

THE FUTURE OF TANGUT (HSI HSIA) STUDIES

by SIR GERARD CLAUSON

Introduction

The study of the Tangut, or as it has hitherto usually been called the Hsi Hsia, language has always been one of the backwaters of Oriental studies. The language is remarkable for being written in one of the most inconvenient of all scripts, a collection of nearly 5,800 characters of the same kind as Chinese characters but rather more complicated; very few are made up of as few as four strokes and most are made up of a good many more, in some cases nearly twenty. It is extremely difficult to remember them, since there are few recognizable indications of sound and meaning in the constituent parts of a character, and in some cases characters which differ from one another only in minor details of shape or by one or two strokes have completely different sounds and meanings.

The language first came to the notice of European scholars as one of the six languages in the multi-lingual inscription of A.D. 1345 on the gateway of Chü-yung kuan in North China, and was first identified as Tangut in 1895, and then only tentatively.

Interest in the subject has revived in recent years, and has been greatly stimulated by the publication of the collected works of Prof. N. A. Nevsky, who died in 1938 after devoting many years to the study of Tangut. The following is a list, for help in compiling which I am greatly indebted to Mr. E. D. Grinstead, of the British Museum, of the more important recent publications on the subject in English, Russian, and German.

(a) The Tangut language

1. Z. I. Gorbacheva, *Tangutskie Rukopisi i Ksilografy Instituta Vostokovedeniya Akademii Nauk S.S.S.R.*, Uchenie Zapiski Inst. Vost. IX, Moscow-Leningrad, 1954, pp. 67 ff.
2. N. A. Nevsky, *Tangutskaya Filologiya*, Moscow, 1960, 2 vols. (hereafter cited as Nevsky *op. cit.*).
3. Tatsuo Nishida, *The Numerals of the Hsi-hsia Language, their Reconstruction and Comparative Studies*, Memoirs of the Research Department of Toyo Bunko, No. 19, 1960, pp. 123 ff.
4. K. Sedláček, *Zur Frage der Etymologie der Tangutischen Partikeln *TA, *NGU, *RI, *KI, *TO, und *HGO*, Central Asiatic Journal, VII, 3, 1962, pp. 153 ff.
5. E. D. Grinstead, *The General's Garden*, The British Museum Quarterly, XXVI, 1-2.

6. M. V. Sofronov and E. I. Kychanov, *Issledovaniya po Fonetike Tangutskogo Yazyka*, Moscow, 1963 (hereafter cited as Sofronov *op. cit.*).
7. Z. I. Gorbacheva and E. I. Kychanov, *Tangutskie Rukopisi i Ksilografy*, Moscow, 1963.

(b) Tangut History and Culture

(all by E. I. Kychanov)

1. *Nekotarie Svedeniya Kitayskikh Istochnikov ob Etnografii Tangutov*, Sovetskaya Etnografiya, 1959, 4.
2. *Gosudarstvo Si Sya*, Leningrad, 1960.
3. *Gosudarstvenny Ustroystvo Si Sya*, Uch. Zap. Leningradskogo Gosudarstvennogo Universiteta 281, Seriya Vostokovedcheskikh Nauk X, pp. 103 ff.
4. *Ob odnom Obryade Religii Bon*, Kratkie Soobshcheniya Instituta Etnografii X, 1960.
5. *Kultura Si Sya*, Vestnik Istorii Mirovoy Literatury, 1961, 6.
6. *K Voprosu o Proiskhozhdenii Tangutov*, Voprosy Istorii i Filologii Stran Vostoka, Moscow, 1961.

Mr. Nishida has also published some papers in Japanese.

The unpublished works of whose existence I have heard are a Tangut Grammar, shortly to be published by Mr. Sofronov, a card index of Tangut characters compiled by Mr. Grinstead and one or two papers in preparation by him, and a skeleton dictionary of Tangut compiled by myself in 1937-38 and now deposited in the School of Oriental and African Studies of the University of London, of which I have sent a microfilm to Mr. Sofronov. Other scholars may acquire microfilms of it if they wish.

Apart from the scholars mentioned above, I understand that Mr. Hashimoto Mantarō of the University of Tokyo, and Miss Mary Ferenczy, a pupil of Prof. L. Ligeti in Budapest, are working in this field. I know of no others, but some scholars in China may be at work also.

The Tangut Language

The general nature of the Tangut language is now perfectly clear. It is an early northern member of the Tibeto-Burman group of the Sino-Tibetan family of languages. It was spoken by a people who are first mentioned under the actual name Tangut in the funerary inscription of the Türkü ruler Bilge Kagan dated A.D. 735.¹ It is clear therefore that by that time they were a recognized people organized in some kind of state, but this state may well have existed at a rather earlier date.

The language became a written one when the Tangut ruler Li Yüan-hao assumed the title of Emperor and, as part of the organization of his government, ordered a certain Ye-li Jên-jung (or, according to another account Yü-chi) to provide him with a script for Tangut as the official language of his Empire. Although accounts of this event differ in detail in the various Chinese authorities, it can be taken as certain that the complete repertory of nearly 5,800 Tangut characters was invented in a single operation in, or about, A.D. 1038.

The "Empire" continued to exist, without apparently playing any significant part in history, until it was finally destroyed by Chinggis Khan

¹ West side, line 24. "In my seventeenth year (A.D. 710/11) I campaigned against the Tangut and routed them."

in A.D. 1227, but this did not entail the immediate disappearance of the Tangut language. Unless its inclusion as one of the six languages of the A.D. 1345 inscription was merely a piece of deliberate Chinese archaism, which is not at all likely, it must still have been written and read over a century later. This is the latest trace of the language in written form, but there is good reason to suppose that it still survives as one of the, now unwritten, Hsifan languages spoken in the border area between Tibet and Ssüchuan.² As Nevsky pointed out (*op. cit.*, I, 39) the Tangut seem to have called themselves, and sometimes to have been called by the Tibetans, by some such name as Mi-ñag and this is still the name of a tribe and language in this border area not very far south of the traditional Tangut country.

The Tangut Script

Any study of the language must necessarily start with a study of the script, of which we are fortunate in having, to all intents and purposes, the complete repertory. It was clearly modelled on the contemporary Chinese script, but unlike the latter was the product of single-stage invention, not long slow evolution. Naturally the long and complicated history of the evolution of the Chinese script, which had already by that time extended over more than two millennia, was unknown to the inventor of the Tangut script; he probably regarded the Chinese script with which he was familiar as being made up of characters of four categories: (1) basic (originally pictorial) characters not susceptible of structural analysis to represent words; (2) similar basic characters used to represent personal, family, tribal or geographical names and sounds occurring as syllables in foreign words (Indian names and *dhāraṇī*'s in Buddhist scriptures and the like); (3) characters which could be analysed as containing two meaningful elements to represent an abstract idea (like mouth and empty for "hungry"); (4) characters which could be analysed as containing a phonetic and a radical. How he set about his task we cannot now tell, but he probably had some dictionary or handbook of Chinese which he used as a model. He certainly did not attempt to produce anything in the nature of a picture-writing, or slavishly to follow the Chinese ductus. The Chinese character for "mouth", for example, a small square, is still vaguely reminiscent of its pictorial origin; the Tangut equivalent, Nevsky's 54th radical (the 6th radical in *op. cit.*, II, 670), which is not used by itself but is used, like the Chinese character, as the radical for such words as "to speak, to eat, to drink, food" does not look in the least like a mouth or anything else. He probably started by devising a certain number of characters to represent Tangut words, probably the same sort of words as are represented by basic characters in

² Regarding the Hsifan languages see the Introduction to F. W. Thomas, *Nam, an ancient language of the Sino-Tibetan Borderland*, London, 1948.

Chinese, and also a certain number of graphic elements for use as radicals, and then completed the repertory by using these elements in various combinations. There was one difference between his practice and the Chinese practice, since, if the essays in character analysis in the Sea of Characters are to be taken as accurate, he used parts of characters, as well as whole characters, as phonetics, and not only as radicals. It seems possible to recognize in the repertory characters of the following categories: (1) basic characters not susceptible of structural analysis representing Tangut words; (2) similar basic characters representing sounds like those listed in (2) of the Chinese list; (3) characters made up of two meaningful elements like those listed in (3) of the Chinese list; (4) characters made up of phonetics and radicals to represent Tangut words; (5) a few characters made up of Nevsky's 52nd radical, connoting "sound", and a phonetic element to represent sounds; and (6) characters made up of elements taken from two phonetic characters to represent foreign sounds which were not Tangut words, for example the character of *tya* which is made up of parts of the characters of *ti* and *ya*. The last class seems to have been an independent invention not based on a Chinese model; it may have been inspired by the compound letters in the Tibetan script with which the inventor was probably familiar. It seems clear that characters in category (1) were also used in the same way as characters in category (2) to represent sounds in foreign words. It also seems clear that the characters which were meaningful were not mere ideographs but represented words with precise sounds. This is no doubt the explanation of the fact that sometimes the same verb is represented by three different characters with different, but usually similar, sounds. It is a regular characteristic of most Tibeto-Burman languages that some verbs have several different forms; in Tibetan, for example, some have as many as four different forms, for the Present, Perfect, Future and Imperative respectively; and it is reasonable to suppose that this was the case in Tangut also.

It seems clear that except in the case of graphic elements used, on the Chinese model, as phonetics in compound characters of categories (4), (5), and (6) the inventor made no correlation between individual graphic elements and sounds; to him the basic phonetic unit of the language was the syllable, not individual consonants and vowels; he made no attempt to analyse the syllable into its component consonants and vowels and represent each of them by a standard graphic component, as is done in the case of an alphabetic script. It is, therefore, quite impossible to reconstruct the phonetic structure of Tangut by a simple study of the script.

Tangut Phonetic Studies

In course of time some Tangut scholars who were familiar with the Chinese language and had learnt something of the Chinese science of

phonetics began to take an interest in the phonetics of their own language. It is a great misfortune that they based their phonetic studies on Chinese and not Tibetan practice for two reasons. The first is that the Tibetan alphabet would have been a far more precise instrument for representing the sounds of Tangut, which was itself a Tibeto-Burman language. The second is that the phonetic structure of Chinese differed in several respects from that of Tangut, more particularly in the sense that (1) at least one sound, the unvoiced labial fricative *f*, and probably more, existed in contemporary Chinese but not in Tangut; (2) at least one sound, the liquid *r*, and possibly more, existed in Tangut but not in contemporary Chinese; (3) some sounds, like plosives, did not exist at the end of words in the Chinese dialect with which these scholars were familiar but probably did exist in Tangut; and (4) certain combinations of sounds like initial, and perhaps final, consonantal clusters did not exist in this Chinese dialect but probably did exist in Tangut. Thus the Tangut works on the phonetic structure of the language, based as they are on quite inappropriate Chinese models, are excessively difficult to interpret correctly. It should be added, however, that there was some discrimination in the use of these models. The Chinese practice was to treat as separate subjects the initial and final sounds of the syllable, with rather inadequate measures for dealing with any medial sounds that existed. Tangut scholars accepted this technique in principle and adopted, lock, stock and barrel, the mediaeval Chinese repertory of the "thirty-six (initial) sounds", and the two broader divisions of these sounds into nine and five categories; but they could not adopt the actual Chinese categories of final sounds or "rhymes", since obviously the shapes of Tangut words differed too widely from those of Chinese words to make this possible.

In the study of the Tangut works on phonetics and other documents from which phonetic information can be extracted, there is a very important time factor to be taken into account. The Tangut script was invented in about A.D. 1038; the earliest Tangut works on phonetics cannot be dated much earlier than the third decade of the 12th century, the remainder come from later dates in that century, and some phonetic evidence exists even in the inscription of A.D. 1345. Considering that during the whole of this period there must have been close personal contacts between Tanguts, and especially educated Tanguts, and Chinese, and that China must throughout this period have been accepted as the fountain-head of education and culture, it is hard to believe that some of the asperities (from the Chinese point of view) of the Tangut language, like initial consonantal clusters and final plosives, were not subjected to a steady process of abrasion and erosion. These external pressures were after all merely a reinforcement of a tendency which was in any case proceeding under its own velocity. In all languages of the Sino-Tibetan family, in Chinese

perhaps more than the rest, but to a marked extent also in the languages of the Tibeto-Burman group, from the earliest period which we can see or infer a steady process of phonetic abrasion has been at work. We cannot tell how fast abrasion was proceeding in Tangut during the 11th to 14th centuries, but it certainly was proceeding. The point is relevant to any study of the Tangut works on phonetics, but even more relevant to the study of the representation of foreign sounds (principally if not entirely Chinese) in Tangut texts of various dates (some of which like the Buddhist sutras may not be much later than A.D. 1038, and one of which is as late as A.D. 1345), the representations of Tangut sounds in Chinese texts (some of which, the historical records, may be earlier even than A.D. 1038, while the principal one, the Chinese-Tangut vocabulary called "The Pearl on the Palm", of which a hand-copy by Lo Fu-ch'êng was published in facsimile in Tientsin in 1924, is as late as A.D. 1192) and those in the transcriptions in Tibetan characters in the margins of Tangut texts, for which so far as I know no date has ever been hazarded. Generally speaking it would I think be true to say that, with a few possible exceptions, all the Chinese and Tibetan transcriptions of Tangut sounds date from a period by which those phonetic peculiarities which would have distinguished Tangut from Chinese had already been severely abraded. There is in fact remarkably little internal consistency about them. The same Tangut character is represented by different Tibetan transcriptions in different manuscripts, sometimes even the same manuscript; the Tibetan transcriptions differ from the sounds which we may conjecture to have been current for the Chinese characters used to transcribe the same Tangut character; and when all is said and done we do not really know exactly how those who wrote these Tibetan transcriptions would have pronounced what they had written. It is clear that a great deal of detailed character by character work still remains to be done on the Tangut transcriptions of foreign sounds, especially those of the earliest period, and on the foreign transcriptions of Tangut sounds, before valid conclusions can be drawn from them. Of these two classes of evidence the first is by far the more important. In this case the end product is likely to be a collection, perhaps disappointingly small, of Tangut characters, the phonetic values of which can be reconstructed with some confidence. This will have both a positive and a negative value. The positive value will be that a reasonably exact phonetic value will have been established for the characters in question and those which can be shown by the processes described below to have been homophonous with them, and also for the initial sounds of characters with the same initial and for the final sounds of characters with the same final sound or "rhyme". The negative value will be that it will have been established that no other final sounds ("rhymes") can be exactly the same as those which have thus been established.

The Native Authorities

The surviving Tangut works on phonetics have been described in greater or less detail in Nevsky, *op. cit.*, I, 95 ff. and Sofronov, *op. cit.*, 13 ff. The earliest is called in Tangut "The Sea of Characters", a name no doubt used in imitation of the Chinese name *wên hai* for similar works in Chinese. For the reasons stated in Sofronov, *op. cit.*, p. 24, it must be earlier, but need not be much earlier, than A.D. 1124. Of about the same date, and complementary to it in the sense that no character is listed in both works, is one called in Tangut "The Mixed Categories of the Sea of Characters", (see Sofronov, *op. cit.*, p. 27). Not much later than these is a work the Tangut title of which can be translated "Homophones". The earlier edition of this is dated A.D. 1132, see Sofronov, *op. cit.*, p. 29, and there is a later, abbreviated, edition which cannot be dated. This is the only work of this class of which there is a modern edition, a hand-copy by Lo Fu-ch'êng, published in facsimile in Port Arthur in 1935, of which there are several copies in this country, one in the Library of the School of Oriental and African Studies. Next in order of date must come the work called in Sofronov, *op. cit.*, p. 30, "The Nameless Manuscript Dictionary", which is a kind of conflation of the Sea of Characters and the Homophones. It is fragmentary and lacks both beginning and end, and cannot be described as an independent authority except to the extent that on some minor points it differs from one or both of its sources. Mr. Grinstead has shown me what appears to be a fragment of the same manuscript in the British Museum, which is important because it contains characters belonging to Chapter IX of the Homophones including two which are missing from the surviving, but damaged, copies of that work. Finally there is a work of which the Tangut name is "The Dissection of the Rhymes of the Five Sounds", usually described more briefly as "The Phonetic Tables". The first edition of this is dated A.D. 1173 and there were later editions, see Sofronov, *op. cit.*, p. 13 ff. Nevsky, *op. cit.*, I, 129, refers to another manuscript dictionary which he says was called "The Precious Rhymes of the Sea of Characters" and is incomplete. Sofronov, *op. cit.*, p. 32, says that he cannot now find any trace of this work, but I rather suspect that Nevsky's description of it is not wholly accurate and that it is in fact the Nameless Manuscript Dictionary.

The Homophones is the only work of this character which is practically complete. The others, even when parts of more than one copy have survived, are all more or less incomplete.

The Interpretation of the Authorities

It is obvious that much more information can be extracted from these works if they are studied in conjunction than if they are taken individually, and I venture to suggest action on the following lines.

The obvious starting point is the Homophones with which a great deal of ground can be cleared and certain lines laid down. This work contains nine chapters each devoted to words starting with sounds belonging to one of the nine categories in the first broad arrangement of the "thirty-six (initial) sounds" referred to above. Each chapter falls into two parts. The first part contains a series of sections, each of which contains a number of characters, from two to nearly thirty, which are pronounced exactly alike. The second part contains a number of characters with "unique sounds"; that is pronounced differently from every other character in the repertory.³ I shall call the latter characters, conveniently but not quite accurately, "unique sounds". The total number of sections in the first part and unique sounds in the second part of each chapter represents the total number of word forms in the language beginning with sounds of the kind to which the chapter is devoted. The number of word forms represented differs widely in different chapters. Seven of them each contain between rather less than 150 and nearly 200 sections and between a little over 100 and nearly 200 unique sounds, say a total of somewhere between 250 and nearly 400 word forms per chapter, but Chapter II contains no more than forty-four sections and thirty-four unique sounds, and Chapter IV contains only twenty characters in all, of which five are unique sounds.

(1) Initials

The reasons for these disparities are to be found in the past history of the "thirty-six (initial) sounds" already referred to, and their partial unsuitability as a basis for classifying Tangut sounds.

Regarding these "thirty-six (initial) sounds" Professor Pulleyblank has very kindly supplied the following note:

"The 'thirty-six initials' belong to a system for representing the sounds of the *Ch'ieh-yün* rhyme dictionary (A.D. 601) in tabular form, which seems to have originated in the late T'ang period (9th century) and was current during the Sung (960-1279). It is clearly ultimately derived from Sanskrit models and was attributed to western monks by Sung writers. The 'thirty-six initials', which belong to a stage in the evolution of the language somewhat later than the *Ch'ieh-yün*, may be reconstructed as follows: (1) velar stops and nasal: *k, k', g, ɣ*, (2) dental stops and nasal: *t, t', d, n*, (3) retroflex stops and nasal: *ʈ, ʈ', ɖ, ɳ* (reconstructed as palatals by Karlgren—the stops later became affricates and coincided with the corresponding terms of (6) below; the nasal, the independent phonemic status of which is somewhat doubtful, became a dental *n*), (4) labial stops and nasal: *p, p', b, m*, (5) dentilabial affricates and nasal: *f, f', v, mv* (these were not

³ In future I shall refer to sections in the first parts of chapters by an ordinal number and to characters in the second parts of chapters by an ordinal number preceded by A.

distinguished from the ordinary labials in the *Ch'ieh-yün* but had developed during the T'ang period), (6) dental affricates and sibilants: *ts, ts', dz, s, z*, (7) retroflex affricates and sibilants: *ts, ts', dz, s, z*, combining two series of initials kept strictly apart in the *Ch'ieh-yün*, retroflex *ts, ts', dz, s*, and palatal *c, c', j, s, z* (the phonemic status of the distinction between voiced affricate and fricative is doubtful and there are other problems about these initials), (8) laryngals, (the glottal stop), unvoiced and voiced aspirates: *h* and *h'*, plus *h'* before *i* and *y* treated as a separate initial, (9) the lateral *l*, and a phoneme which had earlier been a palatal nasal *ny* but which had developed a homorganic closure or friction and become partially denasalized giving *nj*, or *nš*. (Like the other *Ch'ieh-yün* palatals, *nš* became retroflex and, losing its nasality completely, yielded the initial in Pekingese written *j* in the Wade romanization and *r* in Gwoyeu Romatzyh and the new official spelling. It did not become confused with *dz/z* from earlier *j/z* and *dz* because these initials lost their voicing and became *s* or *ts'*.)"

Although this table was in certain respects inapplicable to the sounds of Tangut, which would no doubt have fitted much more neatly into a table on the Tibetan model, the Tangut phoneticians adopted it in its entirety, but for some reason which we cannot now discover rearranged the groups in a different order (see Sofronov, *op. cit.*, p. 15), which is the order of the chapters of the Homophones: (1) Labial stops and nasals; (2) Labial affricates; (3) Dental stops and nasal; (4) Retroflex stops and nasal; (5) Guttural (Velar) stops and nasal; (6) Dental affricates and sibilants; (7) Retroflex or Palatal affricates and sibilants; (8) the glottal stop, etc.; (9) the liquids *l, r* (a Tangut sound unknown in Chinese) and some sound related to *nj/nš*. So far as the first, third, fifth, sixth and, subject to the remarks below, the seventh groups are concerned, there is no reason to suppose that the Tangut repertory of sounds differed in any respect from the Chinese. The only native Tangut sound which can reasonably be brought within the scope of Chapter II of the Homophones, with its seventy-eight different word forms, is the bilabial *w*. In Tibetan less than half a dozen words begin with this sound, but this initial was commoner in Nam (see Thomas, *op. cit.*) and may have been fairly common in Tangut. Some of the seventy-eight word forms, but perhaps not very many, no doubt represent Chinese loan words, names, etc. with initial *f*-. The only obviously possible explanation of the remainder is that in Tangut medial *w* was regarded as the principal sound in an initial consonantal cluster of which it was the second member and that words containing it were included in this chapter. This would of course be different from the practice in Tibetan where medial *-w-* is represented by a subscript and the consonant to which it is attached is regarded as the initial. The Tangut phoneticians were obviously at a complete loss when it came to the fourth group (Chapter IV) which by the 11th century had probably already passed from the plosive to the affricate

stage and so, so far as the first three sounds were concerned, were homophonous with the equivalent sounds in the seventh group. Neither Retroflex nor Palatalized Dentals existed in Tibetan and they probably did not exist in Tangut either. In both languages such combinations as *ty* existed in loan words of Indian origin, and in the unique sounds in Chapter III there are several characters, of the kind described above as category (6) of the graphical categories, representing such sounds, certainly *tya, tye, t'ya, dya, dye* (A117, A118, A132, A113, A112) and probably others. There is good evidence that all the characters in Chapter IV began with the Palatal nasal *n̄*, but some difficulty is caused by the fact that there is also evidence of words with initial *ny*- in Chapter III. Section 138 of that chapter seems to contain words pronounced *nya* and the unique sound A102 seems to be *nyin*. It is uncertain what Tangut sounds were brought within the eighth group (Chapter VIII). There were certainly words with initial vowels in the language, but whether there were two classes with initial glottal stops (·) and smooth vocalic ingress (*h*) respectively is hard to determine. There were also words with initial *y*-, and *h*-, but not, if Tibetan is a reliable guide, *h'*-. As regards the ninth group (Chapter IX) there were certainly words with initial *l*- and *r*- and a third sound, the nature of which is obscure. There was a voiced Palatal sibilant *z* in Tibetan but this is not one of the letters which can take *h*- (the letter used to represent the smooth vocalic ingress, which had a faintly nasal connotation when prefixed to some plosives) as a prefix, and it can therefore be said confidently that there was no such sound as *nš* in Tibetan, and the position was probably the same in Tangut, but the combination *hj* (*nj*) did exist in Tibetan and this may be the third sound in Chapter IX. If not it must have been *z*, and in that case this sound must have been excluded from Chapter VII. To sum up, there were certainly less than thirty-six pure initial sounds in Tangut (excluding Chinese loan words) but how many less is uncertain.

When it was complete, the earlier, longer, edition of the Homophones listed 5,778 characters, plus up to six more, of which 133, plus up to six more, are missing from the existing text owing to damage to two or three pages in Chapters VII and IX. Some, perhaps nearly all, can be recovered from other authorities. This is likely to have been the complete repertory of A.D. 1038, since, as pointed out in Nevsky, *op. cit.*, I, 98, some (in actual fact at least fifteen) are noted in the Homophones as being no longer in use. Some confusion regarding the actual number of characters has been caused by a colophon which gives the total number of characters as 6,132 (see Nevsky, *op. cit.*, I, 96 and Sofronov, *op. cit.*, p. 30), but this figure was reached by adding in the number of characters in the Preface and chapter headings (except oddly enough in the case of Chapter IV), a process comparable to that of adding in the date with the pounds, shillings and pence in one's accounts. This exact figure of 5,778 plus up to six is important in

connexion with the question of "tones" referred to below. It seems likely that the later edition is merely an abbreviation of the earlier. The only indication to the contrary is the statement (Nevsky, *op. cit.*, I, 121) that its Chapter IV contained twenty-three characters as against twenty in the earlier. But this, too, is probably merely a matter of counting. If the chapter heading had been added in in the earlier edition the number there would have been greater than twenty. The figure of 4,719 characters in the later edition (Nevsky, *op. cit.*, I, 121, Sofronov, *op. cit.*, p. 30) is probably as inflated as that of 6,132 in the earlier editions.

When the information in the two editions of the Homophones has been properly classified and collated it will show the category of initial sound of nearly every character in the repertory and also which characters were pronounced in the same way as other characters. If this information is co-ordinated with that derived from a study of Tangut transcriptions of foreign sounds, particularly those of the earlier period, it will be possible to assign an exact sound to some of the groups of characters and some unique sounds, though perhaps disappointingly few, and thus to be sure that no other characters in the repertory can have these exact sounds.

(2) Finals

The next step will be to co-ordinate the information thus accumulated with the information in the other works enumerated above about the initials and also about the "rhymes", that is the final sounds, of which as Nevsky and Sofronov have shown, there were 105 in the language. To the extent that it is available, this information will make it possible to discover something about the final sounds of a great many characters, perhaps nearly all except some of the unique sounds, and in particular what sections and unique sounds in the Homophones have the same final sound. If one character in any section has a particular final sound it follows of course that all the characters in that section have the same final sound. Moreover if one such character is a character used in Tangut transcriptions of foreign sounds of which the phonetic value has been established, the exact sound of the "rhyme" in question will have been identified. In order to explore the possibilities of a co-ordination of this kind, I made a simple test. In my own skeleton dictionary every character is accompanied by a reference to the chapter and section (or unique sound) of the Homophones in which the character occurs. In Nevsky's skeleton dictionary, *op. cit.*, I, 174 onwards, this information is lacking, but against a good many characters he noted that the character ended in one of the 105 rhymes, or rather in a particular *p'ing* or *shang* rhyme. (I shall discuss these terms later.) I collated the first 340 pages of his dictionary, about one-third of the whole. One rhyme about which I obtained a good deal of information was the 37th in the overall numeration, the 36th *p'ing* and 33rd *shang*. The following sections and

unique sounds were shown by my sample to have this rhyme (a complete collation would probably produce more): I 8, 20, 21, 33; III 15, 37, 38, 44, 45, 82, 83, 96, 143, A 22; V 20, 46, 52, 57; VI 2, 26, 73, A 104; VIII 4, A 18; IX 107. I then checked against this list a rather rudimentary list which I found in my old papers of the characters used to represent syllables in *dhāraṇī*'s in two published Tangut Buddhist sutras and was delighted to find that characters used to represent the following sounds all came from one or other of these sections: *ke, te, de, ne, pe, be, me, ye, le, se, se*. A number of interesting conclusions can be drawn from this little experiment. (1) the final sound of the 37th rhyme was *-e*; (2) no other rhyme can have had this exact final sound; (3) it was sometimes the practice of the author of the Homophones to list two consecutive sections with the same final sound (I 21, 22; III 37, 38; 44, 45; 82, 83); (4) apart from this there does not on the present evidence appear to be any correlation between the order of sections in the Homophones and the rhymes; (5) the author of the Homophones distinguished a large number of initials (including initials plus medial sounds) within some chapters. Thus in Chapter III which theoretically contains words beginning with the four Dentals, *t, t', d, n* there are no fewer than nine sections and one unique sound (and on a complete check perhaps even more) which had the same final sound *-e*. In the case of the Dentals the only probable medial sounds in Tangut words are *-r-* and *-w-*, and as pointed out above it is possible that words with medial *-w-* were listed in Chapter II; *t/-* etc. are not possible consonantal clusters in the Tibeto-Burman languages, and *ty-* etc. are equally foreign to them; there are in fact unique sounds representing *tye* and *dye* which have already been mentioned; they probably end in the 37th rhyme, but this cannot at present be proved. After assuming that the full gamut of initials *t-, t'-, d-, n-, tr-, t'r-, dr-* and perhaps *tw-, t'w-, dw-* and *nw-*, and *ny-*, which may be represented in this chapter and not Chapter IV, existed with the 37th rhyme, the known ten sections and unique sounds listed above are only just accommodated, and it seems very probable that in Tangut, as in Tibetan, initial Dentals with prefixed *g-, b-, m-* and *h-* also existed. I shall return to this subject later.

The number of 105 rhymes is an impressive, and rather daunting, one, even though Sofronov, *op. cit.*, p. 77, has shown that they can apparently be broken down into a great cycle of sixty-eight rhymes, and three smaller cycles of 18, 12 and 7 rhymes respectively which seem to contain small selections of similar rhymes in much the same order.

A comparison with the final sounds in Chinese and Tibetan is instructive. In the more or less contemporary *Ch'ieh-yün chih chang t'u* already referred to, about 120 to 130 different final sounds are listed (see Karlgren, *Études sur la Phonologie Chinoise*, Leyde etc., 1915-25, p. 36), but this figure is much inflated since in this work words with the same final

sound in different tones were regarded as having different "rhymes". In his *Grammata Serica*, Stockholm, 1940, pp. 41 ff. Karlgren set out a table of his reconstruction of the final sounds of "Ancient Chinese" (Pulleyblank's "Middle Chinese") and the total number of rhymes (disregarding tones) is only ninety-four, of which twenty-five end in open vowels or diphthongs, thirty-five in nasals (-*m*, -*n*, -*ŋ*) and thirty-four in plosives (-*p*, -*t*, -*k*, for which perhaps -*b*, -*d*, -*g* should be substituted).

The phonetic structure of Classical Tibetan of about the same date (early 7th century A.D.) is very different. There were only five vowels, *a*, *e*, *i*, *u*, *o*, though no doubt these written letters represented a larger number of shades of sound. These five vowels could all end a word, and so could eleven diphthongs. In addition words could end in any one of the five vowels followed by one of ten consonants, -*g*, -*ŋ*, -*d*, -*n*, -*b*, -*m*, -*h*, -*r*, -*l* and -*s*, four consonantal clusters -*gs*, -*ŋs*, -*bs*, and -*ms*, and in pre-Classical Tibetan three others -*nd*, -*rd* and -*ld*. Thus the theoretical total of final sounds, which may be a little greater than the actual, was five, plus eleven, plus five multiplied by fourteen, that is eighty-six in all, with an extra fifteen in pre-Classical Tibetan. It will be seen that the total number of final Tangut sounds (rhymes) distinguished by the phoneticians of the 12th century was appreciably greater than the total number of final sounds which existed in Chinese and Tibetan three or four centuries earlier, and it is clear, for the reasons stated below, that this is not a total inflated by the double counting of the same sound in different tones.

Sofronov, *op. cit.*, p. 83 ff., has made a brave attempt, based on the Chinese and Tibetan transcriptions, to reconstruct these final sounds, but it is pretty clear that his reconstruction cannot really represent the final sounds of Tangut as they were in the 11th century, and if they represent the final sounds as they were when the language had been severely abraded it is probably rather too elaborate. Moreover one technical question immediately arises. In the Chinese system the final sound was the open vowel, diphthong or vowel plus consonant, whether or not these sounds were preceded by a medial -*ɣ*-, -*ʏ*-, or -*w*-; in Karlgren's table of Ancient Chinese final sounds, for example, -*a*, -*ia*, and -*wa* count as one final sound, not three. Sofronov's list contains sixty-five open vowels, thirty-one open vowels preceded by -*ɣ*-, three open vowels preceded by -*w*- and six final sounds of vowel plus -*n*, two of them preceded by -*ɣ*-. On the Chinese system of counting this represents only sixty-nine final sounds. It does seem impossible that a Tangut phonetician, however acute his hearing, could have distinguished sixty-five different open vowel sounds, even if some of these were in fact diphthongs. In Ancient Chinese Karlgren reconstructed only fourteen distinct vowel sounds (seven shades of *a*, four of *e*, *i*, *o* and *u*), and of these only *a*, *ā*, *i*, *o* and *u* occur as absolute finals, the remainder occurring only in diphthongs or before final consonants. It must surely be the case that in its

original state the table of 105 rhymes included a number of sounds ending in consonants or consonantal clusters. It remains to be seen whether they can be reconstructed.

(3) Initial Consonantal Clusters?

The next stage in the co-ordination of the information provided by the various authorities will be the attempt to get greater precision regarding the initial sounds. As Nevsky and Sofronov have pointed out, the Phonetic Tables classify the Tangut characters not merely under the nine categories recognized in the Homophones but under the thirty-six individual sounds. It has been pointed out above that the division of Chapter III of the Homophones into sections proves that in fact there were not merely four Dental initial sounds but a good many more. The same is no doubt true of other chapters. The division of sounds of the eighth category (glottal stop, etc.) and the ninth (liquids) may be particularly illuminating, since at any rate in the first case the existence of prefixes is impossible, except perhaps before *h*-. In this matter some help may be got from the *fan-ch'ieh*'s where they are given by some authorities, and also from the fact that no character occurs both in the "Sea of Characters" and the "Mixed Categories". Only a close study of these two authorities could disclose the reason, but it is possible to hazard the conjecture that the first authority lists characters with simple initials and the second characters with initial consonantal clusters.

There are obviously two important points not touched by the procedure outlined above. The first is that of the medial sounds, if any, existing between the initial and the vowel, that is -*r*-, -*l*-, -*y*-, and -*w*- (unless words with medial -*w*- are assembled in Chapter II of the Homophones). It seems likely from the information given by Sofronov that a good deal of light will be thrown on this subject by the way in which characters are arranged in different lines in the Phonetic Tables and also, of course, by the *fan-ch'ieh*'s.

(4) "Tones"

Finally comes the perplexing problem of "tones" in Tangut. There is no reference to these in the Homophones, and it is clear that the compiler of that work, even if he realized that such things existed, did not regard them as sufficiently significant to require him to put characters of which the pronunciation was identical apart from a difference of "tone" into different sections of his work. In the exercise of collating Nevsky's dictionary and mine described above I found, for example, characters belonging both to the *p'ing* rhyme No. 36 and the *shang* rhyme No. 35 in sections I, 20, 21 and 33 and III, 15, 37 and 96. The number of cases of this kind is

much too great to permit the assumption that this was due to simple error.

The other authorities all refer to "tones", if that is the right translation of the Tangut word used, and in some cases classify characters under different "tones". The question of terminology is extremely important, since there has been a great deal of confusion in regard to it, not only in the case of Tangut but also in the case of other languages. It is essential to distinguish between two phenomena, both of which are sometimes described as "tones". The first is "pitch", that is the level, higher or lower, at which a word is pronounced; the second is voice modulation, that is, normally, a rising or falling modulation or an abrupt interruption of sound. In all modern and some ancient dialects of Chinese both these phenomena occur, and Chinese phoneticians were perfectly well aware of them and described them. Unfortunately their terminology is somewhat ambiguous. They had a word *shēng*, "sound", which might mean pitch or voice modulation (and other things as well). They distinguished, as a general rule two pitches "low" (*hsia*) and "high" (*shang*) and three voice modulations "level" (*p'ing*), "rising" (*shang*), "falling" (*ch'ü*, literally "departing") apart from words which originally ended in plosives described as "abrupt" (*ju*, literally "entering"). It will be noticed with regret that they used the same word, *shang*, both for a high pitch and a rising modulation.

Pitch is a factor of the initial sound of a word, and voice modulation a consequence of an alteration in the final sound of a word. So far as pitch is concerned, the original position in Chinese is quite clear, although it has been somewhat obscured by more recent developments; words with a voiced initial were pronounced in a low pitch and those with an unvoiced initial in a high pitch. The question of voice modulation has been discussed on many occasions and most recently in Pulleyblank, "The Consonantal System of Old Chinese", *Asia Major*, NS, IX, 1962. The position, as I understand it, can be summed up as follows. As far back as we can see the Chinese language was completely monosyllabic. Originally few monosyllables ended in an open vowel, all the rest ended, as a minimum, in a glottal stop or in a consonant or consonantal cluster, and all were pronounced in a level "tone", in a high or low pitch according to the quality of the initial. As time went on there was a steady tendency, which proceeded at different paces in different dialects, for the consonantal final sounds to be abraded, so that ultimately nearly all the monosyllables ended in open vowels, or diphthongs which evolved when the final consonants disappeared. The final consonants which were most resistant to abrasion were the nasals *n*, *m* (which, however, in some dialects became *ŋ*), and, rather less resistant, *ŋ*. The elimination of some consonants left the level tone unaltered, that of some others produced a rising modulation, that of some others a falling modulation, and the elimination of final plosives an abrupt interruption of sound.

In Tibetan, too, there is evidence for the existence both of pitch and of voice modulation.⁴

With regard to Tangut "tones" it is hard to determine with certainty what the Tangut phoneticians themselves, who had adopted the Chinese terminology with its inherent ambiguities, said on this subject, and in particular whether the Tangut word equivalent to *shang* should be taken to mean "high pitch" or "rising modulation", or indeed both, and whether the word normally used in antithesis to it should be taken to mean "low (*hsia*) pitch" or, as it is usually translated, "level (*p'ing*) tone".

In the discussion which follows I shall use the Chinese terms *shang*, which is conveniently ambiguous, and *p'ing* with the reservation that *hsia* might be more correct, for these two words. We can conveniently start with a key passage in the Preface to the Phonetic Tables, of which we have translations by Nevsky (*op. cit.*, I, 134) and Sofronov (*op. cit.*, p. 81). Unfortunately there are fundamental differences between the two. Nevsky's is as follows: "The five tones of the present written characters are *p'ing*, *shang*, *ch'ü* and *ju*; each of them is expressed by the initial *tzü mu* (that is one of the thirty-six sounds G.L.M.C.). The following passage is not altogether clear to me, but evidently it says that the *p'ing* tone is divided into two, 'pure' and 'impure', with light and heavy pronunciation of a higher (Chinese equivalent *kao* not *shang* G.L.M.C.) and a lower (*hsia*) pitch." One error in this can be disposed of straight away; the word translated "tone" is *shēng* in the sense of "sound"; the reference is not to the tones, of which only four are enumerated, but to "the five sounds" in the broader system of categorizing the "thirty-six (initial) sounds", which was mentioned above. These are (1) Labials; (2) Prepalatals; (3) Postpalatals (Guttural or Velar); (4) Dentals; (5) the eighth category in the ninefold list, with the Liquids thrown in as an extra. Sofronov's translation is as follows: "The five sounds of the language are manifested in such a way that they occur as the initial of every word, be it of the *p'ing*, *shang*, *ch'ü* or *ju* tone; according as they are unvoiced or voiced, aspirated or unaspirated a high or low category of pitch is disclosed."

While at first sight this seems clearly to imply the existence of the four Chinese tones in Tangut, it is at any rate arguable that they were mentioned merely because they were mentioned in the Chinese phonetic treatise which the author used as his model, and that this does not necessarily prove that they really existed in Tangut. What is at any rate certain is that the author was aware of the existence of the two pitches in the language and connected them, quite correctly, with the quality of the initial sound of the word.

It could be argued that since the terms *p'ing* and *shang* are consistently

⁴ Professor Simon kindly refers me to the latest contribution to this subject, the *Grundlagen der Phonetik des Lhasa-Dialektes*, by Eberhardt Richter (Berlin, Akademie-Verlag, 1964).

related to the rhymes, that is final sounds, they must refer to voice modulation and not to pitch, since it is an alteration in the final sound that produces a voice modulation, but this does not necessarily follow. It may merely show that in this respect the Tangut phoneticians followed their Chinese model too slavishly.

What is perhaps more significant is that fact that practically all the references to this subject are to *p'ing* and *shang* only. Sofronov (*op. cit.*, p. 32) says that it is stated that in the first part of the Sea of Characters there were 2,600 characters with *p'ing* rhymes, that the surviving pages of the second part contain 2,107 characters with *shang* rhymes and that the surviving pages of the Mixed Categories contain 448 characters (with either *p'ing* or *shang* rhymes) which are not listed in the Sea of Characters. To these figures he suggests the addition of up to 500 characters for the thirty *shang* rhymes not represented in the surviving pages of the second part of the Sea of Characters and up to 250 for the characters in the lost part of the Mixed Categories. If these five figures (2,600, 2,107, 448, 500, 250) are added together, totalling 5,905, it will be seen not only that there is no room in these two authorities for characters with rhymes other than *p'ing* or *shang* but also that the two estimates are a little too high, since the total exceeds the total number of characters listed in the Homophones, 5,778 plus up to six by a margin of more than 100.

Similarly in the second part of the Phonetic Tables (see Sofronov, *op. cit.*, p. 31, ff.) characters are classified only as having either *p'ing* or *shang* rhymes. As he points out, references to *ch'ü* and *ju* rhymes are exceedingly scanty and occur only in other authorities and not in the Phonetic Tables, Sea of Characters, or "Mixed Categories". According to Nevsky, *op. cit.*, I, 129, rhymes described in the Sea of Characters as *p'ing* are described in the lost Precious Rhymes as *p'ing* and *ju*, and according to Sofronov, *op. cit.*, p. 32, rhymes described in the Sea of Characters and the Mixed Categories as *p'ing* are described in the Anonymous Manuscript Dictionary as *ch'ü*. Whether or not these two authorities are different or, as I have suggested above, the same differently described, it does not look as if more than simple questions of terminology were involved. There are also listed in Sofronov, *op. cit.*, p. 33, some very odd double descriptions in the Anonymous Manuscript Dictionary) like *p'ing p'ing*, *p'ing shang*, *p'ing ch'ü*. It is possible that all the characters concerned are Chinese loan words, in which case the double description might be appropriate. It therefore seems to be reasonably established that, apart from a few minor eccentricities, the only differences in the pronunciation of pure Tangut words which were recognized were between *p'ing* and *shang*, and a twofold categorization of this kind almost inevitably applies to pitch and not to voice modulation, more particularly since the sporadic references to *ju* might be taken to apply to words which still actually ended in a plosive or had lost this final

so recently that they were pronounced with an abrupt interruption of sound.

There is, however, one obvious difficulty about interpreting *p'ing* as indicating a word with a voiced initial and *shang* as indicating a word with an unvoiced initial, and that is the fact that words described elsewhere as *p'ing* and *shang* are listed side by side in the same section of the Homophones. It has been pointed out above that in Chapter III there are nine sections containing words with the 37th rhyme, that is ending in *-e*, and that in addition there is one unique sound which certainly, and two (*tye*, *dye*) which probably, ended with the same sound. It is hard enough as it is to find twelve different Dental initials of words rhyming in *-e*, it becomes plumb impossible if we have to assume that in some of these sections words beginning with *t-* coexisted with words beginning with *d-*. It seems to me, therefore, that we must face the fact that a certain amount of phonetic abrasion had already occurred before the 12th century, so that words otherwise pronounced exactly alike were pronounced with different pitches which the author of the Homophones did not regard as significant and the authors of the other works did. It is not easy to see how this could have come about, but one possible explanation is that words with, say, initial *t-* and a high pitch had always begun with a simple *t-*, while words with initial *t-* and a low pitch had originally begun with a consonantal cluster like *gt-* or *bt-* and the voiced prefix had been abraded leaving its own inherent low pitch as the only memorial of its previous existence.

This explanation may seem fairly far-fetched, but any explanation based on the assumption that *p'ing* and *shang* relate to differences of voice modulation would be equally far-fetched and would run into even greater difficulties. In that event, since it is generally accepted that a rising modulation is caused by the loss of a consonantal final, some phonetic abrasion prior to the 12th century would have to be postulated and it would have to be assumed that *te* (*p'ing*) went back to an earlier *te*, and *te* (*shang*) to a earlier *teX*, *X* in this case representing a consonant subsequently lost. The corollary to this would be that in an earlier stage of the language there must have been a good many more than 105 rhymes. It has already been pointed out that it is hard enough to see how there could have been as many as 105 rhymes; it is plumb impossible to see how there can have been a great many more. As between the two possible explanations the explanation of the difference between *p'ing* and *shang* as a difference of pitch is the more plausible.

Moreover, while it is true that in many of the sections in the Homophones *p'ing* and *shang* words coexist side by side, this does not occur with all the rhymes. The global figure of 105 rhymes is made up of seventy-eight rhymes common to *p'ing* and *shang*, nineteen exclusively *p'ing* and eight exclusively *shang* rhymes. A careful section by section analysis of the

Homophones might make it possible to discover a correlation between sections and unique sounds shown in the Phonetic Tables as containing words beginning with voiced initials and exclusively *p'ing* rhymes, and sections and unique sounds containing words beginning with unvoiced initials and exclusively *shang* rhymes, and even to go further and show that in addition to the mixed sections there are some individual sections which contain words known to begin with either voiced or unvoiced initials and to have either *p'ing* or *shang* rhymes. If such a correlation can be established it would greatly strengthen the case for regarding the difference between *p'ing* and *shang* as one of pitch.

(5) Final Plosives ?

Another useful line of enquiry would be to see whether any evidence, such as the sporadic mention of *ju* tones, can be found which would make it possible to reconstruct a final plosive in some of the 105 rhymes and so to correct the unnatural appearance of the finals as reconstructed by Sofronov with their enormous preponderance of final open vowels. There is, indeed, one specific instance, in which a plausible case for a final plosive can be made out. It was pointed out above that Nevsky, *op. cit.*, I, 39, suggested that the Tangut called themselves something like *Miñag* or *Minyag*, a name confirmed both by more or less contemporary Tibetan authorities and by the nomenclature of the modern Hsifan tribes. Nevsky identified the characters representing this name as the first two of the three characters in footnote 22 in I, 51. The first appears in Chapter I, 17 of the Homophones and has the 30th rhyme (27th *shang*/29th *p'ing*). It cannot have been exactly *mi*, since the rhyme indicated by the transcriptions of the *dhāraṇī*'s as *-i* is the very common 11th rhyme (11th *p'ing*, 10th *shang*), and the normal character for *mi* in these transcriptions is the 6th character in Nevsky, *op. cit.*, I, 306 (Chapter I, 30; 11th *p'ing* rhyme), but it must have been something like it. Sofronov, *op. cit.*, p. 92, reconstructs this rhyme as *-ī*. The second character appears in Chapter III, 171 and has the fairly rare 21st rhyme (21st *p'ing*, 18th *shang*), which Sofronov, *op. cit.*, p. 89, reconstructs as a rather low *iæ*. It seems more reasonable to reconstruct the rhyme as *-ag* and the character as *nyag*.

(6) Summary

The end product of this elaborate collation of all the information in the Tangut works on phonetics enumerated above will be a table showing the approximate pronunciation of the words contained in most of the sections in the Homophones, and a good many unique sounds, with a great many queries and cross-references stating that some of these words must have been pronounced in much the same way as other words, and that the exact difference between them is obscure. In preparing this table the most useful

evidence will be that provided by Tangut transcriptions of foreign sounds. The Chinese and Tibetan transcriptions of Tangut sounds will not add very much, except that it may be possible to reconstruct initial consonantal clusters where Tangut characters are transcribed by two linked Chinese characters in the Pearl on the Palm or by Tibetan transcriptions containing initial consonantal clusters, but this latter evidence is very dubious.

Pre-11th and Post-14th Century Tangut ?

The final question in this part of the field is whether further precision could be introduced by calling in aid comparative philology in one direction or the other. The question whether we can find any specimens of the Tangut language antedating A.D. 1038 is as fascinating as it is tantalizing. Among the numerous documents in Tibetan script which have been recovered from innermost Asia, mainly at Tunhuang which was at one time well within the boundaries of the Tangut "Empire", there are, as well as Tibetan, Turkish and Chinese texts, texts in at least two unidentified languages of Sino-Tibetan type, possibly more. One long one is that published by the late Prof. Thomas in the language called by him Nam. Prof. Thomas associated this language with an area in north-eastern Tibet which was almost certainly at the time (8th or 9th century?) when this manuscript was written inhabited by the Tanghsiang or a closely related tribe. It may well in the end turn out that this, or one of the other "unknown" languages, is actually pre-1038 Tangut, but since the interpretation of these documents is at present no more than an exercise in more or less inspired guessing, a great deal more progress will have to be made with the reconstruction of the meanings as well as the sounds of Tangut words before any such identification becomes plausible. This, however, can be said, that the phonetic structures of Nam and of the language represented by the long text in British Museum MS.Or. 8212 (188) are very much like what the phonetic structure of Tangut is likely to prove to have been, with a considerable apparatus of initial and final consonantal clusters and, at any rate *prima facie*, a fairly limited stock of vowels.

Be that as it may, there can hardly be any doubt that since from many points of view Tangut lies squarely between these "unknown" languages, some two or three centuries older than early Tangut, and the modern Hsifan languages, some five or six centuries younger than late Tangut, it must have had a great deal common with both in the fields of phonetics, vocabulary and grammar. Enough has already been discovered about Tangut grammar to make it clear that it closely resembles the grammar of Tibetan and other Tibeto-Burman languages, but the vocabulary is certainly not identical with that of Tibetan. The latter point is perhaps disappointing but not surprising, considering the wide differences between the vocabularies of Tibetan and the Hsifan languages. Nevsky collected

some material from the latter in his skeleton dictionary. It seems to give later forms of some Tangut words, but the vocabulary material of these languages which has been collected to date is scanty and unsatisfactory, the areas where they are spoken are at present inaccessible to European scholars, and the languages themselves, although they retain some early features like a sprinkling of final plosives, seem generally speaking to be in such an advanced state of phonetic decay that they are unlikely to give any very great help for the reconstruction of Tangut.

It is of course possible that a little greater precision might be obtained by one purely internal form of comparative study. It is possible that a close study of those groups of Tangut words, presumably verbs, which are linked together by common meanings and close phonetic resemblances would make it possible to reconstruct some two- and three-form verbs of the Tibetan type.

What I have stated above is, so far as I can see, a reasonable summary of the present position and future prospects of reconstructing the phonetic structure of the language, but this of course is only a part, and perhaps not a very important part, of the general study of the language.

Lexicography

A more important task is that of compiling a definitive Tangut dictionary. In one respect the prospects are good, since as I have said above, we have got on record to all intents and purposes the complete repertory of its written characters. One point immediately arises, that of arrangement. The Chinese language has been studied by Chinese scholars for many centuries and by foreign scholars for a long time now, and there is an intolerably large number of different systems of arranging the characters in a Chinese dictionary. Tangut studies are still in their infancy, the field is almost virgin and the scholars engaged in it are few. It should surely be possible to avoid such a situation arising in Tangut lexicography by reaching now an international agreement on the way in which the characters should be arranged. Only one Tangut dictionary has so far been published, that of Nevsky, and it is certain that the arrangement of the characters which he devised in his pioneering days is neither sensible nor practical. In this dictionary the characters purport to be arranged, on one Chinese model, under 408 radicals or classifiers (not 353 as stated by Prof. Menges in his review of the book in the *Central Asiatic Journal*, VIII, 1, p. 59). There does not seem to be any indication of the principles which he followed in selecting and arranging these radicals. There is, it is true, some sort of progression, in the sense that the first two have one stroke and the next fourteen two strokes (although if the 4th radical is interpreted as having two strokes, the 14th which has an extra one should have three) and that the last seventy or eighty are pretty complicated, although those that have most

strokes do not in fact come last, but there is no intelligible arrangement of those which lie between. It would be an intolerable burden to have to remember these 408 radicals and their position in the list, and even to look them up in the index is a time-consuming process. But finding an individual character under some radicals is an even more time-consuming process. The characters arranged under the radical which opens the second volume cover 128 pages (in my own dictionary there are 635 such characters but the two lists are not identical) and are arranged in no conceivable kind of order. Even Nevsky himself could not operate his own system effectively and over 200 characters are entered more than once (some, Mr. Grinstead tells me, even three times); for example the first character in I, 340, is the same as the last in I, 347, and the fourth in I, 356, as the last in I, 357. These criticisms are not intended to depreciate the value of this work; it is the result of years of labour and is full of valuable information; but it is better compared to a gold mine rich in precious nuggets than to a finished product of the goldsmith's art.

The only other comprehensive Tangut dictionaries of whose existence I am aware (there may of course be others) are Mr. Grinstead's card index and my own manuscript dictionary. Before I began to write it out in October 1938, by an odd coincidence at about the date of Nevsky's death, I had carried out a detailed analysis of the structure of Tangut characters and compiled a list of the components and sub-components which seem to have been used in building up the individual characters. I had arranged these in a logical order starting with the simplest and most regular and proceeding from them to the more complicated and irregular. My arrangement of the characters was based on the assumption that every character could be regarded either as itself a component (few are), or as having been built up of a series of sub-components and components, and that the pre-determined order of these sub-components and components could be used to determine the order of characters in the dictionary. It may not be the best arrangement which could be devised, but it has at any rate stood the test of time to the extent that, coming back to my Tangut studies after an interval of nearly twenty-seven years, I found no difficulty in locating a character in the dictionary in a very brief space of time. It is not of course in its present state suitable for use as more than a working dictionary; like Nevsky's it is a mine not a finished product. It was necessary, in order to provide a basis for cross-references, to number the characters serially and there are some errors, gaps and double entries, in the numeration; some of the characters missing from the surviving text of the Homophones were later discovered and entered under numbers followed by A or bis. The last number in the book is 5,723, but after taking account of these additions the actual number of characters listed cannot be much less than the true total of 5,778 plus up to 6. Unless some other scholar has devised a better system of

arrangement I suggest that it could be taken as the basis for the definitive Tangut dictionary when more missing characters have been located. It would have of course to be renumbered, I should hope for the last time, since any residual strays could be entered under *bis* numbers. There would also have to be a few double entries, the duplicates of each under A, B etc. numbers, since some characters are written slightly differently in different authorities and, under the system adopted, would be placed in different positions. Mr. Grinstead's card index is arranged on a similar principle, but starting at the right-hand bottom corner of the character instead of the left. It might be sensible in the definitive dictionary to arrange the characters in the text on one principle and add an appendix containing the characters arranged on the other principle with cross-references. This would give a double chance of locating badly written, incomplete and cursive characters which are difficult to analyse.

An agreed numeration is no less essential than an agreed arrangement. The characters are so complicated and the total surviving Tangut material relatively so small that there will never be any question of having a fount of Tangut type cut. In any case imagination boggles at the thought of teaching typesetters to set it up. The characters in the definitive Tangut dictionary will have to be hand-written and the remaining material typewritten or printed and the whole then reproduced by some suitable method. Other Tangut texts will have to be published in reproduction of the original or hand copies, and questions will undoubtedly arise regarding the precise identity of some characters. The sensible way to discuss such problems in the notes is to quote the characters concerned by their serial numbers.

The other aspect of compiling the definitive dictionary is of course that of fixing the meaning or meanings of each character. This is simply a matter of sheer hard work. A great many meanings are already known and recorded, more in Nevsky's dictionary than in mine, either because they are given in the Pearl on the Palm, because they have been deduced from Tangut translations of known texts like the Buddhist sutras and Chinese classics and other works, or because they can be worked out from the explanations given in the Homophones, the Sea of Characters and the Mixed Categories. At the end of the road the dictionary will probably be nearly complete, and a fit instrument for trying to link Tangut up with one of the earlier "unknown" languages and some of the modern Hsifan languages, but there will probably still be a few characters, especially unique sounds, which will elude the investigators.

Other Tangut Texts

The other remaining task is that of editing and translating the surviving remains of Tangut literature. This too is simply a question of hard work. Some of the literature, for example the translations of Buddhist

sutras, will not be worth this elaborate treatment; the original texts are well known and the Tangut versions are of little interest except as hunting grounds for the lexicographer, but Nevsky and others have already shown that there is some native literature which may be of great interest, and more is known to exist in Kozlov's collections, in the British Museum and perhaps elsewhere.

Conclusion

Perhaps I might close on a personal note. So far as Tangut studies are concerned this is, I hope, my swan song. When I embarked upon them over thirty years ago I regarded them as a fascinating exercise in cryptography, with a little mathematics thrown in. But I very soon found that a profound study of Tangut was impossible without, if not a profound, at any rate a good knowledge of Chinese, Tibetan and, if possible, some Hsifan languages. To have described my knowledge of any of these languages as better than inadequate would have been gross flattery, and so I retired from the field, I thought for ever, and returned to the Turkish studies on which my interests are now concentrated. I have returned momentarily to this subject only because I am perhaps the only survivor from an earlier and less scientific age, and felt it my duty to put my own experience at the disposal of the new and vigorous generation of young students of this fascinating language before it is too late. I am greatly indebted to my old friends, Prof. Walter Simon and Prof. Edwin Pulleyblank, for their help in compiling this paper.