

The Verbal Entries and Their Description in a Grammatical Information-Dictionary of Contemporary Tibetan

Jiang Di, Long Congjun, and Zhang Jichuan

Dept. of Phonetics and Computational Linguistics, Institute of Ethnology & Anthropology,
Chinese Academy of Social Sciences, Beijing, 100081
jiangdi@cass.org.cn

Abstract. This paper discusses verb information items in the grammatical knowledge dictionary of Tibetan language. The verb information items include morphology, word-formation and syntax, in which the core is the verb classification with syntactic and semantic features. So the influences on clause structures of 12 types of verbs have been discussed. Also the paper designs an information table to compile the verb items and their related knowledge, in which most of the details of verbs have been described. At last some newly-found phenomena have been discussed and the paper therefore proposes that there are still some special tokens or constructs need to be mined out in Tibetan.

1 The Background of Constructing a Grammatical Knowledge Dictionary

Segmentation is one of the important tasks of text-processing in Tibetan. In order to gain satisfactory results we have proposed a suggestion, in which most of chunks can be identified with markers of chunks, and words can be segmented within chunks. Although the proposal achieved well in our sample of small scale data,[1] yet the wrong match and ambiguous phenomenon still obstruct our future work in the real texts. The reasons, we think, are in two aspects. One is lack of a perfect Tibetan grammatical system which is now still not constructed for information processing. Another is deficient in useful resources, especially a Tibetan knowledge dictionary of high quality with morphological and syntactic information.

A grammatical knowledge dictionary means a base with some essential information of the morphological and syntactic grammar, including pragmatic and rhetorical information as well. With such a dictionary, most phenomenon of mistaken word matching and recognizing, to some extent, may perhaps be eliminated during the word-segmenting. As to Tibetan, there are three main types of lexical information. Firstly, the lexical item itself is a kind of grammatical functional markers, such as, the genitive markers; secondly, the lexical item serves as a formal marker of some grammatical sections, such as, in *sku gdung* “Buddha’s bone”, “sku” acted as the prefix morpheme of the honorific words can construct amounts of honorific vocabulary;[2] last, the grammatical function of words may be ascertained after being analyzed, such as the verb *yid ches byed* “trust” belongs to the cognition verbs, which

specify case markers for their subjects and objects. Nevertheless, no matter which information types, they all need previously be tagged as information attributives in a grammatical knowledge dictionary.

It should be pointed out, all the above information of lexical types may regard as the principle to construct a grammar knowledge dictionary. No matter how widely an item is used or how many its function includes, each item is equal in information. That is to say, equal attention should be paid to every lexical item with filling its information contents in every field of the dictionary, no matter much or few phenomenon of items involve.

Finally, the source of information attributives of Tibetan grammatical dictionary should be mentioned. Generally speaking, all the information of lexical items stem from language material or real texts. But the information contents and information attributives of each lexical item should be ascertained through the Tibetan grammatical system. So, we will take verbs as examples and discuss information contents of lexical items in our Tibetan grammatical system, which we call GIDCT (a Grammatical Information-Dictionary of Contemporary Tibetan).

2 Grammatical Features of Verbs

The most important features of Tibetan verbs embodied three dimensions, the morphology, word-formation and syntax. These three dimensions influence each other and condition each other. So the relationship and the forms of grammatical representation among them need to be ascertained gradually.

2.1 Morphological Features

The morphological features only refer to the forms of tense and mood remained from the ancient original forms of Tibetan monosyllabic verbs. That is to say, each monosyllabic verb may have four forms, the present tense form (also named proto-form), future tense form, past tense form and imperative mood form. However, the paradigm of Tibetan monosyllabic verbs, by no means, only adds the affixes or exhibits the regular inflection to construct. From the following examples, we can clearly comprehend the morphological conception of Tibetan verbs in table 1.

Obviously, the paradigm is of irregularity, as is the irregular change of verbs in English. It is quite difficult to analogize for this reason. But on the other hand, there are amount of Tibetan verbs which never change their forms. Related with the syntax, those sentences with prospective aspect markers always use the future tense forms of verbs, and sentences with realis markers always use past tense forms. That means morphology and syntax restrictively interacts each others. In addition, according to H.A.Jäschke's views[3], the present and the future forms always appear together with the negative particle "mi", and the past and imperative forms with negative particle "ma", which also have influences on the orders words. From the above, we know that the morphological features of Tibetan monosyllabic verbs still remain some restrictions on the syntactic function in the contemporary Tibetan, though Tibetan verbal morphemes are the remains of ancient times.

Table 1. Tibetan verbs with inflection forms

states of changes	present	future	past	imperative	
no change for the whole verb	bstod	bstod	bstod	bstod	acclaim
change by affixing and vowel alternation	lta	blta	bltas	ltos	see
change by affixing	srung	bsrung	bsrungs	srungs	abide by
alternation of consonant forms of root	vtshong	btsong	btsongs	tshong	sold
alternation of consonant forms and affixes	vchad	bshad	bshad	shod	say
alternation of affixes and vowels	gcog	gcag	bcag	chog	break off

2.2 Word-Formation Features

Most of the verbs in contemporary Tibetan are polysyllabic verbs, which appears in high frequency, and their meanings trend to be complicated. Tibetan polysyllabic verbs have stronger regularity in word-formation. They consist of nouns and verbs or adjectives and verbs. Among them, most of the verbal morphemes in compounds are common verbs without substantive sense. Please observe the following words in table 2.

Table 2. Compound Verbs in Tibetan

verbal morpheme	noun-verb compound		noun-verb compound		adjective-verb compound	
gtong	jus gtong	make ideas	bstod ra gtong	praise	byams po byed	love
rgyag	glo rgyag	cough	vphya las rgyag	sneer at	yag po byed	be kind
byed	dpang byed	testify	lta rnogs byed	look after	vtshub po byed	act widely
shor	khrag shor	bleed	rgyab vdre shor	fight	dgav po byed	like
vdebs	khrid vdebs	explain	go nor thebs	misunderstand		
gnang	bzo ba gnang	make	thugs khral gnang	be worry about		

One of the most important features of compound verbs is formal markers, namely, the verbal morphemes within words. Because of the productivity of compound verbs, dictionaries always fall behind. So it is possible to resolve the problem of recognition of verbs if verbal morphemes be regarded as morphological markers and rules of compound word-formation be built up in compounds.

Taking *ngal gso rgyag* “relax” as an example, it is always viewed as a phrase, so the compound word and the likes can not enter into *A Tibetan-Chinese Dictionary*[4]. In this case, the segmentation of the word may be *ngal-gsol rgyag* with the dictionary, in which the former is a noun with meaning “relaxation”, the latter is a monosyllabic verb with not-thorough-going meaning “doing, making”. The real reason led to such a grammatical conception is that many compound verbs can be inserted some other grammatical constituents, such as *ngal-gso yag-po rgyag* “take a good rest”, *ngal-gso tog-tsham rgyag* “take a short break”. In our opinion, Tibetan polysyllabic compound verbs are still in progress and not yet fall into a pattern in mind. Anyway, when

engaged in computer processing, we can not give up an undertaking on account of a small obstacle. So while we put compound verbs in the dictionary we make some other rules to process the verbal phrases with inserted constituents.

Quite few are another type of common disyllabic compound verbs in Tibetan, which are non-productive in word-formation. For instance, *ha go* “understand”, *kha nor* “make an indiscreet remark”, *ngo tsha* “be shy”, *blo’gel* “trust in”, *khas’che* “promise”, *ha las* “amaze”, *ya mtshan* “be strange at”. Those words were formed from phrases in history and are unmarked forms. So we can put them all into the dictionary directly without morphological analysis.

2.3 Syntactic Features

The Tibetan verb syntactic features cover many categories. Such as, whether or not they take objects, what kind of objects they take, which types of case markers they require to subject and object constituents, and so on. On the other hand, The attributions of verbs themselves are very important, which determine the quantity of arguments and the frame of sentences, such as attribution of human verbs, of volition verbs, of controllable verbs, causative verbs, and so on. In addition, some other syntactic features are also of importance. e.g. the formal markers of normalization for each verb, and the position relationships between verbs and adverbial modifiers and so on.

In our project, an effective way to analysis Tibetan verb syntax has been found that is classification for verbs according to the types of syntactic features and semantic features of verbs. With this method we can describe those grammatical information for each verb and compile verbs into the knowledge dictionary in different types. The following are some description of 12 types of verbs.[5]

Verbs of possession: This type of verbs indicates those which contain the meaning “to possess, to gain, to come being to, to form” and so on. For example, *yod /dug* “have”, *vtob* “gain”, *byung/ yong* “take place”, *rag* “obtain”, *’dzams* “belong to”, *’chor/ shor* “appear, emerge” and so on. The basic format of verbs of possession is “NP+[POS]+NP+V_(VOP)”, which requires that the subjects take possessive case markers, and that the objects take absolutive case marker (zero form). So the syntactic information items of verbs of possession are like these {V_(VOP), NP_(S)+POS, NP_(O)+Ø}. Also, most of the verbs of possession possess the characteristic of {HMN}(subjects are human being).

Existential verb: This type of verbs expresses the meaning of someone or something being somewhere. For instance, *yod/ ’dug/ yog red* “exist, be”. Existential verbs demand the subjects take absolutive case marker, and the objects take locative case marker. The basic syntactic format of existential verbs is “NP+NP+[LOC]+VP_(EXI)”. Existential verbs and verbs of possession do not obligatorily demand the characteristics of {HMN} and {CTR}(subjects can control actions) for subjects. The grammatical information items of existential verbs include {V_(EXI), NP+LOC}.

Verbs of change: The verbs of change indicate those which take resultant complement constituents. Such as *sprul* “change”, *sgyur* “make to change”, *’gyur* “turn to”, *’gro* “change”, *’dul* “domesticate” and so on. The basic format of the verbs of change is “NP+VP_(adj/ v)+{COP}+VP_(CHA)”. That means verbs of change require complements of adjectives or verbal phrases which express the results of verb

changes and there is a complement particle between complement constituents(COP) and verbs. So in the information dictionary, the characteristics of {V_(CHA), VP+COP} should be described.

Perception verbs: This type of verbs indicates responses and feelings human being perceives from things or events by sight, by hearing and by aesthesia. Such as, *mt hong* “see”, *go* “hear”, *ha go* “understand”, *dran* “remember”, *brjed* “forget”, *ngo shes* “recognize” and so on. Perception verbs possess the characteristic of {HMN} and {INV}(non-volition). Except of a few words such as *brjed* “forget”, the action outcome of the perception verbs can not be controlled by human being with consciousness. So they are non-controllable verbs. The basic format of the perception verbs is “NP+[AG]+NP+VP_(PER)”. That means the subject may take an agentive case marker. So the information items in the dictionary are {V_(PER), NP_(S)+AG, HMN, UCT(=non-control)}, but the information items for verbs like *brjed* “forget” are {V_(PER), NP_(S)+AG, HMN, CTR}.

Directional verbs: This type of verbs indicates those which express moving action concepts or staying status concepts. The common verbs are *'gro* “go”, *phyin* “go”, *yong* “come”, *phebs* “come”, *bsdad* “stay”, *bzhag* “lay up” and so on. The directional verbs can construct directional-predicate structure: “NP+VP_(DIR)”, or construct directional complement: NP+VP+VP_(DIR), in which the verbs are called directional constituents, or construct serial verbal phrases: NP+VP+[TAP]+VP_(DIR). The information items of directional verbs in the dictionary are {V_(DIR), VP+V_(DIR), VP+TAP+VP_(DIR)}. There are no obligatory demands on the characteristics of {HMN}, {CTR}, or {VOL} for subjects in directional verb structures.

Cognition verbs: This type of verbs indicates those which express non-action psychological activity. The common verbs are *brtsi 'jog byed* “respect”, *dga' zhen byed* “like”, *ngo tsha* “be shy”, *sems shor* “be infatuated with”, *mt hong tchung byed* “belitter”, *ngo rgol byed* “be object to”, *dogs* “be suspicious of” and so on. The basic syntactic structure of the cognition verbs is NP+[AG]+NP+[OBJ]+VP_(COG), namely the subjects can take agentive case markers, and objects can take objective case markers. The information items of cognition verbs is {V_(COG), NP_(S)+AG, NP_(O)+OBJ} in the dictionary.

Narrate verbs: The narrate verbs indicate the verbs expressing the meaning of citation and thinking in mind. Such as *bshad* “say”, *zer* “speak”, *'dri* “ask”, *lan slog* “answer”, *bsam* “think”, *dgongs* “think over”, *shes* “know” and so on. The narrate verbs require clausal objects or objects of normalization phrases, the basic structures are “NP+[AG]+[NP+VP]+VP_(NAR)”, or [NP+VP]+NP+[AG]+VP_(NAR), in which the clausal objects of citation verbs unusually attaches clause particle “se”. The information items of narrate verbs in the dictionary include {V_(NAR)}.

Interrelation verbs: This type of verbs indicates those which interconnect two or more things or events on logic, or social rules, or common senses, which correspond or subordinate each other. For instance, *'dre* “mix up”, *'gal* “violate”, *mt hun* “be in consensus”, *kha bral* “separate”, *stun* “fit”, *kha thug rgyag* “meet”. The syntactic format of the interrelation verbs is NP+NP+[ITP]+VP_(REL), namely the interrelation verbs need attach an interrelation case marker “*dang*” to the nouns related. Therefore the information items of interrelation verbs should include {V_(REL), NP+ITP}.

Causative verbs: The causative verbs include the meaning of “cause to do”, such as *bcug* “let to do”, *byed* “let to do, make sth. appear”, *bzo* “make sb. do” and so on.

Generally, the subjects of causative verbs attach agentive case marker to subjects, and sometimes possessive case marker to subjects in object-clause. Usually, there is a causative particle “ru” between the object-clause and the causative verbs. The basic syntactic structure of causative verbs are “NP+[AG]+NP+[POS]+(NP+)VP+[CAU]+VP_(CAV)”. The information items of causative verbs in the dictionary are {V_(CAV), NP_(S)+AG, NP_(PCC)+POS, PCC+CAV}.

Stative verbs: This type of verbs indicates those which express the state of actions. Such as *shi* “die”, *bzi* “drunk”, *snyun* “be ill”, *yal* “disappear” and so on. The basic syntactic structure of the stative verbs is “NP+(NP+) VP_(STA)”, namely the subject and object should be with absolutive case marker. The information items of the stative verbs in the dictionary are {V_(STA)}.

Action verbs: Action verbs include several types. Some express the meaning of rendering and take two objects, some express the meaning of objects changing and take a result object. The objects’ types of action verbs include accusative objects, indirect objects, resultative objects. The subject of action verbs are often attached an agentive case marker. Action verbs can be divided into transitive and intransitive verbs. The basic syntactic structure of intransitive action verb is “NP+VP_(ACT)”, and that of transitive is “NP+NP+VP_(ACT)”, or that of bi-objects is “NP+NP+[DAT]+NP+VP”, or that of resultative objects is “NP+[AG]+NP+NP+[FAT]+VP_(ACT)”.

Copula: Copula denotes verbs *yin* / *red* “be” which express the classes of grammatical subjects. The basic structure of copula is “NP+NP+VP_(COU)”, namely both of the subject and the object use absolutive case marker.[6]

As space is limited, the above description about verbal syntax and semantics is quite simple. We believe more and more grammatical information phenomena will be mined out and be collected into grammatical information dictionary one by one in our future work.

3 Descriptions of the Information Items Related to Verbs and Some Samples

In this section, we will describe the information items of verbs which are in our GIDCT. For the sake of a clear statement, we will illustrate some representative samples as well.

<Number>, it denotes the order location of a word entry in GIDCT.

<Word form>, it denotes Tibetan characters (orthography) of written words, e.g.
ལྷུང་ལངས་

<Transliteration of Word Form>, it denotes transliteration forms of Tibetan words with roman letters according to a scheme designed. E.g. *lung lang*.

<Lhasa pronunciation>, it denotes the phonetic transcription of words according to Lhasa pronunciation. E.g. *lungf langw* or *luŋ⁵⁵ laŋ¹³²*.

<Meaning_(Chinese)>, it denotes the basic meaning of entries with Chinese or other languages, this sample is “anger, rage”.

<Homograph>, it denotes the homographs of entries, namely dividing a polysemous word into different words, then giving their other meanings here. E.g. “blow”.

<Variant>, it provides a different entry forms in spelling. There is no variant for this sample.

<Syllabic Numbers of Entries>, There are about 1300 monosyllabic verbs, and about 100 disyllabic verbs in Tibetan, which are classic forms. A number of modern Tibetan verbs are polysyllabic with a noun and a monosyllabic verb, which shows the value and importance of the syllabic numbers in computer processing. This sample is a two syllabic verb.

<Types of word-formation>, it let choose one item from the four: mono-morphemic words, non-productive compounds, productive compounds, reduplicative compounds. Mono-morphemic verbs are monosyllabic verbs; non-productive compound verbs are classic disyllables, both of which are unmarked. Compound verbs are polysyllabic verbs with verbs serving as markers. Reduplicative compound verbs are too complicated, which have been discussed in another paper.[7] For this sample, word-formation type is a non-productive compound.

<Word-formation structure>, it let choose one from types of subject-predicate, predicate-object (verb-object), modifier-head, complement-verb, and co-ordination constituents. Different structures of word-formation of compound verbs have effects on syntactic structures in processing, as in the case of compound verbs with accusative objects. In the format of “noun +noun+ (monosyllabic)verb” (namely: N1+N2+V), the “N1” may be an object of compound verb “N2+V”, or a part of the compound verb. On the other, the difference of the inner structure of compound verbs may cause different object types. Such as, the subject-predicate compound verbs usually require accusative objects with zero marker, and verbs of predicate-object type generally require objective objects with objective marker. Therefore, the inner structural relationship of compound verbs may produce influence on the objects and their types.[8] This sample is a non-productive compound verb.

<Word-formation productivity>, it denotes the ability of monosyllabic verbs serving as verbal morpheme in word-formation. The chosen items are strong one, weak one, and null. E.g., classic disyllabic compound verbs possess no capability of combining new-type compound verbs, which are null. The sample is null.

<Classic Tense form>, it denotes the forms of classic monosyllabic verbs, some of which show different forms survived from old time. One of the four forms can be chosen from present form, past form, future form, and imperative form. The sample is present form.

<Formative structure>, it denotes the reduplicative form of verbs only. The sample is not reduplicative form.

<Root>, it let write down directly the transliteration forms of verb root or stem. The sample is “rlung”.

<Meaning of verb root 1>, it let write down directly the meaning in Chinese or other language. The sample is “air, wind”.

<Meaning of verb root 2>, it let write down directly the meaning in Chinese or other language. The sample is null.

<Word-formation or formative morpheme>, it let write down the transliteration form of word-formation morpheme. The sample is “langs”.

<Meaning of word-formation or formative morpheme>, it let write down the meaning of the word-formation or formative morpheme. The sample is “happen”.

<Honorific form>, it let write down the honorific forms of some verbs. The sample is null.

<Parts of Speech>, it defines parts of speech of every verb entry, in which there are two main types: transitive and intransitive. The sample is transitive one.

<Verbal Types of syntax and semantics>, it defines the syntactic and semantic types of verbs, and choose one from following: stative verb, action verb, cognition verb, perception verb, verb of change, directional verb, narrate verb, copula, verb of possession, existential verb, interrelation verb, causative verb. The sample is cognition verb.

<Person>, it denotes the agreement between person and copula verbs or existential verbs. The sample is null.

<HMN/UMN>, it denotes whether a verb is a human verb or not. The sample is HMN.

<CTR/NTR>, it denotes whether a verb is a controllable verb or not. The sample is CTR.

<VOL/NOL>, it denotes whether a verb is a volitional verb or non-volitional verb. The sample is volitional verb.

<CAV/NAV>, it denotes whether a verb is a causative or non-causative verb. The sample is NAV.

<Subject case type>, it denotes what type of case markers the verb requires for a subject (including absolutive /agentive/ possessive). The sample is agentive (AG).

<Object case type>, it denotes what type of case markers the verb requires for object (including accusative / objective / resultative). The sample is objective one.

<Indirect object case type>, it denotes what type of case markers the verb requires for indirect objects. The sample is null.

<Cases for noun phrases>, it denotes what type of case markers the verb demands for specific noun phrases (including locative, ablative, allative, instrumental, comparative, exclusive, genitive, factitive and so on).

<Argument value>, it denotes the quantity of arguments the verb refers to. The sample is 2.

<Negative form and the position>, it let choose one from the two negative forms (mi, ma), and ascertain the distributing position in verb phrases, before verbs, or following verbs, or between syllables of verbs. The sample is between two syllables.

<Normalization marker>, it let write down the form of normalization markers of the verb. The sample is “mkhan, dus”.

As to the examples, it is better to collect one or more phrases or clauses as examples from real texts, which can illustrate morphological or syntactic phenomenon.

<Example>, sku rlung ma gngang “Don’t be angry.” (honorific)
pha-ma-s nga-r rlung-langs-song. “My parents are angry with me.”
parents-AG me-OBJ angry-PEF(perfect aspect)

The following are two verbal examples. “#” denotes this item is null, and “√” denotes the function may exists.

Sample 1 <No. xxxx >

<word form>	ངལ ' ག སོ ' རྒྱ ག	<transliteration>	ngal-gso-rgyag
<Lhasa pronunciation>	ngaevsof gyaw	<meaning>	rest
<homograph>	#	<variant>	#
<syllabic number>	3	<types of word-formation >	compound
<word-formation structure>	VO	<word-formation productivity>	strong
<tense form>	present tense	<formative structure>	#
<root>	ngal gso	<meaning of morph 1>	tired , exhausted
<meaning of morph 2>	#	<honorific form >	#
<word-formation/formative morpheme>	rgyag	<meaning of word-formation/formative morpheme>	recuperate
<parts of speech >	vi	<type of verbs>	action
<person>	#	<HMN/UMN>	Y
<CTR/NTR>	Y	<VOL/NOL>	Y
<CAV/NAV>	N	<subject case type>	#
<object case type>	#	<indirect case type>	#
<case for noun phrases>		<argument vaule>	1
<negative form>	mi	<negative position>	postposition
<normalization marker >	#	<as complement>	#
<corresponding prototype>		other features	
<example>: ngal-gso yag-po rgyag “have a good rest”;			
ngal gso tog tsam brgyab “have a short break”			
nga da-lta	ngal-gso-rgyag gi-yod.	nad-pa-tsho	ngal-gso yag-po rgyag-ru bcug dgos.
I now rest	DUR	Patient-PL rest	good V-CAV cause AUX
I am resting now.		Let the patients have a good rest.	

Sample 2 <No. xxxx>

<word form>	བ རྒྱ ག	<transliteration>	bcug
<Lhasa pronunciation>	juh	<meaning>	make, allow
<homograph>	#	<variant>	√
<syllabic number>	1	<types of word-formation >	mono-
<word-formation structure>	root/etyma	<word-formation productivity>	null
<tense form>	past tense	<formative structure>	null
<root>	bcug	<meaning of morph 1>	make, allow
<meaning of morph2>	Insert, install	<honorific form >	bcug gnang
<word-formation/formative morpheme>	null	<meaning of word-formation/formative morpheme>	null
<parts of speech >	vt	<type of verbs>	CAV
<person>	#	<HMN/UMN>	Y
<CTR/NTR>	Y	<VOL/NOL>	Y
<CAV/NAV>	Y	<subject case type>	Agentive

<object case type>	Cl- O+CAV	<indirect case type>	#
<case for noun phrases>	√	<argument vaule>	3
<negative form>	ma	<negative position>	preposition
<normalization marker >	#	<as complement>	#
<corresponding prototype>	'jug	other features	
<Syntactic example>			

rgan-lags kyis nga-tsho-r sbyong-tshan bri-ru bcug-gnang-byung.
teacher AG us-PL-POS homework write-CAU make-HON-PEF

The teacher makes us write /do our homework.

khyed-rang-gis nga mi mang-po 'di-'dras dkyil-la ngo-tsha-ru bjug.
you-AG me people many so among-LOC shame-CAU make
You let me lose face in front of so many people.

4 Conclusions and the Problems to Be Solved

The data in this project is in construction. Although the data bulk is still not large, some special problems of data has appeared which are somewhat difficult to enter GIDCT base. For instance, *ngal rtsol byed gi yod mdog mdog byed* “to pretend to work”, in which *mdog* originally means “color”, after being reduplicated, it means “false appearance”. But it can not be used alone, it is the same with the verb form of *mdog mdog byed* which is constructed by reduplicating form plus verb form. And its object is always a clause (e.g. *grwa pa yin mdog mdog byed* “pretend to be a lama”) or a verb phrase (e.g. *mi shes pa shes mdog mdog byed* “pretend to know when one does not know”). There is not any markers between the verb and clause-objects or verb-phrase-objects. Obviously, when processing this kind of forms, we should add some new information items of morphology to the dictionary, such as verbal bound forms or non-absolute verb phrase structures. In addition, some scattered phenomenon can not be overlooked, because as information items of the dictionary, they are of the same value.

Professor Yu Shiwen has made very careful annotation for Chinese word items in *The Grammatical Knowledge-base of Contemporary Chinese*. [9] By contrast, GIDCT is impossible to reach such depth for the time being. There are many reasons, but the essential one is still lack of extensive and thorough researches on grammatical infrastructures. On the other hand, Tibetan language information processing is quite different from Chinese, it is not suitable to make a dictionary by copying the description method of Chinese grammar simply. Take Chinese resultative verbs for an example, this type of verbs always extend to syntactic resultative phrases, such as *记住* (remember-fixed) “remember” to *记得住* (remember-able-fixed) “able to remember”. Yet the type of verbs only contains some of Chinese verbs, so it is worth to list their information in the dictionary. For Tibetan polysyllabic compound verbs, nearly all of them can accept modifiers or complements between nominal and verbal morphemes, it is difficult to compile them or describe them in the dictionary.

In true texts, there are close connection among verbs and other constituents, especially predicate endings, including aspect markers, evidential markers and mood markers. But these information may list separately in GIDCT, which is quite different from Yu’s GKBC, in which Chinese “着 *zhuo*, 了 *le*, 过 *guo*” are listed as verbs’ information items. [10] Also, some other information, such as information of verbs

modified by adverbs and of the relations with auxiliary verbs, etc, have not yet been described in GIDCT temporarily. All in all, it will be an important task to mine information as much as possible in Tibetan grammatical information dictionary construction.

Now the infrastructure of languages and construction of electronic dictionaries has been attached more and more great importance to, including classifying syntax dictionary and semantic dictionary.[11][12] The GIDCT is also a good beginning to Tibetan natural language processing.

Acknowledgement

This research is supported by NSFC of China. No. 60173024, 60473135.

References

1. Jiang Di: The Method and Process of the Definition to Grammatical Chunks in Modern Tibetan. *Minority Languages of China (Minzu Yuwen)*, Vol.3, (2003) 30-39
2. Zhou Jiwen, Xie Houfang: *A Grammar of Lha Sa Dialect of Tibetan Language*. Beijing: Minzu Press (2003)
3. Jäschke, H. A. (Heinrich August, 1817-1883): *Tibetan grammar*. addenda by A.H. Francke, assisted by W.Simon. [monograph] monograph 1929
4. Zhang Yisun (eds.): *A Tibetan-Chinese Dictionary*. Minzu Press (1993)
5. Huang Xing, Jiang Di: The Lexicographic Principles for Compound Verbs in Tibetan Electronic Dictionary. In: JiangDi, Kong Jiangping (eds.): *Advances on the Minority Language Processing of China*. Beijing: Social Sciences Academic press (2005) 94-107
6. Huang Xing, Jiang Di: Automatic Recognition of the Subject and Object of Linking Verbal Sentences in Tibetan. In: Sun Maosong, Chen Qunxiu (eds.): *Language calculation and content-based text processing*. Beijing: Tsinghua University Press (2003) 167-172
7. Yu Shiwen (eds.). *The Grammatical Knowledge-base of Contemporary Chinese---A Complete Specification*. Beijing: Tsinghua University Press (2003)
8. Yin Yiling, Chen Qunxiu: The Research & Implementation of the machine Tractable Dictionary of Contemporary Chinese Predicate Adjectives. In: Sun Maosong, Chen Qunxiu (eds.): *Language calculation and content-based text processing*. Beijing: Tsinghua University Press (2003) 357-363
9. Zhan Weidong: An Overview of Large Scale Semantic Lexicons for NLP. In: Xu Bo, Sun Maosong, Jin Guangjin (eds.): *Some important Issues of Chinese language processing*. Beijing: Sciences Academic Press (2003) 107-121