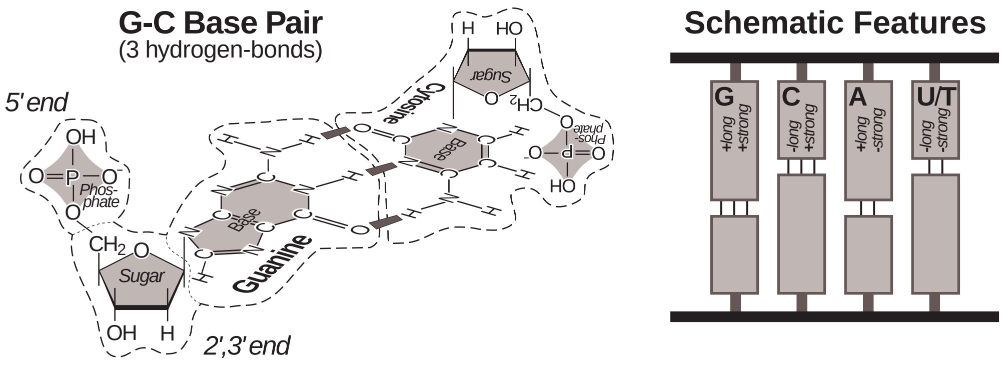
An identical phosphate-sugar unit appears in each segment of a DNA chain. It is not only constant in composition, but also adopts an approximately constant shape. The phosphate-sugar backbone is the chain of strong covalent bonds that holds the DNA chain together. The sidechains are called that because they stick out from the backbone and don’t make any further covalent bonds with other parts of the chain. A break or "nick" in a single stranded DNA molecule immediately disconnects the two parts of the chain; if the DNA is double-stranded, however, the weak but numerous base-pair bonds made with the opposite, un-nicked strand can hold the nicked one together. Not surprisingly, cells are constantly repairing nicks in DNA; but they also routinely nick (and immediately repair) it--e.g., to untwist and untangle. The backbone, however, does not carry any of the explicit sequence information used to control the order of amino acids in proteins.

The constant backbone part of a double-stranded DNA molecule forms the exterior of the double-stranded DNA molecule (in contrast to proteins, where the sidechains stick out); the variable sidechains that contain the sequence information are normally hidden in the center of the double helix. This sequence information is normally read off to make proteins by separating the strands so that the ’distinctive features’ at the ends of the internal sidechains are exposed. There is, however, another fundamentally non-code-like way by which the cell can access DNA sequences. Cells have a lot of machinery to decide which part of the DNA to read in a code-like way. As a bacterial example, if the cell comes across a particular sugar, the cell will rapidly begin to make an enzyme to metabolize the sugar; to do this, it has to find the sugar-break-down gene and immediately copy it into the sugar-break-down protein (enzyme). Because of language production, there is no exact analog to this in humans; it would correspond to a kind of internal tape librarian mechanism that decided which of a giant internal verbatim store of useful speech streams should be played back into the mind at each moment of the day. Humans can certainly retain large verbatim stores--e.g., actors, comics; but even they get most of their language input from listening and reading.

While there is no exact analogue of gene regulation in language, there most definitely is non-code-like access to speech streams; and the contrast between code-like and non-code-like accessing of symbol strings is instructive. In contrast to the reading of the DNA code to turn it into a parallel amino acid sequence, the methods of recognizing sequences in double-stranded DNA for gene control do not involve digital effects; rather, the strength with which a particular DNA recognition protein binds to different sequences varies continuously. Rather than recognizing a few clear binary features of the bases (see next section) double-stranded DNA recognition proteins are quite various and pick up subtle variations in the overall surface shape, charge distribution, and flexibility of different sequences. For example, a recently obtained structure of a eukaryotic transcription factor that binds best to a 9-base sequence showed that it only directly contacted the sides of 2 bases in the major groove (Becker et al., 1998). The task is difficult because the signatures of the different base pairs in the grooves of the DNA double helix are quite subtle. There are some direct but small effects of the variable sidechain sequence (base sequence) on the structure of the backbone itself, however. One effect arises because each base pair, though almost flat, has a tendency to twist a little (always in the same direction), like a propeller, but a propeller with unequal length blades (long purines, short pyrimidines--see below). Thus, if the base pair ’propellers’ are stacked with the long blades all on one side of the helix, there is no clash. But with an alternating long-short arrangement, there is maximum interference between the inner parts of the long ’blades’. The result is that the 6 backbone bond torsion angles in each nucleotide in such a sequence give a small amount to allow the longer blades to stick out of the stack a slight bit, and to slightly reduce the twist of the stack. The resulting local adjustments are predictable (Calladine, 1982; Dickerson, 1983), and may be one aid to non-code-like recognition of DNA sequences by DNA-binding proteins.

Variable aromatic base sidechains

Code-like access of nucleic acids, by contrast, is much easier to understand. The sidechains of DNA are the four aromatic bases--adenine (A), cytosine (C), guanine (G), and thymine (T) (uracil (U) in RNA). The genetic code is based on a very simple system of 4 (by themselves) meaningless segments, each of which can described as having 2 distinctive or binary features. The first feature is the distance the base sticks inward from the external backbone (these differences are much larger than the minor bulges in the backbone described above). There are the long (two-ring) purines, A and G, and the short (one-ring) pyrimidines, C and T. The second feature is the number of possible hydrogen bonds--there are 3 bonds possible with C and G and 2 possible with A and T; for this reason CG or GC base pairs are stronger than AT or TA base pairs. Hydrogen bonds are weaker than covalent bonds but stronger than van der Waals forces (a weak attraction that arises between any pair of nearby molecules--see Chapter 8). The left side of Figure 21 shows a pair of



URE 21: Binary
ures in DNA

nucleotides bound together as a G-C base pair by 3 hydrogen bonds. The right side of the figure shows a more abstract representation of the key features of the four different sidechains. Thus, the two binary features of adenine are [+long] and [-strong], and so on. There are several chemical variants of the bases--e.g., 5-methyl cytosine; but they have the same features as the unmodified bases when encountered in the chain. These modified bases are like molecular allophones, or perhaps ’allobases’. Thus, there would be a more abstract category /C/ that marks off a set of bases that all have the same effect when it comes to coding for meaning; /C/ could be "realized" as [C] or [^{Me}C] but both [GG^{Me}C] and [GGC] stand for the amino acid glycine. This is something like the situation in language that a phonologist calls "free variation"--e.g., the sequences [map] and [map^{h}] (i.e., either unaspirated or aspirated "p") can both be used to mean ’map’.

This is hardly the first time that distinctive features have been ’discovered’ outside of a strictly phonological milieu--such an analysis has been performed on things as disparate as kinship terminology and visual shapes. However, it is interesting to see how much more naturally distinctive features fit with code-like access to DNA (when it is replicated and used to make proteins) compared to the non-code-like recognition of DNA by regulatory proteins. In the cellular case, where we actually have access to the underlying structures, it is not clear that this analysis adds a great deal (though it is a handy mnemonic). Certainly, it doesn’t help understand or explain cases when the binary feature recognition system falters or fails--which requires that we ’really’ understand how the system works in 3-D. Its main use here, rather, is to emphasize what such an alternate, more abstract level of analysis looks like when applied to a well-understood system. Finally, it can be said that distinctive feature analysis is surely at its best in describing the two clearest natural occurrences of pervasive digital effects--the recognition of DNA and speech streams.

Double strandedness

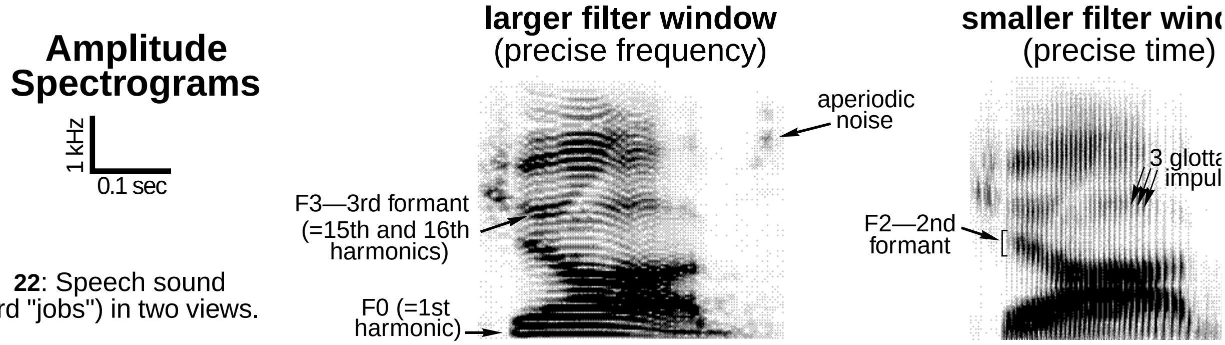
A prominent feature of the molecular symbol chain is that it consists of two complementary strands. This arrangement is nowhere to be found in linguistic level symbol chains (though something analogous could certainly occur in internal, folded representations of symbol chains in the brain). Once again, the difference makes more sense in light of the lack of code production at the molecular level. A complementary double strand has two big advantages for a perception-only molecular system. First, it is much more stable than a single strand. This is important in cells, which "listen" only to their own permanently stored set of symbol chains. If a sequence is lost, so is the meaning that that sequence codes for, since there is no way for cells to generate new meanings *de novo*. The second advantage of a double strand, guessed almost immediately by Watson and Crick, before any of the required enzymatic machinery was understood, is that a complementary double strand can be "semi-conservatively" replicated; the parent double strand is duplicated by splitting it in two and then synthesizing the complementary halves for each single strand. This allows there to be a strong covalent backbone holding together the sequence information at all times. There is nothing quite like replication in language in its native state before the invention of writing (and xeroxing). Spoken language propagation depends on entirely different mechanisms; it requires neither a permanently persistent internal symbol chain, nor direct copies of the whole thing, which is about as long as a 1000 page book at one letter per nucleotide even for a mere bacterial cell.

The notions of backbone and sidechain, however, were for the most part defined in terms of a single strand. Important structural determinants like base stacking and helix formation occur in single strands. Single stranded structures are slightly different from the structure of one of the strands in a double helix--e.g., the single strand would naturally not be expected to show the effect of opposite strand base clashes. Furthermore, during the actual operation of "perceiving" the DNA the double strand is unwound, and only one strand is read. Thus, a number of the properties discussed are at least partly independent of double-strandedness. When we return to discuss the concepts of the linearity and flexibility of symbol chains, however, we will not be able to completely factor it out, and it will remain a point of disanalogy.

What the speech signal looks like

The acoustic spectra of many of the sounds of speech have two distinct parts. At the low frequency end, there is a "fundamental frequency" component (abbreviated F0) that will be compared to the constant phosphate-sugar backbone, and at higher frequencies, there are emphasized spectral bands called "formants" (labeled F1, F2, and so on) that will be compared to the variable information-bearing aromatic base sidechains. These two are considered in turn.

The fundamental frequency is what is perceived as the pitch of a voice. It is generated by the periodic flapping motions of the vocal folds (the glottis). This frequency averages near 100 Hz for an adult male, over 200 Hz for an adult female, and can go over 1000 Hz for a screaming child. Since the volume of air per unit time allowed through the vocal folds varies more like a sawtooth wave than a pure sinusoid, a large series of harmonics that are integer multiples of the fundamental frequency are also generated; for a 100 Hz F0, there will be harmonics as 200, 300, 400, and so on. It is the selective emphasis of certain harmonics that constitute vowel formants. This is the result of the tongue setting the resonant frequencies of the oral (above the tongue) and pharyngeal (behind the tongue) vocal tracts that the glottis-produced sound passes through on it way out of the mouth. Thus, F0 is the first harmonic, but F1 and up are made by whatever harmonic happens to hit one of the vocal tract resonant frequencies. A single word ("jobs" cut out of the sentence "Hi, this is Steve Jobs") is shown in Figure 22 filtered using a wider moving window at the left (which gives better frequency resolution, but poorer temporal resolution), and a narrower moving window at the right (better temporal, poorer frequency resolution). The harmonics of the vocal folds (horizontal striations) are easier to see at the left (they blur together at the right), while transient features (e.g., the vertical striations are the individual impulses of the



RE **22**: Speech sound
vord "jobs") in two views.

vocal cords) are more visible on the right. The auditory system performs sophisticated versions of both of these kinds of filtering.

Fundamental frequency backbone

Many speech sounds contain a fundamental frequency (27 out of a basic list of 36 in English). These include 12 vowels, a total of 7 nasals, liquids, and glides, which are vowel-like, and 8 voiced obstruents, which have a secondary construction in the vocal tract (beyond the glottis) that constitutes an additional but aperiodic (noise) sound source. Nine sounds without a fundamental frequency remain; 8 are unvoiced version of the voiced obstruents, and the last could be called a glottal obstruent. Thus, normal connected speech consists of stretches of higher frequency segment "sidechains" (i.e., formants and aperiodic noise bursts) underlain by an F0 "backbone" that is interrupted from time to time by the presence of unvoiced obstruents. Some unvoiced obstruents in English (e.g., [s]) occur more often than their voiced partners (e.g., [z]); and since there is also a moderate cross-linguistic preference for unvoiced obstruents (Greenberg, 1978), the periodic lack of a "backbone" appears not to be unusual in the speech symbol stream. There is a special case of almost completely "backbone-less" speech--namely whispering. In whispered speech, the vocal cords are lax and do not vibrate; instead, they generate broadband noise which is, nevertheless, filtered in the same manner as the series of harmonics generated while speaking aloud--which means the filtered noise has the same formants as regular speech. Though less emotionally expressive, it is certainly quite understandable. By contrast, stacked aromatic bases without a backbone wouldn’t last long at the molecular level. Once again, this is a difference attributable to differences in production. During perception of either kind of symbol string, the symbol segments have to be presented in the right order. At the molecular level, the backbone is entirely responsible for maintaining the order in a symbol string while at the linguistic level, the other speaker maintains the order using stable, repeatable patterns in their brain. In spite of this difference, there are informative analogies between the different kinds of information carried by the backbone and the sidechains in the two systems, and the methods by which this information is extracted.

By comparison with the formant sidechains, the fundamental frequency backbone usually does not carry what linguistics call "contrasting" phonetic information. We had previously seen this was also the case with DNA where the main things that count in coding are the biochemical features of the particular bases in a sequence (i.e., long or short; 2 or 3 bonds possible); variability in backbone shape rarely has an effect on what amino acids are coded for. As a demonstration of this in speech, it is easily possible to pronounce a vowel (e.g., ’eeee’) at a variety of different levels of voice pitch--i.e., different fundamental frequencies--and always have it sound like the same vowel. The situation is a little more complex in languages that have a small number of different pitch levels or contours (called "tones") as contrastive elements--e.g., in Nupe, an African language, high-tone [ba] means ’to be sour’ while low-tone [ba] means ’to count’ (example from Hyman, 1975). The overall range of an F0 contour for a given sentence in such a language, nevertheless, can be moved up or down without changing the words that are signalled; and a certain amount of non-contrastive variation is allowed even within a single breath group (e.g., lowering at the end of a sentence).

The fundamental frequency contour does, however, carry some information. Under normal circumstances, fundamental frequency changes continuously and is quite complexly patterned; one only need pronounce a sentence or two with a completely steady pitch to see how odd and mechanical it sounds. First, there are some direct local effects of certain segment sequences on F0. Sorenson and Cooper (1980), for example, showed a slight fall in F0 both within and across words during the first 50 msec of a vowel after a prior voiceless consonant. This effect is reminiscent of the local base-sequence dependent modifications of the DNA backbone noted above; interactions between certain classes of sidechains affect the F0 backbone in a predictable way. At a slightly larger scale, there are usually small pitch increases on each stressed syllable, which occur every few syllables, which means about once every 6 or 8 linguistic segments (see e.g., Fujisaka, 1983). Like the regular 10-base-per-turn DNA helix, stress is a locally repeating pattern that accompanies all speech streams. Finally, there are the longer range, large scale changes in the pitch contour that are usually what comes to mind as "intonation". These can sometimes alter the overall meaning of an utterance--e.g., change it from a statement to a question. But intonation can also question the assumptions of the hearer ("it’s not really THAT bad"), or focus on one member of a set mentioned in a sentence, sometimes even affecting things like the scope of negation (e.g., Ladd, 1980). The linguistic functions of large scale pitch contours are more similar to the functions of DNA backbone shape, as sometimes detected by DNA-binding regulatory proteins, than they are to the functions of DNA base sequences, which control the exact sequence of a protein.

Turning away from the disputed area of exactly how much intonation adds to meaning (whispered or written language is certainly understandable without intonation) to the mechanics of how it is perceived, there is a clear parallel with the way the backbone and sidechains are accessed at the molecular level. Normal word meaning gets into the system by sidechain (i.e., formant) sequences based on local cues (see below), while intonational meaning arrives in the form of the overall contour of the F0 backbone, which is only measurable over several words. A similar distinction in input pathways appears at the molecular level. Molecular sidechain sequences are perceived via base pairing (i.e., by detecting the local distinctive features of the bases) as normal amino acid word meaning, while the overall shape of the backbone is perceived more holistically by transcriptional control proteins that detect its overall shape.

Variable formant sidechains

The linguistic symbol segment sidechains have been referred to as a group as formants. The term was introduced in the late 19th century with reference originally to vowels, so will start with them. As described above, the vocal tract is essentially an adjustable resonant cavity with a driver, the glottis, at its base. Formants F1 and up are defined for a particular adjustment of the cavity as its 4 or 5 lowest resonant frequencies, the lowest of which is above the highest voice pitch (F0) in adults. The long series of harmonics generated by the vocal folds are then filtered--the harmonics that happen to land away from a resonant frequency are attenuated while those that land on or near a resonant frequency are passed and appear in the total output spectrum as emphasize bands or "formants". What vowel is perceived is determined mostly by the relative and absolute frequencies of the first two formants. Since the pitch of a harmonic series sounds like the pitch of the first harmonic even when the first harmonic is removed, it is difficult to consciously hear the formants as pitches; instead they affect the timbre of the sound.

Since men, women, children, and worst of all, Alvin and the Chipmunks have different sized vocal tracts, the resonant frequencies for the ’same’ oral gesture differ (speeding up a tape raises voice pitch, but also shrinks the vocal tract--in Alvin’s case to ridiculously small dimensions). Therefore, the relative positions of formants for a given speaker are more important than their absolute frequencies. Later, we will see that bats have solved a related problem (for an entirely unrelated reason--namely, to detect the velocity of flying insects). It is as if people can immediately estimate the size of a speaker’s vocal tract and compensate for it. This is one of a family of similar perceptual problems (e.g., estimating the distance of an object to tell how big is really is). Already, the system is much more complex than with nucleotide bases where only two features are needed and where there is no "sexual dimorphism" in nucleotide chains. A popular, quasi-articulatory based method of describing 12 vowels in English uses 4 features--"height" (high, mid, low), which is inversely proportional to F1 frequency, "backness" (front, back), which is proportional to the difference between F1 and F2, and "rounding" (rounded, unrounded) and tenseness (tense, lax), which have complex acoustic correlates. A more straightforward acoustic specification (Ladefoged, 1980) would involve 6 continuously varying "features"--the amplitude and frequency of F1, F2, and F3. The 12 vowels commonly distinguished in English are found in the words--teak, tick, take, tech, tack, tock, talk, took, toke, "tuke", tuck, turk. International phonetic alphabet symbols for them are, in order--i, ɪ, e, ɛ, æ, a, ɔ, ʊ, o, u, ə, ɚ.

Two-thirds of the contrasting sounds in English--the consonants--remain. Briefly, these can be divided into vowel-like and obstruents. The 7 members of the vowel-like group includes nasals, liquids, and glides. All of these have sets of vowel-like formants that differ, however, in their patterns of intensity, frequency, or rates of change from those in true vowels. The 7 vowel-like consonants distinguished in English are found at the end of "sun", "some", and "sung" (nasals, and at the beginning of "lat" "rat" (liquids) and "yet" and "wet" (glides). Phonetic symbols in order are--n, m, ŋ, l, r, y, w. The largest group of consonants are the 17 obstruents in which there is a secondary constriction in the vocal tract past the glottis that generates a turbulent air flow and thus, unlike most glottally generated sounds, an aperiodic (i.e., noise-like) spectrum. The main motor variable is the "place of articulation", which can range from the lips all the way to the glottis itself. For a sensory point of view, different locations in oral cavity constriction result in a different spectral profile for the noise and often one or more "format" peaks in the noise. Two other distinguishing features of obstruents are whether or not "voicing" (a periodic sound generated simultaneously by the glottis) accompanies the noise, and the degree of abruptness ("manner") of the onset and offset of the noise. There are 17 obstruent consonants distinguished in English. They are listed (as the first sound in each word) in voiced and unvoiced pairs, and in order from most distal (i.e., near the lips) to most proximal (i.e., near the glottis) place of articulation--pin, bin; fin, vim; thin, then; tin, din; sing, zing; Shawn, genre; chin, gin; come, gum; him, (no voiced version possible). The phonetic symbols for these sounds in order are--p, b, f, v, θ, ð, t, d, s, z, š, ž, č, ǰ, k, g, h. This is an oversimplified summary since several speech sounds appear to have multiple cues. Nevertheless, recent work had emphasized the particular importance of the cues listed above. In the case of stop bursts, for example, Repp (1984) showed that listeners are remarkably sensitive to even very weak release bursts (i.e., short noise formants).

To summarize, a total of at least 10 features are conservatively needed to describe the 36 segment "sidechains" listed here, in contrast to 2 features for the 4 nucleotides. The more abstract well-known analysis of Chomsky and Halle (1968), for comparison, uses 13 binary features for English to describe 46 different sounds, while Ladefoged (1980) lists 15 continuously varying strictly acoustic features, and Ladefoged (1982), 20 traditional features to describe over 100 sounds present in one or another of the world's languages. Clearly, there is a major quantitative difference here between the linguistic and molecular systems in the number of sidechains and in the number of ways that they differ (e.g., their features). There is, of course, a proportional difference in the number of word meanings coded for. The interesting similarity is that for a given feature, there are rarely more than 2 or 3 different "levels" of it (*pace* Chomsky and Halle et al., who allowed only two and Ladefoged who prefers the accuracy of continuously variable parameters); even "place or articulation", with 7 possible levels is rarely used by itself to distinguish more than 3 or 4 otherwise similar sounds in a language.

Initially, it might seem perverse to think that there is a common explanation for why molecular level symbol chains weren't based entirely on 4 different sidechain lengths, or alternatively on four different types of base pair bonds, or in the case of language, why speech sounds weren't more "rationally" designed with, say, 6 levels of F0 and 6 levels of F1 (i.e., for 36 possible combinations) or even better, 36 levels of F0. There are plausible enough reasons in each case why this didn't happen. With DNA, for example, 4 different base lengths would probably be hard to fit into a uniform helix, or too similar to allow reliable recognition and base pairing. Remember that code-like recognition depends on very localized features. With language, 6 levels of F0 (much less 36) would be difficult to produce in an arbitrary sequence and hard to reliably distinguish. But stated this way, a certain similarity should be more apparent. A way to phrase the common constraint in the present framework is that the conflicting requirements of being a symbol chain--i.e., the need for units uniform enough to all fit into a uniform symbol chain versus the need to have reliably locally distinguishable units--conspire to make it impractical for there to be more than a couple of levels of a given feature. In the next chapter, we will describe in more detail some of the other defining features of a symbol chain (segmentation, linearity, and sequence arbitrariness).

Word recognition

Cells and persons routinely "perceive" and "comprehend" their code chains. One stage in this process is *word recognition* (marked in bold in Fig. 23). Word recognition is a process by which a continuous chain of symbol segments lacking word delimiters is recognized to contain a series of multi-segment words. It takes place after the symbol chain has been copied into an active message. A key requirement is that the end of one word must be actively recognized in order to locate the beginning of the next word. The mechanistic details of word recognition are better understood at the molecular level than they are with respect to neurophysiology of language

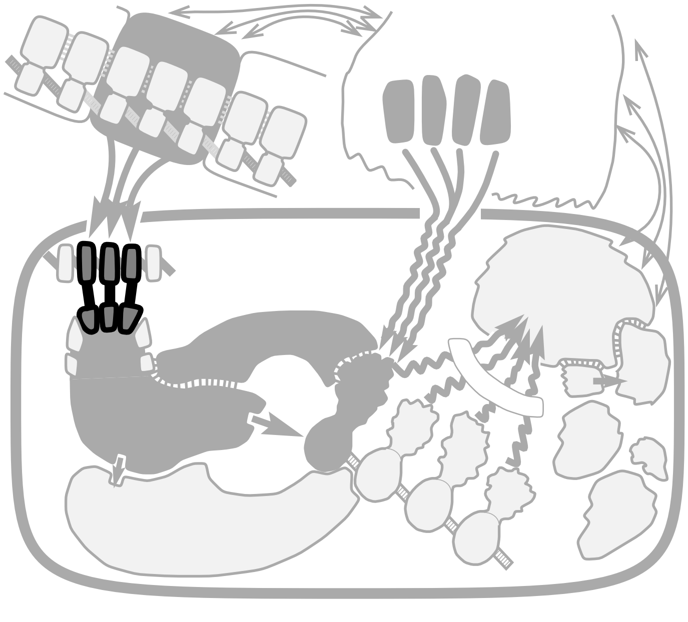


FIGURE 23: Word recognition

comprehension. Nevertheless, what is already known makes word recognition an unambiguous exposed comparison.

In cells, the smallest meaningful group of symbol segments (DNA nucleotides) is the *codon* or *word*--a triplet of DNA nucleotides that stands for one of the 20 different amino acid meanings. Thus, all words in cellular symbol chains are the same length. Four different types of nucleotides taken three at a time make 64 possible codons, of which 61 are used to code for amino acids. There are no systematic chemical differences amongst the 16 internucleotide linkages *within* and *between* codons in both DNA (chain of symbol segments) and RNA (chain of symbol-representation segments); that is, there are no explicit markers indicating where one codon ends and another begins (see Fig. 24). Therefore, each triplet codon in a messenger RNA strand has to be recognized sequentially in order to determine the start of the next codon. This is done by a set of 40 tRNAs, each of which binds, in the context of the ribosome, to a particular mRNA codon or in some cases, a few codons. As soon as the currently bound tRNA adds the amino acid it is carrying onto the growing chain of bound amino acids, it is released from the ribosome. The mRNA is then pulled through the ribosome, exposing the next codon. When by chance, the correct tRNA diffuses into place, the process is repeated. The ribosome actually binds two tRNAs at once; when the second site is vacated, the tRNA in the first site moves into the second site, opening up the first site for a new tRNA. This allows the amino acid chain to remain attached to a tRNA at all times (see Alberts et al., 1994, and the next chapter).

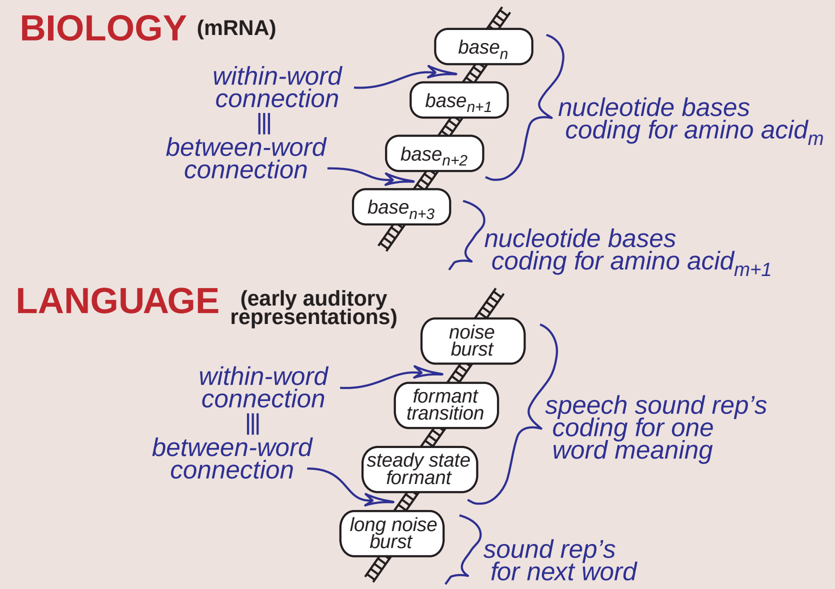


FIGURE 24: Between word same as within word

In persons, an easy demonstration of the lack of explicit word-boundaries can be gotten by listening to a fluent speaker of a completely unfamiliar foreign language; it is impossible to tell where one word ends and another begins. Now human language is considerably more complex than cellular protein synthesis. There are more symbol segments in human languages--typically

30-40 different speech sounds or phonemes *versus* 4 nucleotides in cells. There are many different human languages, while with few exceptions, all cells use the same codons for the 20 amino acid meanings. Word length is variable in human language--there are 1-10 or more symbol segments *versus* exactly 3 per word in cells. And finally, there are more word meanings--10,000 or more in human language *versus* 20 amino acid meanings. Nevertheless, there is still a strong requirement in spoken language that each word be recognized in turn, so that the beginning of the next word can be identified. The process of word recognition has been studied in great detail by psycholinguists (e.g., Frauenfelder and Tyler, eds., 1987). Many theories recognize an early, automatic component that operates one word at a time in a rather context-insensitive manner, and several later processes that are sensitive to syntactic and semantic context. It is known that initial stations in the auditory pathways carry a sequence of activity patterns that closely parallel the rapid spectral changes in speech sounds (e.g., deCharms et al., 1998). The early process of auditory word recognition mentioned above must commence within tens of milliseconds after this stream of activity arrives in secondary and tertiary cortical auditory areas. In spite of our current ignorance of the low level details, the process of word recognition already constitutes a unique exposed parallel between the cellular and human symbolic using systems.

The ratio between the number of mathematically possible symbol segment combinations--i.e., possible symbols--and the number of combinations that are actually used is much larger in human language than in cells. There are billions of possible word-sized speech sound combinations but only tens of thousands are used to code for word meanings in any one human language. In cells, by contrast, there are 64 possible nucleotide triplet combinations and almost all of them (i.e., 61) are used to code for 20 word meanings in cells. Thus, the code in cells is almost entirely overlapping--it makes sense if one starts reading from any nucleotide. Of course, the cell virtually always uses only one of the three possible interpretations of a piece of DNA; the others are only accessed in the event of a frameshift error produced when a nucleotide is somehow skipped or lost or mistakenly. Since there are so many more combinations to work with in human language, it can afford to be less overlapping. Though the probability of an ongoing ’frameshift error’ in human language is low, there are often many alternate parsings on a short term basis (see Cole and Jakimik, 1980). Interestingly, the perceived stringency of sequential word recognition prompted early molecular biologists to look for ways to get around it. The ingenious but ultimately incorrect ’comma-less’ code proposed by Crick et al. (1957), for instance, could only be read in one frame. Human language is certain not completely overlapping like the DNA code, but neither is it comma-less.

Now that we have gotten a sample of some comparisons from several points along the way, we can begin a more systematic trek. The comparisons in this chapter were easier because we proposely glossed over some of the finer points so as not to lose sight of our main point. As the comparison is developed further, we will return to a number of the issues introduced in this chapter to tie up a number of loose ends.[~8,100]