Antrag auf Gewährung einer Sachbeihilfe

Part A

1 General information (Allgemeine Angaben)

New proposal/ Neuantrag

1.1 Applicants (Antragsteller)

Name	Prof. Dr. Manfred Krifka	Dr. Kilu von Prince
Fachgebiet	Allgemeine Sprachwissenschaft	Allgemeine Sprachwissenschaft
Geburtsdatum	26. 04. 1956	01. 12. 1982
Staatsangehörigkeit	deutsch	deutsch
Anschrift	Institut für deutsche Sprache und Linguistik Humboldt-Universität zu Berlin Unter den Linden 6 D-10099 Berlin Tel. dienstlich: +49-30-2093-9670 Email: `krifka@rz.hu-berlin.de` Tel. privat: 030-84719950 Web (persönlich): `amor.rz.hu-berlin.de/~h2816i3x/`	Zentrum für Allgemeine Sprachwissenschaft Geisteswissenschaftliche Zentren Berlin Schützenstr. 18 D-10117 Berlin Tel. dienstlich: 030-20192-422 Email: `prince@zas.gwz-berlin.de` Tel. privat: 030-78001522 Web: `www.zas.gwz-berlin.de/mitarbeiter_prince.html`

1.2 Topic (Thema)

A corpus-based contrastive study tense, aspect, modality and polarity (TAMP) in Austronesian languages of Melanesia (MelaTAMP).

Eine korpus-basierte kontrastive Studie von Tempus, Aspekt, Modalität und Polarität (TAMP) in Austronesischen Sprachen Melanesiens (MelaTAMP).

1.3 Research Area (Fach- und Arbeitsrichtung)

General Linguistics, Austronesian Linguistics, Typology, Semantics

Allgemeine Sprachwissenschaft, Austronesische Sprachwissenschaft, Typologie, Semantik

1.4   Funding Period (Antragszeitraum)

01. 11. 2013 – 31. 12 . 2016

1.5   English summary (Englische Zusammenfassung)

Studies on the TAMP systems (tense, aspect, modality and polarity) of the world's languages have focused on the roles of tense and aspect rather than of modality – especially concerning the veridicality dimension of modal meanings. Moreover, our most detailed knowledge about such systems tends to come from tense-prominent languages (in the terminology of Bhat 1999); and those data from mood-prominent languages that have made their way into large-scale surveys may be biased because elicitations and descriptions were often based on the authors' own, usually tense-prominent, mothertongue.

Only in recent years have large amounts of primary data from mood-prominent languages become available through language documentation. We propose to explore corpus data from seven Oceanic languages of Melanesia, which have been known to be mood-prominent, for an areal typological study of their TAMP systems.

The primary goal of the project is to enrich the field of typological research by opening up new empirical data: First, all the languages under investigation primarily encode mood and aspect rather than tense and therefore belong to a type of language that has been relatively underrepresented in the typological debate on TAMP categories.

Second, the languages of Melanesia are typically spoken by only several hundred to several thousand speakers, and do not have a tradition of written records, which has made them largely inaccessible to researchers, except for individual grammatical descriptions. By exploring existing, annotated corpora, we will increase their accessibility both for studies on TAMP-related aspects, and for other typological research questions.

1.6   German summary (Deutsche Zusammenfassung)

Die bisher umfangreichsten Studien zu TAMP-Systemen (Tempus, Aspekt, Modalität, Polarität) befassen sich wesentlich stärker mit der Rolle von Tempus und Aspekt als mit Modalität – besonders was den Realitätsstatus modaler Bedeutungen betrifft. Dazu kommt, dass ein Großteil unseres Wissens über TAMP-Systeme von Sprachen kommt, die in der Terminologie von Bhat (1999) als *Tempus-prominent* bezeichnet werden können. Insofern Daten von Modus-prominenten Sprachen Eingang in großflächig sprachübergreifende Studien gefunden haben, ist deren Darstellung möglicherweise dadurch verzerrt, da sie oft auf Elizitationen aus den Muttersprachen der jeweiligen Autoren basiert, die ihrerseits wiederum in der Regel Tempus-prominent sind.

Erst seit Kurzem sind große Mengen an Primärdaten von Modus-prominenten Sprachen durch Sprachdokumentation erschlossen worden. Wir haben vor, Korpusdaten von sieben

ozeanischen Sprachen Melanesiens zu untersuchen, die als Modus-prominent bekannt sind, um eine areal-typologische Studie ihrer TAMP-Systeme durchzuführen.

Part B

1 State of the art, earlier work (Stand der Forschung, eigene Vorarbeiten)

1.1 State of the art

1.1.1 Modality, veridicality and the realis-irrealis distinction

General observations In the traditions of formal semantics and of linguistic typology, discussions of modal meanings tended to focus on the dimensions of modal force (necessity or possibility) and modal sources (epistemic, circumstantial, deontic, dynamic etc.) to the detriment of the distinction between the factual, the counterfactual and the possible (cf. for example Nuyts 2006, Hansen & de Haan 2009 and Portner 2009). Modal force and source are most prominently expressed by modal auxiliaries in many languages, especially within the Indo-European family. The distinction between factual, counterfactual and possible modalities plays a more important role in conditional clauses and complement clauses, which are often treated as separate phenomena in both traditions. At the same time, the split between factual utterances on the one hand and counterfactual or potential meanings on the other has been recognized as fundamental for the TAMP systems of some language groups and has been closely associated with the terms **realis** and **irrealis** respectively (e. g. Dixon, 1980; Timberlake, 2007; Foley, 1986). Such languages are also sometimes described as differntiating **future** from **non-future**, thus aligning them with a more familiar temporal distinction. In the terminology of Bhat (1999), they are referred to as **mood-prominent** languages as opposed to tense-prominent and aspect-prominent languages.

This type of system can be illustrated by the language Daakaka, one of the subject languages of the proposed project. The basic split between the actual and the possible in particular is illustrated by contrast between the realis marker, which refers to actual events in the past or present, and the potential marker, which expresses both neutral predictions about the future and possiblities or wishes:

(1) a. *lisepsep **ka** **we** me=ne ada a=tak sa nge=tak*
         lisepsep MOD.REL POT come=TRANS 1D.IN LOC=PROX CM nge=PROX
         'the lisepsep[1] will come for us here now' (sto31:54)

     b. *kana=**m** **dimyane ka** kana=**p** me usi ∅-am wotop*
         1D.EX=REAL want COMP;NFAC 1D.EX=POT come ask CL2-2S.POSS breadfruit
         'we wanted to come and ask for your breadfruit' (sto31:18)

---

[1]A lisepsep is a magical dwarf-like creature.

3

The reality status of a proposition can thus be said to be more central to the meaning of the markers than their purely temporal reference.

Even though Bybee *et al.* (1994: 236ff.) have dismissed the realis-irrealis distinction as being too vague and inconsistent between languages to play a useful role in cross-linguistic comparisons, the debate on its significance is far from settled. Typologists and specialists on individual non-European languages have repeatedly challenged the position of Bybee *et al.* (1994) and maintain that the realis-irrealis distinction is fundamental for describing and analyzing certain types of TAM systems (cf. Chafe 1995; Hofling 1998; Kinkade 1998; Martin 1998; Elliott 2000; de Haan 2006).

In formal semantic theories, the similarity between a non-future/ future distinction and a realis/irrealis distinction is not surprising, as both temporal and modal notions have been analyzed with the tool of index-dependent propositions and quantification over indices. Modal logic and temporal logic only differ in terms of their accessibility relation (Hughes & Cresswell, 1996).

In particular, theories assuming branching times stand out as one approach that places particular emphasis on the interdependence between tense and modality (e. g. Dowty, 1977; Thomason, 1984; Condoravdi, 2002; Kaufmann *et al.*, 2006; Werner, 2006; Laca, 2012). The split between future and non-future references that many languages exhibit, as well as the modal implications of this split, find a straightforward representation in this framework: The past and present differ from the future in that there is only one actual present and past, but several possible futures, among which none is privileged as the actual future. Thus, it is possible to draw a clear distinction between the actual and the counterfactual only with respect to the past and present, whereas the future is divided into the still possible and the no longer possible.

The simplicity and intuitive appeal of this approach lend plausibility to the idea that the split between the actual and the non-actual that has been observed in many languages corresponds to a straightforward and fundamental semantic distinction. At the same time, it suggests that tense and modality are but two dimensions of the same conceptual space, rather than independent notions, and predicts correctly that some linguistic expressions may have both temporal and modal implications.

Furthermore, while it is true that the term *irrealis* in particular has been used cross-linguistically to refer to a wide variety of expressions with very different properties, it is also true that the design of the larger typological surveys so far has not been fit to decide on its significance. Like most traditional typological studies (e. g. Comrie, 1976, 1985; Timberlake, 2007), large-scale comparative surveys such as, Bybee *et al.* (1994), Dahl & Velupillai (2011) and Miestamo (2005) were mostly based on grammatical descriptions. This methodology has long been used to good effect and established numerous interesting correlations, but its potential pitfalls are only too well known:

> "Even in the case of comparatively well-described languages, which constitute a small minority, the information found in reference grammars and more specialized publications tends to be insufficient and often misleading. This is in particular the case for grammatical categories such as tense, mood, aspect, number, definiteness, case etc.,[...]" (Dahl, 2007)

4

Similar problems hold for the less widespread method of elicitation-based typologies such as Dahl (1985), as the author himself is the first to acknowledge (Dahl, 1985: 50).

For the topic of TAMP systems, the kinds of biases that grammar-based and elicitation-based surveys may produce are especially relevant since most of the languages for which mood distinctions have been reported as the most prominent are among the least accessible to linguists: They are spoken in Melanesia and Australia, along with some languages from North and Central America (see also de Haan, 2006: 41); they often have relatively small numbers of speakers, and spoken and written materials have been scarce or not accessible at all. This means that the data that find their way into the typological surveys are often based on a single grammatical description that might or might not do full justice to their TAMP system.As Bhat (1999: 99) remarks:

> "…our inability to consider mood as a verbal category that can be as prominent as tense or aspect (see Bybee, Perkins and Pagliuca 1994:239) is probably a result of our dependence, for the description of mood-prominent languages, upon a tense-prominent language."

The innovative project by Lisa Matthewson and Hotze Rullman on *Cross-Linguistic Expressions of Modality* (SSHRC) that was launched in 2011 seeks to shed more light on modality in a number of less described languages, but they too focus mostly on the dimensions of modal force and source and on expressions like auxiliaries rather than TAMP markers.

In sum, the concentration on modal auxiliaries and on expressions of modal force and source rather than reality status and the resulting biases may have created a blind spot for certain types of modal meanings.

The obvious remedy to this situation is to study natural language data from mood-prominent languages in order to investigate whether the difference between the actual and the non-actual is as basic and as systematic in the TAMP systems of some languages as the difference between necessity and possibility is in others. This will be achieved by the proposed project with its focus on corpus data from seven Oceanic languages of Melanesia.

Complement clauses   The factual/ non-factual distinction may also play a greater role in the complement clauses of our subject languages than has so far been appreciated in the typological literature.

According to the typology by Dixon (2006), a language may differentiate between various types of complement clauses, depending on the formal properties and semantic domain of the embedding verb (such as perception, thinking or liking) and on the meaning expressed by the embedded clause (such as fact, activity or potential). This approach predicts that complement clauses of verbs like *think* should pattern with complement clauses of *know (that sth. is the case)* rather than verbs like *want* or *be able*, as is the case in English: *She thinks (that) Laura left* has essentially the same structure as *She knows (that) Laura left*, in contrast to *I want to leave / I'm able to leave.*

However, our research on the languages of West Ambrym indicates that the encoding of the complement clause in these languages depends primarily on its **veridicality** status (cf. Giannakidou, 1998): a complement clause of *think* will have the same complementizer as a

complement clause of *know*, if the embedded clause is assumed to be true by the speaker; by contrast, if the complement clause of *think* is not supposed to be true – as in *she thought Laura had left [but in fact Laura hadn't left]*, it will pattern with clauses embedded by *want*.

This phenomenon is illustrated below, first with factive complement clauses and then with non-factive complement clauses:

(2) a. *A   nge mo   kuowilye na   ma   sengep mo   nok.*
         but 3s REAL know   COMP;REAL REAL be.open REAL finish
         'And he already knew that it was open.' (rep15:42)

     b. *daa   ne   bató   mon na=m   dimye na   mwe vyan seaa.*
         language TRANS taboo.place also 1s=REAL think COMP REAL go   entirely
         'I think all has been said about the *bató*, too.' [lit. ,I think that the speech of the *bató*, too, has gone completely.'] (exp03:020)

(3) a. *na=m   dimyane ka   na=p   pun usili pun=an ne   myaop*
         1s=REAL want   COMP;NRE 1s=POT tell follow tell=NM TRANS volcano
         'I want to tell a story about the volcano.'

     b. *Na=m   dimye ka   ko=t   penin seaa ∅-ada vis   a   ∅-ada*
         1s=REAL think MOD.COMP 2s=DIST roast.pl every 1D.IN banana and 1D.IN
         vis   mwe pwer kyun
         banana REAL stay just
         'I thought you had roasted all our bananas, but our bananas are just here' (sto37:104)

A detailed comparative study of different types of complement clauses in the region may help to deepen our understanding of the factors that determine how different types of embedded clauses are distributed.

Veridical dependency and specificity of determiners   Another related area to which our proposed project may provide new and interesting data is the behaviour of different groups of determiners and quantifiers. It has been shown that the contrast between polarity items such as *any* as opposed to *some* depends on veridicality and that the manifestation of this difference may differ both between various items of a single language and between languages (Nam, 1994; Giannakidou, 1998, 2001).

We know from our research on the West Ambrym languages that there are different sets of quantifiers whose distribution appears to depend mostly on specificity, but which also show a strong resemblance to polarity items in other languages. Exploring these elements across the subject languages may shed a light on the relation between specificity and veridicality and may help to develop precise criteria to decide between the two cross-linguistically.

1.1.2   Cross-linguistic relations between TAMP categories

General observations   The tendency to focus on modal force and source rather than veridicality has also influenced the investigation of relations between modality and aspect, tense and negation.

Apart from Bybee *et al.* (1994), the most comprehensive typological study of TAM categories is Dahl (1985).[2] While both studies have done a lot to enlighten our understanding of tense, aspect and mood cross-linguistically, they both focus on the relation between tense and aspect far more than on the relation of both these categories to modality.[3] The study done by Dahl does so explicitly, despite its original goal to cover all three categories (Dahl, 1985: 2). By contrast, Bybee *et al.* (1994) set out to investigate all three categories comprehensively, and they do point out some relations between modal and temporal notions, especially in the case of the future. Primarily, however, they emphasize relations between aspectual and temporal notions such as perfective and past, or imperfective and present.

In various studies from both before and after Dahl (1985) and Bybee *et al.* (1994), certain correlations between temporal and modal meanings have been consistently reported cross-linguistically, which were not investigated by those two studies in any detail. Most prominent among the relations between tense and modality is the link between past and counterfactuality (Steele 1975; James 1982; Fleischman 1989; Givón 1994, see also Plungian & van der Auwera 2006). As Palmer (2001: 209f.) observes:

> "[of the tenses, the] future may be thought to be the most 'modal' [...], yet it is the past tense that is in fact mostly interrelated with modality, and in particular with unreality."

In the case of English, this correlation is reflected by the convergence between subjunctive and past forms as in *They wish they were taller*. In Daakaka, the distal marker expresses both distant or discontinuous past (something used to be the case, but is no longer the case) and counterfactuality:

(4)   a.   *meu=an na nenyu   **te**   melumlum, a   meu=an na doma **mwe** yas*
            live=NM ATT yesterday DIST quiet   but live=NM ATT today REAL hard
            'the life of before was easy, but the life of today is hard' (con02:90)[4]
   b.   *saka   w=i   Kaingas, Katolik **te**   dyanga Sesivi*
            SUBCONJ.NEG POT=COP catholic NAME   DIST lack   NAME
            'if it hadn't been for Kaingas, there wouldn't be Catholics in Sesivi' (rep09:57)

While Dahl (1997) argues based on data from European languages that the use of past tense markers in counterfactual conditionals reflects different diachronic stages in a vague development of meanings, researchers continue to explore the semantic and syntactic implications of the correlation between past and counterfactuality in typology and semantics (e. g. Iatridou 2000; Guéron & Lecarme 2008; Hacquard 2010, compare also Comrie 1986: 90).

Other well-established correlations between TAMP categories include an interdependence between imperfective aspect and potential modalities (e. g. Fleischmann, 1995; Bhatt, 2006;

---

[2]Certain aspects of the study in Dahl (1985) have since been considerably extended by some chapters in the World Atlas of Language Structures such as Dahl & Velupillai (2011). But as they focus on basic morpho-syntactic features such as whether or not a language marks the difference between perfective and imperfective aspect, rather than relations between meanings and functions, we do not consider them here.

[3]The same holds for Dahl (2000): *Tense and Aspect in the Languages of Europe.*

[4]All Daakaka data come from von Prince's own research on the language. References refer to positions in the corpus.

Dowty, 1977; Portner, 2009); a link between perfective aspect and root modality (cf. Abraham & Leiss, 2008); and a relation between modality and negation that manifests in whether languages classify a negated sentence as realis or irrealis (Bybee *et al.*, 1994; Miestamo, 2005).

In the proposed project, we will not attempt to sort individual morphemes into neat boxes that correspond to only one dimension of their meaning, such as temporal, aspectual or polarity-related. Instead, by focusing on correspondences between these semantic dimensions, we hope to improve our understanding of how they are related.

A systemic perspective   The interpretation of a TAMP marker may not only depend on its lexical meaning, but also on the availability of paradigmatically contrasting elements (Comrie, 1985: 84).

Such pragmatic effects on interpretation have been analyzed as cases of M-implicatures (Levinson, 2000) or of semantic **blocking** (Blutner, 2002): An expression will receive a more marked, more specific interpretation than its lexical meaning warrants, if there is a less marked, otherwise synonymous expression available in the language.

Ross (1995) and Zeitoum *et al.* (1996) have observed in their studies on the TAMP systems of Formosan languages that the close comparison between neighbouring languages reveals such blocking effects in the interpretation of TAMP markers: They remark that in particular the so-called neutral (zero-marked) form of assertive sentences is interpreted differently in different languages, depending on the inventory of paradigmatically competing markers:

> "[The] functional range [of the neutral form] depends on the availability of other forms in the language, e. g., on whether there is a special durative form or whether the neutral form functions as both punctual and durative." (Ross, 1995: 742)

Thus, comparing closely related systems may reveal aspects of TAMP meanings that are implicatures rather than part of their lexical definitions.

1.1.3   Methodology

Overview   In terms of the quantity and the nature of our data, the study is positioned between the two more established traditions of 1) contrastive studies involving comparisons between only two or three languages based on large, sometimes parallel, corpora with between one million and several hundred million words; and 2) typological studies comparing several dozen to several hundred languages, based on grammatical descriptions and questionnaires. In other words, the amount of data per language in our proposed project is higher than in a traditional typological study, but lower than in a traditional contrastive study, while the amount of languages is lower than in large-scale typological studies, but higher than in contrastive studies.

The data for this type of study in the context of small, resource-poor languages has only become available recently. Alarmed by rising rates of language loss and endangerment, several

scientific foundations have launched international campaigns to document small, resource-poor languages, such as the DoBeS program (Dokumentation bedrohter Sprachen, Volkswagenstiftung, ca. 50 projects since 2000), the HRELP program (Hans Rausing Endangered Languages Project, Arcadia, about 300 projects since 2003), the DEL program (Documenting Endangered Languages, NSF/NEH, about 50 projects since 2005) or the ELPR program (Endangered Languages of the Pacific Rim, funded by the Japanese Ministry of Education). Of the seven languages we propose to investigate, six were documented only during the last few years within the context of the DoBeS and HRELP programs. At the same time, software, formats and standards for cross-linguistic annotations of specific categories in small corpora, have only recently begun to emerge.

Corpus-based typology   As mentioned above, typological studies have traditionally relied on information from grammars and, to a lesser extent, questionnaires. Among those exceptional studies that are based on texts from the original languages instead are Greenberg (1960); Givón (1983); Myhill (1992); Bickel (2003); DuBois *et al.* (2003); Güldemann (2008); and Haig *et al.* (2011). The relatively novel field of text-based typology is now also being explored by several research projects that were launched in recent years, such as Wälchli (2007); Bickel (2011); Mettouchi & Vanhove (n.d.); and Lehmann (2012).

For our study, we will rely on corpus data as a primary source instead of grammars or elicitations primarily for the following reasons:

- Both grammatical descriptions and elicitations may be biased by the authors' expectations. They may not always reflect natural language use.
- Grammatical descriptions may not contain sufficient information to allow for fine-grained comparisons between languages.
- Elicitation-based surveys can only find what they are looking for, they may miss specific markers. For example, the questionnaire by Dahl (1985) would have missed the Daakaka marker *doo*, which occurs exclusively in embedded polarity questions referring to past or present events:

  (5)   *sye swa sa   bwe me te   ka   da=p du ongane   ka **doo** me*
         REAL say something one FOC CONT come DISC say 1D.IN=POT stay hear COMP;NRE
         pwisya-ne ada
         WHETHER come come.out-TRANS 1D.IN
         'something is coming, let's try to hear whether it's coming to us' (sto34:064)

- Grammars and elicitations may not allow to detect phenomena that are specific to certain genres or to certain environments, such as reported speech.
- Grammars and elicitations do typically not reflect how frequently a construction is used for certain functions, or how frequently a given function is expressed by a certain construction. Quantitative differences may however hold important clues to qualitative differences between TAMP systems.

1.1.4 Annotation tools and methods

In recent years, the annotation of temporal and modal meanings in natural language corpora has received increasing attention in the area of large-scale language processing (e. g. Reichart & Rappoport, 2010; Hacquard & Wellwood, 2012; Rubinstein *et al.*, 2013; Baker *et al.*, 2012). The experiences and results of this work are only partially applicable to our project: The frameworks for modal annotation in particular focus mostly on the dimensions of modal force and source, while in our project, the dimension of veridicality will be the main focus of modal annotations. Not only is this category apparently the most prominent in the language family we are working with, but experience from previous annotation projects also indicates that it is far easier to find unambiguous and clear-cut categorizations for veridicality status than for distinctions such as epistemic versus circumstantial (e. g. Carretero & Zamorano-Mansilla, 2013).

Furthermore, the amount of data we are dealing with is only a fraction of what computational linguists usually work with and our main motivation for using corpora is not statistical but qualitative reliability of data. The requirements for working with corpora of millions of words do not necessarily scale down to corpora of several ten thousand words.

On the other hand, annotation tools such as MMAX2 (`http://mmax2.sourceforge.net/`), which have proven useful for the annotation of modal and temporal meanings and dependencies, may well provide feasible solutions to some of our technical needs.

1.2 Earlier work

1.2.1 TAMP systems of Oceanic languages and modal semantics

The proposed project will involve seven Oceanic languages of Melanesia. Both Krifka and von Prince have already done substantial work on the TAMP systems of three of the seven subject languages. Funded by the Volkswagen Foundation, Krifka and von Prince have documented the languages Daakie, Daakaka and Dalkalaen on the island of Ambrym, Vanuatu between 2009 and 2012. Based on large and systematic collections of data from their own fieldwork, they have published detailed accounts of the TAMP systems from Daakie (Krifka, 2012b, 2013) and Daakaka (von Prince, 2011a,b, 2012). Except for the more descriptive approach in von Prince (2012), all of these accounts incorporate a formal semantic approach to the meanings of TAM markers. Krifka has also published on topics related to modal semantics more generally (Krifka, 2012a; Grosu & Krifka, 2008).

1.2.2 Annotation and exploration of corpus data

Through their work with the corpora they created from transcribed recordings in Daakaka, Daakie and Dalkalaen, Krifka and von Prince are intimately familiar with the design and exploration of small corpora. In addition to the creation of an annotated text corpus in SIL Toolobox, von Prince has written an XSLT script, together with Ciprian Gerstenberger (Universitetet i Tromsø), to transform her text database into a human-readable customized XML format. This

has allowed her to thoroughly consolidate previously inconsistent glosses and makes the addition of further layers of annotation easily feasible. The resulting corpus will be transformed into PAULA XML (Zeldes *et al.*, 2013) and be hosted at the online platform ANNIS (SFB 632), in addition to the DoBeS archives. Von Prince has presented this process during a workshop she organized in May 2013 on *Exploring data from language documentation*, together with Felix Rau (von Prince, 2013), which was hosted at the ZAS.

From the context of her master's thesis (von Prince, 2008), von Prince also has previous experience with the exploration of corpora from Hindi (McEnery & Gaizauskas, 2003; Almeida, n.d.), Swahili (HCS, 2004) and Mandarin Chinese (McEnery & Xiao, 2004).

1.2.3   Project related publications

Peer-reviewed articles

Grosu, Alexander, & Krifka, Manfred. 2008. The gifted mathematician that you claim to be: Equational intensional 'reconstruction' relatives. *Linguistics and philosophy*, **30**, 445–485.

Krifka, Manfred. 2012a. Definitional generics. *Pages 372–389 of:* Mari, Alda, Beyssade, Clare, & Del Prete, Fabio (eds), *Genericity*. Oxford: Oxford University Press.

Krifka, Manfred. 2012b. Notes on Daakie (Ambrym, Vanuatu): Sounds and modality. *Pages 46–65 of:* Clemens, Lauren Eby, Scontras, Gregory, & Polinsky, Maria (eds), *Proceedings of AFLA 18 (Austronesian Formal Linguistics Association).* Cambridge, Mass.: Harvard University. online publication.

Other publications

Krifka, Manfred. 2013 (04). *Modal and temporal reference in Daakie (Ambrym, Vanuatu).* Talk given at New York University.

von Prince, Kilu. 2011a. *On the Daakaka modal system: Imperatives and futures.* Talk given at the Chronos 10 conference at Aston University, Birmingham.

von Prince, Kilu. 2011b (June). *On the Daakaka modal system: the distal.* Talk at workshop *Speaking of Possibility and Time II*, Universität Göttingen.

von Prince, Kilu. 2012.   *A grammar of Daakaka.*   Ph.D. thesis, Humboldt University Berlin. Published online: `http://edoc.hu-berlin.de/dissertationen/von-prince-kilu-2012-07-11/PDF/von-prince.pdf`.

von Prince, Kilu. 2013 (May). *Problems with and solutions to consolidating corpus data from fieldwork.* Talk given at workshop *Exploring data from language documentation* at ZAS Berlin.

2   Objectives and work schedule (Ziele und Arbeitsprogramm)

2.1   Expected duration of the project

We expect to complete the project within 3 years, during which we will require funding by the DFG.

2.2 Objectives

The primary motivation for the project is to expand our knowledge about TAMP systems in mood-prominent languages. By focusing on corpus data as our primary empirical source, we hope to minimize the effect of biases typically induced by elicitation- and translation-based methods. Furthermore, we want to be able to identify the role of different genres and specific environments, such as reported speech, on the usage of TAMP markers; and to establish quantitative differences between markers within and across languages.

By comparing several languages from the same family and linguistic region, we intend to establish cross-linguistic correlations between a variety of different TAMP-related meanings; by comparing closely related, entire systems we also hope to differentiate lexical meanings of TAMP markers from interpretational tendencies prompted by blocking effects from paradigmatically contrasting elements.

At the same time, the study will open up new resources from a previously underrepresented language region to wider typological studies.

A detailed list of research objectives is given below:

- probing the relation between tense, aspect and modality: instead of identifying individual markers as primarily modal, temporal, aspectual or polarity-related and discussing them only in the corresponding contexts, we will focus on correlations between these categories;
- exploring the role of veridicality in embedded clauses and determiners or quantifiers;
- identifying the significance of specific genres and environments for the use of TAMP-markers;
- investigating the systemic effect of paradigmatically contrasting markers on the interpretation of individual markers;
- making Oceanic languages of Melanesia more accessible to typological studies of TAM systems and beyond;
- adapting existing questionnaires such as in Dahl (1985) to reflect our findings;
- sustainability of data formats and methods (PAULA XML, ANNIS).

2.3 Working schedule

Table 1 shows the main work units, with references to the principal investigator and associated personnel for each unit.

Table 1: Organization plan

WU	Tasks	Responsibility
WU1	Compiling a questionnaire of distinctive meanings and environments, creating a comparable subcorpus	von Prince, NN$_1$
WU2	Annotation of corpora, compilation of stratified examples	von Prince, entire team
WU3	Preprocessing of corpora, consistency checks, statistical analysis, transformation of corpus data to PAULA XML	von Prince, NN$_2$
WU4	Synthesis and theoretical analysis of results	Krifka, entire team

Table 2: Core Corpora, with more than 3000 sentence units

	Language	tokens (approx.)	Country	Collaborating Author
1	Daakie	62.000	Vanuatu	Manfred Krifka
2	Daakaka	60.000	Vanuatu	Kilu von Prince
3	Dalkalaen	30.000	Vanuatu	Kilu von Prince
4	North Ambrym	> 30.000	Vanuatu	Michael Franjieh
5	Mavea	10.000	Vanuatu	Valérie Guérin
6	South Efate	37.000	Vanuatu	Nick Thieberger
7	Saliba/Logea	50.000	Papua New Guinea	Anna Margetts

2.3.1 Data and Methodology

We already have obtained corpora from the seven languages we intend to work with (see table 2). Three of them come from our own research, the rest has been provided by colleagues. The corpora are fully glossed and annotated for parts of speech. All corpora are in SIL Toolbox format and will be converted to lightly embedded XML for more versatility and stability. They will then be enriched by additional annotations to allow for searches for specific markers (including zero-marked forms) and specific meanings, and for stratification according to genres and environments. This means that we will conduct searches both from meanings to forms (semasiological) and from forms to meanings (onomasiological).

We will then map each marker of a language to its range of meanings and compare these mappings between languages to establish correlations between TAMP-related meanings. We will try to model these correlations semantically. The results will be related to contemporary typological and semantic findings on TAMP categories.

2.3.2 Phase I

As mentioned above, we have received corpus data on seven languages from our associated language experts. Early in the first phase, we will invite them to a workshop in order to discuss their data in the context of the project. Also, we will test appropriate annotation software.

Stratification and comparable corpora We will want to control for different environments and genres and to construct correspondingly stratified samples for each target morpheme. In order to do so, we will annotate each corpus for certain coarse-grained features such as genre, reported speech, and position within a text (beginning, middle, end).

In addition, we will also choose several texts per corpus for the creation of comparable sub-corpora (cf. Johansson, 2003): While the corpora we use are too diverse in their composition to be truly comparable, there are significant overlaps in the types of texts and genres. For example, stories about how the rooster came to be colorful and domestic, while the megapode is black and wild, is widespread throughout the region.

It is therefore possible to create a relatively small sub-corpus for each language such that these sub-corpora have a very similar composition in terms of genre and about the same number of words. These sub-corpora will be annotated completely for the semantic and contextual

distinctions we want to investigate (such as reported speech, conditional clauses, etc. – see below).

These comparable corpora will be a major source for sampling basic TAMP-related meanings (see below) and for zero-marked functions, and they will be used exclusively for establishing statistical correlations.

Identifying and sampling target morphemes For each language, we need to identify the set of morphemes which will be compared in the study. In all the subject languages there is a system of morphemes which are obligatory in most types of sentences, and it is the elements of these systems that we consider the primary targets for our investigation. However, certain other morphemes also need to be taken into account. Thus, the marker for potential mood in Daakaka is part of the same paradigmatic system as for example the realis marker, which marks assertions and questions with present or past reference as finite. But the potential marker can express an assertion only in combination with the preceding morpheme *ka*; otherwise, the utterance is interpreted as a directive speech act. The morpheme *ka* is therefore also considered as a target morpheme.

(6) a. *da=**p**   vyan*
         1D.IN=POT go
         'let's go' (rep15:025)
     b. *barvinye swa **ka** **we** luk   teve-sye   m-ada   em*
         grass   one MOD POT grow side.of-3S.POSS CL2-1D.IN.POSS house
         'a grass will grow next to our house' (sto17:013)

On the other hand, there is a great number of lexemes in each language whose meaning may be related to TAMP categories, such as auxiliaries, adverbs and modal tags, which will not be investigated with the same degree of detail as the obligatory TAMP markers.

We will select those elements as target morphemes that are either obligatory in marking basic semantic and contextual distinctions as described below, or that paradigmatically contrast with such morphemes and have the same basic structural properties.

For each target morpheme, we will create a stratified sample of up to one hundred occurrences and map this sample to the basic meanings we want to distinguish. The sampling process will be made more complicated by the fact that certain basic meanings are not expressed by an overt morpheme in some or all environments. We will therefore have to annotate zero-marked clauses correspondingly.

Identifying and sampling basic semantic distinctions In order to identify a set of contexts and meanings to investigate across the subject languages, we will draw on general typological work as well as on existing analyses of the subject languages and their close relatives (Thieberger, 2006; Margetts, 1999; Guérin, 2011: e. g.). On the one hand, we will consider all those TAMP distinctions which have been established as fundamental cross-linguistically, such as different speech acts, positive and negative polarity, perfective and imperfective aspect, factive and non-factive embedded clauses, different types of conditional clauses, direct and reported speech, etc.

On the other hand, we will also take into account those distinctions which may not be too widespread globally, but have been attested in at least one of the subject languages or appear to be a more frequent feature of languages in the region. A case in point would be embedded factive and non-factive polarity questions, for which at least Daakaka has a special marker, as mentioned above.

2.3.3 Phase II

In the second phase, we identify example sentences of each of the previously established semantic categories in each language.

For each function, we will consider every example from within the comparable sub-corpora and examine which target morphemes are used in each case. For those meanings that are not represented in the sub-corpora, we will both extend our sample to a greater number of texts, and, where applicable, search the free translations of the entire corpus for potential matches.

We will sort the result sentences from the stratified sample into these categories. We will then check for consistency between annotators and discuss our preliminary results. Gaps in the data will be compiled into questionnaires to be used during a fieldtrip to Vanuatu at the end of this phase.

2.3.4 Phase III

We will compare the results from the individual languages and establish cross-linguistic correlations between TAMP-related meanings. We will invite our collaborating language experts to a second workshop to present and discuss our results. We will also explore the theoretical implications of our findings by modeling the core semantic, pragmatic and morpho-syntactic properties of the investigated morphemes.

We will then relate our findings to the wider typological and semantic debate and prepare our results for publication. One publication will be an anthology of contributions to the second workshop, relating individual languages to the rest of the sample. Finally, we will publish our data and make them as widely accessible as the observance of speakers' and authors' rights allows.

References

Abraham, Werner, & Leiss, Elisabeth (eds). 2008. *Modality-aspect interfaces: implications and typological solutions.* John Benjamins Pub Co.

Almeida, Ashis. *CFILT corpus of Hindi.* `http://www.cfilt.iitb.ac.in/~corpus/hindi/.`

Baker, Kathryn, Bloodgood, Michael, Dorr, Bonnie J., Callison-Burch, Chris, Filardo, Nathaniel W., Piatko, Christine, Levin, Lori, & Miller, Scott. 2012. Use of modality and negation in semantically-informed syntactic MT. *Computational linguistics*, **38**(2), 1–28.

Bhat, D. N. S. 1999. *The prominence of tense, aspects, and mood.* John Benjamins.

Bhatt, Rajesh. 2006. *Covert modality in non-finite contexts.* Mouton De Gruyter.

Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language*, **79**(4), 708–736.

Bickel, Balthasar. 2011. *Towards a corpus-based typology of clause linkage.* `http://www.research-projects.uzh.ch/p15656.htm`. Research Project funded by the SNF since 2011.

Blutner, Reinhard. 2002. Lexical semantics and pragmatics. *Pages 27–58 of:* Hamm, Fritz, & Zimmermann, Ede (eds), *Semantics.* Linguistische Berichte, Sonderheft, no. 10. Hamburg: Helmut Buske.

Bybee, J. L, Perkins, Revere, & Pagliuca, W. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* The University of Chicago Press.

Bybee, Joan L., & Fleischman, Suzanne. 1995. *Modality in grammar and discourse.* Typological studies in language. John Benjamins.

Carretero, Marta, & Zamorano-Mansilla, Juan Rafael. 2013. An analysis of disagreement-provoking factors in the analysis of epistemic modality and evidentiality: the case of English adverbials. *In: Proceedings of the 10th international conference on computational semantics (IWCS), workshop on annotation of modal meanings.*

Chafe, Wallace. 1995. The realis-irrealis distinction in Caddo, the Northern Iroquoian languages and English. *In:* Bybee & Fleischman (1995).

Comrie, Bernard. 1976. *Aspect.* Cambridge Textbooks in Linguistics. Cambridge University Press.

Comrie, Bernard. 1985. *Tense.* Cambridge Textbooks in Linguistics. Cambridge University Press.

Comrie, Bernard. 1986. Conditionals: A typology. *In:* Traugott, E.C., Meulen, A.T., Ferguson, C.A., & Reilly, J.S. (eds), *On conditionals*, 2 edn. Cambridge University Press. first edition 1986.

Condoravdi, Cleo. 2002. Temporal interpretation of modals: Modals for the present and for the past. *Pages 59–88 of:* Beaver, David, Casillas, L., Clark, Brady, & Kaufmann, Stefan (eds), *The construction of meaning.* Stanford, CA: CSLI Publications.

Dahl, Östen. 1985. *Tense and aspect systems.* Blackwell.

Dahl, Östen. 1997. The relation between past time reference and counterfactuality: a new look. *Pages 97–114 of:* Athanasiadou, Angeliki, & Dirven, René (eds), *On conditionals again.* John Benjamins Publishing Company.

Dahl, Östen. 2000. *Tense and aspect in the languages of Europe.* Empirical Approaches to Language Typology. Mouton De Gruyter.

Dahl, Östen. 2007. From questionnaires to parallel corpora in typology. *Sprachtypologische Universalienforschung*, **60**, 172–181.

Dahl, Östen, & Velupillai, Viveka. 2011. The past tense. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online.* Munich: Max Planck Digital Library.

Dixon, R.M.W. 1980. *The languages of Australia.* Cambridge University Press.

Dixon, R.M.W. 2006. Complement clause and complementation strategies in typological per-

spective. *In:* Dixon, R. M. W., & Aikhenvald, Alexandra (eds), *Complementation.* Oxford Univeristy Press.

Dowty, David. 1977. Toward a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and philosophy*, **1**, 45–77.

DuBois, John W., Kumpf, Lorraine E., & Ashby, William J. (eds). 2003. *Preferred argument structure: grammar as architecture for function.* John Benjamins.

Elliott, Jennifer. 2000. Realis and irrealis: Forms and concepts of the grammaticalisation of reality. *Linguistic Typology*, **4**, 55–90.

Fleischman, Suzanne. 1989. Temporal distance: a basic linguistic metaphor. *Studies in language*, **13**, 1–50.

Fleischmann, Suzanne. 1995. Imperfective and irrealis. *In:* Bybee & Fleischman (1995).

Foley, William A. 1986. *The Papuan languages of New Guinea.* Cambridge University Press.

Frawley, William (ed). 2006. *The Expression of Modality.* The Expression of Cognitive Categories, vol. 1. Berlin, New York: Mouton de Gruyter.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency.* John Benjamins.

Giannakidou, Anastasia. 2001. Varieties of polarity items and the (non)veridicality hypothesis. *Pages 99–127 of:* Hoeksma, Jack, Rullmann, Hotze, & Sánchez-Valencia, Víctor (eds), *Perspectives on negation and polarity items.* John Benjamins.

Givón, Talmy (ed). 1983. *Topic continuity in discourse: a quantitative cross-language study.* John Benjamins.

Givón, Talmy. 1994. Irrealis and the subjunctive. *Studies in language*, **18**, 265–337.

Greenberg, Joseph H. 1960. A quantitative approach to morphological typology. *International journal of American linguistics*, **26**, 178–194.

Grosu, Alexander, & Krifka, Manfred. 2008. The gifted mathematician that you claim to be: Equational intensional 'reconstruction' relatives. *Linguistics and philosophy*, **30**, 445–485.

Guérin, Valérie. 2011. *A grammar of Mav̌ea.* University of Hawai'i Press.

Guéron, Jacqueline, & Lecarme, Jaqueline (eds). 2008. *Time and modality.* Springer.

Güldemann, Tom. 2008. *Quotative indexes in African languages: A synchronic and diachronic survey.* Mouton De Gruyter.

de Haan, Ferdinand. 2006. Typological approaches to modality. *In:* Frawley (2006).

Hacquard, Valentine. 2010. On the event relativity of modal auxiliaries. *Natural Language Semantics*, **18**, 79–114.

Hacquard, Valentine, & Wellwood, Alexis. 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and pragmatics*, **5**, 1–29.

Haig, Geoffey L. J., Schnell, Stefan, & Wegener, Claudia. 2011. Comparing corpora from endangered language projects: Explorations in language typology based on original texts. *In:* Haig, Geoffrey L. J., Nau, Nicole, Schnell, Stefan, & Wegener, Claudia (eds), *Documenting endangered languages. achievements and perspectives.* Mouton De Gruyter.

Hansen, Björn, & de Haan, Ferdinand (eds). 2009. *Modals in the languages of Europe.* Mouton De Gruyter.

HCS. 2004. *Helsinki corpus of Swahili.* Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC – IT Center for Science.

Hofling, Charles Andrew. 1998. Irrealis and perfect in Itzaj Maya. *Anthropological Linguistics*, **40**(2), 214–227.

Hughes, George E., & Cresswell, Maxwell J. 1996. *A new introduction to modal logic.* Routledge.

Iatridou, Sabine. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, **31**(2), 231–270.

James, Deborah. 1982. Past tense and the hypothetical: a crosslinguistic study. *Studies in Language*, **6**, 375–403.

Johansson, Stig. 2003. On the role of corpora in cross-linguistic research. *In:* Zanettin, Frederico (ed), *Corpora in translator education.* Manchester: St Jerome.

Kaufmann, Stefan, Condoravdi, Cleo, & Harizanov, Valentina. 2006. Formal approaches to modality. *In:* Frawley (2006).

Kinkade, Dale. 1998. Is irrealis a grammatical category in Upper Chehalis? *Anthropological Linguistics*, **40**(2), 234–244.

Krifka, Manfred. 2012a. Definitional generics. *Pages 372–389 of:* Mari, Alda, Beyssade, Clare, & Del Prete, Fabio (eds), *Genericity.* Oxford: Oxford University Press.

Krifka, Manfred. 2012b. Notes on Daakie (Ambrym, Vanuatu): Sounds and modality. *Pages 46–65 of:* Clemens, Lauren Eby, Scontras, Gregory, & Polinsky, Maria (eds), *Proceedings of AFLA 18 (Austronesian Formal Linguistics Association).* Cambridge, Mass.: Harvard University. `http://westernlinguistics.ca/afla/proceedings/afla18/index.htm`.

Krifka, Manfred. 2013 (04). *Modal and temporal reference in Daakie (Ambrym, Vanuatu).* Talk given at New York University.

Laca, Brenda. 2012. On modal tenses and tense modals. *Pages 171–194 of:* Nishida, Ch., & Russi, C. (eds), *Building a bridge between the linguistic communities of the Old and the New World. current research in tense, aspect, mood and modality.* Cahiers Chronos, no. 25. Amsterdam: Rodopi.

Lehmann, Christian. 2012. *Referential project.* `http://www2.uni-erfurt.de/sprachwissenschaft/referentiality/bilder/proposal_full.pdf`. Research proposal, funded by the Volkswagen Foundation since 2012.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature.* MIT Press.

Margetts, Anna. 1999. *Valence and transitivity in Saliba.* MPI.

Martin, Laura. 1998. Irrealis constructions in Mocho (Mayan). *Anthropological Linguistics*, **40**(2), 198–213.

McEnery, Anthony, & Gaizauskas, Robert. 2003. *Emille corpus.* http://www.emille.lancs.ac.uk/manual.pdf.

McEnery, Anthony, & Xiao, Richard. 2004. *The Lancaster Corpus of Mandarin Chinese.* `http://www.lancs.ac.uk/fass/projects/corpus/LCMC/`.

Mettouchi, Amina, & Vanhove, Martine. *The typology and corpus annotation of information*

structure and grammatical relations. `http://llacan.vjf.cnrs.fr/TCA-ISGR/Home.html`. Research project funded by the French Ministry of Research.

Miestamo, Matti. 2005. *Standard negation: the negation of declarative verbal main cclause in a typological perspective.* Mou.

Myhill, John. 1992. *Typological discourse analysis: quantitative approaches to the study of linguistic function.* Blackwell.

Nam, Seungho. 1994. Another type of negative polarity item. *Pages 3–15 of:* Kanazawa, Makoto, & Piñón, Christopher J. (eds), *Dynamic polarity and quantification.* Center for the Study of Language and Information.

Nuyts, Jan. 2006. Modality: Overview and linguistic issues. *In:* Frawley (2006).

Palmer, F.R. 2001. *Mood and modality.* 2 edn. Cambridge University Press.

Plungian, Vladimir A., & van der Auwera, Johan. 2006. Towards a typology of discontinuous past marking. *Sprachtypologische universalienforschung*, **59**, 317–349.

Portner, Paul. 2009. *Modality.* Oxford University Press.

von Prince, Kilu. 2008. *Attributive linkers in three languages.* M.Phil. thesis, TU Berlin.

von Prince, Kilu. 2011a. *On the Daakaka modal system: Imperatives and futures.* Talk given at the Chronos 10 conference at Aston University, Birmingham.

von Prince, Kilu. 2011b (June). *On the Daakaka modal system: the distal.* Talk at workshop *Speaking of Possibility and Time II*, Universität Göttingen.

von Prince, Kilu. 2012. *A grammar of Daakaka.* Ph.D. thesis, Humboldt University Berlin. Published online: `http://edoc.hu-berlin.de/dissertationen/von-prince-kilu-2012-07-11/PDF/von-prince.pdf`.

von Prince, Kilu, & Rau, Felix. 2013 (May). *Exploring data from language documentation (EDLa).* `http://www.zas.gwz-berlin.de/workshop_edla.html`. Workshop hosted at ZAS Berlin.

Reichart, Roi, & Rappoport, Ari. 2010 (October). Tense sense disambiguation: A new syntactic polysemy task. *Pages 325–334 of: Proceedings of the 2010 conference on empirical methods in natural language processing.* Association for Computational Linguistics, Cambridge, MA.

Ross, Malcolm. 1995. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. *Pages 727–792 of:* Li, Paul Jen-kuei, Tsang, Cheng-hwa, Huang, Ying-kuei, Ho, Dah-an, & Tseng, Chiu-yu (eds), *Austronesian studies relating to Taiwan.* Symposium Series of the Institue of History and Philology, no. 3. Taipei: Academia Sinica.

Rubinstein, Aynat, Harner, Hillary, Krawczyk, Elizabeth, Simonson, Daniel, Katz, Graham, & Portner, Paul. 2013. Toward fine-grained annotation of modality in text. *In: Proceedings of the 10th international conference on computational semantics (iwcs), workshop on annotation of modal meanings.*

Steele, Susan. 1975. Past and irrealis: just what does it all mean? *International Journal of American Linguistics*, **41**(3), 200–217.

Thieberger, Nick. 2006. *A grammar of South Efate: an Oceanic language of Vanuatu.* Oceanic Linguistics Special Publications, no. 33. Honolulu: University of Hawai'i Press.

Thomason, Richmond H. 1984. Combinations of tense and modality. *Handbook of philosophical logic*, **2**, 135–165.

Timberlake, Alan. 2007. Aspect, tense, mood. *Pages 280–333 of:* Shopen, Timothy (ed), *Language typology and syntactic description: Grammatical categories and the lexicon*, 2 edn., vol. 3. Cambridge University Press. first published 1985.

Werner, T. 2006. Future and non-future modal sentences. *Natural Language Semantics*, **14**, 235–55.

Wälchli, Bernhard. 2007. *Original-text typology. exploiting an underexplored data source in linguistic typology. three case studies: nominal number, predicative possession, and motion verbs.* `http://www.isw.unibe.ch/unibe/philhist/isw/content/e4229/e4285/e4500/e4504/e4505/BW-Projekt_ger.pdf`. Research proposal, funded as SNF Förderungsprofessur at Universität Bern since 2007.

Zeitoum, Elizabeth, Huang, Lillian M, Yeh, Marie M, Chang, Anna H, & Wu, Joy J. 1996. The temporal, aspectual and modal systsems of some Formosan languages: a typological perspective. *Oceanic Linguistics*, **35**(1), 21–56.

Zeldes, Amir, Zipser, Florian, & Neumann, Arne. 2013 (February). *PAULA XML documentation: Format version 1.1.*

3 Beantragte Mittel

3.1 Eigene Stelle

Für von Prince beantragen wir eine eigene Vollzeitstelle. Sie wird nicht nur an der Koordination und Auswertung der einzelnen Arbeitsschritte, sondern auch bei deren Durchführung maßgeblich beteiligt sein, was ihre volle Arbeitszeit in Anspruch nehmen wird.

3.2 Personalkosten

Eine Doktorandenstelle mit 65 Prozent Arbeitszeit wird für eine Person mit einem Master in Linguistik, mit Schwerpunkt in der Typologie, ausgeschrieben (NN$_1$). Diese Person wird gemeinsam mit von Prince für die inhaltliche Erschließung der Korpora verantwortlich sein.

Eine weitere halbe Stelle werden wir für eine Person mit Master-Abschluss in der Korpus- oder Computerlinguistik ausschreiben (NN$_2$). Die Hauptaufgaben werden in der Konvertierung der Datenformate, der teilweisen Automatisierung bestimmter Annotationsaufgaben, statistische Analysen und Unterstützung bei der Stratifizierung der Stichproben bestehen. Die detaillierte Aufgabenverteilung ist in Tabelle 1 aufgeführt.

Zusätzlich werden wir eine studentische Hilfskraft mit einer Arbeitszeit von 80h/Monat benötigen, vor allem zur Unterstützung bei Annotationsaufgaben.

Arbeitszeit	Vergütung	Qualifikation	Aufgaben
66%	TV-L E13	LinguistIn (Typologie)	Annotation und Auswertung eines Teils der Korpora
50%	TV-L E13	Korpus-/Computerlinguist	Datenkonvertierung, teilweise Automatisierung, etc.
80h/Monat (32.000 €)	Hilfskraft	StudentIn	Unterstützung bei der Annotation

3.3 Reisen

Das Projekt sieht vor, dass drei der Mitarbeiter einmalig für mehrere Wochen nach Vanuatu fliegen, um die Korpus-Studie vor Ort durch Elizitationen zu ergänzen. Zusätzlich sollen alle Mitarbeiter Gelegenheit haben, auf mindestens zwei Konferenzen pro Jahr zu reisen. Da relevante Konferenzen zu Austronesischen und Ozeanischen Sprachen (z. B. ICAL, COOL, AFLA) regelmäßig im pazifischen Raum stattfinden, müssen auch hierfür höhere Kosten veranschlagt werden.

Manfred Krifka wird einen Großteil der Kosten für Konferenzreisen auch aus anderen Mitteln bestreiten können deshalb veranschlagen wir hier nur Reisen für drei der Mitarbeiter.

Anzahl	Zweck	Gesamtkosten
3	Reise nach Vanuatu für Elizitationen	12.000 Euro
9	Nationale und Internationale Konferenzreisen	20.000 Euro
Gesamte Reisekosten		32.000 Euro

3.4 Projektspezifische Workshops

Für die beiden Workshops zu Beginn und zum Ende des Projekts werden wir die assoziierten Sprachexperten Michael Franjieh, Nick Thieberger, Anna Margetts und Valérie Guérin einladen. Mindestens Thieberger und Margetts werden aus Australien anreisen. Guérin arbeitet derzeit in den USA, Franjieh in England.

Wir veranschlagen pro Workshop 7000 Euro, insgesamt 14.000 Euro.

4 Voraussetzungen für die Durchführung des Vorhabens

4.1 Angaben zur Dienststellung

- Krifka, Manfred: Professor, HU Berlin. Für die Mitarbeit im Projekt sind 4 Stunden/Woche vorgesehen.
- von Prince, Kilu: Wissenschaftliche Mitarbeiterin, ZAS Berlin;
  Förderung durch die Volkswagenstiftung bis 30. 06. 2013;
  weitere Anstellung am ZAS (PB3) vorgesehen bis 31. 12. 2013;

4.2 Angaben zur Erstantragstellung

Erstantragstellerin: von Prince, Kilu

4.3 Zusammensetzung der Projektarbeitsgruppe

Mitarbeiter, die nicht aus dem Projekt finanziert werden: Professor Manfred Krifka (Lehrstuhl am Institut für deutsche Sprache und Linguistik)

4.4 Zusammenarbeit mit anderen Wissenschaftlerinnen und Wissenschaftlern

4.4.1 Wissenschaftlerinnen und Wissenschaftler, mit denen für dieses Vorhaben eine konkrete Vereinbarung zur Zusammenarbeit besteht

Wir werden zum einen mit den Autoren der untersuchten Korpora zusammenarbeiten:

Thieberger, Nick	University of Melbourne, Australia
Franjieh, Michael	School of Oriental and African Studies, UK
Margetts, Anna	Monash University, Australia
Guérin, Valérie	North Carolina State Univeristy, USA

Darüber hinaus bestehen Vereinbarungen zur Zusammenarbeit mit Julie Barbour (Waikato University, Neuseeland), die vor kurzem ein Projekt zu Modalität in den Sprachen Vanuatus in die Wege geleitet hat. Die Gefahr der Redundanz durch die beiden Projekte besteht nicht, da sich Barbours Projekt auf grammatische Beschreibungen aller bisher erschlossener Vanuatu-Sprachen konzentriert, und da Tempus, Aspekt und Polarität in ihrem Projekt keine spezifische Rolle spielen. Wir erwarten im Gegenteil, dass sich die beiden Projekte sehr gut ergänzen werden.

Anke Lüdeling (HU Berlin) hat zugesichert, uns gegebenenfalls als Ansprechpartnerin bei korpus-linguistischen Fragen und Problemen zur Seite zu stehen und die Konvertierung der Daten nach PAULA XML, sowie die Veröffentlichung auf ANNIS zu betreuen.

4.4.2 Wissenschaftlerinnen und Wissenschaftler, mit denen in den letzten drei Jahren wissenschaftlich zusammengearbeitet wurde

- Mitarbeiter des ZAS Berlin
- Prof. Alexander Grosu, Tel Aviv University
- Prof. Ariel Cohen, Ben-Gurion University of the Negev
- Prof. Hans-Martin Gärtner, Hungarian Academy of Sciences

5 Interessenkonflikte bei wirtschaftlichen Arbeiten

Nicht zutreffend.