Chapter 5: Context-Free Languages: DNA and RNA

Complexion

"The term 'complexio' [231, pp. 101, 102] was, from the twelfth century, the Latin commonly used for the Greek *crasis*, or temperament, that is to say, the balance of the qualities of hot, wet, cold, and dry resulting from the mixture of the elements in the human body. Since it served as a fundamental concept, not only in physiology but also in pathology and therapy, complexion theory provided important support for the idea that medicine constituted a unified and rational body of knowledge. The general theory, already quite fully developed in the works of Galen, underwent considerable further elaboration during the Middle Ages.

"Complexion also varied among different peoples or geographic regions; Sythians, who lived in a cold climate, were supposed to be colder and moister in complexion than Ethiopians, who lived under the hot sun."

The Humors

"The concept of humors [231, pp. 104-106]–that is, specific bodily fluids essential to the physiological functioning of the organism–originated at a very early stage of Greek medicine. Various Hippocratic treatises mention one or more of the fluids; *On the Nature of Man* presents what was to become the standard set of four: blood, phlegm, bile (also termed choler, or red or yellow bile), and black bile (or melancholy). The theory of the humors probably developed because bodily fluids of all kinds played a large part in ancient, and subsequently medieval and Renaissance, physiology, diagnosis, and therapy.

"Like all body parts, they (the humors) were themselves complexionate.

"Hence, the balance of humors was held to be responsible for physiological as well as physical disposition, a belief enshrined in the survival of the English adjectives sanguine, phlegmatic, choleric, and melancholy to describe traits of character.

"Medical theory asserted that the human body exists in either health, sickness, or a neutral state between the two. Deviations from health were classified into congenital malformations (in medieval Latin, *mala compositio* of the body), complexional imbalance (*mala complexio*), and trauma (*solutio continuitatis*, or break in the body's continuity). This classification placed almost all internal illness in the domain of complexional imbalance. Relatively little attention was paid to the first of these three categories, and when surgery emerged from medicine as a separate occupation and discipline in the West during the twelfth and thirteenth centuries, the management of trauma became the characteristic task of the surgeon."

Palindromes

A special example of context-free languages are the languages composed of palindromes. Palindromes are of special importance in their relationship to DNA and RNA. Before going any further, a good idea of what a palindrome is can be obtained through examining examples of palindromes.

Definition: A palindrome over V_T is words or strings composed of terminal symbols in V_T but where the strings are spelled the same way in either direction.

Some examples follow.

Example 1:

desserts stressed

$\rightarrow$ / $\leftarrow$ } same spelling, either direction

Note that in example 1, $V_T = \{d, e, s, r, t, \Delta\}$ where Δ signifies a "blank" or "space".

Example 2:

able was I ere I saw elba

$\rightarrow$ / $\leftarrow$ } same spelling, either direction

Example 3:

NIYONANOMHMATAMHMONANOYIN

$\rightarrow$ / $\leftarrow$

same spelling, either direction

(Greek inscription on the sacred font in the courtyard of the Hagia Sophia in Constantinople. The inscription means "wash your sins not only your face")

Example 4:

abbcccdcccbba

$\rightarrow$ / $\leftarrow$ } same spelling, either direction

Example 5:

$\{w \cdot w^R \mid w \in V_T^*\}$

$\rightleftarrows$ } same spelling, either direction

Example 6:

each of the three words: racecar, kayak, level

Context-Free Grammars: DNA and RNA

Constructing context-free grammars that can generate palindromes is an easy task. To obtain the palindrome "desserts stressed", we see that $V_T = \{ d, e, r, s, t, \Delta \}$ (where Δ will signify a space character). Note that the palindrome has an odd number of characters. The grammar follows.

$$G^{pal} = \left(\{S\},\ \{d,\ e,\ r,\ s,\ t,\ \Delta\},\ \mathcal{P}^{pal},\ S\right) \text{ where:}$$

$$\mathcal{P}^{pal} = \{S \Rightarrow dSd \mid eSe \mid rSr \mid sSs \mid tSt \mid \Delta\}$$

A sample derivation:

$$\begin{aligned} S &\Rightarrow dSd \Rightarrow deSed \Rightarrow desSsed \Rightarrow dessSssed \Rightarrow desseSessed \\ &\Rightarrow desserSressed \Rightarrow dessertStressed \Rightarrow dessertsSstressed \\ &\Rightarrow desserts\,\Delta\, stressed \end{aligned}$$

It is an easy task to construct context-free grammars that can generate palindrome-like languages for DNA and RNA.

$$G^{DNA_1} = \left(\{S\},\ \{a,\ c,\ g,\ t\} \cup \{\bar{a},\ \bar{c},\ \bar{g},\ \bar{t}\},\ \mathcal{P}^{DNA_1},\ S\right)$$

$$\text{where: } \mathcal{P}^{DNA_1} = \{S \Rightarrow aS\bar{a} \mid cS\bar{c} \mid gS\bar{g} \mid tS\bar{t} \mid \lambda\}$$

However, using the Watson-Crick complementary relationship, we know that the following is true.

$$\begin{aligned} \bar{a} &= t \qquad & \bar{c} &= g \\ \bar{t} &= a \qquad & \bar{g} &= c \end{aligned}$$

Thus a better grammar for DNA is the following.

$$G^{DNA} = \left(\{S\},\ \{a,\ c,\ g,\ t\},\ \mathcal{P}^{DNA},\ S\right)$$

$$\text{where: } \mathcal{P}^{DNA} = \{S \Rightarrow aSt \mid cSg \mid gSc \mid tSa \mid \lambda\}$$

The question could be asked, "How is the language $L_{DNA} = L\left(G^{DNA}\right)$ composed of palindromes?"

Let us examine a derivation in $L_{DNA_1} = L\left(G^{DNA_1}\right)$; then we will examine the corresponding string in $L_{DNA} = L\left(G^{DNA}\right)$.

In $L_{DNA_1} = L\left(G^{DNA_1}\right)$,

$$
\begin{aligned}
S &\Rightarrow t\,S\,\bar{t} \Rightarrow t\,a\,S\,\bar{a}\,\bar{t} \Rightarrow t\,a\,t\,S\,\bar{t}\,\bar{a}\,\bar{t} \Rightarrow t\,a\,t\,a\,S\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \\
&\Rightarrow t\,a\,t\,a\,c\,S\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \Rightarrow t\,a\,t\,a\,c\,g\,S\,\bar{g}\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \\
&\Rightarrow t\,a\,t\,a\,c\,g\,t\,S\,\bar{t}\,\bar{g}\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \\
&\Rightarrow t\,a\,t\,a\,c\,g\,t\,a\,S\,\bar{a}\,\bar{t}\,\bar{g}\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \\
&\Rightarrow t\,a\,t\,a\,c\,g\,t\,a\,\lambda\,\bar{a}\,\bar{t}\,\bar{g}\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t} \\
&= t\,a\,t\,a\,c\,g\,t\,a \cdot \bar{a}\,\bar{t}\,\bar{g}\,\bar{c}\,\bar{a}\,\bar{t}\,\bar{a}\,\bar{t}
\end{aligned}
$$

While the derived string is not a palindrome, it is palindromic in an obvious sense.

Let us examine the string that would have been derived in $L_{DNA} = L\left(G^{DNA}\right)$:

$$t\,a\,t\,a\,c\,g\,t\,a \cdot t\,a\,c\,g\,t\,a\,t\,a$$

but if we view this as a double-stranded DNA helix:

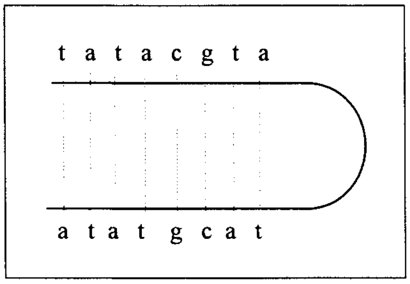

Figure 5.1 RNA folded into a doble-helix like DNA

RNA is typically single-stranded, but the same energy-stabilizing forces that act on a double-stranded DNA helix cause RNA to curl around itself into a double helix, thus RNA is also palindromic, like DNA. The big difference is that in RNA, "t" is replaced with "u". Review the discussion in Chapter 1.

$$G^{RNA_1} = \left(\{S\}\ ,\{a,\ c,\ g,\ u\} \cup \{\bar{a},\ \bar{c},\ \bar{g},\ \bar{u}\},\ \mathcal{P}^{RNA_1},\ S\right)$$

$$\text{where: } \mathcal{P}^{RNA_1} = \left\{S \Rightarrow aS\bar{a} \mid cS\bar{c} \mid gS\bar{g} \mid uS\bar{u} \mid \lambda\right\}$$

Similarly, using the Watson-Crick complementary relationship (but for RNA, now), we know that the following is true (sometimes g and u are paired).

$$\bar{a} = u \qquad \bar{c} = g$$
$$\bar{u} = a \qquad \bar{g} = c$$

Thus a better grammar for RNA is the following [160, pp. 586-588].

$$G^{RNA} = \left(\{S\},\ \{a,\ c,\ g,\ u\},\ \mathcal{P}^{RNA},\ S\right)$$
$$\text{where:}\quad \mathcal{P}^{RNA} = \{S \Rightarrow aSu \mid cSg \mid gSc \mid uSa \mid \lambda\}$$

Both DNA and RNA may have recursive instances of palindromic structures. However, a grammar supporting recursive or repeated palindromic structures is still context-free. Examples of recursive or repeated palindromes:

1. repeated palindromes: desserts stressed · eve
2. recursive palindromes: desserts eve stressed

DNA and RNA grammars that support repeated and recursive palindromic structures [160, p. 589]:

$$G^{DNA_r} = \left(\{S\},\ \{a,\ c,\ g,\ t\},\ \mathcal{P}^{DNA_r},\ S\right)$$
$$\text{where:}\quad \mathcal{P}^{DNA_r} = \{S \Rightarrow SS \mid aSt \mid cSg \mid gSc \mid tSa \mid \lambda\}$$

$$G^{RNA_r} = \left(\{S\},\ \{a,\ c,\ g,\ u\},\ \mathcal{P}^{RNA_r},\ S\right)$$
$$\text{where:}\quad \mathcal{P}^{RNA_r} = \{S \Rightarrow SS \mid aSu \mid cSg \mid gSc \mid uSa \mid \lambda\}$$

More complicated structures, such as supporting copies of substrings, would require a grammar more complex than context-free, such as context-sensitive. This will be discussed at length in the next chapter.