OLDŘICH ŠVARNÝ
(Prague)

Rhythmical Features of Spoken Chinese: Quantitative and Grammatical Analysis (Methodology)

The research of such a kind as mentioned in the title of this paper presupposes some kind of previous clarification of all the grammatical as well as prosodical features of the language studied. Some 30 years ago I had to tackle the problem of the basic smallest grammatical units of Modern Chinese and I had also to invent phonetic transcription, that would somehow reflect also the prosodic features of Spoken Chinese in their complexity (including rhythmical features). During my research of the problem of the basic grammatical units in Chinese (my starting point in this field of study was Prof. L u   Z h i w e i' s work *Beijinghua danyinci cihui*, Peking, Kexue chubanshe 1956) I succeeded with the help of four Chinese informants to amass a representative body of exemplificatory sentences of the extent of about one hundred and fifty thousand characters. This material illustrates the usage of the most frequent monosyllabic sememes of Modern Chinese, represented in the script by about 1800 most frequent characters (cf. the list of the so called Changyongzi frequently used characters in the periodical "Zhongguo yuwen", July 1952). Since middle seventies this material is prepared for publication and it is also the text (further referred to as: the text), the rhythmical features of which I am analyzing and quantifying at the present time.

The distinction between the concepts of (monosyllabic) sememe and morpheme I defined in my CSc dissertation *K otázce morfému a slova v moderní hovorové čínštině* (On the problem of morpheme and word in Modern Colloquial Chinese), Prague 1963. In this work sememe was designated as "s-morpheme" (s-: semantic), morpheme was designated as "k-morpheme" (k-: component), i.e. component part of disyllabic or more than disyllabic lexical units. The concept of (monosyllabic) sememe corresponds to the concept of *yusu* as it is used by Y i n   W u y o n g in his article *Hanyu yusu dingliang yanjiu*, published in "Zhongguo yuwen" 1984,5. A similar concept is also suggested in the dictionary *Xiandai Hanyu cidian* (XHC), Peking 1979, by

subdividing the treatment of certain characters into two or more separate entries. At last, I should like to call attention to the concept of the functional characteristic (usually many functional characteristics) of a sememe in Chinese. These functional characteristics (abbr.: f.ch.) are for the most part designated by terms usually designating units (not functions) as N (noun), V (verb), A (adverb) etc. The functioning of a sememe as morpheme (component part of a disyllabic or more than disyllabic morphological or semi-morphological unit) is only one among these characteristics; in case of many sememes traditionally termed as "bound morphemes" this is the only possible functional characteristic of sememe.

As far as the interactions of tones as phonological distinctive features of language with sentence intonation, emotional expressivity, contrastive and other types of stresses, speech tempo and other prosodic features are concerned, it is of crucial importance to train and check one's ear by comparing one's own auditive impressions with the data obtained by help of various instrumental methods. I had the possibility to process various samples of speech as spontaneous as possible recorded in Peking in the year 1955 and afterwards in Prague by three native Chinese from Peking through the pitch-analyzing equipment at the Phonology laboratory of the University of Berkeley (director Prof. W.S.-Y. W a n g) in the years 1969–70. The graphs obtained were then analyzed and carefully compared with the auditive impressions.

The text of about 150 000 characters mentioned above was recorded by Mrs. Tang Yunling R u s k o v á in the year 1969 during a period of about 4 months. The speaker was instructed to get acquainted with the material beforehand, but then to speak as spontaneously as possible (definitely not in a "reading" manner or as if putting a model for pronunciation; lapsus linguae were tolerated, even welcomed). She had to pronounce each sentence two or several times, especially when the semantic content should allow several interpretations. The speaker kept to this instruction (especially during the 1st two months of recording) and by the variants of pronunciation obtained in this way a study of the modifications provoked by emotive expressivity, stresses and similar phenomena was made possible. On the other hand the speaker, as the recording proceeded, got a little bit "annoyed" and gradually hastened the recording. A slight tinge of "model pronunciation", present at the beginning of the recording, was gradually dropped, and her speech, remaining altogether vivid, became perfectly fluent. All these changes were at first scarcely perceptible, but they were clearly discovered by later quantitative analysis. In this way, by chance, an interesting study of the influence of the varying speech tempo upon the rhythmical structure was made possible.

By the rhythmical structure of language it is understood:

a) differentiation and hierarchization of the prominence of the syllables of an utterance, i.e. vertical hierarchy (with the neutral line in the middle: the 2nd, 1st and 4th tones according to the degree of their prominence are pushed higher above this line, but the "inversed" 3rd tone is pushed below this line and simultaneously, there is a tendency to push—by contrast—the neighbouring syllables of this 3rd tone syllable to somewhat higher position);

b) differentiation and hierarchization of the closeness of joining syllables (thus organizing them into sequences of various length and various rhythmical structure), i.e. linear hierarchy;

c) in the case of utterances of more than two syllables the organization of various basic rhythmical sequences of identical or differing length and of identical or differing prominence hierarchy into major rhythmical patterns.

The degree of prominence of a syllable in Chinese means the degree of more or less clear preservation of the distinctive features of the etymological tone of the syllable. This hierarchization is made possible especially through the varying "prominence" quantity of the syllables. (We have to distinguish occasional tonally neutral lengthening of the atonic syllables before spaces and breaks; there exists also tonally neutral prominence of e.g. interjections or, sporadically, sentence particles; in these cases the pitch is governed by the sentence intonation.)

In order to facilitate the study of the rhythmical structure we need a suitable system of prosodic transcription of our recordings. In the final stage of my preparatory work after a long practice of comparing auditive impressions of arses ("ictus" syllables) and theses ("weak" syllables) with the mingograms and comparing these impressions also objectively with the impressions of other listeners (cf. O. Š v a r - n ý, *Variability of Tone Promienence in Chinese*, "Dissertationes Orientales" 34, Prague, Academia 1974) I set up a transcription system giving full information about the rhythmical features of Chinese without being too cumbersome.

The linear hierarchy

The speech is organized into sections, terminated by perceptible pauses (duration from a fraction of a second to unlimited extent of time), i.e. terminal breaks marked by specific sentence intonations, marked in the transcription by /./?/!/. Within these sections there are often found non-terminal breaks, characterized by non-terminal intonation and marked in the transcription by /,,/ or /,/; the duration of the pause marked by /,,/ is approximately more than half of a second, the duration of /,/ is less. The rhythmical unit delimitated by breaks (without doing any distinction between them) is designated as "colon". (There is certain correlation between the occurrence of the non-terminal /,,/, /,/ and the inner syntactical structure of sentences; nevertheless it is by no means of the same type in every case.)

The colon is organized into rhythmical segments: it may comprise one segment only (segment colon) or there may be more segments; a segment may have one or more arses. The "tone sandhi", i.e. mutual accomodation of the tones of the neighbouring syllables within a segment is stronger than that between the syllables at the boundary of segments. In the transcription the syllables joined by junctures into segments are written together or hyphenated (this distinction has bearing on the grammatical structure of the text and has no phonetic or rhythmical implications; therefore all the syllables joined by junctures—whether written to-

gether or hyphenated—in respect to the rhythmical analysis are designated as hyphenated).

The boundaries between segments in our transcription are marked by (blank) spaces and are designated as spaces. At spaces, i.e. between the segments the "tone sandhi" is less conspicuous than within the segments. The syllables joining the space on both sides may be eventually bearer of arses, whereas within the frame of segments in such cases one of two possible arses has to be subordinated to the other.

Within the frame of segments we have to distinguish between close and loose junctures as well as within the frame of cola between close and loose spaces. There is a tendency for certain grammatical divisions to appear by slow speed of speech as (close) spaces, but by fast speed of speech as (loose) junctures. In accordance with this the average length of segments varies from about 2,6 syllables by slow speed of speech to about 4 syllables by fast speed. The average length of cola (about 6,5 syllables in our text) is practically independent from the speed (it varies a little dependent upon the length of the sentences).

The (tone) prominence hierarchy

Atonic syllables are marked in the transcription by the form /ba/ (the form /ba/ represents also any other syllable; in case of tonal syllables the form /bā/ represents the forms with any other tone as well). They are subdivided into toneless (the number of the toneless non-lexical monosyllables is a fixed one; in the case of disyllabic lexical units the toneless syllables are indicated in the dictionaries) and neutralized ones.

Weakened tone syllables at the beginning of a segment are written /ba̩/ or more often /ba/, as at this position toneless syllables are excluded (cf. below); within the segment and at its end they have to be written /ba̩/ in order to distinguish them from the atonic syllables.

Full tone syllables are marked /bā/.

Salient syllables are marked /'bā/.

The salience of the syllables serves other purposes than primarily helping to mark the rhythmical structure; all the existing rhythmical patterns (except for the rarely occurring descending sequence 3-2, 3-2-1, 3-2-0, 3-2-1-0, cf. below) may be realized in speech without salience syllables and noted in the transcription without using the salience mark'. It was made possible by setting up the so called "positional" rules of differentiating arses and theses, which apply in two cases:

a) if there is an immediate sequence of two or more tonal syllables (of the same degree of prominence) surrounded by hyphenated less prominent syllables, by spaces or breaks, the arsis rests at the e n d of such a sequence; in case of trisyllabic or tetra-syllabic structures there is another equally prominent arsis on the fi r s t syllable (the rule holds good in these sequences also when any of the two arses is explicitly marked as salient). In sporadic instances of a series of 5 undifferentiated full tone syllables, the arses are put at the odd syllables.

This positional rule applies very often to sequences of full tone syllables but sporadically also to sequences of the weakened ones.

This "acronymic" positional rule is based on the frequency in Chinese of such a rhythmical pattern, which is found also in European languages in the case of acronymic abbreviations, provided that the letters are pronounced as they are pronounced in spelling (cf. abbr. UK, USA, USSR)

b) the rule of leaving weakened tone syllables at the beginning of a segment unmarked (cf. above) is based on the asymmetry of one-one and two or more tones sequences: the two and more tones sequences are of both crescendo and decrescendo types, whereas the one-one sequences, whether binary or consisting of more syllables (maximum 5 syllables) are only of the type arsis-thesis (theses). Sequences: atonic (i.e. neutralized) thesis + arsis do not occur but w i t h i n the segments and they are conditioned here by the rhythmical pressure (in a sequence of two weakened syllables *ba-ba* ... at the beginning of a segment the second syllable is considered more prominent).

Marking the degrees of prominence by figures (0-1-2-3: atonic, weakened, full tone and salient) and using indexes: a (arsis), t (thesis), t + (the second thesis) for syllables differentiated by the acronymic positional rule, we obtain a clear picture of the rhythmical flow of any text, e.g. (the numbering of the exemplifications is that of our text of 150,000 characters):

1490-8

zui-jǐn zhèngfu guǐdingle you-guān-guómǐn-bǎoxiǎn-d 'xiángxi-tiáoliè.

| 1-2 | 2-0 | 2-0-0 | 1-2a-2t-2a-2t-2a-0 | 3-0-2t-2a |

1640-6

ni-bạ-zhę-ge-wénjian jiāochuqụ, děi-yào yi-ge-huì'zhì

| 1t-1t+-1a-0-2-0 | 2-0-1, | 2t-2a | 1-0-2-3 |

1932-2

wàiguo-jiàoshòu, 'yě-zai-women-d héping-hūyushūshang qiānle-míng

| 2-0-2t-2a | 3-0-1-0-0 | 2-0-2-0-2-0 | 2-0-2 |

With the data provided by this type of transcription we proceed to state the percentage of the salient, full tone, weakened and atonic syllables of the text; we discover the maximum length of various types of sequences (e.g. 4 syllables for the crescendo as well as decrescendo sequences); we discover the operation of the positional rules of ascertaining arses and theses in the crescendo sequences but not in the decrescendo ones etc.

Nevertheless the rhythmical features bearing also on the grammatical structure of language cannot be quantified unless we take into consideration also various types of grammatical relations within the text.

In this way, we ascertain first of all the f.ch.'s of the monosyllables forming segments by themselves (in our text these monosyllables amount to 3,5% of the total number of syllables in a sample taken from the beginning of the text—the average duration of the syllables in this sample being about 0,19 sec., whereas in a sample taken from the end of our text—with the average duration of the syllables of about

0,16 sec.—they amount merely to 0,5%) at the beginning, in the middle and at the end of the colon.

In the remaining segments, which are formed by or consist of binary sequences (abbr.: b. seq.) combined between them and/or with "odd" monosyllables, we ascertain first of all the percentage of the various (grammatical) types of b. sequences occurring in various types of rhythmical realizations (one tone sequences and two types of two tones sequences: rising and falling).

Grammatically binary sequences are of three types:

a) lexical b. seq. (i.e. morphological or semi-morphological)

b) quasi-morphological, e.g. seq. with sememes of the "unit" f.ch. (classifier, measure unit, etc.) at the end, nouns with postpositons, seq. verb-noun, sememes of the attributive function followed by noun etc. Binary seq. of this type are to certain degree labile (thus e.g. the sequence numeral—measure unit is disrupted by the insertion of an attributive sememe to the measure unit, the sequence verb-noun is disrupted by insertion of an aspect-temporal marker etc.)

c) contextual, arising by more or less accidental joining of sememes of various f.ch.'s. The forming of these sequences is rather labile as the rhythmical structure of an utterance is subject to constant reorganization under influence of the variability of the "linear context" (adding or deleting various units, substituting lexical disyllables for monosyllables and vice versa etc.) as well as the variability of the "prosodic context" (changing neutral intonation for emotionally expressive one, stressing different units, changing the speed of speech etc.). Most often occurring contextual b. sequences are personal pronouns (as subject) followed by adverb, preverb, prepositional verb, verb /shi/ (copula); (b. seq. of this type arise by encliticization of otherwise proclitic elements); sememes of varied f.ch. followed by sentence particles etc.

The way in which the odd syllables are joined to b.seq. within the frame of a segment is designated as "inclination". It is of two types:

a) apposition, if the odd syllable joins the a r s i s of the b. seq.

b) affixation, if the odd syllable joins the t h e s i s of the b. seq.

The inclination occurs on the right, on the left or on both sides of the b. seq.

In both types of inclination the odd syllable may be thesis or arsis: in the case of apposition the arsis of the b.seq. is subordinated to the arsis resting on the odd syllable.

The syllables in more than disyllabic lexical units (the total number of syllables organized in these units in various samples of our text doesn't exceed 4.5% to 5.5% of the total number of syllables of the samples) are also analyzed into b. seq. and odd syllables. In some specific cases there may occur two inclined syllables (joined on the right side of the b. seq.). Finally, there occur also b. sequences with supressed (i.e. tonally neutralized) arses (always resting on the first syllable of the b. seq.), which cannot preceed, but always follow either monosyllables or b. sequences.

By the processes of inclination there arise by side of the "simple" disyllabic sequences the following basic types of rhythmical patterns:

a) centrical pattern with arsis at the center

b) acronymic pattern with two arses: one arsis at the end, the other at the beginning

c) trisyllabic or tetrasyllabic ascending or descending patterns, with the prominence of the syllables continuously rising or falling

d) one-tone sequences of more than two syllables (maximum 5 syllables).

Binary sequences and basic rhytmical patterns (either occurring as independent segments or joined together to form more complicated segments: the average length of segments varies from about 2.6 syllables in a sample taken from the beginning of our text to about 4 syllables in a sample taken from the end of our text) are coordinated to form rhythmical sequences within the frame of a colon. At the beginning and in the middle of a colon these sequences are typically characterized by the recurrence of decrescendo rhythmical sequences (especially the one-tone ones; recurrence of crescendo sequences is rare); at the end of a colon we find typically—though not always—crescendo, i.e. rising sequences, very often preceded by decrescendo ones which results in the acronymic rhythm).

The quantifying of the rhythmical features of our text on hand of suitably chosen samples (the most important features being analyzed on hand of four samples of the extent of about 8 230 syllables) is a process rather complicated in itself, but the results may be conveniently summed up as to clarify many interesting points of the typology of Modern Spoken Chinese. Equally or even more important is the possibility of setting up a few relatively simple rules helping the student, when speaking Chinese, to avoid substituting the rhythmical patterns of his own mother tongue for those of Chinese.

Appendix

1490-8
最近政府规定了有关国民保险的详细条例。

1640-6
你把文件文件交出去得要一个回执。

1932-2
外国教授也在我们的和平呼吁书上签了名。