

# An Annotation of What Is Not There: Empty Arguments and Cross-Clausal Reference in Spoken and Written Tibetan Texts

Bettina Zeisler

SFB 441, Universität Tübingen  
E-mail: zeis@uni-tuebingen.de

## 1 Introduction: purpose of the annotation

Like in many other Asian languages, speakers of Tibetan languages avoid to express information that is already given or can be derived by the hearer from the context. Practically, the verb is the only obligatory constituent of a clause, while all nominal constituents can be deleted, as in the second clause of (1).

- (1) *khon gñiskas de žiηpola sabon btabste o o rmos* | (Francke [8]: 1)  
The two sowed the seeds on the field and ploughed [it].

In particular, there is no syntactic restriction for the deletion of arguments. While the deleted argument in (2) could only be a subject (agent) in a syntactically accusative language and only a patient (direct object) in a syntactically ergative language, in Tibetan this sentence

- (2) *žiηbatpas yokpo kolte o žiη rhmos*. (Field data 1996)  
The farmer hired a servant and o ploughed the field.

can be interpreted in three ways: either the farmer or the worker or both ploughed. A competent speaker resolves the ambiguity mainly along pragmatic considerations (eg why should a farmer hire a worker if not for ploughing). By contrast, students with the background of a European language have considerable problems to find out who is acting upon what. Nevertheless, the phenomenon has not received much attention in Tibeto-Linguistics. Preliminary reflections can be found in Andersen [1] and Zeisler [16]. Earlier, Zimmermann [19] had pointed to the fact that a patient often becomes the subject of a following intransitive clause in Classical Tibetan. A similar relation was termed ‘ergative coreference’ by Cooreman et al. [4], but their classification for Indonesian does not account for the reference relation in sentence (2) where a patient is continued as an agent of a ‘transitive’ clause.

Our research project aims at the formulation of rules or at least statistical preferences for the identification of antecedents of empty arguments for different Tibetan varieties. We expect different frequencies of deletion according to the saliency of an argument corresponding to its semantic role and its

position in the verb frame, but we also expect changes in these frequencies diachronically. To provide an empirical basis for this research, we are building a syntactically annotated corpus of written and spoken texts from different periods: Old Tibetan (8<sup>th</sup>–10<sup>th</sup> century), Classical Tibetan (11<sup>th</sup>–19<sup>th</sup> century), and contemporary West Tibetan as spoken in Ladakh (India).

## 2 General design of the annotation

All texts are annotated in XML. The annotation is done semi-automatically (by incrementally using chunks of already annotated text, see Wagner and Zeisler [15]) and provides rich syntactic information about phrasal and textual structures as well as information about the argument structure of each verb and the realisation or non-realisation of the arguments in the text. The information is encoded as embedded annotation (ie the markup is placed locally at or around the corresponding text).

The basic unit to be annotated is a clause, which typically contains a verb and possibly other elements. The internal structure of a clause is represented as an XML hierarchy tree. One or more clauses constitute a sentence. Higher textual units, tagged as divisions (<div>), consist of a number of sentences. Hence, sub-clausal and super-clausal hierarchies do not overlap and can be captured within a single document hierarchy. Concurrent hierarchical units occur only marginally and are not of primary importance; such boundaries are marked by empty XML elements (eg <pb/> for a page break), which do not violate the well-formedness of the document.

The lowest level of annotation marks the tokens (ie ‘words’) (<tok>) with their orthographic realisation (<orth>) and part-of-speech classification (<pos>). The phrase level is encoded by <ntNode> (non-terminal node) elements. An <ntNode> spans an inflectional group: a noun phrase if this group forms an argument and an adverbial phrase otherwise. This distinction is marked by an element <ntNodeCat>, which contains the category NP or AvP, respectively. A clause (<clause>) encompasses a verb token (always at final position), associated arguments or adverbials, and, if present, embedded clauses. Participle clauses may also be part of an <ntNode>. An element <clauseCat> specifies the type of the clause (simple, chained, embedded, etc). Tokens, phrases, and clauses may receive a further linguistic description (<desc>) specifying the case for sub-clausal phrases and the reference for anaphoric elements. For verb tokens, the corresponding argument structure and its realisation in the text is encoded. Above the clause level, sentences (<s>) are marked. The annotation of the textual level specifies discourse units such as direct or indirect speech, poems or songs, and text segments.

## 3 Annotation of empty arguments

It would be in contradiction with a naturalistic representation of the clause structure to assign slots within a clause for the representation of what is not

there. Furthermore, due to a comparatively free word order, we cannot predict where such slots should be placed. Neutral word order is SOV, or more precisely: Agent > Recipient > Patient > Location > Verb and similarly for inagentive verbs.<sup>1</sup> The verb always comes last, but the order of the arguments can be changed for topicalisation or focussing. We thus chose to represent empty arguments in a second layer: the argument structure of the verbs.

Each verb token receives a serial ID number as an attribute and a special description of its subcategorisation frame (within <desc>). This description comprises (a) the ‘canonical’ argument structure as listed in the lexicon (a list of <complement> elements within a <frame> element), and (b) the ‘real’ frame, ie the realisation of the arguments in the clause (a list of <realComplement> elements within a <realFrame> element). For each canonical and real complement, a <role> element is specified, and each canonical complement receives a specification of its case, as does each real complement whose case deviates from the canonical assignment. The redundant structure is motivated by a higher convenience for annotation, XSLT transformations, and queries. Since the frame is to be found identically in the verb lexica accompanying each text, we might delete it from the final annotation as soon as the corresponding lexicon is linked with the annotation.

To encode cross-clausal reference, each <realComplement> receives an ID based on the verb number. The order of the arguments in the <realFrame> corresponds to the order of the ‘canonical’ frame. Inverted word order is indicated by an attribute on the <realFrame>. Empty arguments receive an attribute marking emptiness and a pointer to the antecedent in the text, which in most cases will be a <realComplement> specified in the argument structure of some previous clause. In general, a pointer is encoded as a reference tag (<ref>) with an attribute ‘target’ that points to the ID number of the corresponding referee, see example (3), the annotated version of (1). In case, no antecedent can be located, the reference tag will remain without target attribute, but might contain additional contextual information. Reference tags are also used for all other anaphoric elements, such as personal (line 9) and demonstrative pronouns (line 29) or referential adjectives.

(3) *khon gñis·kas de žiŋ·po·la sa·bon btab·ste* (v13) *rmos* | (v14)

```

    <s>
    ...
      <clause>
        <ntNode>
5      <tok>
        <orth>khon</orth>

```

---

<sup>1</sup> Tibetan is commonly classified as ergative language (cf Tournadre [14] for modern Central Tibetan), because the (semantic) subject of an intransitive sentence receives the same (zero) case marking as the direct object of the transitive sentence, whereas the (semantic) subject of a transitive sentence receives a different (overt) case marker. For Tibetan this definition poses quite a few problems (Zeisler [18]).

10 <pos>**PRON3:pl**</pos>  
 <desc>  
 <ref target="v10c1"></ref>  
 </desc>  
 </tok>  
 <tok>  
 <orth>**gñis·kas**</orth>  
 <pos>**NUM**</pos>  
 15 </tok>  
 <ntNodeCat>**NP**</ntNodeCat>  
 <desc>  
 <case>**Erg**</case></desc>  
 </ntNode>  
 20 <tok>  
 <orth>**yaŋ**</orth>  
 <pos>**CONJ**</pos>  
 </tok>  
 <ntNode>  
 25 <tok>  
 <orth>**de**</orth>  
 <pos>**DEMfar**</pos>  
 <desc>  
 <ref target="v10c3" type="precDEMfar"></ref>  
 30 </desc>  
 </tok>  
 <tok>  
 <orth>**žiq·po·la**</orth>  
 <pos>**NOM:inan-df**</pos>  
 35 </tok>  
 <ntNodeCat>**NP**</ntNodeCat>  
 <desc>  
 <case>**DatLoc**</case>  
 </desc>  
 40 </ntNode>  
 <ntNode>  
 <tok>  
 <orth>**sa·bon**</orth>  
 <pos>**NOM:inan**</pos>  
 45 </tok>  
 <ntNodeCat>**NP-inc**</ntNodeCat>  
 <desc>  
 <case>**Abs**</case>  
 </desc>  
 50 </ntNode>  
 <tok id="v13">  
 <orth n="2">**btab·ste**</orth>  
 <pos>**V**</pos>  
 <desc>  
 55 <frame>

```

        <complement>
          <role>eA3</role>
          <case>Erg</case>
        </complement>
60    <complement status="omissible">
          <role>DIR</role>
          <case>~Loc</case>
        </complement>
        <complement status="inc">
65    <role>P3</role>
          <case>Abs</case>
        </complement>
      </frame>
      <realFrame>
70    <realComplement id="v13c1">
          <role>eA3</role>
        </realComplement>
        <realComplement id="v13c2">
          <role>DIR</role>
75    </realComplement>
        <realComplement status="inc" id="v13c3">
          <role>P3</role>
        </realComplement>
      </realFrame>
80    </desc>
    </tok>
    <clauseCat>chained</clauseCat>
  </clause>
  <clause>
85    <tok id="v14">
      <orth>rmos</orth>
      <pos>VFIN</pos>
      <desc>
        <frame>
          <complement>
90    <role>eA2</role>
          <case>Erg</case>
        </complement>
          <complement>
95    <role>P2</role>
          <case>Abs</case>
        </complement>
        </frame>
      <realFrame>
100   <realComplement status="empty" id="v14c1">
          <role>eA2</role>
          <ref target="v13c1"></ref>
        </realComplement>
        <realComplement status="empty" id="v14c2">

```

```

105          <role>P2</role>
            <ref target="v13c2"></ref>
          </realComplement>
        </realFrame>
      </desc>
110 </tok>
     <clauseCat>endchain</clauseCat>
   </clause>
 <punct>|</punct>
</s>

```

In clause 14 (starting with line 84), the verb ‘plough’ canonically selects two arguments, one with *ergative* case for the *agent* (eA2) and one with *absolute* case for the *patient* (P2). None of the arguments is realised in the clause, and each realised counterpart can be found exactly in the preceding clause. Therefore, a reference tag with the attribute: target="v13c1" (line 102) points from the first empty <realComplement> (id="v14c1") to its antecedent, the first <realComplement> of the previous clause (line 70). A second reference tag with the attribute: target="v13c2" (line 106) points from the second empty <realComplement> (id="v14c2") to its antecedent, the second <realComplement> of the previous clause (line 73).

The detailed annotation of argument structure and cross-clausal reference allows to evaluate the distances between empty arguments and their antecedents and to observe the changes of role and/or case in chains of corresponding empty arguments referring to the same antecedent. While the target attribute has to be set manually (via a selection tool), the implicit information about reference distance, intervening textual boundaries, and changes in role or case can be retrieved automatically and made explicit via XSLT transformations. The information is stored in a list of antecedents (<antecedList>) for each empty argument or other anaphoric element and a list of anaphors (<anaphList>) for each element that has been referred to. The number of entries in the list is indicated by an attribute ‘size’. Each element <antecedent> or <anaphor> contains the pointer (<ref>), the specification of the reference type <refType>, the source of reference, ie its complement ID <refCompID>, various tags counting intervening textual boundaries, and two tags specifying the matching or mismatching between antecedent and anaphoric element with respect to role and case, see example (4) for the list of anaphora. The list of antecedents has the same structure.

```

(4) <realFrame>
    ...
    <realComplement id="v13c2">
      <role>DIR</role>
      <anaphList size="6">
        <anaphor>
          <ref target="v13c2"></ref>
          <refType>empty</refType>

```

```

    <refCompId>v14c2</refCompId>
    <clauseDist>1</clauseDist>
    <sDist>0</sDist>
    <qDist>0</qDist>
    <seqDist>0</seqDist>
    <finDist>0</finDist>
    <divDist>0</divDist>
    <roleMatch congr="n">
      <antRole>DIR</antRole>
      <refRole>P2</refRole>
    </roleMatch>
    <caseMatch congr="n">
      <antCase>~Loc</antCase>
      <refCase>Abs</refCase>
    </caseMatch>
  </anaphor>
  ...
</anaphList>
</realComplement>
  ...
</realFrame>

```

From this redundant information we can now easily derive detailed statistics, eg concerning the role relation between empty argument and antecedent, its total frequency and its frequency with respect to the clause distance. While the overall picture derived by this process corresponds fairly well to our expectations, the surprises lie in the detail. Eg we expected that active or neutral coreference would be dominant in Ladakhi. Accordingly, our annotation yields a preference for congruent reference relations (31%) and agency (all agent-agent relations independent of valency >50%). One would also expect that reference relations depend on the reference distance, thus agency relations would become even more prominent over long distances.

What we did not expect, however, was that neutral coreference between two patients would be the preferred reference relation (23%) within a distance of only one clause, followed by active (two-place←one-place: 9%, one-place←two-place: 7%) and neutral agentive coreference (two-place: 6%). Since clause chains with agent continuity are much more common than such with patient continuity, chains of the type *X cut an apple and o ate o* must be typically embedded in longer chains with a deleted agent, thus *X verbed ... and o cut an apple and o ate o (and ...)*. The annotation also proved very helpful to identify marginal reference relations, particularly in complex embedded structures where they are no longer evident. Thus we believed that the typologically not predicted relation between a patient and a transitive (ergative) agent could not be found at all in the Ladakhi Kesar epic (Francke [8]). While this relation is predictably rare and apparently restricted to contexts of social interaction or exchange, we could locate four instances alone in the first chapter (583 clauses out of ca. 8500).

## 4 Some particularities of Tibetan and their impact on the annotation

The annotation merely serves a tool, although a quite useful one, for our empirical research. The annotation scheme thus is not derived from theoretical reflections or existing models, but has developed and is still developing out of the needs of our project and along with the growing understanding of what exactly we are dealing with. Thus, the most important prerequisites are practicability, a structure that focuses on the essential, transparency and flexibility. Eg in order not to lose the overview over the complex Tibetan clause and phrase structures we chose to specify the node category after the node content in contradiction to the common linguistic practise.

For the same reason, morphological information is encoded on different levels, depending on its transparency, necessity for the understanding of the text, and saliency for the research project. The non-obligatory use of plural or definiteness markers, eg, is integrated into the part-of-speech tag (3), line 34. Verbal polarity items (negation and question markers) and the use of modal auxiliaries are specified with the help of <feature> tags within the description of the verb token.

Tense (and aspect), however remain unspecified. The reason for this seemingly inconsistent decision is simple: there is no agreement among Tibetologists about what exactly is encoded: absolute tense according to the Tibetan grammatical tradition, ‘aspect’ (certainly not of the slavic type) according to the present mainstream, or relative tense according to a minority view (Zeisler [17]).<sup>2</sup> As for Old and Classical Tibetan, our unknown entity is basically expressed through (up to four) derivational stem forms, which may vary according to the orthographical tradition. We decided to indicate these stem forms in the accompanying verb lexicon, particularly also to be able to compare the spellings of different texts with each other and with the idealised forms of the dictionaries. A mere stem can have a finite or non-finite function, ie it typically terminates a chain of clauses (a final sentence particle may be added, but is not necessary), but it may be also used in particular contexts in place of a non-finite verb form, such as a verbal noun or a converb.

Besides the mere stems, Old and Classical Tibetan also show a great number of finite periphrastic constructions, based on stems, nominalisers, auxiliaries, and additional material. The morphological material itself is functionally intransparent, ie it is not possible to derive the meaning of the whole expression from its parts; it can only be deduced from its usage in discourse. Research into this topic is still missing, and it would be impeded by the fact that Classical Tibetan obviously integrated various regional varieties using different constructions or similar constructions with different meanings. It is practically impossible to know the exact function of most of these constructions.

---

<sup>2</sup> Tibetan languages are still poorly described. The best resource for Classical Tibetan (although with some conceptional errors) is the textbook of Hahn [9]. For contemporary Ladakhi see Francke [7] and Koshal [10], for Balti Read [12] and Bielmeyer [2].



The situation is different for the modern varieties and thus for Ladakhi, but we decided to keep the same annotation scheme and to concentrate on the more relevant issues. Nevertheless we hope to develop an annotation tool that would allow to add this information automatically.

The most salient feature for the project is the encoding of syntactic-semantic relations between a verb and its arguments. These relations are encoded by what we call ‘case markers’ and ‘postpositions’. Postpositions are typically lexically transparent inflected nouns joined to their head via the genitive case, eg *khanpa-ḥi naṅ-la* (house-Genitive interior-Dative/Locative) ‘into the house’ lit. ‘to the interior of the house’. Like the European prepositions, postpositions are mainly used for the more specific spatial relations. Case markers are more evidently integrated into the intonational unit ‘word’ as most of them show assimilation features, and they express relations that are less semantically transparent. Old and Classical Tibetan has the following morphological markers: *{kyi}* Genitive, *{kyis}* Instrumental, *na* Locative, *la* Dative/Locative (Allative), *nas* and *las* Ablative, *{tu}* Locative/Purposive (Allative), and *daṅ* Comitative. Ladakhi has only one Ablative marker /nas/~na/~ne/ and only one productive Dative/Locative marker /(l)a/; the Instrumental as a peripheral argument marker has been replaced by the Comitative.

Except for the Genitive, all case markers can be replaced by a corresponding postposition. Therefore, we use variables in the argument structure, but specify case marking and use of postposition in the annotation of the text. We might change the design when we can predict the use of the locative markers and postpositions. According to our approach, different syntactic slots have different functional weight, conditioning the possibilities of variation in the markers or exchange of position. We thus differentiate between subject and non-subject markers, reaching thus at a set of eight case variables: Erg[ative] (=Instrumental), Aes[thetive] (=Dative/ Locative), ‘~Loc’ (ie the functionally overlapping locative or allative case markers and postpositions), Abl[ative], Instr[umental], Com[itative], Gen[itive], and Abs[olutive] or zero-marking.

Generally, we attempt to keep the annotation as flat as possible. Thus we reduce non-terminal nodes to tokens wherever possible: elementary parts of speech such as conjunctions, cf (3), line 22, are annotated as token, not as phrase nodes. Similarly, we do not annotate the argument structure of verbal nouns if they are lexicalised. On the other hand, since Tibetan shows nominal group inflection, ie only the last element of a noun phrase is modified by a case marker or postposition, we annotate non-terminal nodes for each inflectional group, even if it consists of a single noun. But although an inflectional group may show internal subgrouping through relational case marking, either Genitive ‘of’ or Comitative ‘and’, we leave the internal structure unspecified, adding only a description for the case marker to the respective token.

Based on the syllabic appearance of the script, Old and Classical Tibetan has traditionally been treated as being monosyllabic. It is true, that many elementary lexemes consist of only one syllable, such as *ṛta* ‘horse’, or could be analysed as consisting of one lexemic syllable and one or more derivational morphemes, eg *ṛtapa* ‘rider’. In this case, the morpheme *pa*, typically

functioning as a nominaliser with verbs, can be described as a derivational suffix, expressing a relation of possession or attribution thus *rta-pa* ‘someone who has a horse as his attribute’. But in the case of *khaṅpa* ‘house’, there is no meaningful word *\*khaṅ* and the function of the second element is completely intransparent. Furthermore, Tibetan is very productive in forming compounds such as *rta-mgo* ‘horse head’, *rta-mchog* ‘best of horses’, *rta-bskrags* ‘clattering (train of horsemen)’, etc. Derivations as well as compounds form intonational units that also extend over following syntactic morphemes. Intonational units can be defined as displaying assimilation features (including tone contour in the modern tonal varieties) and prefix dislocation at the lexeme or morpheme boundary, such as Ladakhi /stap-raks/ for the above *rta-bskrags*. Some classical spellings such as *bud-med* ‘woman’ probably from *bu-dmad* ‘of low birth’ indicate a certain tendency for treating compounds as lexical units.

Instead of annotating functionally intransparent particles with dummy designations for no other purpose than to unify them with their lexemes to higher units, we are annotating lexical units and (hypothetical) intonational units or ‘words’ as the basic units (<token>). Following our observations in the spoken language we treat postpositions as being part of the intonational unit (this is also done by Bielmeier [2]), and we likewise treat complex verbal expressions as one ‘word’, although we might revise this in the future for modal constructions. We cannot avoid that such decisions are sometimes debatable and thus in conflict with our attempt to keep the annotation theory neutral and open for other users. But we think that it is legitimate to concentrate our limited resources on those features that are relevant for the project.

## 5 Frames and ‘semantic’ roles

We started our project with the basic assumption that ergative languages, by definition, could not have the syntactic categories subject and object. We understood that the concepts of ‘agent’ and ‘patient’ were introduced as their semantic substitutes, in the discussion of ergativity at least. Accordingly, we initially thought that Tibetan case markers and postpositions would refer to semantic roles. But, actually, how semantic are ‘semantic roles’? An, in principle, unlimited number of truly semantic, verb specific roles was certainly not what we were looking for. We were thus attracted by the promising idea of a limited set of semantic role types, only to find this reduced to a merely binary set of “thematic proto-roles” (Dowty [5]), not so different from the notions of subject and object, not enough for ergative languages, where two different types of agent are encoded, and even less convincing in the view of yet another case marking for experiencer-subjects as in Ladakhi (Zeisler [18]).

We finally realised that (case, thematic, or semantic) roles are neither purely semantic nor purely syntactic categories. In order to cope with the interaction of these two layers, we hope to further develop the Indian concept of *kāraṅka*-relations as a syntactic-semantic interface (Cardona [3]). For the

time being we conceive of our roles as ‘macro-roles’, possibly associated with properties such as [ $\pm$ animacy] or [ $\pm$ control]. They fill particular syntactic slots, but in doing so, the semantic component might be either bleached (eg when an instrument is treated as agent) or might be open for further specification or ‘micro-roles’: Verbs of the ‘(un)load’ type may have either ‘container’ and ‘content’ argument in the two-place patient slot, but only the ‘content’ argument in the three-place patient slot. Our roles are thus necessarily differentiated with respect to valency and further specified according to their specific position in the frame, and their syntactic behaviour. Presently, we work with an inventory of 34 roles.

Semantic considerations lead to a terminological distinction of ‘patient’ (the second or third argument of an agentive verb) and ‘undergoer’ (the sole argument of an inagentive verb). The terminological identification of these two roles without consideration of valency, as in the most neutral designation “Arg<sub>i</sub>” (eg Palmer et al. [12]) might be useful for languages like English with no [ $\pm$ control] distinction and plenty of ambitransitive verbs or for other types of valency raising, such as the ‘experiencer derivation’ in Ladakhi (Zeisler [18]). But Tibetan [ $-$ control] one-place verbs like *ḥbye*, *bye* ‘open’ or *ḥchag*, *chag* ‘break’ exclude the possibility of an intentional agent controlling the event (an external force argument might be added). Whereas [ $+$ control] two-place verbs like *ḥbyed*, *phyes*, *dbye*, *phyes* ‘open’ and *gcog*, *bcag*, *gcag*, *chog* ‘break’ inevitably presuppose the presence of an agent (or an agent-like force or agent-related instrument), even if this agent is not mentioned.

Our concept of ‘frame’ is based on a bottom-up description of how verb arguments are encoded in an utterance, rather than on intuitions about event structures or conceptualisations of events as eg in the FrameNet approach of Erk & Padó [6]. We do not have the empirical base for such top-down analysis: the dictionaries are often silent about the argument structure, and we will hardly find a native speaker of Classical Tibetan, not to mention of Old Tibetan. With respect to contemporary Ladakhi, we have sampled a relatively complete set of verbs for one dialect. In this database, which is going to be enlarged with data on other dialects, we have classified each case pattern (eg 03 Abs~Loc; 06 Aes~Abs; 07 Erg~Loc; 08 Erg~Abs). These schemes allow us to identify verbs of similar behaviour (including frame variation) and establish something like Levin’s [11] verb classes. More basically, however, they help to reflect our guesses about Old and Classical Tibetan argument structure and may serve as a model for similar work on other Tibetan varieties.

## References

- [1] Andersen, Paul Kent (1987) Zero-anaphora and related phenomena in Classical Tibetan. *Studies in Language* 11, 279–312.
- [2] Bielmeier, Roland (1985) *Das Märchen vom Prinzen Čobzan. Eine tibetische Erzählung aus Baltistan. Text, Übersetzung, Grammatik und westtibetisch vergleichendes Glossar*. St. Augustin: VGH Wissenschaftsverlag.

- [3] Cardona, George (1976) *Pāṇini. A survey of research*. The Hague, Paris: Mouton.
- [4] Cooreman, A., Fox, B. and Givón, Talmy (1984) The discourse definition of ergativity. *Studies in Language* 8, 1–34.
- [5] Dowty, David (1991) Thematic proto-roles and argument selection. *Language* 67.3, 547–629.
- [6] Erk, Katrin and Sebastian Padó (2004) A powerful and versatile XML format for representing role-semantic annotation. In *Proceedings of LREC-2004*. Lisboa, Portugal.
- [7] Francke, August Hermann. 1901. Sketch of Ladakhi grammar. *JASB* 70, 1–63. Reprint 1979 as *Ladakhi and Tibetan grammar*. Delhi: Seema Publications.
- [8] Francke, August Hermann (1905-41) *Gšamyulna bšadpaḥi Kesargyi sgruṅs bžugs. A Lower Ladakhi version of the Kesar saga*. Calcutta: Royal Asiatic Society of Bengal.
- [9] Hahn, Michael (1985) *Lehrbuch der klassischen tibetischen Schriftsprache*. Bonn: Indica et Tibetica.
- [10] Koshal, Sanyukta. 1979. *Ladakhi Grammar*. Delhi, Varanasi, Patna.
- [11] Levin, Beth (1993) *English verb classes and alternations: a preliminary investigation*. Chicago: Univ. of Chicago Press.
- [12] Palmer, Martha, Mitch Markus, Scott Cotton, Kate Forbes, Daniel Gildea, Paul Kingsbury, Joseph Rosenzweig (2002) Proposition Bank: Automatic Content Extraction (ACE) at the University of Pennsylvania. <http://www.cis.upenn.edu/~ace/>
- [13] Read, A.F.C. (1934) *Balti grammar*. London: The Royal Asiatic Society.
- [14] Tournadre, Nicolas (1996) *L'ergativité en tibétain. Approche morphosyntaxique de la langue parlée*. Paris: Louvain.
- [15] Wagner, Andreas und Bettina Zeisler (2004) A syntactically annotated corpus of Tibetan. In *Proceedings of LREC-2004*. Lisboa, Portugal.
- [16] Zeisler, Bettina (1994) Ergativ, Passiv und Antipassiv. Entwicklungen im Tibetischen. *Zentralasiatische Studien* 24, 45-78.
- [17] Zeisler, Bettina (2004) *Relative Tense and aspectual values in Tibetan languages. A comparative study*. Berlin, New York: Mouton de Gruyter.
- [18] Zeisler, Bettina (to appear), Case patterns and pattern variation in Ladakhi: a field report. In: Bielmeier, Roland and Felix Haller (eds.) *Himalayan Linguistics and Beyond. Papers submitted to the 8th Himalayan Languages Symposium, held at the University of Berne, Switzerland, Sept. 19-22, 2002*.
- [19] Zimmermann, Heinz (1979) *Wortart und Sprachstruktur im Tibetischen*. Wiesbaden: Harrassowitz.