

## 1. INTRODUCTION

This thesis is a regionally oriented study which aims at mapping the various kinds of formally repetitive constructions found in descriptions of a set of East-Asian and Southeast Asian languages. These constructions turned out to be of three kinds: (i) derivations of explicitly stated productive repetitive<sup>1</sup> processes, (ii) repetitive items presumably representing outputs of such processes and (iii) compounded lexical entities containing repetitive elements. The items under (ii) are probably of various degrees of lexicalisation, from items where the input form of the derivative process has, for some reason, been left out in a source, to more lexicalised entities. According to one interpretation, the term 'lexicon' refers to a list of items that have been stored in the memory of a native speaker and are not produced by means of general linguistic rules<sup>2</sup>. Consequently, in addition to repetitive compounds, those repetitive forms in the languages of the corpus which are not explicitly mentioned as being derived from a non-repetitive form in the sources, are regarded as probable entries on such a list in a given speech form. The distinction between words and phrases among the exponents of the set of repetitive patterns in the data, a subject with a long and controversial history in Chinese linguistics, for example<sup>3</sup>, is not considered to be an issue of primary importance in this study.

A general delimitation of the subject matter of the study is adapted from Moravcsik (1978: 300) and is the following:

Form 1: ... A ... = ... X ...

Form 2: ... B ... = ... Y ...

---

<sup>1</sup> Since the term 'reduplication/-ive' is rather infelicitous, as it refers to an already repeated entity, which is doubled again, we prefer to use the term 'repetition/-ive' or 'duplication/-ive', or even 'triplication/-ive' if needed, instead. I thank my colleague Morteza Elmolhoda for bringing to my attention the mismatch between the commonly accepted meaning of the term and its actual reference.

<sup>2</sup> Katamba (1993: 297) formulates the probability of an entity of entering the lexicon in the following way: the less compositional a form becomes, the greater is the likelihood of its being listed. This means that all morphemes must be listed since they are the smallest non-decomposable semantic entities in the language, and most complex words containing several morphemes are listed, except the semantically quite transparent ones. Compounds, on the other hand, are not so easily memorised and only those of syntactic phrases, which are idiosyncratic in meaning, must be listed. Evidently, most sentences need not be listed as they are composed of other units according to general rules.

<sup>3</sup> Quite recently it has been proposed that concepts of phonological, morphological and syntactic words, found to be valid for other languages, have validity for Chinese, too (Dai 1998: 103-134), and that converging metrical and tonal evidence seems to indicate a clear

where A and B stand for semantic representations of linguistic forms with some elements in common, while X and Y are their formal representations whose relation is such that Y includes a part or all of X repeated  $n$  times. If translated into Moravcsik's terms, the data in the study is of the kind that Y as a repetitive form does not necessarily have a corresponding X. A more specific delimitation of the subject matter restricts the data to consecutive repetitive forms unless a disjunctive form is specifically labelled by a source as repetitive.

Geographically, the corpus consists of languages spoken mainly within the territorial limits of the People's Republic of China, the Republic of China (Taiwan) and Vietnam, the most notable exception being Burmese.

In genetic terms, the languages contained in the data of the study comprise representatives from the Sino-Tibetan, Tai and Miao-Yao language families as well as the Mon-Khmer branch of the Austro-Asiatic family, with Sinitic speech forms and Vietnamese providing most of the items in the total corpus. A broader diachronic view relates the Tai and Miao-Yao groups to the Sino-Tibetan family, the former being also connected with the Austronesian stock, with which the Austro-Asiatic languages have been linked in particular. The genetic classification in this part of the world has not been settled, as Karen, for example, is treated either as one of the primary divisions of Tibeto-Burman or is regarded as standing outside Tibeto-Burman proper, so that Tibeto-Burman and Karen form a higher-order taxonomic unit labelled Tibeto-Karen. In this view, Sinitic and Tibeto-Karen constitute the two primary branches of Sino-Tibetan. The final classification of Karen is, however, still uncertain (Matisoff 1995: 41). The membership of the Austro-Asiatic family and its subdivisions is not uncontroversial, either (Crystal 1997: 311).

However, disputes concerning genetic classification of the languages at hand is not the main issue and will not be discussed in detail in the present work, though the diachronic aspect concerning repetition will be touched on in a separate chapter. The introduction of terms connected with genetic studies as titles of chapters does

---

distinction between words and phrases (San 1998: 135-196). As the authors' conclusions are based mainly on investigations of Standard Chinese, it would be methodologically questionable to extend them analogically to concern the many other Sinitic speech forms in the corpus without the support of specific studies with confirmed parallel effects. I think that more care should be taken in applying Chao's (1970: viii) rather blunt assumption that "in terms of grammar, most of what is said... about Mandarin is true of all Chinese", in spite of its being shared by many researchers. Yue-Hashimoto (1995: 3) notes how findings in Yue syntax in recent years have gradually eroded the credibility of this traditional belief in diminishing diversity concerning all Chinese speech forms when moving from phonetics/phonology via vocabulary to grammar. A similar conclusion was reached earlier by Y. C. Li (1986: 394), who in his article on distinct grammatical features in Taiwanese stated that "although they (Chinese dialects) share a common ancestor, a 'universal' writing system, and also the claim of a homogeneous grammar, it is now time to call a spade a spade – especially since their grammatical systems can no longer be maintained to exhibit uniformity". Thus it is not improbable that discoveries in other dialectal areas will bring to light further elements that are in disagreement with the assumption of the overall unity of Sinitic grammar.

not constitute an attempt at validating a specific historical hypothesis, but it simply aims at clarity in that a set of terms familiar to everyone involved in the study of these languages is utilized.

Typologically, the languages of the corpus can be roughly divided into two groups: the Sinitic, Tibeto-Burman, Tai, Miao-Yao languages and Vietnamese, genetically a Mon-Khmer language, tend to be monosyllabic and tonal, while other Mon-Khmer languages are more typically disyllabic and non-tonal.

The term 'formal' is used in the specific sense of referring to the constituent structure of repetitive forms, analysed in terms of a general formal notation, introduced in the next chapter. The description of semantic properties, besides using a tripartite framework and intratextual comments, aims at suggesting a rationally acceptable final set of meaningful categories induced from the sources used.

Finally, a major classification of the repetitive entities as defined above concerns their treatment, under separate chapters, as exponents of nouns, adjectives and verbs, with a chapter on pronouns and measure words as well. Onomatopoeic items have been excluded. The transcription applied in the sources has in most cases been retained. For clarity's sake, original non-spaced syllabic strings have often been provided with spaces, due to the centrality of the syllable as a functional unit.

The metalinguistic lexicon used in the study is based on traditional linguistic terminology. A conservative attitude in this respect is recommendable in the sense that if the terms contained therein have already proved appropriate and useful in earlier descriptions, it is probable that they will behave similarly in other descriptions. Furthermore, the use of terms familiar to participants in linguistic discourse is naturally liable to expedite effective communication.