CHAPTER 15

ANALYSIS OF SYLLABARY SPELLING

15-1. Identification of Syllabary Spelling

The key to breaking into codes and syllabary ciphers is to identify and exploit syllabary spelling. If possible, try to locate instances where the same word is spelled in different ways by combining the syllables and letters in different combinations each time. This situation can be exploited fairly easily.

a. Identifying repeated syllabary spelling in syllabary squares was demonstrated in Chapter 5.

b. In codes, only certain groups represent letters and syllables, but these tend to cluster together. With code charts, if begin spell or letter shift groups are used, identifying these special purpose groups serves to point right to groups used for spelling. Often begin spell-end spell groups or letter shift-word shift groups are the highest frequency groups and tend to alternate in the text. This makes them quite easy to spot.

c. In codes where no shift groups are used, the code groups that represent letters and syllables tend to cluster together, just as code groups that represent numbers do. If necessary, computer produced indexes of code groups and the code groups they appear with will help to isolate those used for spelling.

15-2. Recovery of Syllabary Spelling

By comparing different spellings of the same word, you can often figure out which groups represent single letters and which represents syllables. Then, the groups which represent syllables can be replaced by groups that represent single letters. Reduction to single letter terms, in turn, enables recognition of word patterns. This approach to

recovery of syllabary spelling applies equally to syllabary squares, code charts, and book codes. The segments below, each of which represents the same plaintext, illustrates how spelling can be recovered.

A: 81 35 25 74 60 60 11 54 88 88 14 28

B: 83 29 60 60 11 59 88 14 28

C: 81 35 29 60 60 11 59 88 11 60 25 35

D: 83 25 76 60 11 59 88 14 25 35

a. The first three segments all include the text 60 60 preceded by two, three, or four dinomes. If we suppose that the four dinome spelling is all single letters because it is longer than the others, then the two dinomes in segment B must each represent digraphs. Segment C with its three dinomes helps to confirm this breakout.

b. Similarly, segments A and B end with 88 14 28. Segment D ends 88 14 25 35; therefore, 28 must equate to 25 35.

c. Similar comparisons show that 14 equates to 11 60, 59 equates to 54 88, and 76 equates to 74 60.

d. We now take the first segment, for example, and replace all the dinomes that equate to two other dinomes with the single letter equivalents.

Segment A: 81 35 25 74 60 60 11 65 88 88 14 28

Replacement: 81 35 25 74 60 60 11 54 88 88 11 60 25 35

e. Reduced to single letter terms, the word pattern for the replacement segment is -ABCDDEFGGEHBA. This word pattern equates to the word *RECON-NAISSANCE.*

f. These recoveries can, in turn, be used to recover additional plaintext. Whether the system is a syllabary square, a code chart, or a book code, the initial entry is the hardest part. Once the first confirmed recoveries are made, follow-on recoveries are easier.

g. The example above depended on finding sufficient repeated text to reduce the segments to single letter equivalents. This will not always be possible, but it is only one of the approaches an analyst can use to aid in recovery of the system. Anything that provides clues to the plaintext can help solve the system. Information from other sources such as traffic analysis and direction finding can help. Traffic passed in

other systems may provide isologs or clear clues to the content of the text. If the code is a one-part or uses an orderly matrix, the orderliness itself is a major aid in recovering plaintext. Encoded numbers may also help.

15-3. Recovery of Numbers

Another vulnerable point of entry in syllabary squares and codes is encrypted numbers, as has been demonstrated with other systems. Numbers, whether spelled out or encrypted by direct equivalents tend to occur with each other. Grid coordinates will typically occur in groups of four or six digits. Times are usually four digits, and tend to be rounded off into multiples of 5, 10, or 15 minutes. Times always begin with 0, 1, or 2. The third digit of a time is always 5 or less. Because of these characteristics, it is often quite easy to recognize the equivalents of 0, 1, 2, 3, 4, and 5. Even when variants are used, they tend to stand out. Given these six values, others readily follow. Recovered grid coordinates, in turn, give major clues to the rest of the text. Numbers like 7.62 (millimeter), 47 (AK-47 rifle), 45 (caliber), and 72 (T-72 tank) all provide clues to surrounding text.

15-4. Recovery of Words

Initial entry into code systems is often made through the elements that are most like a cipher. Spelled out words and encoded numbers are the weakest points in a code. Once these cipher-like groups are recovered, making further recoveries depends on recognizing the meaning of code groups that represent words and phrases. Slightly different skills are required to recover the vocabulary of a code than are required for ciphers. Cipher analysis tends to be more mathematical in nature.

a. Code recovery is more related to language skills, particularly when the text is not in English. Although words can be recovered as their English equivalents, the actual foreign language words must be known to take advantage of any alphabetic structure in the code. In languages where the sentence structure varies from English, the characteristic structures must be familiar to make sense of the code.

b. Codes are less apt to be fully recovered than ciphers. Code groups cannot be recovered until they are used, and large codes may contain many groups that remain unused for a long time. Each code group must be observed in use several times before its plaintext value can be confidently assigned. Errors are very common in encrypted traffic, and a group must be reused several times just to be sure it is not in error. It also takes repeated usage, in many cases, to be sure which of several words with similar meanings represent a particular code group. Recovery of book codes may never be completed, even when most text becomes readable at an early stage.