MAGICAL NUMBER THREE OR FOUR
IN
ORAL REPRODUCTION DRILL:
Syllabic Structure of a Learner’s Mother Tongue
and
Its Influence on His Aural Perceptions
of a Foreign Language
by
MICHI SASKI
(Received September 22, 1975)

Magical Number Three or Four

I have always felt it queer as a non-native teacher in English at a Japanese college that a student cannot repeat more than three or four words at once when he is assigned to reproduce an English sentence he has just heard from a tape. Strangely enough, in spite of my career of more than twenty years in teaching English, I still find myself trapped now and then in the same numerical threshold when I try to make a text for students out of a tape newly arrived from abroad. It is interesting that it happens even when a decoder can actually grasp the meaning of a given message so tangibly that he can give its simultaneous interpretation in his mother tongue.

Of course, it is not difficult for us to see that even if the surface structure of a given message cannot be analyzed or synthesized exhaustively, we can take out its semantic content from a few fragmentary conceptions and their logical relations suggested by a vague impression of the rapidly flowing string of sounds, though often supported by a verbal or situational context. Even in the communication in a first language, a hearer usually understands what he heard from others without examining its structure and vocabulary in detail unless it is extraordinary in its diction. That is, an exact reproduction is also difficult even in a decoder’s mother language as often revealed by a serial reproduction test. Whether an original message is in a first language or not, what can be reproduced always seems to depend on the decoder’s semantic interpretation at first hand (Sasaki 1968: 57–8, 1975, Sasaki et al. 1971), only it is very hard for a foreign language learner to manipulate syntactic rules and vocabulary sufficiently enough to reach an assigned target because of the lack of intuition.

In view of these, I have untiringly assigned my students to be busied with reproduction drill in order to transfer their grammatical knowledges to practical
skills through their efforts at approximating to an aural model presented by tape (Sasaki 1975). I have been reluctantly convinced, however, that the poor capacity of the initial information processing in average has not been improved at all. They never seemed to stop stumbling at the same threshold in spite of improvement in scores of various tests. The untiring recurrence of the interruption phenomena has gradually driven me to suppose that there might be a numerical threshold for the input that can be taken in by a decoder at the first stage when he cannot predict what to come.

Perhaps, once a decoder has got a clue about the meaning of a message within the range he is able to see with his own storage of skills, he begins to produce its equivalent in his mind. But, the input still remains partially meaningless at this first stage no matter how hard he is really trying to seek its meaning, because it is inherently beyond his intuition especially for a learner with a mere transitional competence in the language. His perceptions also have to remain prepared only for a passive recognition at a stimulus-response level with a little possibility of overcoming the incomprehensibility to so high a degree that he can actively make a simultaneous reproduction. The amount of retrievable signals at this stage, therefore, must be pretty near to that of absolute judgement or immediate memory.

Where is the Threshold?

The numerical threshold, however, could not be located so easily as had been expected, since the magical number three or four in word count, though fluctuated from zero to six by the difference of ability in organizing an English sentence, was undeniably smaller than Miller's seven plus or minus two (Miller 1956 AB). We have already noticed that when a student knows the constituent words of a message by heart, he can easily achieve success in repeating a sentence composed of less than five or seven words in his pattern practice even when he cannot grasp its synthesized meaning at all. That is, a word often makes a chunk for our information processing. It didn't seem to work in this case, though. On the other hand, the threshold would be abnormally increased beyond an appropriate span when the same input had been measured by the number of phonemes or phones.

There was of course another intermediate unit, namely syllable, that was already defined uniquely as a non-emic unit only for the sake of numerical chunking, without which we cannot manage to handle even a lengthy word jammed with too many sounds (Sasaki 1967). Considering that the problem here is deeply concerned with our immediate perception of verbal sounds, the unit of this level is the most appropriate to measure the span of initial processing of serial sounds in a foreign language, of which the decoder should have known all the lower units below morphemes. Although he could not have a practical vocabulary after six years of learning English at high schools, he should supposedly have exhausted the whole list of possible English syllables from the drills given to him in the period. Then, there had to be no difference between this and a result of retention test on a
meaningless string of syllables in his mother tongue (Cf. Sasaki et al. 1971, Kanno et al. 1973). But, it didn't seem to work here, either.

Everyone can see that so far as English language is concerned, the change of measuring unit does not bring out a significantly larger amount of retrievable signals than was measured by number of words unless the message contains too many polysyllabic words. As a reasonable supposition, the retrievable number of units would be a little larger than a fixed threshold for a meaningless string of signals, because however unprepared for detecting a system in a given message, a college student may have some petrified surpluses of meaningful storage such as "He's a...." or "I'd like to..." in spite of his intuitive inability. Although they do not always make a net gain to his retrieval of meaningless portion, often marking a limit of possible reproduction at a first trial, the whole amount of retrievable signals never can be lower than that of meaningless ones. Thus, the attempt to find a threshold already seemed to fail at this preliminary stage.

**Peculiarity of Janglish**

There was only a way left to solve the problem, though. I am no exception to the teachers who feel that the students are always clumsy at arranging syllables in their pronunciations, perhaps because they cannot get rid of their own habit of pronouncing Japanese that is largely occupied with syllables of CV and V kinds with a few exceptions of syllabic consonant. There are neither monosyllabic diphthongs nor consonant clusters. Therefore, even an aging interpreter often shows his nationality by putting an unnecessary vowel after a final consonant: "Ah, you speakku beri good Japanneezu. Ha, ha, ha. Weah you rearn? (Seward 1967:53)" Besides, the Japanese language has a double-decked structure in which syllables also make a finite system of alphabets that are faithfully reflected on its orthographic system. That is, the minimum unit of letters is syllabic. Then, when English language is transcribed in Japanese syllabic characters, it has to undergo some change in phonemic or syllabic measure.

John Manjiro, a fisherman who had learned living English in U.S.A. and also in an American whale catcher for ten years since 1841, was abruptly promoted to be a diplomat by the Government in 1854 when Matthew Galbraith Perry asked Japan to open the country to trade with U.S.A. by the portholes of his fleet. Then, he wrote "A Shortcut to English Conversation" in 1854 as a professor at the Naval Academy, in which he also transcribed English sounds by the Japanese syllabic characters (Sasaki 1969: 25-30). There, so far as we follow the way we read Japanese, "Can you speak English?" becomes "Cann you sueparker Engelless?", that shows the same peculiarity with the oral version of the interpreter mentioned above. Although the orthographic system may have prevented him from expressing an accurate description, it is not quite impossible that his two years and a half at Bartlett's Academy in Fairhaven, Mass. and practical experiences in English did not sweep off his native peculiarities from his English.

If so, it would not be a leap in argument to suppose that the students take in English sounds through the filter of Japanese syllabic system. That is, they
unconsciously add an unnecessary vowel after a consonant so frequently that the amount of whole signals increases to a degree that it easily exceeds a possible span of immediate retention for speech sounds. So far as their pronunciations are concerned, not all of the constituent sounds are Janglicized in their syllabic measure. A simple test, however, revealed that an Englishness achieved in a learner’s pronunciation does not always seem to substantiate the equally efficient processing at the level of his perception. Besides, it is impossible to predict what is most likely to remain Janglicized except for some English loans with their firm roots in Japanese that are usually very obstinate in preserving Japanese syllabic patterns because of their familiarity. In short, it is more coherent, therefore, to suppose that every student unanimously Janglicizes syllables in what he heard in English at his perception level than try to take out a particular difference in syllabication at the level from his apparent pronunciations where even one and the same word is often pronounced differently from time to time. Thus, even the one who can articulate English syllables is supposed to linger at the Janglicizing level in his perception in this hypothesis.

Experiment

Then, it necessarily followed that the span should be measured experimentally, because it was one of the issues that could be examined up to a certain point. The experiment was conducted with 69 freshmen at Iwate University who had already been trained in Lab once a week for more than three months. They were unanimously assigned in their booths to hear and write ten English sentences composed of six to eleven syllables respectively, most of which were in simple conversational style that would make the students get at the target conditions after three trials except for a few who would be discouraged from making the attempt. The informant who had made the tape was the same native speaker they had already got used to through the tapes for their weekly drills.

In order to control the amount of signals the students would retrieve, three utterances of Mod+S type were mixed with those of usual NP+VP type, since they would perceive the latter more easily than the former simply because of their past trainings from junior-high in which materials were arranged unexceptionally according to a syntactic regularity. The students were allowed to make three trials to attain the targets, of which only the results of first trials were collected to be examined. Thus, 690 responses were obtained there.

Since the magical number three or four in daily experiences often showed a flight over a syntactic consistency that could be counted in a recency effect, the whole number of retrieved syllables were marked as the significant score for a response without regard to its orderliness as an English sentence, measured once as a string of English syllables, then as that of Japanese syllables. If retrieved completely, the average number of syllables in Janglicized version should have amounted to 2s+(1~4) for these six to eleven syllabic sentences, where s is the number of English syllables in the authorized version. There are no stable conversion formulae, however, because a conversion factor differs on what kind of
syllable has to be Janglicized. Even the monosyllabic words in English are virtually converted into various number of Japanese syllables as follows.

<table>
<thead>
<tr>
<th>Word</th>
<th>Sound When Converted into Janglish</th>
<th>Syllabic Measure in Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>[a]</td>
<td>1</td>
</tr>
<tr>
<td>eye</td>
<td>[a i]</td>
<td>2</td>
</tr>
<tr>
<td>cock</td>
<td>[ko k ku]</td>
<td>3</td>
</tr>
<tr>
<td>speak</td>
<td>[su pi i ku]</td>
<td>4</td>
</tr>
</tbody>
</table>

Therefore, each of the retrieved portions had to be determined one by one. The results after a simple statistical treatment are shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th>SM</th>
<th>$S^2$</th>
<th>$S$</th>
<th>CI</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng.</td>
<td>4.310</td>
<td>4.8342</td>
<td>2.1987</td>
<td>4.310±0.1641</td>
<td>4.1416≤M≤4.4742</td>
</tr>
<tr>
<td>Jap.</td>
<td>9.1841</td>
<td>28.9328</td>
<td>5.3789</td>
<td>9.1841±0.4013</td>
<td>8.7823≤M≤9.5843</td>
</tr>
</tbody>
</table>

It seemed fairly validated that a Japanese student takes in English sounds through the filter of his own syllabic system, where he can recall syllables that amount to an upper limit of absolute judgment. The mean obtained as that of English syllables also seemed to give substance to my empirical “magical number three or four” in number of recalled words. There were, however, a few more problems still left to be examined. One was that we might possibly control the whole amount of retrievable signals and its mean by increasing or decreasing the verbal difficulty of an input. So, another test was conducted with a different message that seemed a little harder for an average student in its vocabulary and syntactic complexities than the preceding one. But, so far as every syllable that had been retrieved was admitted to make a score, the result did not show any significant difference. Another was the lack of a comparison with the responses that might be shown by native-speaking subjects to some meaningless strings of English syllables, because they would expectedly show a native system for the retrieval of syllables. It was beyond our reach to get a sufficient number of English-speaking people in this district of Japan, though.

**Observation**

Thus, not only for the lack of other possible validations, the result of my experiment may also be no exception to those of other pseudo-psychological testings that mostly remain dubious especially because they usually stand on the negligence of some potential correlative that cannot always be foreseen. Nevertheless, if we consider the obtained numerals only as an available evidence for an empirical finding, we are possibly allowed to put the hypothesis about
intuitive difference of syllabic system softly on the result.

Although I had already recognized the overwhelmingly strong valence of meaning gripped with a learner’s mother tongue even at his reproduction activity in a foreign language (Sasaki 1975: 46–7), I did not expect at all that the syllabic structure established in the same intuition really affected his perceptions of another language with such a quantitative regularity. Empirically, we saw the students always stumbling in pronouncing a lengthy polysyllabic word like “environmentalists” because of their failure in chunking sounds into an adequate string of syllables (Sasaki 1973), but did not realize that the failure was actually that of perception. Perhaps, they exhausted the capacity of their temporary memory at the threshold we have just examined whenever they received a lengthy word or phrase through their own syllabic system.

If so, we can probably predict where the students will probably stumble at. Of course, as a pathological explanation cannot always promise an effective remedy, we cannot help ourselves going on in ordinary ways of teaching by giving them some understandings and drills for their skills. Indeed, there must be no other way to build a system for appropriate chunking at the perception level except to train learners in articulation of English syllables sufficiently enough to make an effective transfer in their receptive skills. At least, however, it is almost obvious now that we have to prepare a strategy against the perception difficulty of the non-emic unit as well as that of emic units that have long been given a pedagogical importance especially by the followers of Fries-Lado. In order to devise the strategy, the quantitative predictability we have just fabricated here as a working hypothesis may be of use in a degree of weather forecast.

Another interesting glimpse of a linguistic culture occurred to me in connection with the quantitative regularity. A popular comedian in Japan, Chu Arai, often made his audience burst into laughter by uttering a Janglicized version of “This is a pen.” whenever he had to speak English in the scene. They usually laughed at this because everybody knew the sentence that had invariably been contained in the beginning of Book 1 of every English reader on compulsory education level, though they could hardly remember all the rest of it. In other words, it is no more than a symbolic remnant of the ineffective teaching of English in Japan that guaranteed them a complete inability in hearing or speaking English. What is interesting enough is that the epitaph to a forgotten language also consists of seven syllables in the popular Janglicized version.

Every learner would concentrate his effort so hard to memorize a given key sentence at least at the first stage of English learning that the classics of teaching material in a primary reader cannot reasonably be reduced only to “This is a pen.” There must be many simpler sentences that were painfully supported by the same effort. What makes it a sole classic is a prosody that has been traditionally favored by Japanese people who still like a five-or-seven-syllabic oral line to compose a song or poem that are deeply relevant to the magical number seven plus or minus two (Sasaki 1968: 62–3). Who knows, however, that the stage pun in Janglish actually satisfies a metric measure that composed many classical songs
and lyrics famed in the literary history of Japan.

REFERENCES


4) Sasaki, M. "A hypothesis on psychological reality of syllable," *Annual Report of Iwate Medical University, School of Liberal Arts and Sciences* 2, 41–6 (1967)


