Universität Zürich[UZH]

Masterarbeit

zur Erlangung des akademischen Grades

Master of Arts UZH

der Philosophischen Fakultät der Universität Zürich

A quantitative analysis of grammaticalization in Chintang the case of benefactive compound verbs

Verfasser: Rolf Hotz

Betreuer: Prof. Dr. Balthasar Bickel

Institut für Vergleichende Sprachwissenschaft

Abgabedatum: 31. Dezember 2019

Abstract

Several complex predicates in Chintang are ambiguous in the sense that the second verb (v2) may express the lexical, etymological meaning of the recruited verb (also term additive meaning, e.g. *to V and bring*), as well as to operate as a grammatical device. A large part of the literature on grammaticalization predicts both a double change (phonological erosion and semantic bleaching) and an abstraction of the meaning. The coexistence of the lexical and the grammaticalized form is usually considered a transitory phase that will end with the prevalence of the latter.

In order to test these assumptions with quantitative data, we analyzed three benefactive v2s in Chintang, all of which come from independent verbs that are still in use in the language. Our aim was thus to measure the transition from an independent verb to a grammatical device. Of these three benefactives, two had an ambiguous meaning and none of them had undergone phonological reduction.

Our first approach was to compare the translation to Nepali of each benefactive. The data was taken from the Chintang corpus. The second approach was a cloze test, in which native Chintang speakers were asked to *complete the sentence* with the first thing (noun) that came to their minds upon hearing a conjugated verb with: a) one of these three independent verbs, e.g. *I give it to you* and b) the same three lexemes as benefactives (with different host verbs), e.g. *I cook it for you* (lit. I cook give to you).

The first approach suggested a connection between one benefactive (*khutt*) and its etymological origin (*khutt* 'to bring') due to the use in Nepali of one particular construction approximately every third time in a randomly selected sample. The second approach suggested that two benefactives become more abstract (measured with entropy) as a grammaticalized benefactive while a third (*khutt*) does not. However, after merging semantically similar nouns (types) in groups and excluding outlier participants, the results of *khutt* displayed a higher abstraction as a v2 than as an independent verb, similarly to the other two lexemes. These results seem to confirm the claim from the literature that grammaticalization entails an abstraction of meaning. However, they do not suffice to explain ambiguous complex predicates.

Concerning the methodology, entropy and PMI proved to be versatile and effective tools. Their interplay also revealed a positive correlation between polysemy and abstraction that should be further explored. However, these measurements also had their limits, as they do not discriminate between lexical and grammaticalized v2s and thus the augmentation of abstraction cannot be precisely assigned to the effect of the former or the latter. Finally, the use of semantic groups and the exclusion of outlier participants hindered possible misinterpretations.

Keywords

grammaticalization, Chintang, abstraction, benefactive, complex predicate

Contents

Abbreviations

1	first person
2	second person
3	third person
A	agent
ABS	absolutive
ACRSS	metaphorical movement across
ACT.PTCP	active participle
ADD	additive
AFF	affirmative
AFFECT	affect adversely
AGT	agentive
AMB	ambulative
ANAPH	anaphoric demonstrative
ASS	assertive
ATTN	attentional
AUTOBEN	autobenefactive
AUX	auxiliary
AWAY.ITR	intransitive move away
AWAY.TR	transitive move away
BEN	benefactive
CAUS	causative
CELER	celerative
CIT	citative
COM	comitative
COMPL	completive
COMPL.TR	completive transitive
CON	conative
CVB	converb
DAT	dative
DIR	direction
DU	dual
DUR	durative

DEM.ACRSS	demonstrative pointing across
DEM.DOWN	demonstrative pointing down
DEM.UP	demonstrative pointing up
DESID	desiderative
DIST	distal
DIR	direction
e	exclusive
EMPH	emphatic
ERG	ergative
FOC	focus
GEN	genitive
i	inclusive
IMP	imperative
INC	incompletive aspect
IND	indicative
INF	infinitive
INSIST	insistive
INST	instrumental case
INTENS	intensification
IPFV	imperfective
ITR	intransitive
LNK	linker
LOC.ACRSS	locative pointing across
LOC.DOWN	locative pointing down
LOC.UP	locative pointing up
LOC	locative
MAL	malefactive
NEG	negation
NMLZ	nominaliser
NRL	non-relational prefix
NOM	nominative
NPST	non-past
ns	non-singular
NTVZ	nativiser

O	object (argument linked to O-AGR)
OBL	oblique case
OPT	optative
P	patient
p	plural
POSS	possessive
PRF	perfect
PROX	proximative
PST	past
PTCL	particle
PTCP	participle
PURP	purposive
PVB	preverb
Q	question tag
REAL	realis
s	singular
S	subject
SEQ	sequential
SIM	simultaneous converb
SUBJ	subjunctive
SURP	surprise
TEL	telic
TERM	terminative
TOP	contrastive topic
TR	transitive

1 Introduction

1.1 Ambiguous compound verbs in Chintang

Multi-verbal constructions are pervasive in many languages of the world. The verboid elements following the host-verb may carry either a grammatical device that alters the meaning of the main verb or a lexical one, which adds a new action. It is not uncommon in these constructions for these grammatical devices to have their origin in full-fledged, independent verbs that acquired a grammatical sense through a process known as grammaticalization. The literature on grammaticalization dealing with multi-verbal constructions describes the recruitment of an element for a complex predicate as an in-between stage: before an independent verb is fully integrated (phonological reduction and semantic bleaching) as a grammatical device to the morphological paradigm, it may still relapse into its original meaning, i.e. convey two senses: the lexical and the grammaticalized one (cf. Hopper, 1991; Coupe, 2008: 300; Anderson, 2006: 305f.). Thus a state of ambiguity is created, since two meanings cohabit the same suffix.

In this regards, Chintang[1] (Kiranti, Nepal) seems odd, as ambiguous verbal suffixes abound in the language (cf. Bickel & Zúñiga, 2017: 178), cf. the following examples:

(1) Chintang[2]

ʄo=go=yaŋ *na-khutt-i-**ca**-i-hatt-i-bir-i.*

whatever=NMLZ=also 3>2-steal-2pP-**eat**-2pP-AWAY.TR-2pP-BEN-[SUBJ.]2pP

'It (a cat) may steal everything from you and eat it!' (story_cat.204)

(2) *Kina dasai* *a-man-e* *a-numd-o* *kina a-kaʔ-na-**ca**-no?*

SEQ Dasain 2S/A-observe-V.NTVZ 2[s]A-do-[SUBJ.NPST.]3[s]O SEQ 2[s]S-come.up-LNK-**COMPL**-IND.NPST

'So when you are to observe Dasain (a festival) you come up?' (DR_exp.0762)

[1]I would like to thank the Swiss Society for Endangered Languages for their generous support of my fieldwork in Nepal. I am also thankful to my supervisor Prof. Dr. Bickel for having suggested to me this topic as well as Dr. Sauppe, for his help in the development of experiment 2. I am further grateful to Dr. Schikowski, who took the time to teach me Chintang and to the family of Dāl Bahādur Rāī, with whom I stayed in Chintang.

[2]The source of the language examples is always at the bottom right. For Chintang, most of the examples were taken from the corpus (cf. Section 2.1) and each code (here e.g. **story_cat.204**) comes therefrom.

(3) *Yo lo lo luŋtak kopt-a-ŋ-**khutt**-a-hã na!*
DEM yes yes stone pick.up-IMP-1sP-**bring**-IMP-2>1 TOP

'Yes, yes, pick up that stone and bring it for me!' (adapted from CLLDCh4R02S02b.0479)

(4) *Mai-khu-na-**khu**ʔ-no mahãʔ naŋ?*
1nsiP-carry-LNK-**BEN**-IND.NPST no but

'It (a bus) carries it for us, right?' (adapted from chintang_now.0325)

In example 1. and 2., *-ca* can be paraphrased as *to V and eat*, which is equivalent to its meaning as an independent verb *cama* 'to eat', whence it was recruited. In the second example, *-ca* has an completive aspectual value, which affects the host-verb *ka* 'to come up'. These two meanings of the same compound verb coexist in the language. Furthermore, in example 3., *khutt* stands for 'to do V and bring' whereas it is a benefactive in example 4. These four sentences should help exemplify the versatility of verbal suffixes in the language.

In the present research, the aim is to contribute to the understanding of such ambiguous verbal suffixes. For that aim, we shall cover a standard assumption in grammaticalization theory, namely that a verb has to become more abstract once incorporated as a complex predicate. This idea is part of the doxa of grammaticalization, that considers that an element (in this case a verb) that is recruited as a grammatical device is bound to lose its lexical properties and eventually be assimilated into grammar. Taking Chintang's benefactives as an example, we conducted two tests to comprehend the relation between independent verbs and their grammaticalized counterparts. The first test is a comparison of how Chintang's four benefactives are rendered when translated to Nepali in a corpus. The second test is an experiment of free associations, in which participants had to come up with a noun upon hearing a) sentences with one benefactive (in a complex predicate) and b) sentences with the same benefactives but as independent verbs. Th answers (nouns) were measured in terms of entropy.
The main implications of the present investigations are that, i) even though a benefactive may be perceived (through translations) as having much of the semantics from its etymological origin left, being recruited as a compound verb and acquiring a grammaticalized meaning seems always to convey an abstraction of the meaning, the latter understood as an expansion of contexts (in the form of nouns) where it can occur or be associated. This is measured using entropy. As such, the notion of abstractness is insufficient in making sense of ambiguous complex predicates, while the explanation from grammaticalization theory, which describes ambiguity as an in-between stage is also not satisfactory, at least in

the case of Chintang. ii) The second implication is that, when dealing with free associations, semantic groups have to be accounted for, as the lack thereof may mask the results to a considerable degree. Lastly, iii) the third insight is that the augmentation of entropy goes along with an augmentation of high values in PMI. These two measurements suggest that an abstraction of meaning entails a higher number of strong associations.

The remainder of this investigation is organized thus: Section one will introduce the language and its complex predicates, as these are the main topic of this investigation. Furthermore, it will describe two issues, which are central to follow the argument: grammaticalization theory and the concepts of abstraction and polysemy. Section two and three build the main part. There, the two experiments that were conducted are described and the results presented and discussed. Section four mentions the obstacles, alternative explanations and the broader relevance of the study. Section five closes the paper.

1.2 The language

Chintang is a Kiranti (Tibeto-Burman) language from Eastern Nepal spoken by approximately four thousand people in and around two villages in Dhankuta District, Eastern Nepal. Most speakers (if not all) of Chintang are also fluent in Nepali (Indo-Aryan), Nepal's national language and Bantawa (Kiranti), formerly a lingua franca in the area. Within the language family, Chintang belongs to the Eastern branch, along with Athpare, Belhare (Bickel et al., 2010: 382), Yakkha (Schackow, 2015: 32) and probably also Chiling. It should be noted that the classification of the larger language family is a largely unsettled debate (cf. van Driem, 2014).

Most inhabitants of the Chintang area practice subsistence farming; rice, oranges and ginger are the most common products. Most households will also have goats, pigs and oxen. The latter are used to plough the steep fields. Some of these products are sold on a regional market, albeit at a small scale.

People in Chintang practice a set of rituals and traditions, which are akin to other Kiranti groups in Nepal and are strongly influenced by local Hinduism. These practices are sometimes referred to as animism (van Driem, 1987: 12), others as Kiranti religion (cf. Gaenzle, 2016) and it can be connected with a larger continuum of religious manifestations in Asia, sometimes subsumed under the broad term *shamanism*. In Nepal, these practices survive today in small, scattered areas (van Driem, 2002: 104; Hitchcock, 1967), Chintang among them.

There is strong on-going trend to stop passing Chintang on to the newer generation and Nepali has become the main language to most children below the age of ten. This situation is not uncommon in Nepal, where linguistic diversity is fading away at an accelerated pace (cf. Turin, 2007; Borchers, 2007: 12; van Driem, 2002: 600f.). In the present, however, Chintang still remains the main language among adults.

Chintang has been the focus of much linguistic research due to its complex verbal system (especially its agreement system also known as biactantial verbal agreement, perhaps the hallmark of the Kiranti language family), its deictic system (Dirksmeyer, 2008), interchangeability of prefixes (Bickel et al., 2007), object-conditioned differential marking (Schikowski, 2013) and there are also studies on language acquisition (Stoll et al., 2012). There is one printed dictionary (Rai et al., 2011) and a larger non-printed version. Lastly, there is a large corpus, which will be described at length in Section 2.1.

1.2.1 Complex predicates in Chintang

Multiverb constructions, i.e. strings of more than one verb, are a widespread phenomenon in the world's languages. For Aikhenvald (2012: 304f.; 2006: 5f.), the one feature which holds the whole concept together is their monoclausality, that is to say, each string of verbs has to make up a single predicate to fit the definition. This makes multiverb construction a broad term which encompasses a great deal of variation, e.g. in morphology (inflection), the presence of a linker between individual verbs or some sort of subordination.

Within this category, serial verb constructions and converbs have become common concepts when dealing with combinations of two or more verbs. The main distinction between these two is usually explained in the following terms: converbs take a sort of linkage and they cannot appear without the main verb (nonfinite), while the individual verbs of a serial verb construction can appear on their own (cf. Haspelmath, 1995: 3f.; Bisang, 1995: 139f.; Croft, 2012: 346f.; Drude, 2011: 220 for Awetí).

Some authors have questioned the meaningfulness of the distinction between converbs and serial verb construction (Shibatani, 2009: 259f.) and the usefulness of the converb as a valid category for comparison between languages altogether (Zúñiga, 1998). For the following section, we will use the more neutral *multiverbal construction* and *compound verb* for the Kiranti context.

For the rest of this thesis, we will refer to the first (left-most) verb in a multiverbal construction as v1 (verb1), while the last one (right-most) is henceforth v#. In a construction with three chained verbs, the first will thus be v1, the second v2 and the last one v# (or v3).

Multiverb constructions are present in all members of the Kiranti language family. According to Doornenbal (2009: 473f.), they have enough similarities to describe them under one single term: Kiranti compound verbs. For the author, who has written by far the most complete comparative analysis of multiverb constructions in this language family, Kiranti compound verbs are an asymmetrical sequence of two or more verbs where the first verb adds both the main semantic content and the argument structure while the following verb or verbs adjust the meaning by providing additional information. Doornenbal takes notice of the fact that despite considerable variation in phonology and lexicon displayed in the language family, compound verbs are remarkably similar. In terms of morphology, the author takes the

appearance of overt morphology on v2 and v1 as the main parameter to divide the Kiranti language in three groups. On the bottom are those languages where v1 takes either no suffix (described as strong suffix reduction) or an invariable form. As an example we could cite the case of Thulung (example 5.), but this could also be applied to Koyi (cf. Lahaussois, 2009: 16f.) or Jero (cf. Opgenort, 2005: 203).

(5) Thulung (Kiranti)
Krib-∅-jöl-ci.
cut-∅-put.down-PST.DU

We/you/they (dual) cut down. (adapted from Doornenbal (2009: 477))

On the top of the scale there is Limbu, which, even though v1 does not take the whole suffixal chain (as verb# does), prefixes are present on all verbs of the compound, not only on v1 (Doornenbal, 2009: 475f.; Tumbahang, 2017: 154f.), cf. the following examples:

(6) Limbu (Kiranti, Nepal)
Kɛ-dza-m kɛ-sur-u-m-aŋ ta-ʔɛ.
2-eat/3P-pA 2-finish-3P-pA-and show.up-1SPS/NPST

I'll show up when you have finished eating. (adapted from van Driem, 1993: 119)

(7) *Ku-dhge:k-ʔin mɛ-bhund-u mɛ-bu:r-u.*
3POSS-head-ABS NSAS-dislodge-3P nsAS-give-3P

They lopped its head off. (adapted from van Driem, 1993: 128)

In these examples, we observe verbal prefixes being shared between v1 and v2. This is the only language in the Kiranti family to allow this.
The Chintang compound verb is very similar to the cognate constructions in the Kiranti family. They are complex verbs in which each element shares part of the final suffixal chain on verb#. All affixes apply to the whole compound even if they are not present on all elements. Some suffixes are repeated on each element of the compound while prefixes only appear once. Chintang only allows clitics to intervene between the elements of the compound, as in example 8., where *=yaŋ* is attached to the interverbal domain:

(8) Chintang

Mai-ca-na=yaŋ-bi-na-hai-kha *aŋ.*

1nsiP-eat-LNK-ADD-BEN-LNK-COMPL.TR-NMLZ Q

'He might eat us'. (CLLDCh1R01S04.394)

All elements of the compound agree in TAM, polarity and valency. This point is not trivial, since, as noted by Schikowski (2018: 71), each verb in a compound may trigger a different valency-pattern as v1. In example 9., v1 *kop-* 'pick up' is monotransitive and its P is linked to O-agreement. Conversely, v# *khutt* 'bring' is ditransitive and its G is linked to O-AGR, i.e. the one affected by the act of being brought something. In example 9., instead of each verb operating with its own valency agreement, the valency of *khutt*, the final suffixal chain, is imposed over the whole compound:

(9) *Yo* *luŋghek* *kob-a-ŋ-khutt-a-h-ã=na.*

DEM.ACRSS stone pick.up-IMP-1sO-bring-IMP-1sO-IMP[.2sA]=INSIST

'Pick up that stone and bring it to me'. (Schikowski 2018: 71; CLLDCh4R02S02b.419)

In the scale of overt morphology suggested by Doornabel (2009: 475), Chintang can be placed in the middle along with Bantawa, Athpare (Ebert, 1997b: 63f.), Chamling (Rai, 2012: 127; Ebert, 1997a: 34f.) and Dumi (van Driem, 1993b: 201), since prefixes are never repeated and part of the final suffixal chain is repeated in interverbal position.

On what concerns prefixes, Chintang allows free permutation of inflectional prefixes, which does not alter the semantic reading of the compound verb. In example 10., two prefixes may change order while in example 11. they are three. More than three prefixes are not allowed in the language (cf. Bickel et al., 2007: 44):

(10) a. *a-ma-im-yokt-e.*

2-NEG-sleep-NEG-PST

b. *ma-a-im-yokt-e.*

NEG-2-sleep-NEG-PST

Both: 'You didn't sleep.' (Bickel et al., 2007: 44)

(11) a. *u-kha-ma-cop-yokt-e.*
3NS.A-1NS.P-NEG-see-NEG-PST

b. *u-ma-kha-cop-yokt-e.*
3NS.A-NEG-1NS.P-see-NEG-PST

c. *kha-u-ma-cop-yokt-e.*
1NS.P-3NS.A-NEG-see-NEG-PST

d. *ma-u-kha-cop-yokt-e.*
NEG-3NS.A-1NS.P-see-NEG-PST

e. *kha-ma-u-cop-yokt-e.*
1NS.P-NEG-3NS.A-see-NEG-PST

f. *ma-kha-u-cop-yokt-e.*
NEG-1NS.P-3NS.A-see-NEG-PST

All: 'They didn't see us.' (Bickel et al., 2007: 44)

This phenomenon has interesting consequences for compound verbs, since v2s may also host prefixes, which are usually found before v1. Cf. the following example:

(12) a. *u-kos-a-gond-e.*
3-NS.S-walk-PST-AMB-PST

b. *kos-a-u-gond-e.*
walk-PST-NS.S-AMB-PST

Both: 'You didn't sleep.' (Bickel et al., 2007: 51)

In example 12., the prefix *u-*, which affects the whole compound, may be attached to v#, after the interverbal element *-a*. This example also allows us to describe a further component of complex verbs in Chintang, namely the morphophonological processes at interverbal level.

As described by Bickel et al. (2007: 49f.; also Zimmermann, 2012: 354f.), more than one verbal stem can be joined in a compound verb (v1 plus up to three added stems). The stem of v2 is described as having a prosodic subcategorization constraint that requires a disyllabic as a host. Since regular stems are largely monosyllabic, v1 needs to be augmented with an inflectional suffix or, in the absence of such a device,

with an epenthetic element, which in turn creates a disyllabic foot. Now a v2 stem may be added which will host all required inflectional suffixes. The stem of v1 is partially inflected, hosting only one suffix, which is also present in the whole (final) suffix string hosted by v2.
It should be noted that several phonotactic rules and allophony takes place, which blur the idealized string of affixes. For this reason, and for the sake of clearness, some glosses have an extra level displaying isolated elements, as in the following examples:

(13) *Hana akhause.*
hana a-khaŋ-u-ŋs-e
2s 2sA-see-3P-PRF-IND.PST

'You have seen it'. (CLLDCh3R10S03.017)

(14) *To coptoho to.*
to copt-u-khaŋ-u-a to
DEM.UP look.at-3P-CON-3P-IMP[3S>3S] DEM.UP

'Look up there'. (adapted from CLDLCh2R01S01.001)

In example 13. and 14., we observe the reduction of some consonants and vowels. Even though some phonotactical rules are always respected (such as the deletion of the second consonant in a row of three consonants), it must be noted that there is a large degree of interspeaker and generational variation with older speakers being more conservative and thus displaying less fusion. There is also variation concerning the addition of a nasal element to cover the 3sP suffix *-u*, which is *-ŋ* for 1sA and *-m* for 1plA, as in the following example:

(15) *Dui tin taukha numduŋbiduŋcuhẽ gona.*
dui tin thaũ-kha numd-u-ŋ-pid-u-ŋ-ce-u-e gonei
two three place-NMLZ do-3P-1sA-BEN-3P-1sA-3nsP-3P-IND.PST ATTN

'I did (them) at two or three places'. (pear_3-1.0153)

In example 15., the speaker included a nasal element (1sA) after 3sP *-u*, which is regularly dropped by younger speakers. In a similar fashion, after the past suffix *-a*, there is always a nasal morpheme in inter-verbal position.

The next point about reclusive inflection is that morphological material is recruited from v# to the inter-verbal domain. In most cases, it is the first suffix to appear in the last element of the compound verb, i.e. the first suffix of the final suffix string of the verb complex that appears in the inter-verbal domain. Cf. the following example, where *-u* after the benefactive also appears between the benefactive and v1.

(16) *Hui jamma umeĩ ukoptubidoko.*
hun jammai u-meĩ u-kopt-u-pid-u-ko
DEM all 3s.POSS-thing 3nsS/A-pick.up-3P-BEN-3P-IND.NPST

'Four people collect all his things'. (pear_1-1.0024)

Another relevant rule is that a nasal element is always included after the past tense suffix *-a*, as in the following example:

(17) *Aseĩ suptuŋpiduŋsuhẽ.*
aseĩ supt-u-ŋ-pid-u-ŋs-u-ŋ-e
last.time clean-3P-1sA-BEN-3P-PRF-3P-1sA-INS.PST

'I cleaned last time'. (adapted from CLDLCh2R06S03.266)

There are, however, two exceptions to the rule of recursive structure described above. The first is the non-past *–no*, which is directly attached to the stem. The second is the form of the third person singular subjunctive, for which there is no suffix at all. In both cases, we encounter an epenthetic element *-na* (also *-naya*), which only appears in this context:

(18) *Epnalonno.*
ep-na-loĩs-no
stand.up-LNK-TEL-IND.NPST

'S/he stands up'. (CLDLCh1R01S02.0032)

(19) *Duda thuŋ-na-yak-lok=yaŋ khic-e u-numd-o-ko.*
milk drink-LNK-IMPF-[SUBJ.NPST.3sS.]CVB=also take.photo-V.NTVZ 3[p]A-do-3[s]O-IND.NPST

'They even record him while he's drinking milk (from my breast)'. (adapted from Schikowski, 2018: 60)

Schikowski (2018, 71f.) adheres to this analysis but adds one change. If we look the following example, it seems as if the speaker would have skipped the expected *-ŋ* (as in *-u-ŋ-ku-ŋ-ta*) in the final suffixal string, a nasal consonant that we find in interverbal position:

(20) *Khur-u-ŋ-gond-u-ku-ŋ-ta.*
carry-3O-1sA-AMB-3[s]O-IND.NPST-1sA-IPFV

'I am carrying it around'. (Schikowski, 2018: 71)

The author suggests that such forms are not an exception, as it is the first syllable of the suffix chain of verb without the indicative marker, in this case *-u-∅-ŋ*, that is copied. Therefore, the first suffix of the suffix chain in v# is not simply repeated in each interverbal post domain; it is the first syllable without the indicative marker. It comes, therefore, as no surprise that *-no* (IND.NONPST) triggers the same form as the third person singular subjunctive, a form without a suffix (-∅).
Furthermore, he adds other suffixes which trigger the linker *-na* instead of suffix copying, namely *-ning* (PST.SUBJ), *-lok* (CVB) and -ne (OPT), as in the following example:

(21) *Thapi-be* *tha=go* *arkha-ko* *u-sam*
lid-LOC come.down[.SUBJ.NPST.3sS]=NMLZ alcohol-GEN 3s.POSS-steam
chop-na-haʔ-niŋ
dry-LNK-AWAY.ITR-[SUBJ.]NEG.NPST.[3sS]

'The alcohol steam condensing on the lid should not dry'. (adapted from Schikowski, 2018: 72)

Finally, there are three v2-like elements that do not quite behave like it has been described above. The first element is the negative past *-yokt* (also *-yakt*), which co-occurs with the prefix *ma-*. It does not trigger the expected post domain in the preceding verb, as normal v2s would do in Chintang. However, it can still host prefixes and clitics (Schikowski, 2018: 78f.). In example 22., *-yokt* is directly attached to v1:

(22) *Uhũ* *akka* *makhemyoktaŋsehẽ*
ahã akka mai-khems-yokt-a-ŋ-ŋs-e-h-ẽ
INTJ 1s NEG-listen-NEG-PST-1sS/P-PRF-PST-1sS-IND.PST

'I haven't heard anything'. (chintang_sahid.129)

(23) *Khoi* *umapitayoktaŋsaŋnɨhẽ*
khoi u-mai-pid-a-yokt-a-ŋs-a-i-hẽ
I.don't.know 3nsS/A-NEG-give-PST-NEG-PST-PRF-PST-1/2pS/P-PST

'But, they have not provided me yet'. (Schikowski, 2018: 78; kothari_talk.012)

The second element is the negative *-t*, which does not host prefixes and endoclitics. However, it may trigger an interverbal post-domain on the preceding verb, cf. example:

(24) *Meĩ* *boŋkoloŋ* *yaŋ* *maiucotokẽ*
meĩ boŋkoloŋ yaŋ mai-u-ca-u-th-u-k-ẽ
thing peach ADD NEG-3nsS/A-eat-3P-NEG-3P-IPFV-IND.PST

'They did not eat that peach also'. (CLLDCh4R04S06.1548)

The last element is *-watt*, which is used with sudden events and motion. Even though it follows v1 and modifies it as a standard verb2, it does not elicit suffix copy on the preceding verb, but the whole suffixal string, which is then itself repeated after *-watt*, as in the following example:

(25) *Gorce* *ghãsa* *bhuktuce* *wattuce* *kina*
goru-ce ghãs-a bhukt-u-ce-e watt-u-ce-e kina
ox-NS grass give-3O-[3sA.]3nsO-IND.PST CELER-3O-[3sA.]3nsO-IND.PST SEQ
lasadaŋse
las-a-d-a-ŋs-e
return-PST-AWAY.ITR-PST-PRF-IND.PST[.3sS]

'(He) quickly gave grass to the oxen and returned'. (CLDLCh3R04S07.125; Schikowski, 2018: 79)

However, *-watt* is rather uncommon and further research may elucidate its syntactic and morphological properties.

The origin of reclusive inflection

Concerning the historical origin of recursive inflection, Ebert (1997a: 64) suggested that suffix copying was the result of an historical development in which the inflection of v2 was gradually imitated in v1. The preference for inflection in both elements over converbs and participles in subordination would be

related to an areal phenomenon (also cf. Steever, 1988: 71f.; Hock, 1991 and Anderson, 2006: 144f. for an areal perspective).

A different approach is held by Jacques (2018; also Jacques et al., in press: 9), who argues that Kiranti compound verbs have their origin in full-inflected serial verb constructions that fused together. To sustain this claim, he draws on examples from Japhug, a Rgyalrong language that has four different constructions that are relevant for his analysis. These are the following:

1) Bipartite verb – lexical units comprising two non-integrated verb stems, both of which can be (at least partially) conjugated. The verbs that can take the second position make up a small class with only ten verbs. The author describes a gradient of allowed conjugation, ranging from the full set of affixes on both verbs, through the same prefix on both verbs but one suffix on v2, through one prefix on v1 but the same suffix on both verbs to only v1 hosting prefixes and only v2 hosting suffixes. A typology of bipartite verbs is described in Table 1. for the case of the imperative second dual "try hard (both of you)":

Table 1: Bipartite verbs (Jacques, 2018: 175)

Type	Example	v1 suffix	v2 prefix
A (quasi-SVC)	*tɤ-stu-ndʑi tɤ-mbat-ndʑi* IMP-V1-DU-IMP-V2-DU	✓	✓
B (right-dominant)	*tɤ-stu=tɤ-mbat-ndʑi* IMP-V1=IMP-V2-DU		✓
C (left-dominant)	*tɤ-stu-ndʑi=mbat-ndʑi* IMP-V1-DU=V2-DU	✓	
D (semi-compound)	*tɤ-stu-mbat-ndʑi* IMP-V1-V2-DU		

2) Compound verbs (not to be confused with Kiranti compound verbs) are similar to bipartite verbs but they do not allow multiple indexation (as in inflectional morphology between the both roots) along with other constraints not relevant for the present work. Most compound verbs in Japhung are described as additive.

3) Serial verb constructions - In these constructions, both v1 and v2 have the same arguments, transitivity, TAM, polarity and associated motion marker, i.e. they have the same affixes. They are constructed mainly with deideophonic verbs and action deixis verbs.

4) Two joined verbs share the same prefix and suffix, which are not repeated on each verb. There is thus no intervening suffix between verbs.

The author argues that bipartite verbs 1) in Japhug are transparently derived from serial verb con-

structions. They would still be in the process of grammaticalization, i.e. these combinations are being lexicalized, which explains the variation found concerning overt affixal morphology, as sometimes v1 loses its suffix and/or v2 its prefix. Compound verbs may also be a possible origin for bipartite verbs of type D.

The author then argues for the same development in Kiranti, which only has bipartite verbs akin to the Japhug type C and D (i.e. Kiranti compound verbs) but with only part of the entire suffixal chain in interverbal position. In this light, suffix variability of prefixes is taken to be a remnant of this forgoing serial verb construction, as two serial fully inflected verbs fused together preserving, in the case of languages like Bantawa or Chintang, the prefix of verb2. The integration of both verbs hindered, however, the appearance of both prefixes in most Kiranti languages (cf. Jacques, 2018: 184f.). This early construction has had different developments in the language family. Limbu, which retains the prefixes on both verbs, seems closer to this early fusion of serial verbs, while languages like Bantawa are in an intermediate position, for their compound verbs may split up in two phonological words, but the replication of the affixal strings is only partial. Khaling would be more innovative in this respect, since it has a fixed order of affixes, affixal copying is only partial and compound verbs always form one phonological word (Jacques, 2019: personal communication).

To conclude, complex predicates in Chintang can safely be framed as Kiranti. There is, however, a high degree of fusion and allophony in its morphology. These processes are on-going in Chintang due to language change.

Reclusive inflection (suffix copy) is a pervasive phenomenon in the language, however not unknown in the larger area. Concerning its historical origin, the examples from Japhug suggest a possible path of development.

1.2.2 Benefaction

In order to approach the main topic of this research, we have to take a detour in form of a typological introduction to benefaction.

As defined by Kittilä and Zúñiga, the benefactive marks "a participant that is advantageously affected by an event without being its obligatory participant (either agent or primary target, i.e. patient). Since normally only animate participants are capable of making use of the benefit bestowed upon them, beneficiaries are typically animate" (2010: 2; cf. Lehman et al., 2010: 68 and cf. Van Valin & Lapolla, 1997: 382f. for a similar analysis). This broad definition allows the authors to incorporate many formally different constructions and to sort out similarities for a meaningful comparison. From their analysis follows a set of generalizations that are also relevant for our topic, namely: a) typically, benefactives are optional, they can be omitted; b) beneficiaries are neither i) agents nor ii) primary targets of events,

as events usually affect their participants in a neutral way and the i) agent may also benefit from them (e.g. I brushed my teeth). This is, however, not the case in a canonical benefactive construction (as *in X does something for Y*), where the agent is usually different from the beneficiary. Some languages (Chintang among them) have a special construction for autobenefactives. On the same line, beneficiaries are not ii) primary targets of the event from which the benefit, as they benefit rather from the result of the event (compared to patients)[3]; c) many languages do not formally distinguish between beneficiaries and maleficiaries, as it the case with Chintang's benefactives, which encode both and thus they rely on context; and lastly d) recipients of benefaction are usually animate (2010: 4–7).

For the context of Kiranti languages, we could follow Schackow (2019: 492f.) and also highlight intentionality and agentivity, in other words, the intention of creating a desired effect on somebody as a semantic component of benefactive constructions. These characteristics largely apply to benefactives in Chintang,

Benefaction is expressed in a plethora of ways in the world's languages. On the formal mechanisms for coding beneficiary, the most prominent ones according to the authors are: case, adpositions, applicatives and serial verb constructions. The last strategy corresponds to our four benefactives, as serial verb constructions are related to (if not the hypernym of) compound verb. The verb 'to give' is usually recruited for this purpose and it may lose its original semantics and become fully grammaticalized (Kittilä & Zúñiga, 2010: 3f). How exactly benefaction is expressed in Chintang will be the topic of the following section.

A relevant aspect is the variation found regarding the semantic specificity of the devices, a non-trivial issue for Chintang where we are dealing with four of them. Benefactives can be general and have a meaning that varies depending on the constant construction or they can be specific, as they can only appear under certain circumstances and have a reduced distribution. In this sub-section, we will explore some examples of how different benefactives are distributed and on what lines can the specificity of certain benefactives be explained. Let us look at some examples.

Our first example comes from Zo Chin (Teddim), where there is a distinction between an associative-benefactive (to do with someone) and a causative-benefactive, as in the following examples:

(26) Zo (formal register) (Kuki-Chin, Burma and NE India)

[3]For the authors, beneficiaries are usually coded by non-core cases or adpositions with non-applicativized verb, as in English for (Kittilä & Zúñiga, 2010: 4). This seems not to apply for Kiranti (and other unrelated languages from the larger region), where affected participants and beneficiaries are marked alike (cf. Schackow & Peterson, 2011: 13)

Làjip-suŋa a-loːŋpa làj t^hun-pí hiː
mailbox-LOC 3POSS-friend letter put.inside-ASS REAL

'His friend is putting the letter into the box (for him)'. (author's notes)

(27) *Ama-u ka-va-hɔːn-sák u-hiː*
3s-p 1A-DIR-cook-BEN pl-REAL

'We helped them to cook'. (author's notes)

In example 26., the benefactive construction is semantically specific and only encodes beneficiaries that are directly adjacent to the agent. In the next example, *-sak*, which is also a causative, implies that the beneficiary is not present when the action takes place, even though it is performed for his sake (or detriment). The connection between benefaction and causation is well established (Shibatani & Pardeshi, 2002: 114f.).
A next example comes from Shipibo, where syntactic constrains alter the distribution of male/benefactives:

(28) Shipibo (Panoan, Peru)

a. *Pexe Piko-n-ra ea-∅ ka-xon-ke*
P. P.-ERG-ASS 1-ABS go-BEN-COMPL

'Pexe Piko left (and this affects me either possitively or negatively).'

b. *ea-∅ bimi-∅ pota-xon-we*
1-ABS fruit-ABS throw-BEN-INC

'Throw me fome fruits.' (Beneficiary only)

c. *ja-n-ra ea-∅ ochíti-∅ bo-on(aan)-ke*
3-ERG-ASS 1-ABS dog-ABS carry-MAL-COMPL

'He took the dog from me (to my detriment).' (Maleficiary only) (Valenzuela, 1997: 117, 127)

In example 28., *xon* is neutral to the male/benefactive distinction when used with intransitive verbs. The picture changes when the host verb is transitive, as *-xon* only marks beneficiaries and there is a specific device for malefactives in transitive verbs, namely *-on(aan)* (also present in Kittilä & Zúñiga,

2010: 12).

Our last example comes from non-standard English, where there is a malefactive construction built with the preposition *on*, as in *to lie on someone* (in Emonds, 2011: 306). *Lie* already entails a negative sense and the use of *on* seems to emphasize its malefactive nature. This example is interesting as it shows how style, register and/or dialectal variation may also be an axis to explain the distribution of male/benefactives.

In this section, we have observed some typological features that are common to benefactive constructions. Furthermore, in these examples, we have seen three patterns for the distribution of benefactives: semantics, syntax and register. We may now continue to a description of benefactive v2s in Chintang.

1.2.3 Benefactive v2s in Chintang

This section will describe the four benefactives involved in the present research. These are: *pid*, *chokt*, *dhett* and *khutt*. Furthermore, we will also mention the influence of Nepali in terms of benefaction.

pid

Pid is by far the most common benefactive of the four, occurring more than one thousand times in the Chintang corpus (cf. Section 2.1). It can be transparently derived from the independent verb *pid* 'to give'. It operates as a standard benefactive by attracting only O-agreement, as in the following example:

(29) Chintang

A-nisa-ŋa *saŋwai-ko* *u-chala* *u-loĩ-ŋa-bi-ŋa-ʔã.*

1POSS-younger.sibling-ERG buffalo-GEN 3POSS-skin 3[s]A-take.out-1sO-BEN-1sO-IND.NPST

'My brother flayed the buffalo for me'. (author's field notes)

This benefactive is mostly used with transitive verbs but Schikowski (2018: 198) describes a case with an intransitive verb. This has not been tested for the other benefactives:

(30) *Durga-ŋa* *lain-be* *u-eb-a-bid-e-h-ẽ.*

a.name-ERG line-LOC 3sA-stand-PST-BEN-PST-1sO-IND.PST

'Durga stood in the line for me'. (Schikowski 2018, 196)

The verb *pid* 'to give' has also grammaticalized into another construction, namely a permissive, cf. example 31.:

(31) *Pheri kaiʔ-ma mai-pi-th-aʔ o gakkaŋ.*
again come.up-INF NEG-give-NEG-IMP ATTN after.a.while

'Don't let him come up again in a moment!'. (CLLDCh1R08S05.420)

Example 31. is a raising construction in which 'to give' adds permissiveness and the main verb takes a non-finite form. A parallel construction is found in Yakkha (cf. Schackow: 2015, 473f.; also Wambule Opgenort, 2004: 385).
The benefactive *pid* is common to other Kiranti languages as well (cf. Michailovsky, 2017: 35). Cf. the following example:

(32) Chiłiŋ (Kiranti, Nepal)
A-kaka-ʔa chithi hakt-aŋ-bid-aŋ-hẽ.
1S.POSS-uncle-ERG letter send-1SP-BEN-1SP-IND.PST

'My uncle sent me a letter'. (author's field notes)

Dhett

Dhett is the second most common benefactive in the corpus, however, lagging far behind *pid* with only 68 attestations. The origin of this v2 is unknown and we have not found cognate constructions.
Let us look at some examples:

(33) Chintang
Ba yaŋ cett-u-dhett-u-khaŋ-u hun-ce-iʔ.
DEM.PROX ADD feed-3P-BEN-3P-CON-3P DEM-NS-LOC

'Feed this one also, that one'. (CLLDCh4R04S03.0490)

(34) *Phak os-u-dhett-a.*
pig throw-3P-BEN-IMP

'Give that to pig'. (CLLDCh1R05S05.0290)

Chokt

This benefactive is the third most common in the corpus with 60 attestations. Here are some examples:

(35) *Cham u-loiʔ-a-choʔ-ŋa-ʔa.*
song 3[s]A-sing-1sO-BEN-1sO-IND.NPST

'He sings a song for me'. (Author's field notes)

(36) *Tato cuwa tat-e numd-u-chokt-u-ku-ŋ.*
hot water hot-NTVZ do-3O-BEN-3O-IND.NPST-1sA

'I boild water for him'. (Author's field notes)

This benefactive is derived from the verb 'to pass' *chokt*. For Schikowski (2018: 81), this v2 still retains the concrete meaning of its etymon.
Apart from Chintang, some dialects of Bantawa, including the one spoken in the Chintang area itself, also have a benefactive *chokt*, cf. examples 37.

(37) Bantawa (Chintang dialect) (Kiranti, Nepali)
Iŋka-patti weis-aŋ chokt-aŋ.
1s-DAT throw-1sO[IMP] BEN-1sO[IMP]

'Throw (it) over to me'. (Author's field notes)

Chokt does not appear as a benefactive in Doornenbal's grammar but as an independent verb meaning 'to give, pass' (2009, 421; also Winter & Rai, 2013). According to N. K. Rai (2019: personal communication), this benefactive also exists in other dialects of Bantawa, outside the Chintang area. It cannot be ruled out that it is a loan from Bantawa, whence many a lexeme have been borrowed[4].

Khutt

Khutt is the least common benefactive from the corpus with only 45 attestations. It is derived from the verb *khutt* 'to bring'.

[4]As mentioned in Section 1.2, there is an asymmetrical relation between Bantawa and Chintang, as the speakers of the latter are usually fluent in the former, while speakers of Bantawa do not speak Chintang. Furthermore, in the Chintang area there is much intermarriage between these two linguistic communities.

(38) Chintang

Gadi-ce-ŋa *ba-iʔ* *ta* *mai-khu-na-khuʔ-no.*

vehicle-ns-ERG DEM.PROX-LOC FOC 1nsP-carry-LNK-BEN-IND.NPST

'Buses bring (it) here'. (adapted from chintang_now.0301)

(39) *kondaŋkhuttahaʔ*

Kond-a-ŋ-khutt-a-ŋ-a

search-IMP-1sS/P-BEN-IMP-1sS/P-IMP

'Ask (him) to search for someone'. (CLLDCh1R04S01.671)

All benefactives have a malefative reading, cf. the following examples:

(40) *Gakkaŋ* *i-taŋ* *thup-na-pi-na.*

after.a.while 2sPOSS-head hit-1s>2-MAL-1s>2

'I will hit you at your head in a moment'! (CLLDCh1R08S06.316)

(41) *Thek* *mo* *thok-u-yaŋ-dhett-e* *ba-go.*

exactly CIT peck-3P-ADD-MAL-IND.PST DEM.PROX-NMLZ

'The owl pecks the boy'. (frog_story_1.419)

(42) *Lo* *cikne* *cuwa* *le* *lukt-a-chokt-a-kh-a-ŋ* *pid-na.*

okay fucker water RESTR pour-IMP-MAL-IMP-CON-IMP-1sS/P give-1s>2

'I will hit you if you spill the water towards me'. (CLLDCh4R05S04.1210)

(43) *Asa* *caĩ* *asa* *lo* *i-taŋ* *beŋ-khutt-e.*

a.name TOP a.name SURP 2sPOSS-head be.twisted-MAL-IND.PST

'Asu, your head is bent'. (CLLDCh3R11S09.078)

Grammatical and lexical meaning

As it has been mentioned in the Introduction, several v2s in Chintang may elicit both a grammaticalized (e.g. a benefactive) and a lexical, additive meaning, i.e. 'to V1 and V2'. This applies for all benefactives except for *pid*, where there is no attested meaning 'to V and give'.

There are further v2s that have a non-grammaticalized meaning, which is similar to the v2s mentioned above, e.g. *tat*[5] (from *tat* 'to bring'). Differently from our benefactives, *tat* triggers a neutral 'bring' that is not especially done for the sake (or against) someone, i.e. it is not a benefactive.

(44) *Mo* *phakcilek* *ta* *las-u-ŋ-tat-u-h-ẽ.*
DEM.DOWN piglet FOC return-3P-1sA-bring-3O-1sA-IND.PST

'I returned the piglets'. (adapted from CLLDCh2R02S08.102)

Phonological erosion

Benefactives do not undergo phonological erosion in Chintang. In example 29., we saw *pid* with an initial voiced consonant (*bi*). This is regular for suffixes in intervocalic position with non-aspirated initial consonants, except if the voiceless consonant goes back historically to a geminate.
Phonological erosion, however, is attested in other v2s, such as the imperfective marker *-yakt* (from a homophonous v1 'stay overnight') and the in/transitive pair *-hat/t* 'move away'(trans. / intrans.). These have developed a variant without the initial consonant that no longer triggers onset prothesis (suffix copy) and cannot host endoclitics nor prefixes. In the dialect of Sambuteĩ, the imperfective is *-k* and it also does not host endoclitics nor prefixes (cf. Bickel & Zúñiga, 2017: 181; Schikowski, 2018: 78).
In sum, grammaticalized v2s in Chintang can be divided by their phonological erosion: while some v2s have undergone erosion, benefactives have been excepted from such changes.

Influence from Nepali

Finally, it should be mentioned that Nepali has had a strong influence on Chintang and this is also visible in the case of benefactives, as it is not uncommon, especially among younger speakers, to add Nepali case markers to the recipient of the action in the same fashion as Nepali *-ko lagi* (-GEN + postposition 'for'), cf. the following example:

[5] Another similar v2s is *thapt* 'bring across', from a homophonous v1.

(45) *A-nisa-ŋa* *a-mai-ko* *lagi* *duda* *phil-u-bid-e.*
1SPOSS-younger.sibling-ERG 1SPOSS-mother-GEN for milk squeeze-3[s]O-BEN-IND.PST[.3sA]

'My younger brother provided (squeezed) milk for my mother'. (author's field notes)

This is known in other Kiranti languages as well (also Rai, 2017: 221; and beyond King, 2009: 294):

(46) Bantawa (Chintang dialect) (Kiranti, Nepal)
Khos-aʔa *oŋ-ko* *lagi* *cham* *lus-u.*
3s-ERG 1s-GEN for song sing-3s[IND.PRES]

'He sings a song for me'. (author's field notes)

(47) Yakkha (Kiranti, Nepal)
Hoʔi *ak=ka* *lagi* *iya-ca* *tuʔkhi* *n-jog-a-n.*
enough 1SPOSS=GEN for what=ADD trouble NEG-do-IMP-NEG

'No, thanks. Do not bother about me (at all)'. (Schackow, 2015: 148)

In example 46., we observe that the use of Nepali benefactive devices may substitute the Kiranti v2 benefactives. In example 47., this construction has also a purposive meaning, as it has in Nepali.

1.3 Preliminary considerations

Before we proceed to the main part of the present investigation, we have to briefly introduce two theoretical discussions that are key to follow the main arguments. These are i) an overview of grammaticalization theory with focus on complex predicates and ii) a linguistic account of abstraction and polysemy. This section will handle these two topic in that order.

1.3.1 Grammaticalization

In this section, we will go through the main claims associated with grammaticalization theory. Most of them consider semantic bleaching and phonological reduction as a core feature, not to say a *sine qua non* of this process. We will mention the most canonical claims on this topic and then go over to multi verbal constructions. This is not a review of the literature available but rather an attempt to

introduce the reader to certain aspects of grammaticalization theory that will be relevant for the rest of this research.

The first definition of grammaticalization is usually attributed to Antoine Meillet, as he used this term to signify the process by which a independent word acquired a grammatical meaning (1912: 131–132; Kiparsky, 2012:16f.). For Campbell & Janda (2000: 95), Meillet also introduced the idea that an independent element that is recruited for grammaticalization has to undergo some sort of weakening of both its meaning and its phonetic form. This idea was reproduced by a large number of linguists dealing with grammaticalization.

The double change brought about by grammaticalization has been perhaps best depicted by Hoenigswald, as he described it as an "emptying of lexical meaningful morphs and a transformation into functional elements" (1963: 34; Campbell & Janda, 2000: 95).

In his influential paper of 1971, Givón argues for a cyclical understanding of morphology and syntax where loose material (or a loose group of elements) tends to couple with stems and create new morphology. For him, the attrition of grammaticalized elements is tantamount to the reproduction of this cycle, as these are absorbed by the stem. As an example, he mentions the case of English *going to* paraphrase that went from being a motion verb to a immediate future construction. Furthermore, the reduced *gonna* pronunciation would attest to this double change (1971: 22f.).

On the same line, Hopper and Traugott have argued that grammaticalization entails the reanalysis of the meaning of the form recruited for this purpose and that after the reanalysis has taken place, phonological reduction sets in (2003: 2f.; McMahon, 1994, 168f.). As for phonological change, it is described as being by definition concomitant of morphologization (fusion) and that it usually takes the form of change or drop of a vowel, consonant, tone or stress, i.e. phonological attrition (2003: 154). As for semantic change, perhaps the authors' most relevant claim for the present investigation is that there are unidirectional tendencies that make grammaticalized elements be more abstract (2003: 33, 84f.). Here they rely on the analysis of Heine et al. that describes e.g. the already mentioned change of *going to* as a literal meaning being transferred to a different, more abstract domain, a phenomenon called metaphor. Furthermore, grammaticalized elements lose their initial semantic specificities and are described as being reduced to either their core meaning (termed generalizing abstraction) or to a particular part of the meaning (termed isolating abstraction). In both cases, the initial form loses on intensional (of intension) meaning but gains on range, on the amount of contexts it can be applied to (1991: 42f.). As described by Kuteva et al. grammaticalization begins with a concrete form and ends ideally in zero, i.e. losing both its initial semantic and phonetic content (2019: 5f.).

In other publications dealing with grammaticalization, the former claim seems to absent from the core definition of the phenomenon. In their Handbook of Grammaticalization, Narrog & Heine take a very

cautious move to the matter, as they highlight the wide range of approaches and the diversity of views that make a uniform definition impossible (2011: 2). Other contributors to the handbook such as Traugott (2011: 21, 29), however, hold on to the pairing of form and meaning in grammaticalization. She describes this process as a subtype of language change and these two particular changes (semantic bleaching and phonetic reduction) as being not only present in grammaticalization specifically. A similar position is argued by Lamiroy & De Mulder in the same handbook, as they propose a sequence of parameters, namely: extension > desemanticization > decategorialization > phonological reduction. In other words, semantic change will affect the morphosyntactic properties (wordhood) and will eventually lead to the phonetic reduction of the grammaticalized item (2011: 303). The implication of such an approach is that the degree of grammaticalization can be compared e.g. between two related languages and thus a cognate construction can be said to be more or less grammaticalized in one language or the other (e.g. Kinn, 2018).

A double compendium of articles by López-Couso and Seoan sheds a more critical light to what defines grammaticalization. For the present investigation, the most remarkable fact is that they do not postulate neither semantic nor phonological change as tantamount of grammaticalization and that they are also critical about the inherent directionality of language change (2008a: 1f.; 2008b: 1f.). Some contributors, such as Frajzyngier, mention several processes that are sometimes involved in grammaticalization but they are by no means sequential nor mandatory (2008: 65f.). This idea is, however, not shared by all contributors of the volumes, who rely on the Meillet-Hopper model. E.g. Brinton, for whom grammaticalization is associated with a new meaning being more abstract and with the reduction of the initial phonological form (2008: 42f.).

A more critical view of the term grammaticalization theory altogether is perhaps best exemplified by Campbell, who has questioned the usefulness of this concept. The author has analyzed the two elements that are relevant for this section, namely semantic bleaching and phonological reduction. As for the former, he has critically observed that describing grammaticalization as going from lexical to grammatical almost implies that there is a *loss* of semantic content (2000, also Heine et al., 1991: 109; Coussé et al., 2018: 4). However he presents counterexamples, such as postpositions in Balto-Finnic that became case suffixes maintaining the meaning 'with' (Campbell, 2000: 118; also cf. Vitti, 2015: 411). In the case of phonological reduction, Campbell argues that it is not symptomatic of language change and that it might not even be necessary for grammaticalization. The classical example is the German (also in other Germanic languages) auxiliaries used for building the perfect, as these are identical with the verbs *be* and *have*. Their recruitment as auxiliaries for the perfect past tense occurred at an old stage of the language, but their form and lexical meaning have remained so far unchanged (Campbell, 2000: 121f.; 1998: 238f.).

Another dissident voice comes from Bisang, who has questioned the assumed, in his terms, coevolution of form and meaning in grammaticalization theory, at least from the perspective of South Asia. Due to certain characteristics of the languages of that region, namely the property of indeterminateness and a weak correlation between lexicon and morphosyntax, the author describes these as having a reduced likelihood of an emergence of a morphological paradigm, what thus hinders the two changes which are often associated with grammaticalization (cf. Bisang: 2004; 2008; Wiemer & Bisang, 2004: 3f.).

Grammaticalization of serial verbs

The process by which an independent verb becomes a grammatical device has been the focus of many a study and the grammaticaization path proposed by many authors does not differ from the classical Hopperian theory: in a sequence of verbs, one of these is gradually analyzed as contributing functional instead of semantic content to the construction and eventually loses part of its original semantic content and its syntactic and morphological properties while the other remains unaffected (Anderson, 2006: 332f.; Lehmann, 1995: 36f.). In this sense, it is assumed that this process transpires in two stages: a shift from lexical to grammatical and a subsequent increase of grammaticalization as this new device becomes entrenched in the grammar.

On a similar note, Anderson (2006: 338f.) suggests that there is a particular link between the semantic content of an independent verb and its grammaticalized outcome. In other words, despite the important variation found in grammaticalization processes, some verbs are more prone than others to become a specific kind of grammatical device. For the author, this is related to a cognitive event schema, i.e. discourse-pragmatically and semantically grounded constructions that would facilitate the association between the lexical content of an independent verb and a grammatical function (as auxiliary verb, in his terminology). As an example he mentions positional/postural verbs that have frequently been recruited as TAM markers in many language families.

But, going back to the process, how does this transition exactly occur? What does it mean that one verb is more embedded in grammar than others. For Anderson, this is related to abstractness, as the grammaticalized verb, which started as being purely lexical (concrete) and becomes more functional (abstract). Furthermore, this process presupposes that there is a stage of potential ambiguity, as the two forms of a verb (the abstract and the concrete) coexist. It should be noted that grammaticalization-induced ambiguity is by no means restricted to complex predicates (cf. Bourdin, 2008: 40f.).

The coexistence of the lexical and the grammaticalized meanings of the same lexeme is the situation we encountered in Chintang in the Introduction, where v2 *-ca* may trigger the lexical 'to V and eat' or the grammaticalized meaning as autobenefactive. According to Coupe (2008: 297f.), several verbal suffixes may coexist while they are on the grammaticalization path. One example from Mongsen Ao

would be the suffix *-lak*, which means both 'to V down' and a terminative. These two meanings can co-occur in the same verb: *thə̀p-lak-lak-əɪ* (throw-DESCEND-TERM-SEQ) 'after finishing throwing it down'. These forms have their origin in a relational noun *ta-lak* meaning 'last, end' and do not exist as an independent verb.

A further example is the suffix *-sèt*, that can be derived from the PTB root (<**g-sat* 'kill') and it also does not exist as an independent verb in Mongsen Ao (Coupe, 2008: 316f.). The suffix can be expressed as 'to V to death', i.e. the action of the verb results unambiguously in death, as in *tsang-sèt* 'to spear to death'. This sense can only be achieved through verbs affecting animate referents and it reflects the etymological origin. In a second sense, it means that the affected participant is subjected to extreme physical discomfort, however without necessarily resulting in death, as in the following example:

(48) Mongsen Ao (Naga, NE India)

Á-uk tsə la tsə̀ŋi na phu-sə̀t-a a-lú təmáŋ si pa tə-naʔ

NRL-pig DEM TOP sun INST expose.to.sun-AFFECT-SIM NRL-field all ANAPH 3s RL-nose

na mukuŋ-a ili.

INST root-SIM wander[PST]

'As for Pig, burned to the point of death by the sun, she wandered all over the field rooting up the earth with her snout'. (Coupe, 2008: 318)

The last meaning of *-sèt* is an intensifier, as in the following examples:

(49) *Tsəpha-sə̀t-əkə pa nə hmapaŋ tə-tsəʔ à mənu-tʃuk.*

fear-AFFECT-SIM 3s AGT time NMLZ-be.short one be.late-PFV.PST

'Being scared to death, she was late for a short time'. (Coupe, 2008: 318)

(50) *Pi tʃhi-mì-sə̀t-a áŋ li-pàʔ utà.*

PROX take-DESID-AFFECT-SIM just be-NMLZ PTCL

'This is just [the one I've] been wanting!' (adapted from Coupe, 2008: 318)

In example 49., the subject is enormously affected by fear, i.e. a psychological state is exaggerated. The connection to death is only induced by the English translation. That is best exemplified in example 50., where we observe that the death element is completely absent, being a mere intensifier. In sum, we observe several meanings in v2s that have not been victims of phonological reduction, so that the more

and less grammaticalized meanings coexist. The author frames this pattern in Hopper's (1991) theory. For Hopper, new layers of grammaticalization are continually emerging and therefore the lexical and the grammaticalized form coexist. Furthermore, certain restrictions or semantic affinity that go back to the original, semantic meaning may still be present in the grammaticalized form. Grammaticalization is thus described as a dynamic phenomenon with possible scenarios of ambiguity since two senses may coexist (Hopper, 1991). However, it is still described as a process with a clear direction, going from more lexical to grammatical (also Coupe, 2018: 189, 299f.).

Lexicon and grammar

The change from lexicon to grammar entails an assumption which is the *quid* not only of grammaticalization but of how do we conceive the mental representation of language. Bluntly said, the underlying issue is whether or not we consider the lexicon to be separated from grammar. Furthermore, this separation assumes a difference in terms of concreteness, i.e. grammar as being a set of abstract regularities and exceptions and lexicon as being concrete words (cf. Stefanowitsch & Gries, 2003: 210). This is one article of faith that sets apart generative grammar from constructivist and cognitive approaches (Boas, 2010: 54). For the present discussion, however, it is most relevant to mention the theoretical basis of a) the Meillet-Hopper tradition (also termed functional linguistics (cf. Kasper & Boye, 2011: 57f.)) and b) the constructivist model.

In an one-dimensional scale having at one end the unity of lexicon and grammar and in the other their separation, the a) functionalist tradition would be somewhere in the middle. This is due to the assumption that there is a gradual dichotomy between these two. As for grammaticalization, lexicon and grammar are thought of as making up a continuum and when an element acquires a grammatical or a lexical value, it is described as being pushed from one end to the other (Kasper & Boyer, 2011: 58; Lehmann, 2004; 1995: 119).

For b) the constructivist model, speakers retrieve linguistic expressions from complex meaning-form patterns (constructions) (cf. Boas, 2010: 55). It is not always possible to separate the lexical from the grammatical component from concrete constructions, as these two are intertwined and may even have a construction-specific syntax. Constructions are thus nourished from lexicon and syntax, as it is best exemplified in the literature available on idioms (Croft, 2001: 15f.; Booij, 2012; Diessel, 2015: 301; Harder & Boye, 2011: 59f.).

To summarize, the initial theory of grammaticalization entailed a double change (semantic bleaching and phonological erosion) and it also implied unidirectionality. Lately, this theory has been questioned from different angles but some (if not most) of its claims still live on. As for the main interest of this investigation, the assumption that the change from being less grammatical to more grammatical has

to go hand in hand with phonetic erosion and semantic bleaching has been questioned in the literature. However, opinions on grammaticalization are a long way of being unanimous. Lastly, the claim that elements wander from lexicon to grammar when they are grammaticalized entails an important assumption about language, namely that grammar is somehow separated from lexicon. An idea that has not found unanimous consensus.

1.3.2 Abstraction and polysemy

Having reviews the main claims concerning grammaticalization, especially in relation to multi verbal construction, we may turn to the next issue that should be addressed before we go over to the experiment, namely what do we mean when we read that a grammaticalized form is more abstract. Within all the fields that belong to linguistics, abstractness has perhaps been most discussed in psycholinguistics. There, it remains a delicate topic, as the debate on what characteristics define concreteness and abstractness is not settled. In this section, we will begin by describing abstraction and how it has been addressed in linguistics, especially in the form of tests. Next, we will discuss a related topic, namely polysemy.

Abstractness is a common concept both in academic publications as in everyday speech. In the latter usage, it mostly refers to a broad concept that encompasses several specific features but that does not exist as such in reality. In the academic parlance, there seems to be little consensus on the nature of abstractness and what sets it apart from concreteness. Iliev & Axelrod (2017) have pointed out that the mere concept of abstractness entails a paradox from the point of view of cognitive processing, since two different ways of measuring abstractness seem to have similar results. In one approach, one may define abstractness as a) a lack of physicality or distance from sensorial experience. This presupposes that concrete entities are material and that they exist in space and time while abstract materials cannot be perceived by the senses. This line of thought sets apart concrete from abstract entities on account of their physical properties; while both have a mental representation, only concrete objects have materiality. The cognitive consequences of having these two properties, in the case of concrete entities over abstract ones, are, among others, being more easily recalled, more easily associated, recognized faster in lexical decision tasks plus they seem to appear earlier in vocabulary development (also Burgoon et al., 2013: 503).

Other authors that work with this definition point out the importance of background and context for the retrieving or representation of abstract, since the meaning of an abstract word does not stand in isolation (Barsalou et al., 2005: 129f.).

The second approach is termed b) precision and it is defined as the inclusiveness of a concept. Abstractness stands now for inclusiveness and generality, i.e. the amplitude of concepts that fall under this idea.

As an example, *mammal* would be more inclusive (more abstract) than *dog*. The example also implies that there is a chain of abstractness, as in the following taxonomy: mammal > dog > beagle. For this approach, the middle level of the chain (basic-level category; in this case *dog*) would have an advantage over the others in cognitive processing, as it would generate more tokens in listing tasks and would be activated faster in picture naming tasks. The major issue remains how to set the parameter that defines what is in the middle. This being said, inclusiveness remains a strong point in defining abstractness, as it is related to the amount of information associated with a concept.

This definition of abstractness has also been conceptualized as focusing on basic categories, i.e. knowledge which is *abstracted* from experience and allows generalizations, therefore, the abstract CHAIR derives from different (or perhaps just one) chair. For Barsalou (2003, 1182f.), cognitive representation is largely an interpretative process, so that multiple inferences are possible and the context plays an important role, since it is bound to the objects whence they exist.

Other fields of linguistics also agree on defining abstractness as inclusiveness. Most relevant for this research are descriptions of grammaticalization, as these describe a lexical element acquiring a grammatical feature as being more abstract (cf. Sapir, 2004: 40; Heine et al., 1991: 21f.). What is usually meant, is that e.g. a body part, when it is used as a preposition, it will be able to appear in a higher number of contexts; thus be more abstract. Whether grammaticalization always entails abstractness will be the topic of the second experiment (Section 3).

In sum, there are two approaches to conceptualize abstraction. The first approach aims at the type of information while the second (precision) deals with the quantity of associated information. In order to avoid the paradox of having similar cognitive implications for the same category (abstractness) depending on how it is measured, the authors argue for keeping these two measures of abstractness apart (Iliev & Axelrod, 2017: 726f.; also Wiemer-Hastings et al., 2001). Barsalou (2003), on a different note, also adds the importance of context and personal interpretation for the creation of mental representation. It should be stated that most theoretical claims have not always been tested and thus remain speculative (cf. De Deyne & Storms, 2015: 475).

Tests to measure abstractness

Apart from the theoretical approaches to abstractness, some studies have directly asked people how they would rate a large scale number of noun according to how concrete they are (cf. Brysbaert et al., 2014). Wiemer-Hastings et al. (2001; also Nelson et al., 2004) asked people to rank the abstractness of several words. Their findings, replicating older tests, describe a distribution with two pronounced clusters, each with a different mode. In other words, people perceive different degrees of abstractness. This is a remarkable finding since lack of physicality alone cannot explain the observed variation. The

authors explain this variation with a theory of contextual constrains, in which each concept requires a specific number of entities to be meaningful (e.g. *comparison* requires at least two elements to be compared). Studies have also shown that abstract objects (here in the sense of lack of physicality), such as *love* or *thought*, are perceived as being more frequent in texts where the number of both concrete (as *door* or *water*) and abstract objects is the same (cf. Galbraith & Underwood, 1973). The consequence, so the authors, is that abstract objects are related to more variable contexts (also described by Tverky & Kahneman, 1974: 1127). In this sense, the authors relate both definitions of abstractness.
From these studies it seems that nouns are employed much more frequently than verbs in studies dealing with abstractness. Nouns are learned earlier than verbs and have been thought of as being more basic and concrete (Gentner, 1982: 301f.), a claim that has not gone unchallenged (cf. Tomasello, 2009: 43f. for an overview). However, it seems that the measurement used for abstractness in *verbs* is seldom addressed and that the results from large-scale tests what rank the abstractness of verbs are used without much discussion (cf. Naumann et al.: 2018). As an example, Colla et al. (2018) measured the abstractness on verbs depending on the abstractness of the nouns they usually appear with. However, they do not explain their approach further nor the theoretical implication.

Having reviewed the literature on abstractness, we think that the definition where more abstract entities refer to more different contexts and have therefore less restrictions for their use is most helpful for our study (cf. Barsalou, 2005: 131). Concretely for our research, we assume that more abstract lexemes should be related to a wider array of other concepts compared to a more concrete lexemes, since the latter would be semantically bound or associated with a smaller group of contexts (cf. Galbraith & Underwood, 1973). The logic path to this conclusion will be brought up again in the Interpretation (Section 3.6).

Polysemy and vagueness

Abstractness is understood here as semantic inclusiveness, i.e. the width of a concept. However, this does not tell us anything about the depth, i.e. the strength or tightness of the relation between two lexemes. For this reason we have to introduce two new concepts: polysemy and vagueness. Polysemy is usually defined as a lexeme having two or more senses associated with it; it is thus a form of ambiguity (cf. Gries, 2015: 471; Evans, 2005: 95f.). Vagueness is used for lexemes that have one single but non-specific i.e. broad meaning encompassing perhaps different senses but that can be subsumed in this one general meaning (Tuggy, 2006: 168). For the purpose of this research, we will use these two concepts in one particular ways, namely in relation to the strength of the lexeme-lexeme relation. Furthermore, as the distinction between these concepts may be problematic, we will assume that these form a gradient,

going from strong connections to weak connections[6]. This is perhaps best exemplified in the case of collocations (cf. Williams & Milton, 2011): A polysemic lexeme will be able to form a large number of pairs with other lexeme, each with a different meaning. An example would be phrasal verbs; let us take a prepositional-like element, such as *up*, as it conveys a wide array of meanings (cf. Garnier & Schmitt, 2015): aspectual-telic *eat up*, directional *jump up* and lexicalized meanings *give up* (as 'renounce') (cf. Dehé, 2015: 611). We will thus assume that a polysemic lexeme has several meanings that come to be only within a collocation, i.e. it builds strong connections to other lexemes. Conversely, a vague lexeme may also have a wide array of possible collocations, but the meaning will remain in one broad sense. This is best exemplified in adjectives such as *tall* and *big*, as they can refer to any noun but it only modifies its original meaning and does not create new ones (cf. Van Rooij, 2011: 125). We argue that vague lexemes thus form weaker connections to other words. The strength of the connections will be the topic of Section 3.5.2.

To conclude this section, we have described two axes of meaning: inclusiveness of concepts and depth of the lexeme-lexeme connection. All in all, this description remains highly superficial and inconclusive, as the topic is much more complex. However, the instrumentalisation of these two axis of meaning serves solely the purposes of the present analysis and thus it does not claim to question or confirm the foundations of semantics.

For the present research, we will take a look at the interplay of abstractness and polysemy in two ways: the first one is polysemy (measured with PMI) as a tool for controlling the measurement of abstraction. The second one is the shared effect in abstractness and polysemy when a lexemes is recruited as a v2. These point will be presented in Section 3.5 and commented in Section 3.6.

[6]The traditional descriptions mention a gradient between vagueness and ambiguity, polysemy being somewhere in-between (cf. Deane, 1988). Since we are not dealing with ambiguity in that sense, we are not mentioning the whole model.

2 Experiment nr.1 - Evidence from translation

In this section, we will describe the first approach towards understanding the relation between benefactives and independent verbs. As it was described in the Introduction, Chintang's v2s are known to be versatile in the sense that they may carry indistinctly a lexical and a grammatical meaning. Through a study of translations from Chintang to Nepali, we expect to find a first indication on the relationship between benefactives and their etymological origins, namely independent verbs[7].

This section is organized in the following fashion. The first part is a description of the Chintang corpus. From that if follows a brief account of benefaction in Nepali. Going over to the experiment, we will describe the hypothesis, methods and results. The latter will be commented on the discussion, which also closes this section.

2.1 The Chintang Corpus – description of the corpus

The Chintang Corpus was built between 2004 and 2013 by a large group of researchers headed by Prof. Dr. Sabine Stoll and Prof. Dr. Balthasar Bickel, both from the University of Zurich. It includes more than one million glossed tokens or 250 hours of speech. The content of the corpus can be roughly divided in two groups: unsupervised (naturalistic) recording of child speech of six children (and their relatives) and supervised recordings of adult speech. The second group is made up primarily of different settings with two people discussing a given topic. Other glossed speech recordings include a religious register (e.g. *mundhum*, a ritual language, typical of that language family (cf. Rai et al., 2009)), monologues describing different things (e.g. cooking recipes), songs, stories and experiments, such as *the pear story* and *the frog story*, which were carried out with one or more people.

The corpus includes information on age of the speaker, genera of the recorded speech, time of the recording as well as a translation to English and Nepali. The glosses are in Toolbox format. Figure 1. displays an extract from the corpus:

[7] An initial version of this experiment was produced during a summer school organized by Prof. Samaržić from the University of Zurich, Prof. Puskas from the University of Geneva and Prof. Milićević Petrović from the Universiy of Belgrade in July 2019.

```
\ref CLDLCh3R06S05.081
\ELANBegin 00:02:09.414
\ELANEnd 00:02:11.872
\ELANParticipant Sapana
\tx khoi thaʔta bimala akka tina pina
\gw khoi   thaʔta          bimala akka tinapina
\mph khoi   thapt      -a    bimala akka tis      -na   -bid  -na
\mgl what   bring.over -IMP  a.name 1s   put.into -1s>2 -BEN  -1s>2
\lg N      C          -C    N      C    C        -C    -C    -C
\id 4622   2260       -1236 KK917  608  2326     -2327 -6679 -2327
\ps interj vt         -gm   n      pro  vt       -gm   -v2   -gm
\eng Bimala bring that I put there for you
\nep ि्वमला त्यो ल्याउ त म लगाईिदन्छु ।
\dt 14/Feb/2013
```

Figure 1. Extract from the Chintang corpus.

In Figure 1., we observe the original Chintang sentence, its different glosses, an indication on the origin of the lexemes (C for Chintang, N for Nepali etc.) and English and Nepali translations.
The corpus also provides information on language acquisition, as the setting of children interacting with their parents or people they know was repeated cyclically in order to observe the development of speech.

The corpus has served as a fertile ground for various studies on Chintang. Just to name a few, Stoll et al. (2012) have studied the noun – verb ratio in child speech and Bickel et al. (2012) have described the flexibility of prefix-ordering in complex verbs. This has only been possible to asses due to the size of the corpus. This being said, there are some limitations for corpus studies. One of these is of especial concern for our research, namely a reportedly low referential density in Chintang (cf. Schikowski et al., 2015: 2f.; Bickel, 2003 for Belhare, another Kiranti language). This is to say that referents are not overtly mentioned by default. In practical terms, a typical sentence in the Chintang corpus will look like the following example:

(51) *Thup-u-m* *kina* *pid-u-m* *naŋ.*
break-3P-2NS.A[SUBJ.NPST] SEQ give-3P-2nsA[SUBJ.NPST] but

'(You) should give (it to him), by crunching (it)'. (adapted from CLDLCh2R07S02.191)

In sentence 51., neither the subject nor the objects of the verb are overtly mentioned and thus depend on the context to be understood. Such sentences are the norm in the corpus. The research on benefactives in Chintang through the corpus has thus to battle both against the small number of attestations of three out of four benefactives and against non-overt referents.

Nevertheless, the corpus is a treasure for the study of Chintang and linguistics in general due to its size, the variety of registers and its versatility. The approach in this study will be to use the Nepali translations of Chintang sentences to obtain a clue on the benefactives: Given that the number of attestations in the corpus of *chokt*, *dhett* and *khutt* is too small for a study on collocations, we decided to analyze the Nepali translations of Chintang sentences. This was possible because all sentences from the corpus are already translated to Nepali by Chintang-Nepali bilingual speakers. Analyzing a phenomenon through translations is in every sense indirect and thus it does not provide unequivocal evidence for a claim *per se*, since multiple factors can have an effect and these cannot be controlled for. Conversely, it can support evidence and even point towards a tendency, especially if the translations come from naturalistic speech.

Meta-languages are a *sine qua non* in several settings (cf. Sakel & Everett, 2012: 193f.) and they are sometimes also part of the analysis of the target language, especially if certain constructions, uttered by non-natives or bilingual speakers, differ from the normative usage, as this can point to linguistic divergences between the analyzed language and the metalanguage (cf. van Driem, 1987: 109f. for Nepali and Limbu). Furthermore, comparing these two languages (the analyzed and the metalanguage) has been a fruitful source for linguistic research.

Since we will analyze translation to Nepali, benefactive constructions in that language require a brief introduction, which is provided in the following section.

2.2 Benefaction in Nepali

The most common form to express benefaction in Nepali is by the use of dative case *-lai* or a fixed combination of genitive *-ko* plus the postposition *-lagi* 'for', as in the following example[8]:

(52) Nepali (Indo-Aryan, Nepal)
U dinbhʌri tʌpaĩ-ko lagi pʌkau-dʌi-chhʌ.
3s all.day 2s-GEN for cook-PROGR-3.NPST

'S/he is cooking for you all day'. (personal knowledge)

These devices also mark dative/accusative as well as the experiencer in certain constructions.

It is usually the verb *dinu* 'to give' which is recruited to express benefaction as a v2 in complex predi-

[8]The examples in Nepali are transliterated in a fashion that tries to resemble the spoken language rather than the, perhaps more common, transliteration based on the written language. This has as a main consequence that there are orthographic distinctions (i vs. ī) that are not kept in most modern spoken varieties of the language and thus not rendered in the present investigation.

cates, as in the following example:

(53) *MʌI-lai ramro upʌhar lyai-din-chhʌ bhʌnne asa lag-i-rakh-eko chhʌ.*
1s-DAT good present bring-BEN-3s.NPST COMPL hope attach-DUR-PRV.PTCP AUX.3s

'I keep on hoping that he brings me a nice present'. (adapted from Schackow & Peterson, 2011: 7)

In example 53., we observe a complex predicate headed by the verb *lyaunu* 'bring' and the v2 *-di* (< *dinu* 'to give') that adds a benefactive reading. Apart from benefaction, v2s in Nepali can also add an aspectual reading to the main verb in addition to other specific meanings, which are related to the verbal origin of the v2. Other common v2s are the resultative/durative *-rakh* (< *rakhnu* 'to put/keep') and the telic *-sak* (< *sʌknu* 'to finish/be able to') (also in Schackow & Peterson, 2011; Pokharel, 1991). More than one of these vector verbs can follow the host verb, as in the following example:

(54) *MʌI-le timro khelʌuna kinni-di-rah-eko chhu.*
1s-ERG 2sPOSS toy buy-BEN-put-PRF.PTCP AUX.1s

'I have bough a toy for you (and I have put it somewhere, i.e. it is not here)'. (Author's field notes)

In example 54., v2 carries a benefactive meaning while v3 displays the lexical meaning and not the expected resultative/durative. Ambiguity in Nepali complex predicates has not been addressed in the literature.

The literature on Indo-Aryan refers to these constructions as compound verbs and they are described as an innovation of New Indo-Aryan, since they are not attested for Sanskrit (Masica, 1993: 326; Hock & Bashir, 2016: 549; David, 2015: 46f.; Liljegren, 2016: 244f.). We can safely frame these constructions in the South Asian (and even beyond) areal phenomenon of multi verbal construction, as it was mentioned in Section 1.2.1.

In these complex predicates, the main verb (v1) ends in an *-i*, which could be related to an homophonous non-finite ending. This ending, which may be deleted in fast speech and is not triggered in all verbs, can be though of as stem alternation.

2.3 Hypothesis

The main idea behind this approach is that different benefactives in Chintang may trigger different benefactive constructions in Nepali. All sentences in the corpus are also rendered in English, but since

the translators were not native speakers, the results may be an artifact or at least influenced by potentially incorrect translations. We will thus only analyze translations to Nepali.
We expect specific benefactives in Chintang to be translated to specific constructions in Nepali. If the difference of the benefactives can be grasped in terms e.g. of semantics, we could be able to observe that in the translations. In this sense, if these benefactives in Chintang are perceived as being transparently different to each other for native-speakers, this might be observable in translations. Obtaining a signal would help us asses the relationship between Chintang's ambiguous v2s and independent verbs.

2.4 Method

For each benefactive, we extracted 20 sentences (therefore 80 in total) from the corpus. The sentences were randomly chosen using R (RStudio Team, 2015) without replacing. The extraction consisted in isolating all occurrences of each benefactive and randomly picking 20 sentences from each of these four pools. We only took 20 sentences on account of the few attestations we have of some benefactives in the corpus (33 for *khutt*; 48 for *dhett*; 48 for *chokt* but 903 for *pid*). In the next phase, we manually analyzed each sentence one by one to sort the Nepali translations and to control for possible errors. In the case of *chokt*, one of the sentences had no verb, as the benefactive is attached to a demonstrative and thus it is not rendered with a verb.
After the extraction and sorting of the sentences it was all a matter of manually counting the different strategies used in each language to render our benefactives.

2.5 Results

All sentences from our sample had a translation to Nepali. Since we are only interested in the equivalence of Chintang benefactives, we only counted benefactive constructions in Nepali as valid tokens. Translations with no benefactive construction were excluded and thus not further analyzed. In Table 2., we observe the total number of valid answers for each benefactive:

Table 2: Valid answers

benefactives	valid Nepali
pid	18
chokt	15
dhett	13
khutt	17

Out of 79 sentences (one sentence with *dhett* did not have a verb), we obtained for a) *pid*: 18 valid translations ; b) *chokt*: 15 valid; c) *dhett*: 13 valid translations; d) *khutt*: 17 valid translations.

As mentioned in Section 2.2, Nepali has a rich system to express benefaction with both case marking and v2 constructions. In our sample, however, we only encountered the latter. v2 *-di* was always used to express benefaction in the translations from all benefactives. In Figure 2., we observe the distribution of Nepali benefactive constructions depending on each of the four Chintang benefactives:

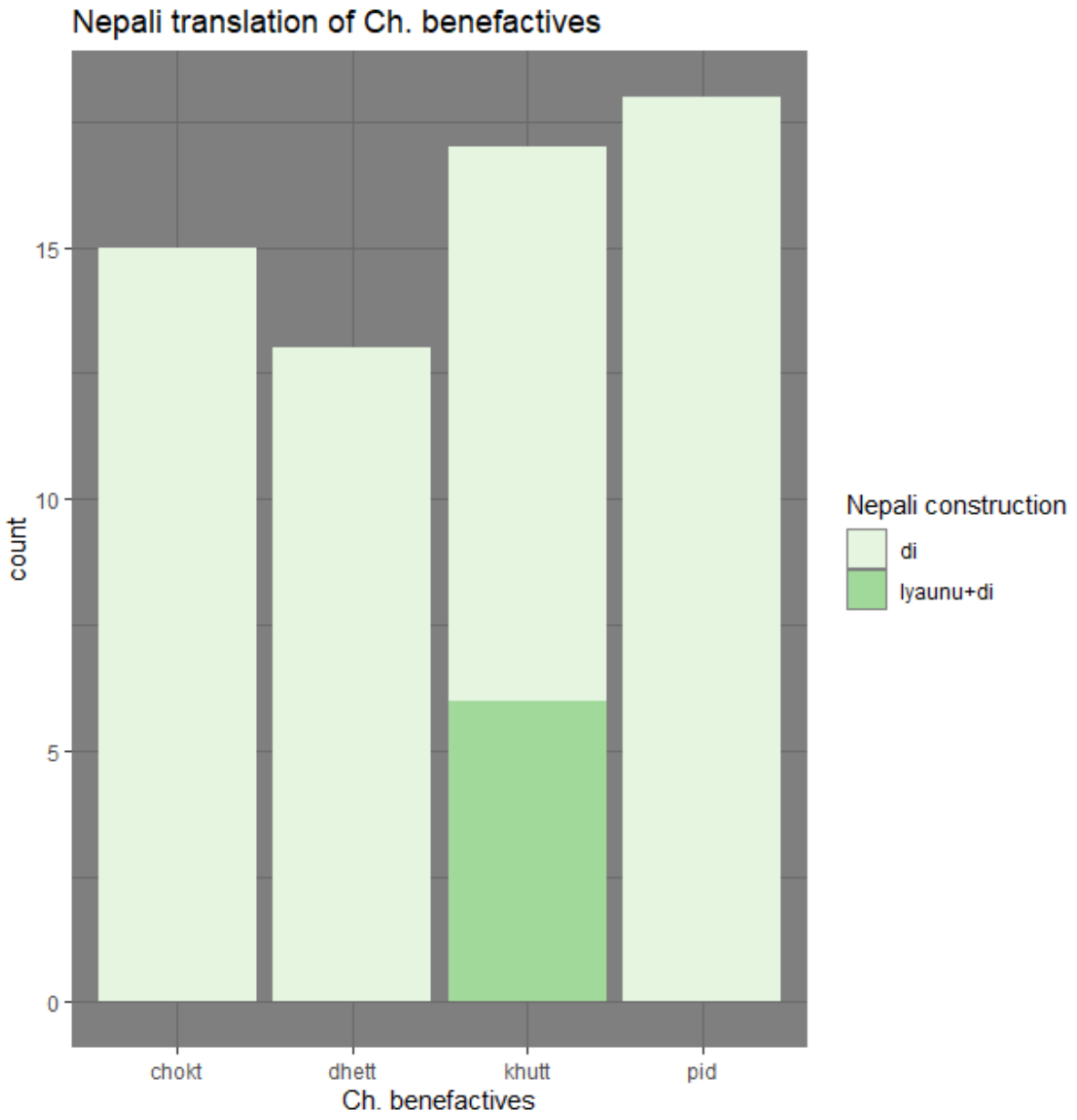

Figure 2. Complex pred. in Nepali according to each benefactive

In Figure 2. we observe that *-di* is the most common v2 employed to translate all benefactives. In the case of *khutt*, however, we found six cases of one particular construction: the v1 from the Chintang original was rendered in Nepali with a non-finite ending (usually termed converb) *-era*, followed by the

verb *lyaunu* 'to bring' with *-di* as v2. Cf. the following example:

(55) *Ama mʌ tʌlʌ-ko ghãsʌ kaʈ-erʌ lyai-di-ũ.*
mum 1s DEM.UP-GEN fodder cut-CVB bring-BEN-1S.OPT

'Mummy, do I bring that grass up there by cutting ?' (adapted from CLLDCh3R02S02.241)

In example 55., we observe that *kaʈnu* 'cut' takes an *-erʌ* ending and *lyaunu*, a benefactive v2. Such a sentence is grammatical in Nepali.

2.6 Discussion

Having four different benefactive v2s in one language is a rare phenomenon and, considering similar patterns found in other languages, it is only natural to assume that there is some sort of distribution and that they are not exactly the same. The Chintang corpus allows us to compare these benefactives through their Nepali equivalences. The most relevant finding from this experiment is that *khutt* differs from all other benefactives in the translation to Nepali, since we find the same combination of a converb followed by a complex predicate 6 out of 17 times. The latter construction is composed of the translation of the Chintang host verb along with *lyaunu* 'bring' and a v2 *-di*, a benefactive, at the end. What we could deduce from this distribution is the tendency of speakers to associate *khutt* with transitive towards-the-speaker motion verbs ('to bring'/*lyaunu*), which is also consistent with the origin of *khutt*, namely the independent verb *khutt* 'to bring'. It would seem as if *khutt* has still much of the semantics of its v1 counterpart, compared to the other benefactives. We know that some v2s may trigger either their lexical or their grammaticalized meaning. These results could thus be read as some benefactives (*khutt*) being more ambiguous than others (*chokt*), for the former are rendered with a semantic equivalent in translations to Nepali. In other words, the lexical detachment, i.e. the semantic bleaching or the abstraction of the meaning that is stipulated in the literature on grammaticalization once a v1 is recruited as a v2, seems so far to have operated in a differential fashion and, through the evidence of translations, *khutt* has been affected the least, since native speakers resort (every third time in our sample) to an extra 'bring' v2 when translating to Nepali.

Due to the small sample size and the fact that many other factors might have intervened when these sentences were translated, we may take these findings as a (albeit weak) sign that does not prove much on its own, but that points towards how benefactives in Chintang may differ in respect to the similarity they bear to their full-verbal counterpart. These results will be further discussed again in Section 3.6.

3 Experiment nr. 2 - Cloze test

3.1 Introduction

Having observed the benefactives in Nepali translations of the Chintang corpus, we decided to perform a cloze test in order to obtain a more fine-grained picture of the differences between *pid*, *chokt* and *khutt* as a v1 and as a v2. The literature on grammaticalization predicts a broadening of meaning for grammaticalized verbs compared to their independent v1 counterparts. This process is usually described as an abstraction of the original meaning. We aim at quantifying this shift using a test of free association with both the independent v1s and the benefactive v2. The rationale behind is producing two sets of associated nouns (one for the independent v1 and one for the grammaticalized benefactive v2, which has to follow a host verb, i.e. a complex predicate) that would allow a comparison. Being able to compare the associations (in terms of nouns) that trigger an independent verb compared to its grammaticalized counterpart will allow us to corroborate or rule out with empirical and quantifiable data the theoretical predictions of a large part of the literature available on grammaticalization. Benefactives in Chintang provide a suitable condition for such an experiment, as three benefactives exist as independent verbs in the language and thus the difference in terms of abstraction between each v1-v2 pair can be measured and compared.

A test of free association asks the participant to come up or fill a gap with the first thing that comes to his/her mind given a particular stimulus. A cloze test, which is the test used in this experiment, requires the participant to "complete the sentence", i.e. the executor of the test provides an incomplete sentence and the participant has to fill the blank, again, with the first thing that comes to his/her mind.
Tests of free association are well established in psychology, neurolinguistics and computational linguistics (cf. Douglas, 2004; De Deyne & Storms, 2015; Huettig & McQueen, 2007). Cloze tests, more precisely, have been used to test language proficiency (cf. Stubbs & Tucker, 1974) and anticipatory cognitive processes (DeLong et al., 2005). It can be said, however, that it is not a standard method in linguistics.

3.2 Hypothesis

As we have seen in the introduction to this investigation, complex predicates in Chintang have been described as having v2s that may either express the lexical content of one particular form or a grammaticalized (usually termed abstract) device. This situation is not unique to Chintang and the literature on grammaticalization describes it as an in-between stage in the grammaticalization path.
Our impression is that there is no evidence to foretell that the lexical meaning of the grammaticalized

v2 will disappear in the long run in Chintang and, for this experiment, we assume that both meanings (the lexical and the grammaticalized one) coexist in a stable way and thus a comparison between these two is most meaningful.

Given that v2s in Chintang are known to carry two meanings, we expect a) to find little difference between v1s and v2s in terms of the distribution of nouns each constructions triggers for *khutt* and *chokt*, but not for *pid*, for it is not an ambiguous v2. This would be an argument against the traditional grammaticalization path proposed in the literature. Furthermore, as we have seen in experiment nr. 1 on Nepali translations from the corpus, benefactive *khutt* is perceived as having much semantics of its etymological origin, namely v1 *khutt*, and hence we expect b) *khutt* to have a particularly low degree of difference between its independent (v1) and its recruited (v2) form, as we assume that the v2 keeps much of the semantic content of v1. These two hypotheses will be tested using the results from the cloze test.

3.3 Materials and methods

3.3.1 The test

The aim of the experiment was to compare the associations in terms of nouns that trigger *pid*, *chokt* and *khutt* as a) independent verbs and b) as v2 in a complex predicate, i.e. following a v1. We thus have two sets of verbs a) and b), each with the same three lexemes (*pid*, *chokt khutt*) that can be compared to each other.

The first set a) is composed of the independent verbs *pid*, *chokt* and *khutt*. For each of these we created 10 conjugated forms, which adds up to 30 different simple v1s with different agreements, as in the following examples taken from the questionnaire that was used in the experiment.

(56) Chintang
U-khutt-e-h-ẽ.
3A-bring-PST-1sO-IND.PST

'S/he brought (it) to me'. (Appendix)

(57) *Na-pid-e.*
3>2-give-IND.PST

'S/he gave (it) to you'. (Appendix)

Even though Chintang has little dialectal variation, some of the tested verb forms were conjugated in a manner which is distinctly from Sambuteĩ, where the prefix of the first person non-singular distinguishes exclusive *ma-* from inclusive *mai-*, cf. the following forms:

(58) *Abo hun-ce-ŋa=le sahayog **mai**-pi.*
now DIST-NS-ERG=RESTR help 1ns**i**O-give[.SUBJ.NPST.3A]

'Now only they will help us'. (Ganesh_talk.201; also Schikowski 2018, 54)

(59) ***Ma**-pid-e.*
1ns**e**P-give-IND.PST

'They gave us (that)'. (CLDLCh2R06S03.061)

(60) *Nob-ma **kha**-pi-nuʔ-nuŋ ta ni.*
touch.lightly-INF 1nsP-allow(give)-IND.NPST-NEG FOC EMPH

'S/he doesn't allow us to touch it'. (CLDLCh1R01S02.0715)

Examples 58. and 59. are in the dialect from Sambuteĩ, for the inclusive and exclusive markers are distinguished. Example 60. is from Multeĩ, where the prefix is not sensitive to inclusiveness. In our test one participant was from Multeĩ, but he knew about this dialectal variation and was not hindered by it.

The second set b) is composed of complex predicates having a host verb (v1) and *pid*, *chokt*, *khutt* and *dhett* as benefactive v2s. The latter benefactive *dhett* has no v1 counterpart and thus does not serve the purpose of our comparison. However, it was included in the test as it is also a benefactive.
The final list of complex predicates was a combination of 30 different v1s and our four benefactive v2s. These had the following distribution: 10 different sentences for *pid*, 11 for *chokt*, 10 for *dhett* and 9 for *khutt*. All in all, there were 40 different complex verbs with a benefactive v2. The following are examples of the employed sentences:

(61) *Oŋs-u-dhett-u-c-e.*
peel-3[s]O-BEN-3[s]O-3nsO-IND.PST[.3sA])

'S/he peeled (it) for them'. (Appendix)

(62) *Loĩs-a-chokt-a-ŋs-e.*

take.out-PST-BEN-PST-PRF-IND.PST[3>3]

'S/he has taken (it) out for him/her'. (Appendix)

The final list, which was read aloud to the participants, contained both sets a) and b); it added up to 70 (30 + 40) different verbs. All of these were tested with a native speaker prior to the execution of the experiment and both sets of verbs were interleaved. The list of verbs is available in the Appendix.

3.3.2 Participants

The cloze test was carried out with native speakers of Chintang, all of whom grew up in a Chintang-speaking area. There were eleven participants, all male[9] and between 23 and 42 years old[10]. As it the case in this linguistic context, all speakers were bilingual in Chintang and Nepali (cf. Bickel et al., 2007: 44). Furthermore, they are also fluent in Bantawa and some of them even in English. The level of formal education was not asked to the participants. There were two other potential candidates but they were excluded since one did not understand the exercise and the other answered with the same noun to all sentences.

The tests were carried out in the villages of Multeĩ, Sambuteĩ and Panchakanne during January 2019. One test was conducted over the phone with a native Chintang speaker living in Dhankuta.

3.3.3 Execution

Once the potential participants accepted to undergo the test, the instructions were explained to each one with two examples: one with the verb *to drink* and the other with *to cut*. The participants had to come up with the first noun that came to their mind upon hearing these verbs in a conjugated form. We chose these verbs for the instructional part since we expected them to elicit two different, non-coinciding sets of words (to drink *water, arkha*[11] *, bier*, i.e. liquids and to cut *hair, grass, wood* etc., i.e. non-liquids) and thus the participants may not always answer with the same noun. Furthermore, the participants were encouraged during the instructional part not to repeat the same word all the time.

The experiment was carried out in the following fashion: Each participant had the prepared list of verb strings (both with V1+V2 and individual V1s in a shuffled order) read aloud. After each verb,

[9] In the Nepalese context, it is not uncommon for men and women to have separate lives and thus recruiting female participants as a male is especially difficult.

[10] The exact ages are: (ascending) 23, 24, 25, 25, 27, 28, 30, 30, 30, 42 and 42.

[11] A home distilled alcoholic beverage made from millet, *rʌksi* in Nepali.

the participant had to name the first thing (a noun) that came to his mind. If the participant did not understand the verb (e.g. due to the pronunciation of the executor of the test), then it was read aloud a second time. In some occasions the participants took too long to come up with something and then the verb string was skipped and the experiment continued. In two occasions, the participants had to take a short break to attend other matters and the experiment was resumed thereafter.
Even though each participant was tested individually, we were seldom alone going over the sentences and other people would look and even try to answer the questions or suggest to the participants what they should say. All those answers suggested by observers were not considered for this test and every time there was an intrusion (either because an observer suggested a possible noun or when the observer asked a third party for a possible answer), that noun was excluded from analysis.

3.4 Results

3.4.1 Distribution of tokens

The answers (nouns) to the list of verbs were uttered either as an individual noun (e.g. *cuwa* 'water') or they were embedded in a sentence which repeated the verb (i.e. *he gives me water*). In some cases, the participants would change the agreement of the verb to first person singular acting on third person singular (i.e. *I give water to him*). Two speakers used the filler *meiʔ* (< *meiʔ* 'thing') before their answers, but this was not taken into account in the analysis.
The elicited nouns were mostly uttered in Chintang, while some were said in Nepali (*paisa* 'money', *phekwa* in Chintang) and a small number in English (e.g. *orange*). Most of the nouns that were not in Chintang are well-known loanwords, however some speakers uttered Chintang words in Nepali (e.g. *suntala* for *sontolong*[12]). The relevance of the chosen language was not further investigated.

When analyzing the responses, all nouns were translated into Chintang, leaving the neologisms in the original language (e.g. *football*). The types (occurrences of individual words, regardless of the frequency) were counted after the translation to Chintang, and thus in the set {*money, paisa (Nepali for 'money'), phekwa (Ch. for 'money')*} the type is one (tagged as *phekwa*) and the frequency three (three occurrences).
In the case of morphological derivation, each variation was treated as an independent type, therefore *phak* 'pig', *phakce* 'pigs', *phakcilek* 'piglet' and *phakcilecek* 'piglets' are four different types.
There was one case of a dialectal variation in the answers, namely the plural *-ce* or *-ca*. Occurrences of *-ca* were counted as *-ce*; this small dialectal difference was collapsed for this study.
Some of the provided answers were also excluded from the analysis: there were some cases of nouns

[12]This in turn is probably an old loanword from Indo-Aryan

which could not be understood in the recordings of the sessions either because of external noise or due to the pronunciation and, lastly, some participants changed the provided verb, when repeating the verb-string with a noun, for another more of their liking. All these cases were excluded from the pool of answers.

Eleven speakers produced 602 eligible nouns (tokens): 287 for a) the simple verbs and 317 for b) the complex predicates with benefactives. There is an unbalance in the number of tokens for each independent verb/ benefactive: In group a) with independent verbs, we obtained 94 for *pid*, 95 for *chokt* and 98 for *khutt* while for group b) of complex predicates with benefactives the numbers are: 84 for *pid*, 104 for *chokt*, 60 for *dhett* and 74 for *khutt*. The uneven distribution of nouns is mainly due to the high amount of answers that had to be omitted. These results are summed up in the following table:

Table 3: Valid answers from the cloze test

v1 *pid*	v1 *chokt*	v1 *khutt*	v2 *pid*	v2 *chokt*	v2 *dhett*	v2 *khutt*
94	95	98	84	104	60	74

The list of nouns follows, unsurprisingly, a classical Zipfian distribution. This means that the frequency of a word is inversely proportional to its rank in the frequency list and thus, there is a large number of nouns that appear only once, termed hapax, and a small set of words with a very high frequency (Levshina, 2015: 64). In more concrete terms, the group with a) a simple verb has 21 hapax values and the group with b) complex predicates 41. Furthermore, the three most repeated types in group a) are *saman* 'goods' 87 times, *phekwa* 'money' 40 times and *kok* 'rice' 19 times and in group b) *saman* 'goods' 53, *ghasa* 'grass, fodder' 24 and *cuwa* 'water' 22. The word *saman* 'goods/thing' has the higher number of repetitions in both groups, which might be related to its broad, unspecific meaning, much like English *thing*. Within each group, the highest value with the largest number of repetitions is not always the same: in group a) *saman* has the higher number of repetitions for *chokt* and *khutt* and for *pid* it is *phekwa* 'money'. As for group b) it is *saman* for *pid* and *chokt*, *ghasa* 'grass' for *dhett* and *cuwa* 'water' for *khutt*.

Outliers, in this case nouns with an excessive number of repetitions, may affect the results and blur the overall picture, a phenomenon known as *masking*, and the data therefore has to be inspected in order to try and explain extreme values. They may be, furthermore, taken as a sign of possible errors (cf. Rousseeuw & Hubert, 2011: 73).

3.4.2 Distribution of types

As commented before, the distribution of types, i.e. different words regardless of their number of repetitions, is different in each group: for the groups of independent verbs it is 46 while it is 76 for the

benefactive v2s. When plotting the types in each verb and benefactive, we obtain the following distribution:

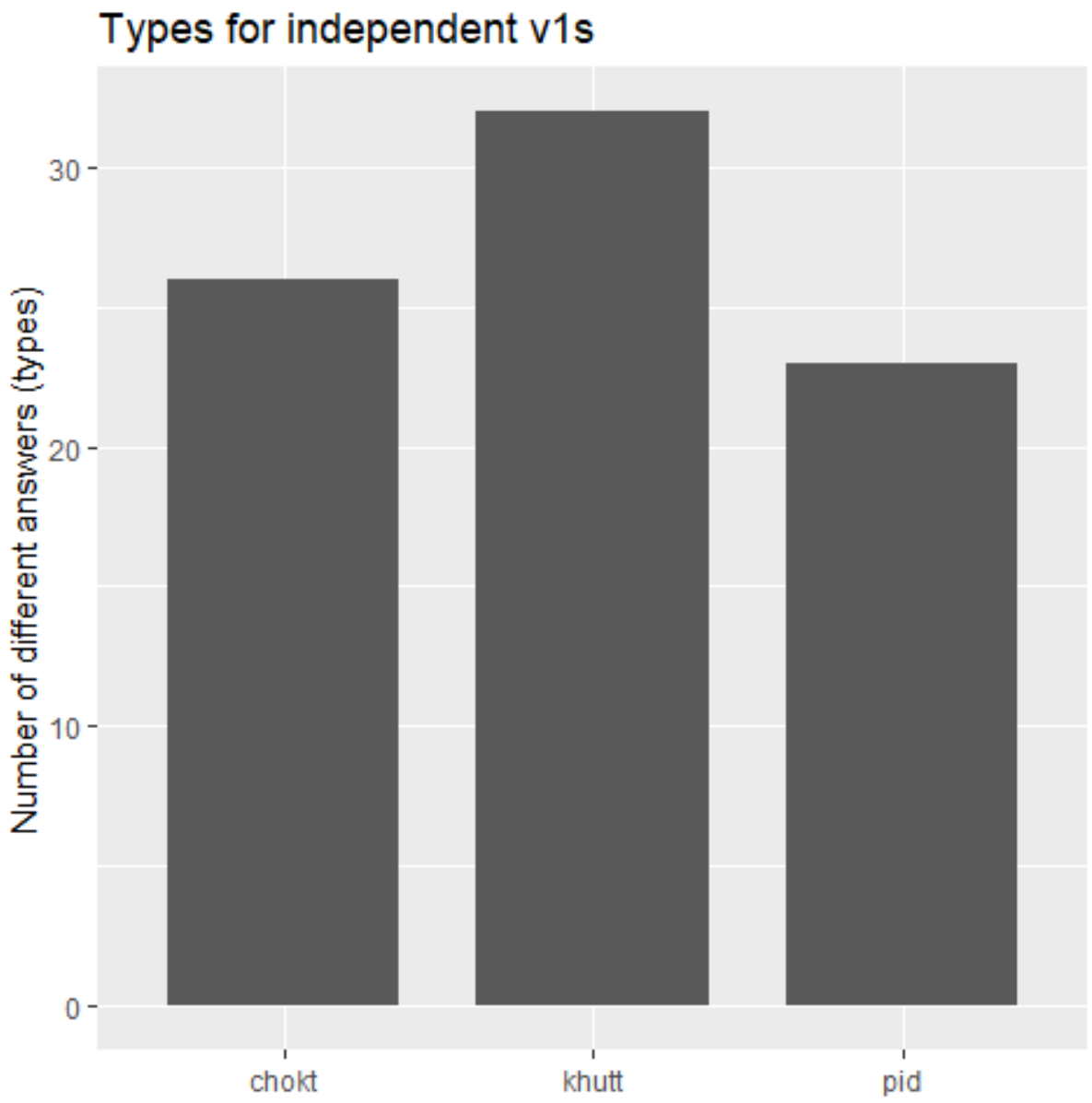

Figure 3. Number of types fora each v1

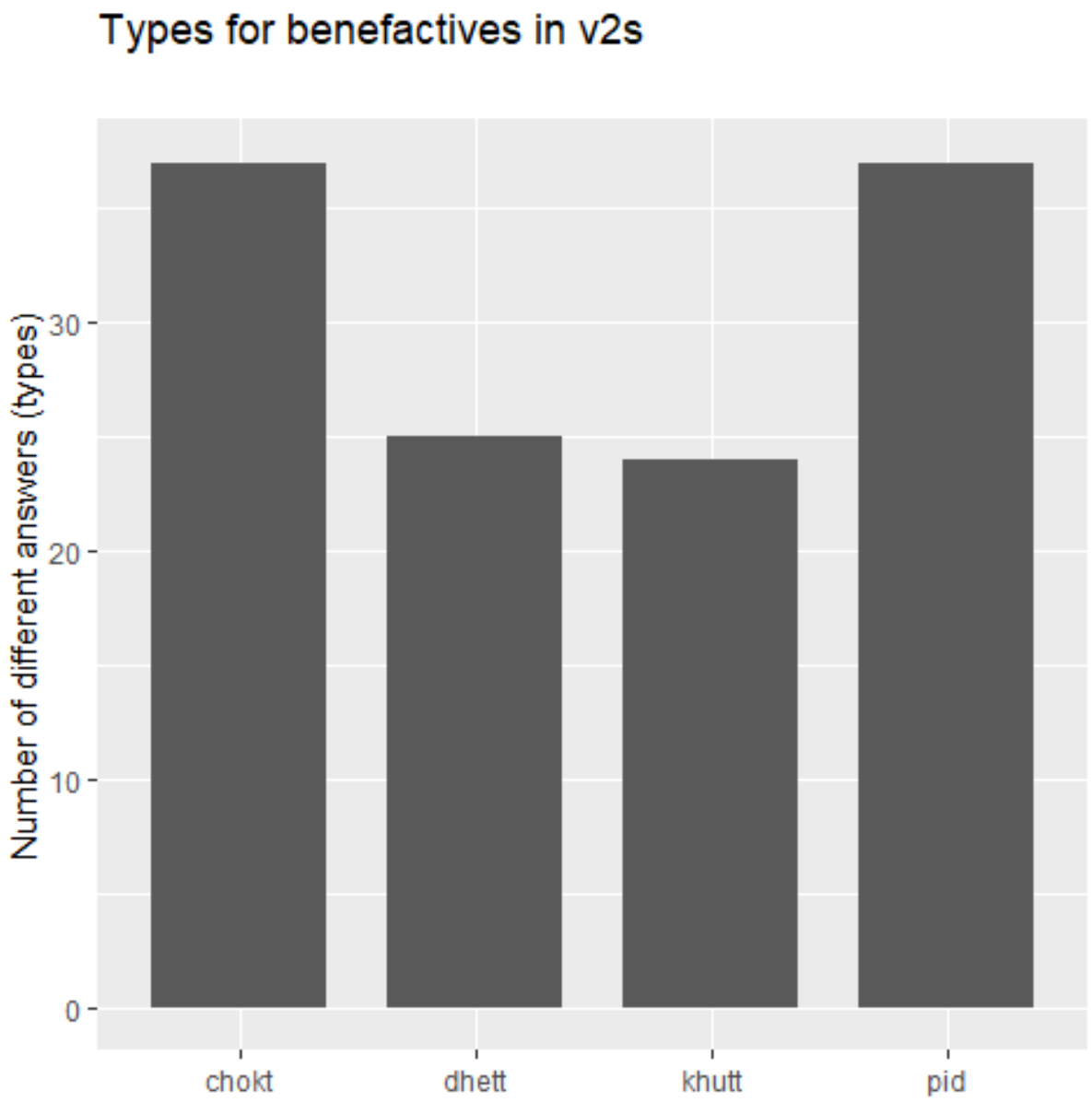

Figure 4. Number of types fora each v2

Image 3. and 4. give us an impressionistic distribution of types. While *pid* and *chokt* have the smallest amount of types as v1s, they have the upper hand as v2s. This plot, however, does not display crucial information on the distribution of tokes: what type is shared by which benefactive and by whom (which participant). The following plot presents the number of repetitions of each noun on each group and each benefactive/v1:

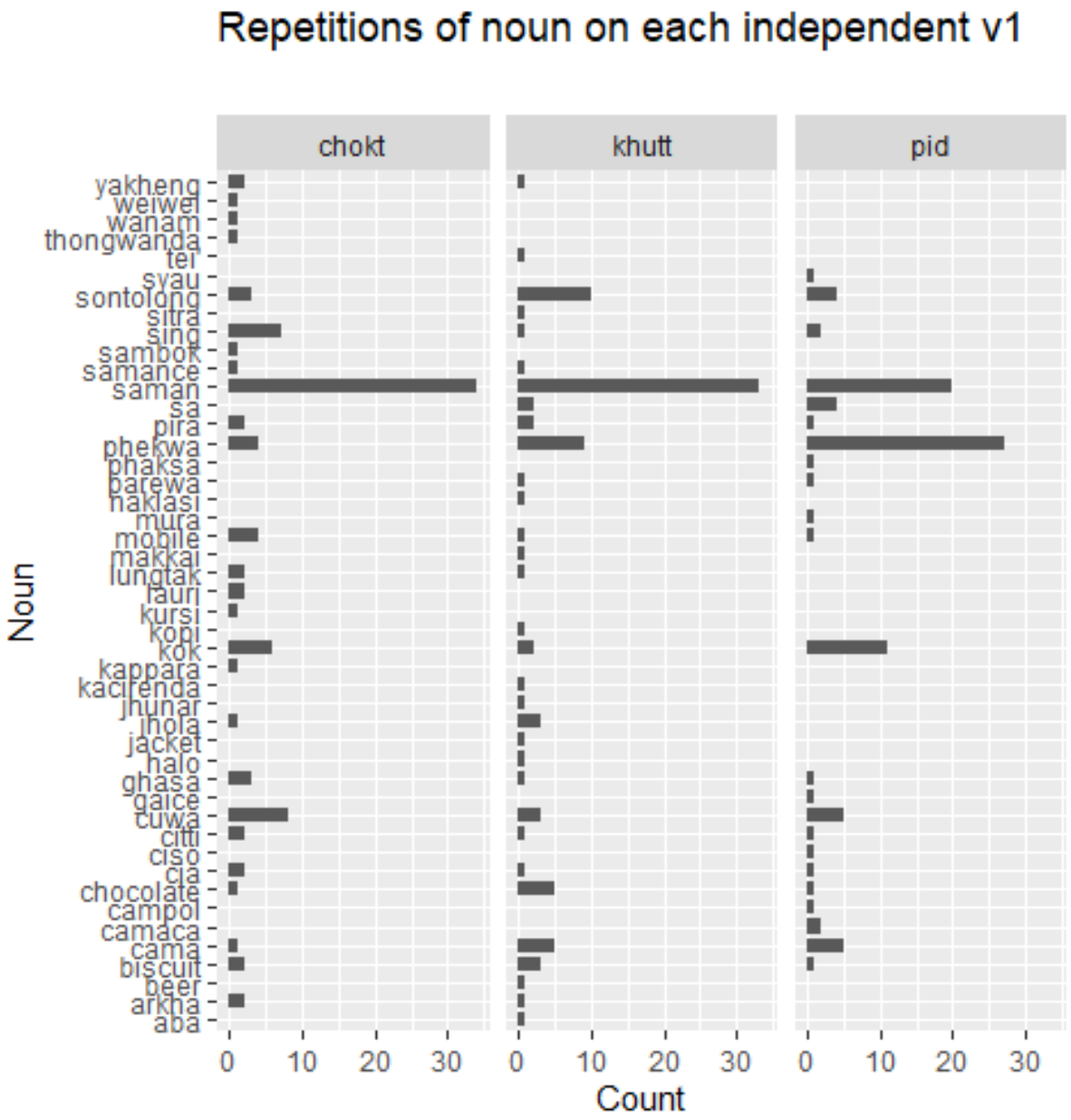

Figure 5. Number of types fora each v1

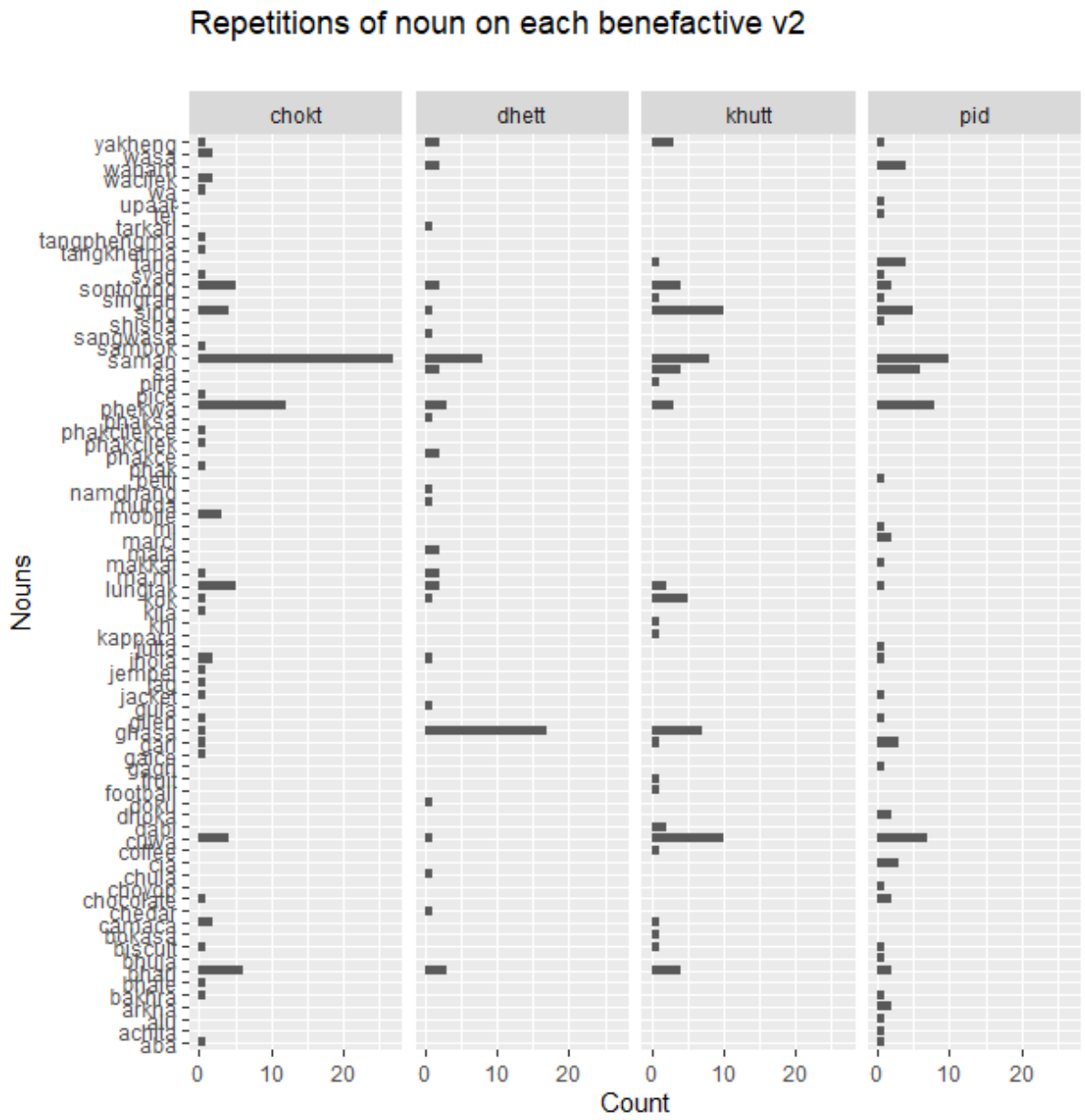

Figure 6. Number of types fora each v2

In Figure 5. and 6., we observe the number of repetitions (Y-axis: Count) for each noun on each verb/benefactive. Again, the Zipfian distribution of data is here most evident, as the majority tokens appear but once and only a handful of nouns have a large number of repetitions.

These plots, however, do not allow us to asses patterns or tendencies because of the large number of nouns. Consequently, the next section will pursue the task of finding a more telling form of visualize the present distribution mostly using component analysis.

3.4.3 Visualization - Correspondence Analysis

In this section, we will go through a further analysis of this data in order to account for possible axis of variation. It should be noted that the plots displayed in this section are an artifact of the visualization method and they are therefore not to be read as descriptive, but rather as an analysis of the data that aims at demonstrating or falsifying a hypothesis that could hardly be assessed by looking at the raw data alone.

The main tool in this section for visualizing the relation between benefactives and nouns will be Correspondence Analysis (henceforth CA). In the social sciences, this method was popularized by Pierre Bourdieu (1979), who famously plotted habits, hobbies, consumption and diverse trends (subsumed as *taste*) along the axis of economic and cultural capital. The filler points were his Parisian informants. In our field, CA is well established as a tool for describing correlations, perhaps most prominently in cognitive and sociolinguistics (cf. the work of Plevoets et al., 2008 and Krawczak & Glynn, 2012 as an example).

This method creates a contingency table of individuals and observations that collapses the dimensions of the matrix and plots the coordinates. The nearness between individual column-values and row-values accounts for correspondence between these points. Concretely, CA applies a Chi-square test to assure that rows and columns are not independent, i.e. it tests the significance of the deviation of the table. The Chi-square statistics will confirm whether there is indeed a relationship between rows and columns or benefactives and nouns.

CA takes the frequencies of co-occurrence and transforms it to distances. In order to plot the correlations on a two-dimensional plane, CA divides the value of each cell to the row total (a value known as profile) and the same operation is done to columns. To avoid overrepresenting infrequent values, CA compensates the values with weighted averages of the profiles to bias overrated scores. Each column has a weight of 1 distributed between the cells according to the given numbers. These values are then used as coordinates in our plane. The distances between these values are assessed geometrically, which accounts for how spread the cloud of data is. The smaller the differences, the more compressed the points will be. Similarly to Principal Component Analysis, this method extracts the crucial dimensions which are responsible for the most variation in the data (cf. Levshina, 2015: 376; Glynn, 2014: 134f.).

The following figures present the results for our two groups (v1s and nouns and v2s and nouns).

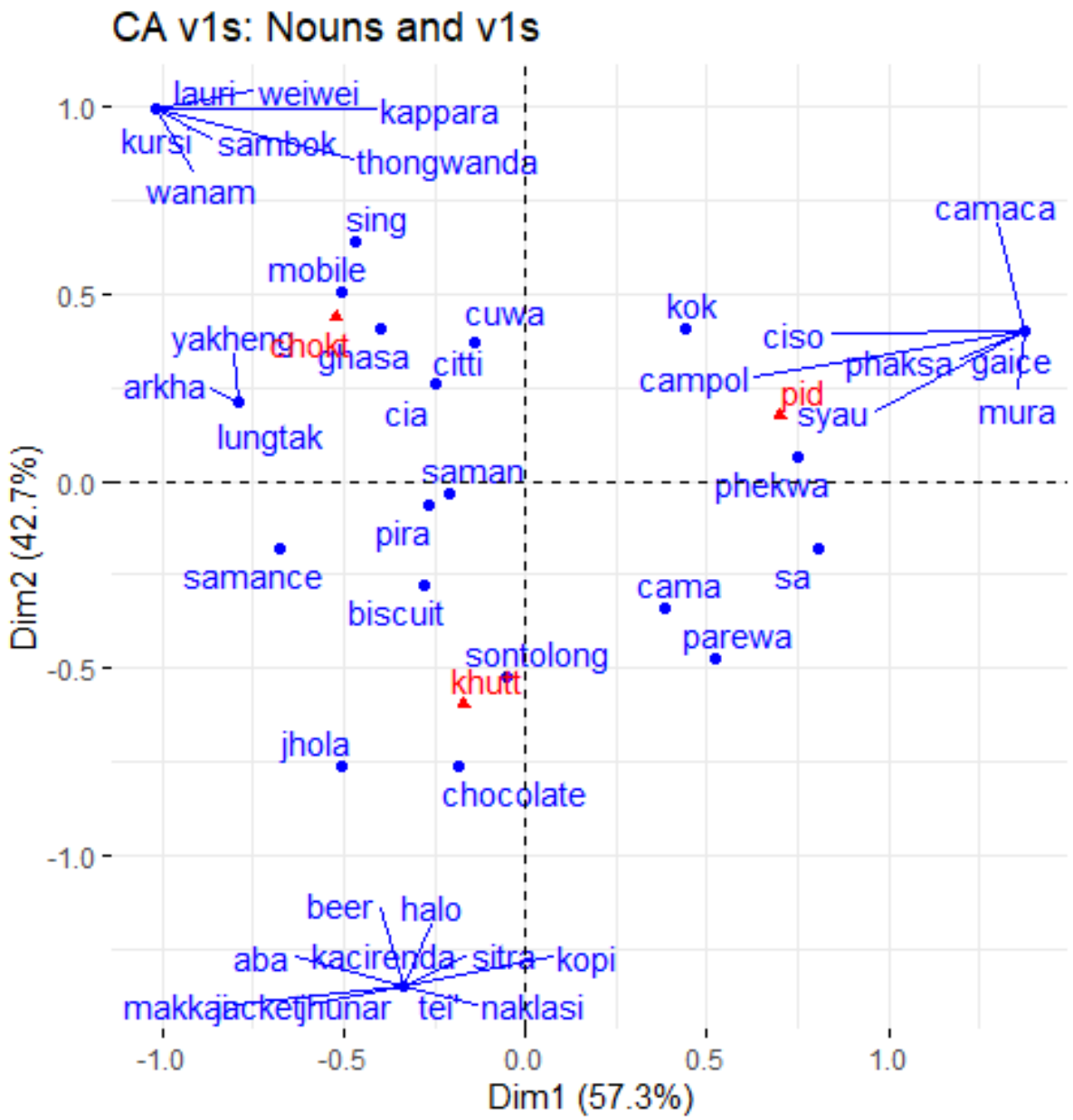

Figure 7. CA of independent v1s with nouns (types)

In Figure 7., the nouns are in blue and the independent verbs in red. The nouns that are clustered on the edge of the plot correspond to the types that only appear with one specific verb. The rest is shared by two or three verbs. The fit of the model (taking the best two dimensions) is high, as both dimensions (here as axis) account for more than 90% of the distribution.

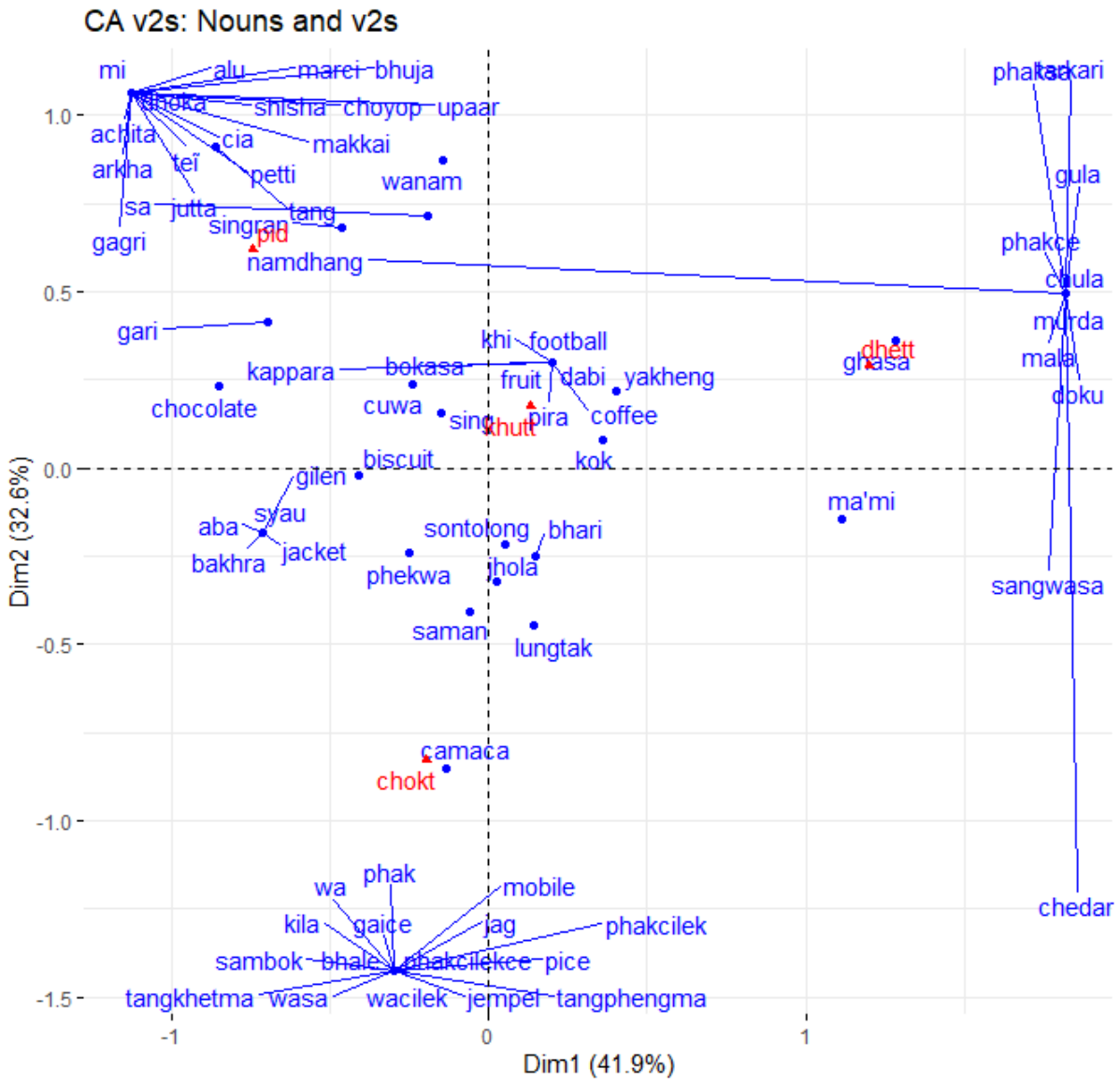

Figure 8. CA of benefactive v2s with nouns (types)

In Figure 8. we see in blue our nouns and in red our four benefactive v2s. The nouns are highly spread out forming four clusters. The benefactives form a triangle with *khutt* in the center. Since rows and columns are treated symmetrically, the proximity of particular nouns and benefactives should be interpreted as the latter having an important weight over the former. Furthermore, the fit of the model is barely bellow 75%, which is the threshold for explained inertia (cf. Glynn, 2014: 154). In should be noted that this figure does not take the host verb of the complex predicate into account. These will be addressed in the following section.

As mentioned before, individual values given in a row (in our case, the number of repetitions of one individual noun in our four benefactives) only have an influence over that particular row. Outliers, such as *saman* 'goods' or *ghasa* 'grass', have a smaller value (therefore a more central position) compared to the majority of other nouns which appeared only once (hapax). This is displayed in Table 4.:

Table 4: Example of CA				
nouns (types)	*pid*	*chokt*	*dhett*	*khutt*
ghasa	0.0000	0.0400	0.6800	0.2800
achita	1	0	0	0

In Table 4., the value of *achita* 'special rice' (upper left corner in Figure 8.) is bigger in the CA plot than *ghasa* 'grass/fodder', even though *ghasa* occurs 25 times while *achita* only once. In other words, the CA plot rewards hapax values. This is helpful to spot groups of nouns that appear with one particular benefactive but it also magnifies their importance and does not account for nouns which were repeated several times.

As for the first group of independent verbs, we observe a cluster of nouns on each verb and a spread distribution of nouns shared by other verbs. In the second group, all benefactives have their hapax-entourage except for *khutt*, which is in the middle. There are, however, seven hapax nouns that only occur with *khutt* but the plot does not pull them apart as it does with the hapax nouns of the other benefactives. Instead, these are mixed up with other nouns that are shared by other benefactives as well.

In sum, the CA plot distributes our four benefactives according to the number of repetitions of each noun. It also down-plays the importance of outliers and highlights hapax values, which are clustered around each v1/v2.

Other possible explanations for the results of the cloze test

Influence of participants

In this section, we will explore the possible influence of individual speakers in the results of the experiment. As a start, the following figures represent the number of repetitions of nouns in each v1 and v2 according to which participant uttered them:

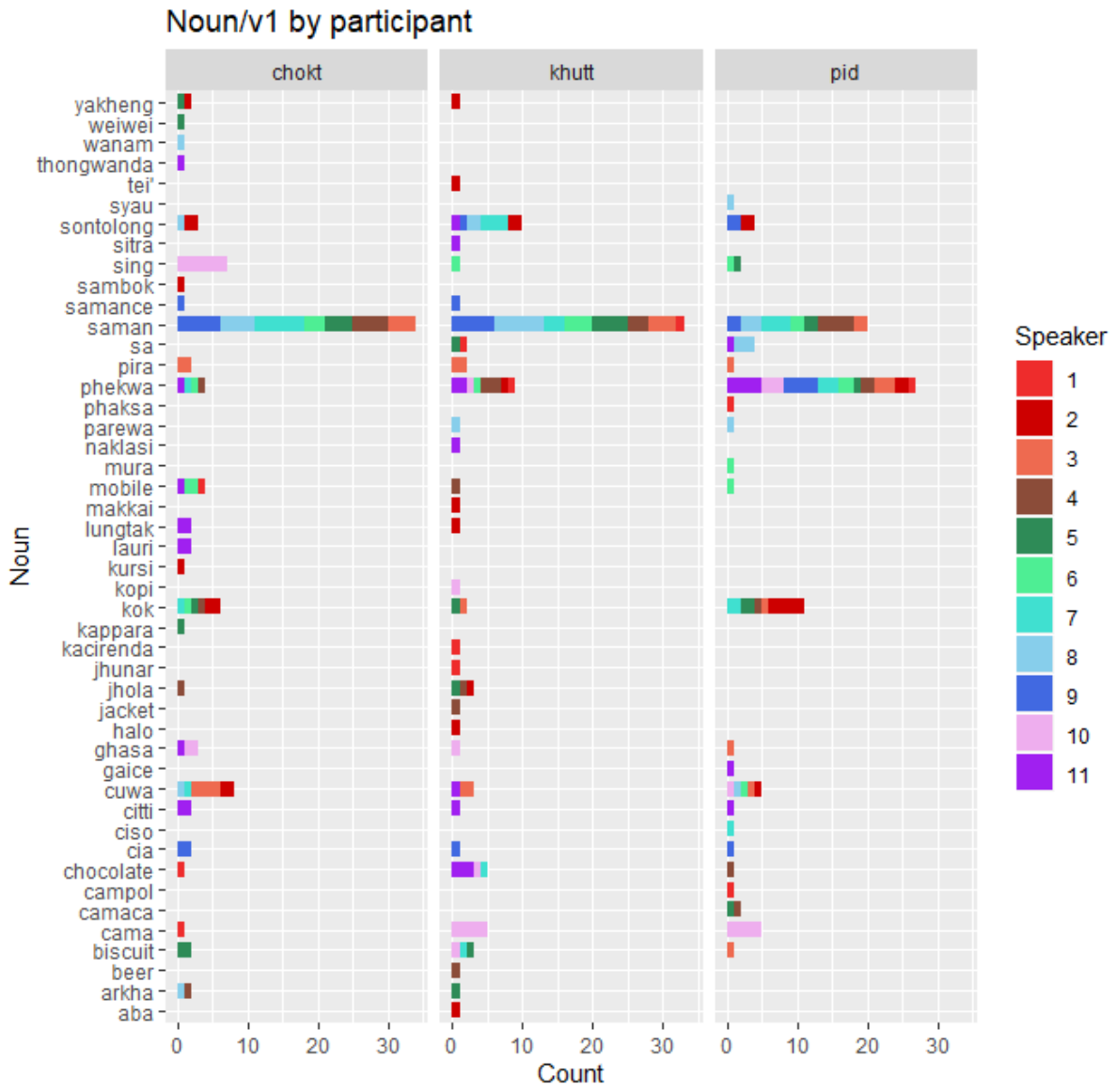

Figure 9. Repetition of nouns in v1s on each participant

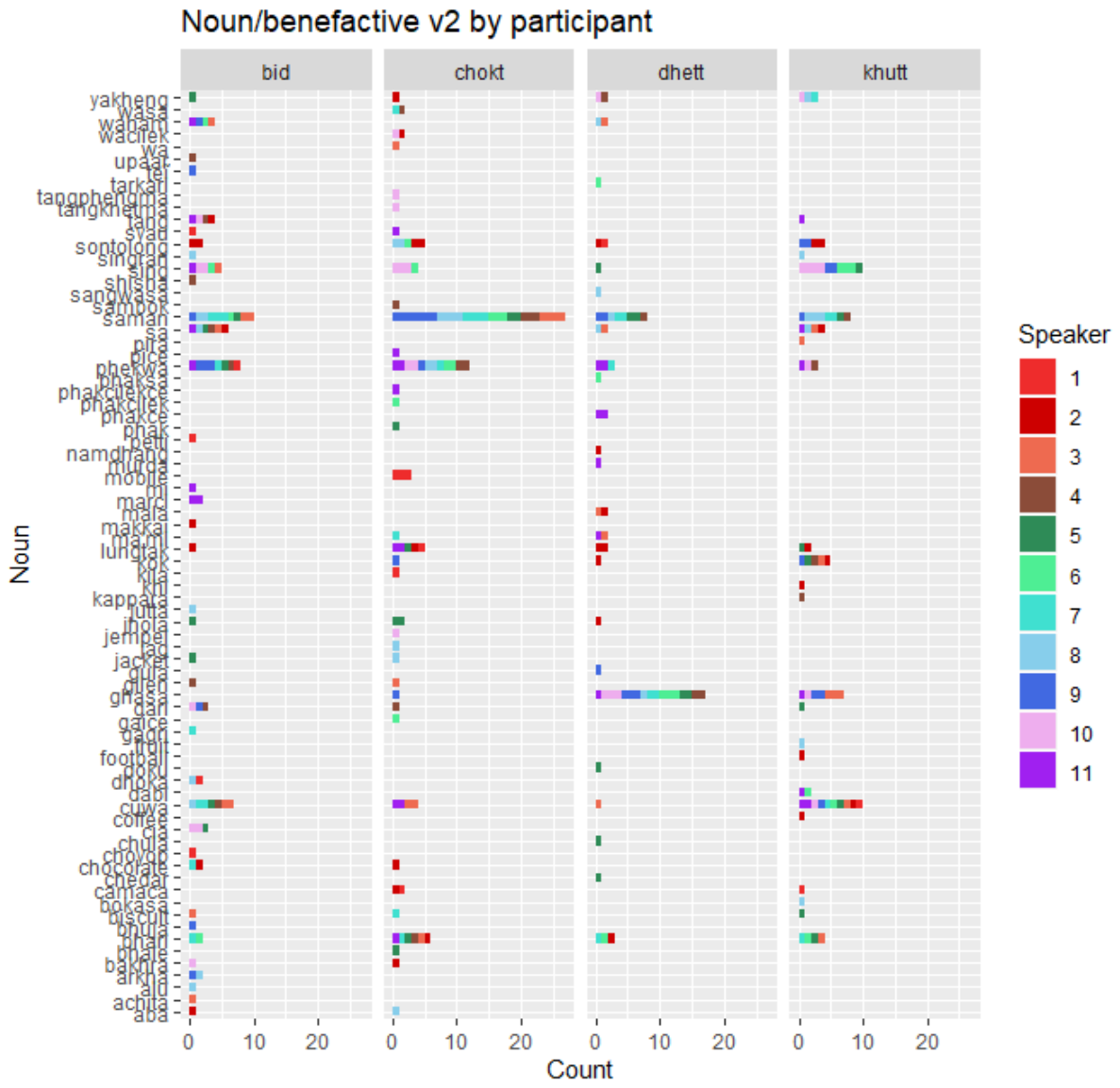

Figure 10. Repetition of nouns in v2s on each participant

Figure 9. and 10. present the same data as Figure 5. and 6. but with colors depending on which participant (from 1 to 11) used a particular noun. This is also helpful to spot repetitions of nouns by the participants. The first plot contains the all types that were uttered in each benefactive and the color stands for individual participants. In the second plot, the four bars stand for the four benefactives. A quick overview reveals that most words appear only once and are seldom uttered by a different speaker: they are truly hapax values. Furthermore, outliers are composed of several speakers; it does not look like as if one speaker would have repeated e.g. *saman* 'goods' 27 times. In Table 5. and 6., we zoom into types with 5 or more occurrences in at least one participant. These types are displayed for all participants:

Table 5: Highest values in v1s											
Type**subject**	1	2	3	4	5	6	7	8	9	10	11
cama	1	0	0	0	0	0	0	0	0	10	0
cuwa	0	3	7	0	0	1	1	2	0	1	1
kok	0	7	2	2	4	1	3	0	0	0	0
phekwa	2	3	3	6	1	4	4	0	5	4	8
pira	0	0	5	0	0	0	0	0	0	0	0
saman	1	0	10	13	11	9	14	15	14	0	0
sing	0	0	0	0	1	2	0	0	0	7	0

Table 6: Highest values in v2s											
Type**subject**	1	2	3	4	5	6	7	8	9	10	11
cuwa	1	1	6	1	2	1	3	1	1	1	4
ghasa	0	0	3	2	2	3	2	1	6	4	2
lungtak	1	5	0	0	2	0	0	0	0	0	2
phekwa	1	0	0	4	1	2	3	2	4	3	6
saman	0	0	6	5	6	4	11	10	11	0	0
sing	0	0	1	0	2	5	0	0	2	9	1
sontolong	1	7	0	0	0	1	0	2	2	0	0

Table 5. and 6. display the sum of repetitions of each type according to the subject (by number) *without* considering the individual verbs and benefactives; all these values (from three verbs in the first plot and from four benefactives in the second one) are summed up.

In the first table, we observe that there is a large number of high values (eg. 10 for *cama* 'food'), but these are not echoed by all participants. *Saman* 'goods' and *phekwa* 'money' are an exception to that pattern. As for the second table, we observe a larger number of ones compared to the first plot.

The next step to account for the influence of participants is to perform a correspondence analysis, as it was done profusely in this section, in order to find frequency-based associations. The following plots present the results of this method:

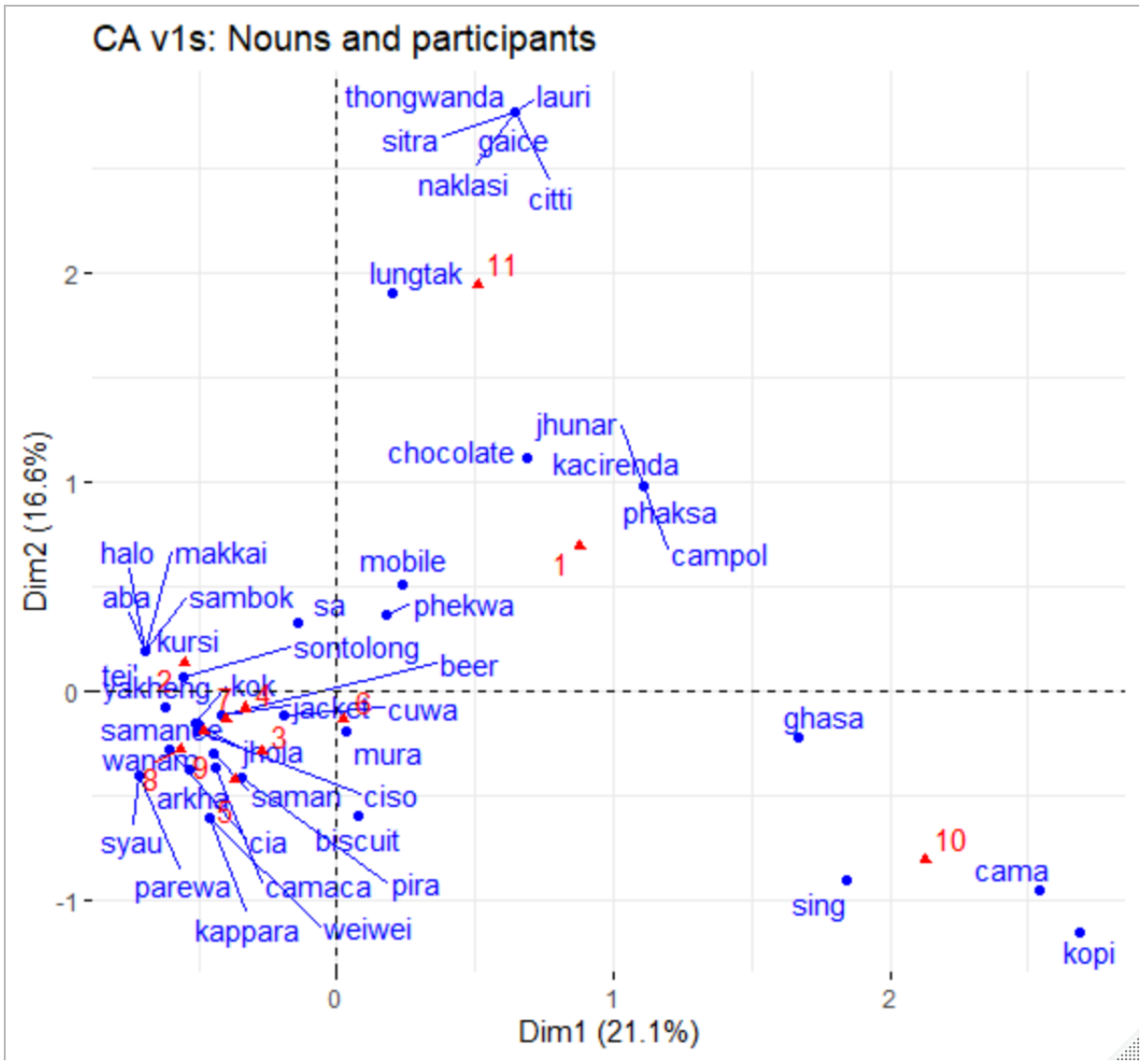

Figure 11. CA of participants and nouns (types) from v1

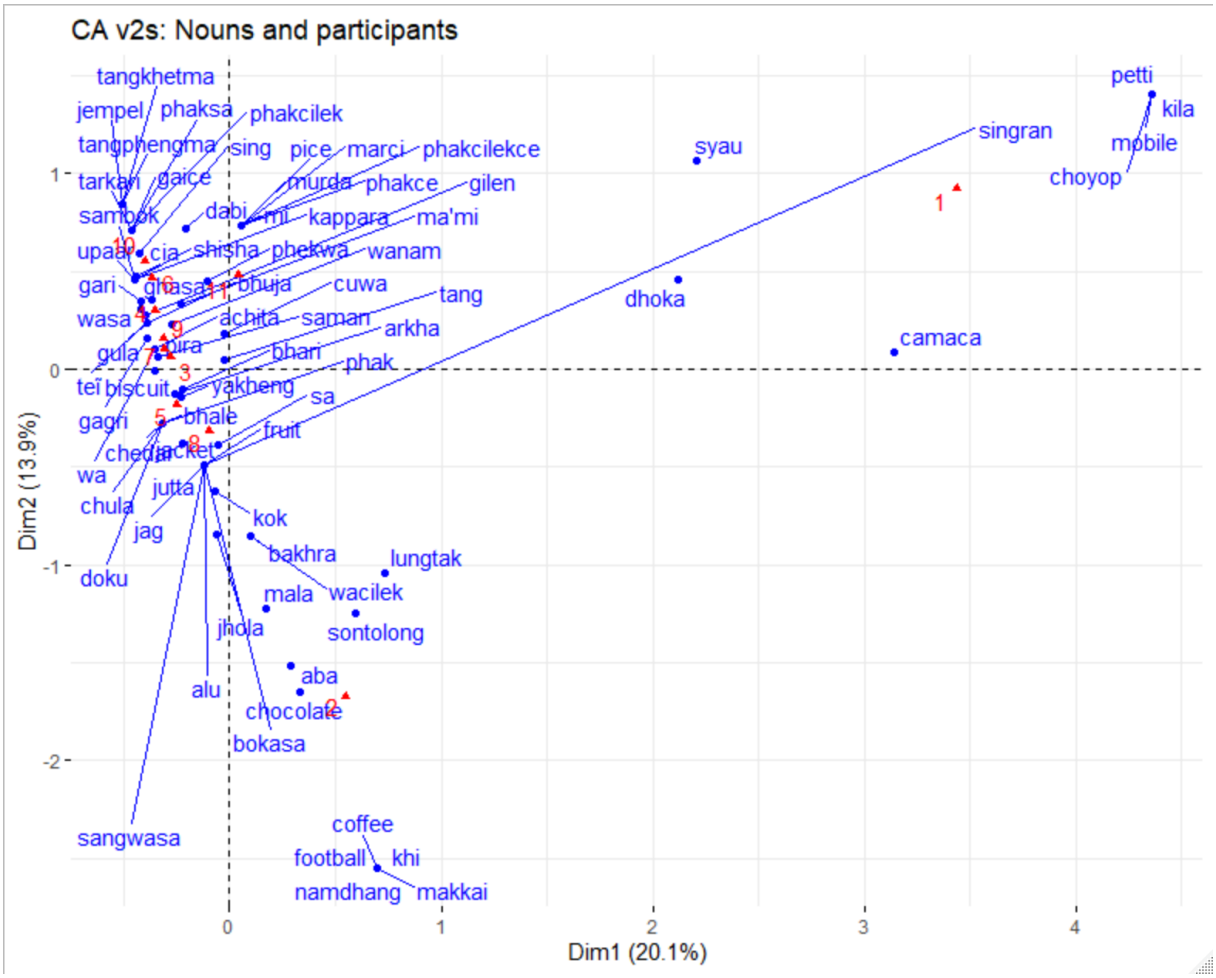

Figure 12. CA of participants and nouns (types) from v2

In Figure 11. and 12., we observe that most of the values are clustered around the left of the center. In Figure 11., three participants are spacially far away from the rest while in Figure 12., there are only two of them. In both cases, participant nr. 1 seems to deviate from the norm. The percentages for each dimension in both plots sum less than 38%, which may be interpreted as a bad fit for the model.
In sum, most of the participants seem to have provided similar answers, while some participants display a more aberrant behaviour. The influence of participants will be taken up again in Section 3.5.

Influence of v1 over nouns in complex predicates with a benefactive v2

So far we have compared one set of simple verbs with the benefactive v2 of a complex predicate. As for the latter group, the influence of v1 over the noun (i.e. without considering the benefactive v2) is a potential obstacle for the whole investigation, since we usually assume that grammaticalized v2s tune or specify v1 but leave its semantic specificities largely unchanged. One way of reducing the influence of the v1 group is by having many of them. In our case, there were 30 types of v1s in a total of 40 different sentences. In the following CA, we plotted the distribution of nouns according to v1s:

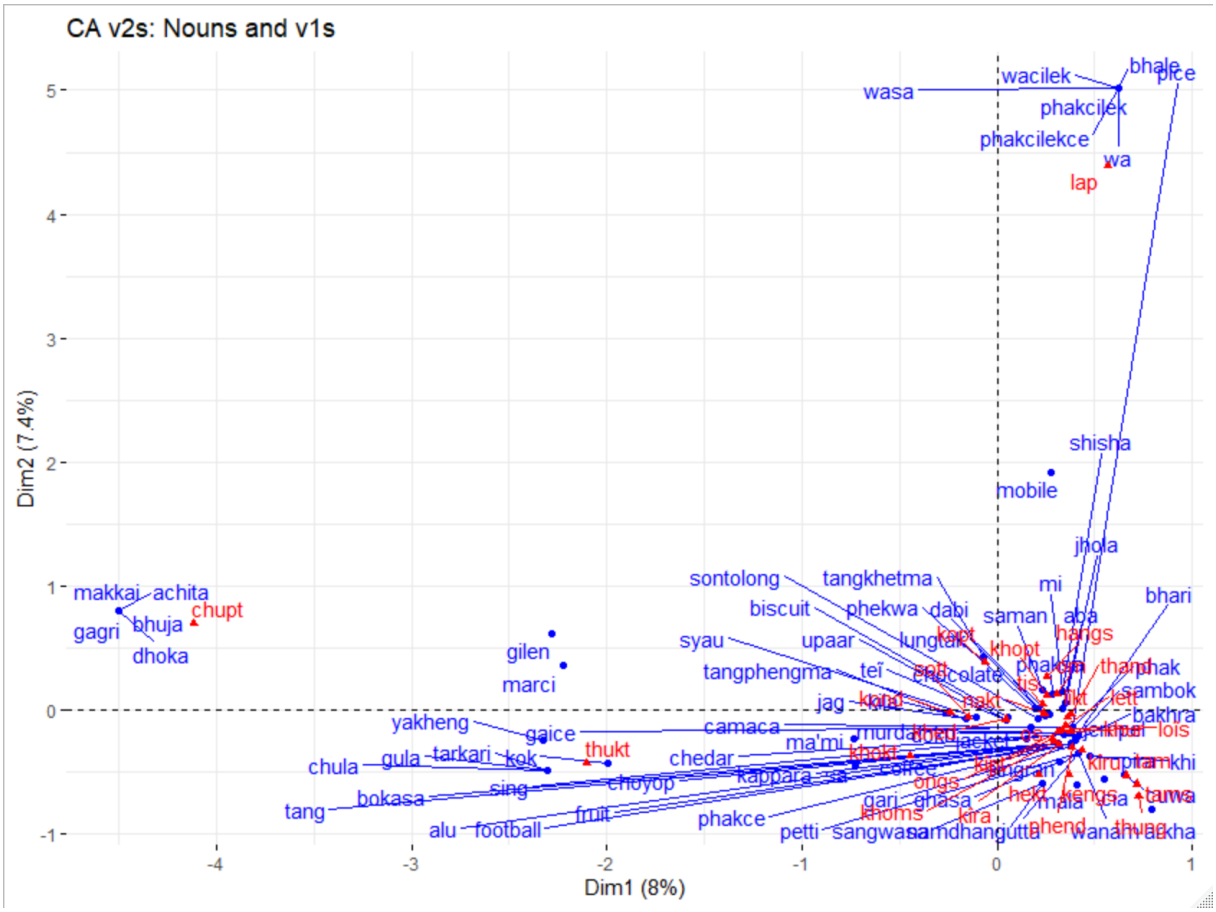

Figure 13. CA of v1 and nouns (types) from complex predicates

As it is visible in Figure 13., most v1s (red) are clustered in the lower right corner except for *lap, chupt* and *thukt*. The fit of the model is also low (less then 16%). The three verbs which are not clustered together with the others are the following: *lap* 'to catch', *chupt* 'to grab' and *thukt* 'to cook'.

Based on the results from this plot, we may assume that the influence of v1s over the response to complex predicates to be low. We may now continue to the next step in the analysis by introducing two concepts from information theory.

3.5 Information theory

The last approach of the present investigation will be to apply some notions of information theory to the results of our cloze test, namely entropy and pointwise mutual information (PMI). In this section, we will go through an introduction and the application of these two concepts.

3.5.1 Entropy

In his famous publication of 1948, Shannon set a cornerstone of how to measure the information potential of a given system of symbols. Very generally said, any string of values can be adapted to be

measured in term of how informative it is or how much uncertainty it conveys. A set of coins, strings of letters put together or even a real language are means by which we can measure uncertainty, i.e. we can measure the amount of information required to describe the variable. This is termed *entropy* and it is expressed in bits. The mathematical definition of entropy looks thus:

$$H(X) = -\sum p(X) \log p(X) \tag{1}$$

The probability of each event has to be calculated. Events with zero probability do not affect the equation as 0 log 0 is equal to zero. The higher the entropy, the bigger the uncertainty. In this sense, the entropy of a biased coin that always ends on heads is lower (no uncertainty) than a fair coin, which has a 50% chance of ending on heads or tails. Similarly, a fair dice has a higher entropy than a biased dice having a 40% chance of ending on 1. The entropy is higher as it is more difficult to guess the outcome of the fair dice over the biased one (Cover & Thomas, 1991: 12f.).

Entropy has been much used in linguistics given that *event* could mean anything: morphemes, phonemes, letters etc. In his initial publication, Shannon (1948: 4f.) even mentions the use of entropy in measuring the information of telegraphic messages in English, where the sequence of letters is obviously not random. As described by Bentz (2018: 54), the more elaborate a system of symbols (i.e. the more choices it has to its disposal), the more information it can express. As entropy increases, the system entails more uncertainty and interpretation also requires more time. Furthermore, since it is a measurement of dispersion and uncertainty, it allows to compare two systems with differing number of objects.
As an example of entropy being used in linguistics, Gibson et al. (2017) tested the uncertainty related to naming objects in Tsimane (probably an isolate, Bolivia). After assessing that Tsimane speakers show a greater variability in a color-naming experiment (except for red!) compared to English speakers and speakers of Bolivian-Spanish, the authors had to make sure that this higher variability was not just on account of the unfamiliarity of the Tsimane with the stimulus used in their experiment (Munsell cards). In order to rule out this option, the authors asked all Tsimane participants to perform a memory test: they asked them to indicate what colors they associate with particular objects. Using entropy (based on types and tokens), the authors concluded that the colors of the objects that present high entropy, and therefore higher uncertainty, are the same as the colors that were shown in the first experiment to have higher variability, i.e. where Tsimane speakers seemed to have less agreement on. In other words, having high variability when describing colors seems to be related to Tsimane (which was compared to Spanish and English) and not an error of the employed methodology.
A further example of the linguistic use of entropy actually involves Chintang. Stoll et al. (2012: 14f.), in trying to account for the influence of the noun to verb ratio in language acquisition (both child speech

and adult speech, the latter being analyzed as an input to the former) and morphological complexity, used entropy to determine proficiency in the following manner: Entropy has been described as a measure of uncertainty and and thus to grow with the amount of choice and variation available in a system. Applied to verbal morphology, the authors argue that language proficiency positively correlates with the degree of uncertainty of one particular form to be chosen from a paradigm of forms (in that case biactantial verbal agreement). Therefore, a fluent speaker should make a larger and more variable use of forms in a given paradigm compared to a learner of a language, who might stick to a few trusted forms and neglect the rest. Entropy is higher when both predictability is low, i.e. several forms have a chance of being used and when the active paradigm is large, as many forms come into play (also cf. Moscoso del Prado et al., 2004).
In what follows, we tried to apply these conceptions to our data from the cloze test.

Values of entropy for independent verbs and for benefactives

Table 6. presents the values of entropy for each verb/benefactive along with number of tokens, types and hapax values from our cloze test. Additionally, there are also two extreme sets of values termed *simulations*, which help understand what the range of these numbers is. In order to avoid calculating the logarithm with base 10 of zero[13], we smoothed the dataset by adding 0.00001 to all zero values.

Table 7: Values for v1					
	pid	*chokt*	*khutt*	Simulation 1	Simulation 2
Type	23	26	32	98	1
token	94	95	98	98	94
Hapax values	14	10	21	98	0
Entropy	2.4	2.6	2.7	4.58	0

Table 8: Values for v2						
	pid	*chokt*	*dhett*	*khutt*	Simulation 1	Simulation 2
Type	37	39	25	24	84	1
token	84	99	60	74	84	60
Hapax values	22	27	13	13	84	0
Entropy	4.7	4.8	4.4	4.1	4.6	0

In Table 7. and 8., we observe our three verbs and our four benefactives plus two other simulated sets

[13]Our data has many zero values because e.g. if one noun only occurs with one verb or benefactive in the tabulation, then the number of repetitions for that noun in the other verbs or benefactives is going to be zero.

of numbers. For all elements, there is information on the number of types, tokens, number of hapax values and the calculated Shannon entropy. As for types, we observe a rather even distribution among the verbs and a larger difference between *chokt* (39) and *khutt* (24) in the benefactives. We had already seen this distribution in Section 3.4.2. In the case of tokens, have an even distribution among the verbs and less so among the benefactives. The hapax values are nouns that appeared only once for each benefactive. Most of them appeared only once in each set of verbs and benefactives. *Chokt* has the largest hapax values among the benefactives and *khutt* among the verbs (the longest tail in statistical parlance).

The two simulated values show extreme numbers for entropy, i.e. the highest and the lowest possible score for the available values. In the first group (of independent verbs), the highest possible entropy is achieved by having the same number of repetitions (in this case 1) in each token. The smallest possible entropy is achieved by having all tokens concentrated in one type; thus entropy is zero, as there is no uncertainty. The same was conducted for the second group of benefactives. The aim of these simulations is to have the boundaries of the possible values for entropy, which gives us a parameter to compare the observed vales in verbs and benefactives.

One methodological question about entropy remains unanswered: does the number of tokens strongly influence entropy? Bentz (2018: 66f.) has investigated this issue using the European Parallel Corpus and came to the conclusion that, for English and 21 other languages represented in that corpus, the value of entropy stabilizes at around 25000 tokens. Before that number, entropy grows exponentially with the number of tokens. This might explain part of the distribution, but then we would expect *chokt* to have a much higher value for entropy than *dhett*, as it has almost 40% more tokens.

What are the implications to our experiment? The number of tokens is certainly below 25000, but we could argue that, if the benefactives have a similar number of tokens, this bias should affect all benefactives in a similar form. More than that, this inherent bias will remain unattended.

Comparison of entropy between v1 and benefactives

Having assessed the entropy of each group, the next step towards a comparison is putting all values together and then calculate again the entropy of each verb and its corresponding benefactive. The benefactive *dhett* can thus not be compared to anything, as it has no known v1 counterpart.

In order to compare verbs and benefactives, both datasets were merged using rbind (RStudio, 2015). Thereafter, the entropy of each benefactive and verb can be measured straightforwardly: Since entropy is a measurement of uncertainty, two systems can be compared even if the number of elements (in our case tokens) is not the same. The value of each of benefactive and verb was turned into a percentage, taking the highest possible value for entropy as 100%. The highest possible value was calculated taking

the highest number of tokens available, which is in our case the benefactive *chokt* with 99 tokens. A new simulation was created with 99 types and the measured entropy (4.58) in that group was used as 100% for all other verbs and benefactives. As an example, the entropy of the benefactive *pid* is 3.26, which is a 70.2% of 4.58. Image nr. 14. displays the comparison of all values:

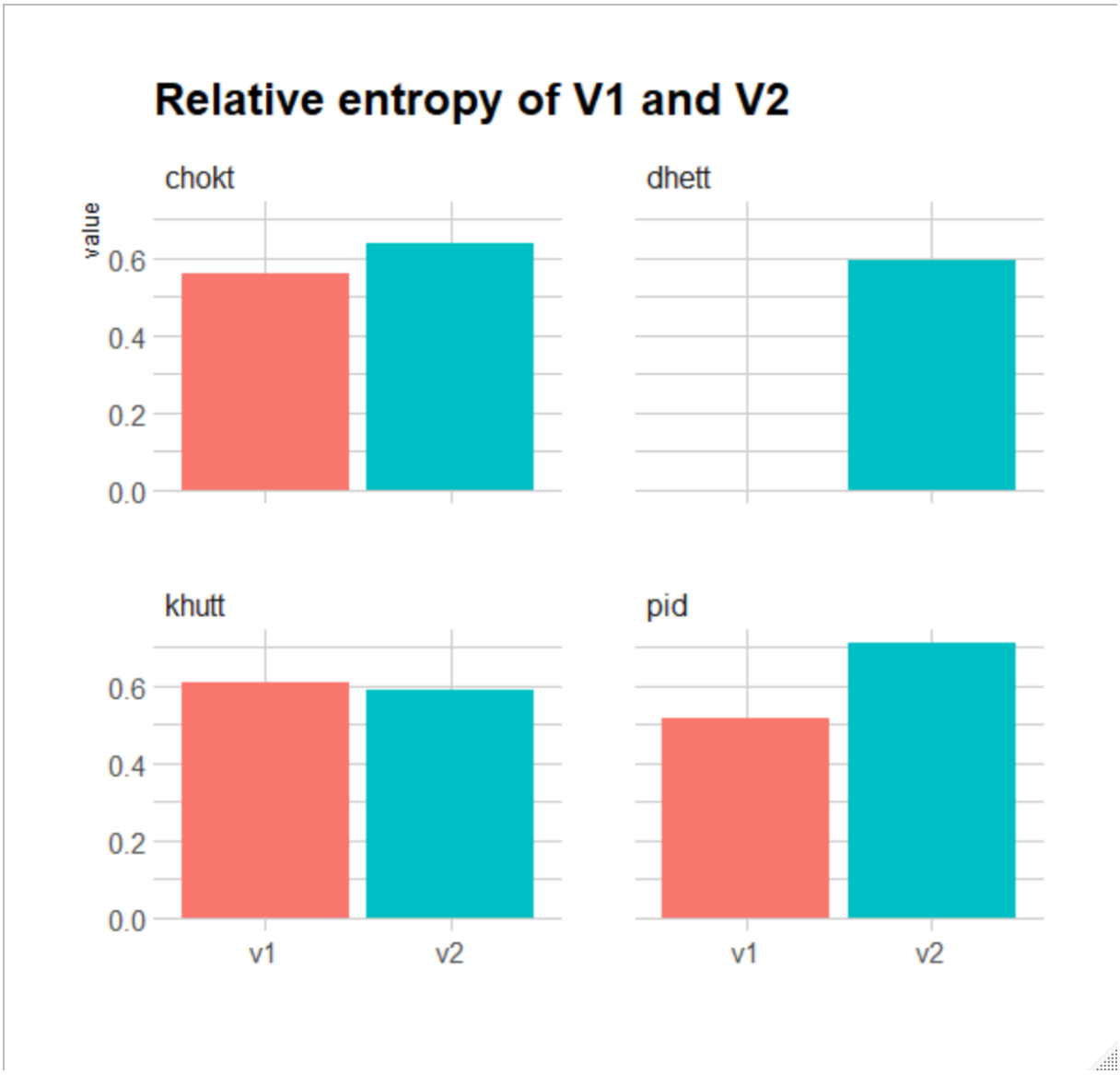

Figure 14. Relative entropy

In Figure 14., we observe the entropy displayed pairwise for independent verbs (termed v1) and benefactives (termed v2). The scale is set to one, which stands for the highest possible entropy (i.e. the 100%). *Dhett* has no v1 counterpart. While there is an increase in entropy from v1 to v2 in *pid* and *chokt*, the opposite is true for *khutt*, albeit to a much smaller extent.

Influence of participants

Even though the order of the sentences in the questionnaire containing both groups (independent verbs and compound verbs with a benefactive) was interspersed, the distribution observed in Figure 14. could have been influenced by a small set of participants who might have acted differently, as we have mentioned in Section 3.4.3. In order to discard that option, Table 9. display the values for entropy as v1 and v2 (without *dhett*) for each participant:

Table 9: Values of entropy for each participant

participants	1	2	3	4	5	6	7	8	9	10	11
v1	2.3	2.3	1.6	1.8	2	1.6	1.6	1.5	1.4	1.7	2.3
v2	2.2	2.6	2.3	2.5	2.6	2	1.9	2.4	2.2	2.2	2.6

In Table 9. we observe the individual score (of each participant) for entropy achieved in each group (v1 and benefactive v2). Except for participant 1., the score of entropy for all other subjects increases from v1 to v2 in varying degrees.

One possible next step is to inspect each verb-benefactive pair in each participant. However, the values for each participant's verbs and benefactives are far too little for a meaningful comparison.

As we have seen in Section 3.4.3, the nouns produced by some participants did not cluster together with the rest: they can be thought of as outliers. This is the case of participants 1, 10 and 11 in the set of v1s and 1 and 2 in the set of v2s. Their diverging answers could have masked the whole picture, to some degree. For this reason, we have replicated Figure 14. but without their (of these specific participants) answers. The following figure presents the results, each set (independent v1s and benefactive v2s) without these outlier participants. As an example, the nouns provided by participant nr. 2 in the v2 set were excluded while the answers for v1 were kept.

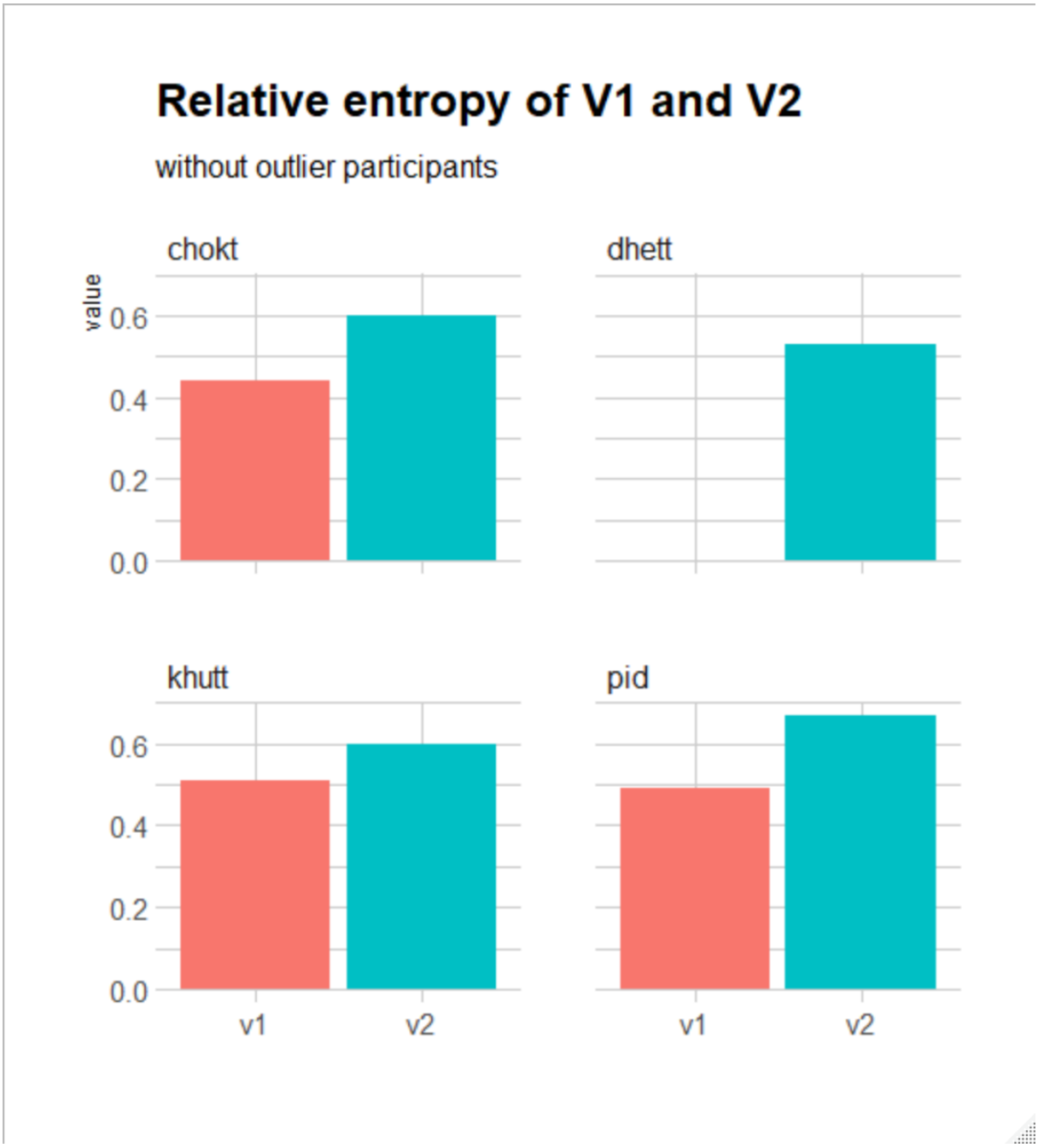

Figure 15. Relative entropy (without part. 1 and 2)

In Figure 15., we observe two bars for each lexeme except for *dhett*. The green bar stands for the entropy obtained by v2s while red stands for v1s. The results were converted to percentages, as it has been already done earlier. There is, however, one difference: a lower number of tokens (due to the exclusion of some participants) means that the maximum possible entropy (the 100%) is slightly lower than in Figure 14. and thus, the percentages of both plots should not be taken as being exactly equivalent, for the number taken to be 100% is slightly different in each plot. The difference between each v1-v2 bar in each plot, however, can still be compared.

In sum, knowing that v1s and benefactive v2s were interspersed in the questionnaire and having seen that the values for entropy for all participants seem to be similar (except for participant 1 in both groups and two other participants, one in each group), we may assume that the influence of outlier individual scores had an influence over the the distribution described in Section 3.4. This will be again considered in Section 3.5.

Influence of individual words (semantic groups)

One further issue that has to be accounted for is the influence that a large range of similar words may have on the larger picture[14]. Even though some words have a rather vague sense (e.g. *saman* 'goods/thing') which may apply to anything, other words are much more specific and even derivations of other words, as in *phakce* 'pigs' and *phakcilecek* 'piglets'. Similarly, many words with related meanings (such as *gaice* 'cows' and *phak* 'pig') may push the whole group and thus also mask the results. One way to counter that effect is to merge similar words into semantic groups and thus reduce the number of types and observe how this affects the values of entropy. Since there is no unified procedure for making these groups, the criterion rests solely on the researcher.

For the present investigation, we created seven broad semantic groups based on the types available. The groups are the following: 1) fruits; 2) animals; 3) body parts; 4) meat - this group has meat from different animals; 5) drinks (beverages) - this group has all drinkable types; 6) clothing - this group encompasses all types that can be worn; 7) devices - this has all man-made gadgets and machines. Table 10. displays these semantic groups and the amount of tokens and types from our v1s and v2s that fall under these categories.

[14]I am grateful to Dr. Lester from the University of Zurich for suggesting this approach to me.

Table 10: Semantic groups					
V1			V2		
	token	type		token	type
animals	3	2	animals	13	10
body parts	0	0	body parts	7	3
clothing	2	2	clothing	6	4
devices	18	8	devices	22	9
dishes	19	6	dishes	15	5
drinks	26	6	drinks	28	4
fruits	21	5	fruits	24	9
meat	7	2	meat	17	5
sum	96	31	sum	132	49
percentage	33.4%	67.4%	Percentage	41.6%	64.5%

In Table 10., we observe the semantic groups for both the independent verbs (v1) and the benefactive v2s along with the amount of types that were collapsed for each semantic group, plus the number of tokens it has. In the first group (individual v1s) we observe two zeros: this is due to a semantic group that was only perceived in the types triggered by the benefactive v2s.

It should be stated, without entering into the analysis of the data yet, that making these groups does not follow any other rationale but that of a trivial sense that two or more words are similar. This method makes the results difficult to analyze, as the implications in terms of entropy of collapsing types and tokens can be substantial.

The following plot presents the results of these semantic groups. Again, all results were thrown together and the entropy was estimated for each verb and benefactive. These absolute numbers were transformed intro relative percentages based on the highest possible entropy from the dataset.

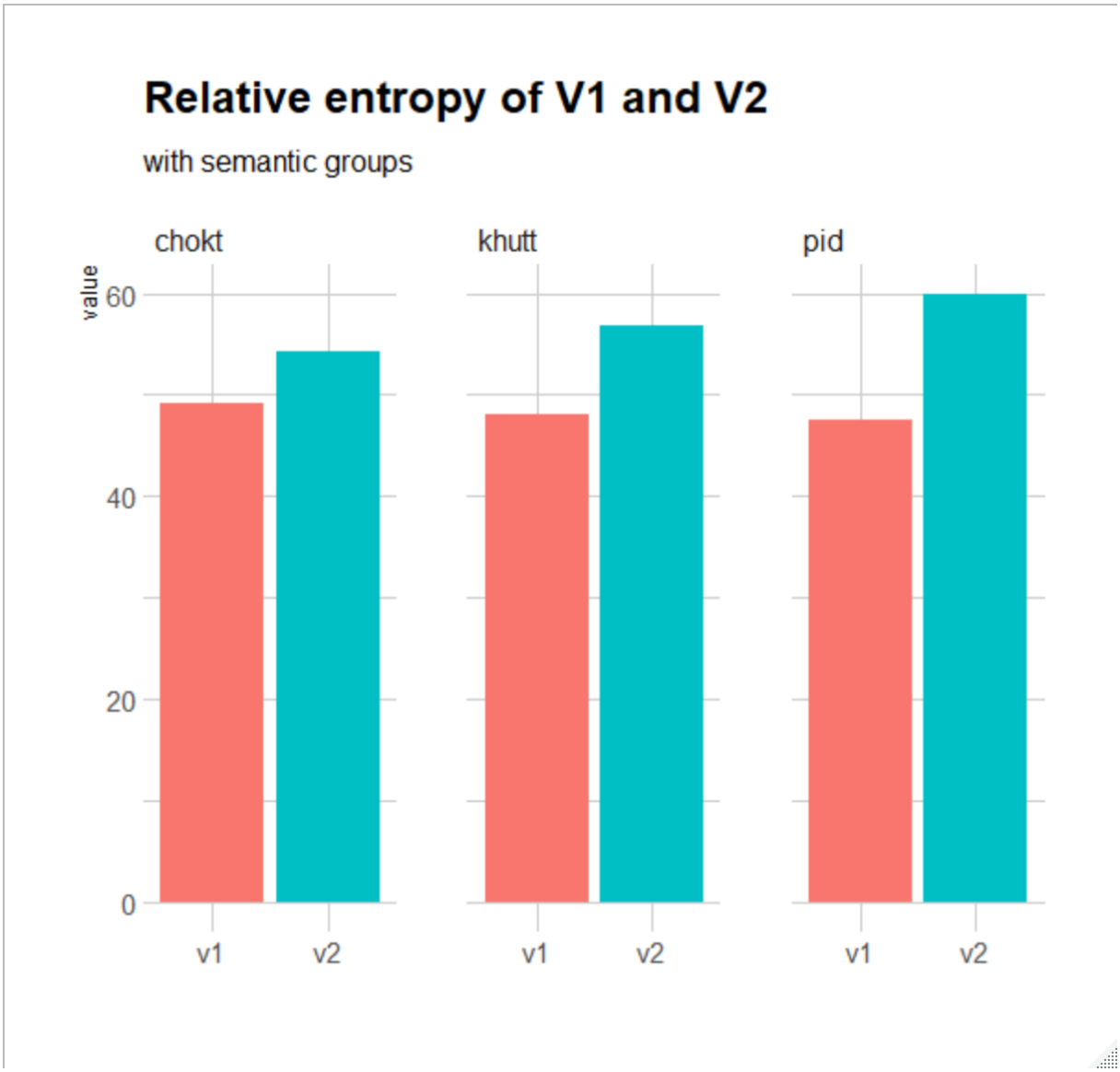

Figure 16. Relative entropy (with semantic groups)

In Figure 16., we observe three double barplots, one for each verb-benefactive. The values are percentages, i.e. that 100 is the maximum.

Bringing similar words into one category and thus collapsing a large number of types had a differential effect on entropy. Reducing the number of types (and keeping the total number of tokens) of any system that carries information results in a decrease of entropy. As an example, the set of numbers {1, 2, 3, 4, 5, 6} has an entropy of 2.54 while {1, 2, 3, 4, 4, 4} has an entropy of 2.25. In our case, the reduction of entropy was not uniform, since the distance between *khutt* as v1 and v2 widened while that of *pid* shrunk. As we have taken the same maximum of entropy, which was calculated based on the highest amount of tokens and not types, we can compare the impact of having semantic groups for all v1/v2 groups. This is displayed in the following table:

Table 11: Effect of semantic groups in entropy			
Without groups	*pid*	*chokt*	*khutt*
V1	0.51	0.56	0.61
V2	0.71	0.64	0.59
With groups	*pid*	*chokt*	*khutt*
V1	0.47 (-0.04)	0.49 (-0.07)	0.48 (-0.13)
V2	0.59 (-0.12)	0.54 (-0.1)	0.56 (-0.03)

In Table 11., we observe the values for entropy with and without the semantic groups. In the group *with* semantic groups, the number between parenthesis shows how much the values changed from their initial number. The main message from this image is that the implementation of semantic groups has had a differential effect on the benefactives; while all groups underwent a reduction in their entropy, this was mostly important for *pid* v2 and *khutt* v1.

One of the shortcomings of entropy is that it does not tell us much on how each value is composed because it is but an average of values. Thus the entropy of the sequence {1, 1, 1, 1, 1} is the same as {5, 5, 5, 5, 5}, namely 1.5. Looking at the individual noun-verb/benefactive score or at the distribution of these could help asses whether the value for entropy is an addition of vague or polysemous datapoints. To go after this question, we turn to the next approach: pointwise mutual information.

3.5.2 Pointwise Mutual Information

If entropy was the measure of uncertainty of one random variable, pointwise mutual information (PMI) is the amount of information that one random variable contains about another random variable[15]. cf. the following formal definition:

$$pmi(x; y) = \log \frac{p(x, y)}{p(x)p(y)} \tag{2}$$

The equation above is straightforward: pointwise mutual information I(X; Y) is the joint distribution divided by the product of both individual distributions. This means that if there is little information shared between the variables, the resulting number will be small. Another way of looking at it is that mutual information I(X; Y) is the relative entropy (i.e. the distance between two distributions) between the joint and the product distribution of two variables (here x and y). Therefore, mutual information is the reduction in the uncertainty of one variable given the knowledge of the other variable. Under this

[15]I am very thankful to Prof. Samaržić from the University of Zurich for suggesting to me this approach.

logic, as an example, the mutual information between two fair dices is zero, for they are independent from each other. Dependent variables have higher mutual information (Learned-Miller, 2013; Desagulier, 2017: 205f.; Manning & Schütze, 1999: 66f. and cf. Cover & Thomas, 1991: 18f. for a complete description).

Perhaps the best-known use of mutual information is the analysis of co-occurrences of matching words in similar contexts. Certain words (random variables) seem to appear often together, even more often than each word independently. These words may come together in fixed expressions or for many other reason, such as their semantic composition. Collocations can be thus ranked based on their distribution as a combination compared to their distribution of its constituents (cf. Bouma, 2009: 31–36). Corpuses provide an ideal setting for such applications. Other uses of this measurement include semantic representations based on the statistical significance of co-occurrences (collocations), especially in the case of language learning and lexicography (cf. Bullinaria & Levy, 2007; Church & Hanks, 1990).

Mutual information has also been used for machine translation with a view to reduce wrong translations in ambiguous settings. The automatic translator may use mutual information to choose the right context and thus reduce the entropy (uncertainty) of the translation (cf. Brown et al., 1991).

It should be noted that it is standard practice to transform negative values obtained in PMI to zero, as these may arise from poor coverage of some variables. For this reason, PMI is also known as Positive Pointwise Mutual Information PPMI (cf. Bullinaria & Levy, 2007: 514; Levshina, 2015: 327).

As for the results of our cloze test, this tool can be used to find meaningful correlations for each benefactive based on the overall token/type distribution of nouns. Similarly to the extraction of collocations, PMI will evaluate the dataset and rate its values according to which combinations make some sense together. This information will give a score for each verb-benefactive/noun relation based on the distribution of each variable. The higher the result for each verb-benefactive/noun relation, the more shared information is available. More shared information means a stronger relation between these two points. Translating this to our point about abstractness (cf. Section 1.3.2), or rather polysemy and vagueness, a sum of high scores stands for polysemy, as that particular element can build many strong connections, i.e. collocations or compositions that make sense in that language. Conversely, a high number of low values stands for vagueness, for this element may appear with others, but without making these strong connections.

We can thus add all the scores of these verb-benefactive/noun relations according to each benefactive and produce four sets of scores. The following image presents the values for each group:

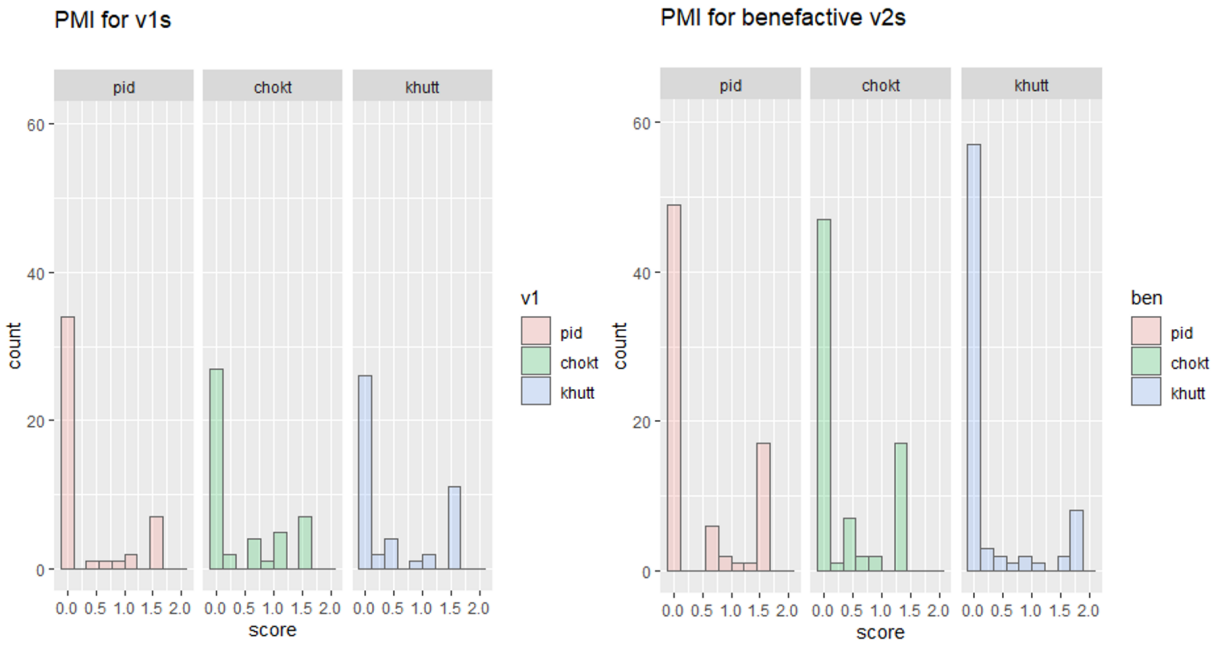

Figure 17. PMI

In Figure 17. we observe two plots: PMI scores for the group of independent v1s and for the groups of benefactive v2s. The numbers on the X-axis indicate the values in PMI while the Y-axis indicates the number of lexeme-lexeme combinations with that particular value.

The first impression that we obtain is that, in both groups, the amount of zero values is substantial. Furthermore, the distribution of zero values follows the order *pid*, *chokt*, *khutt* in the first plot and almost an inverse order in the second plot. Conversely, higher values in PMI (around 1.5) display a different distribution: while the increase in *pid* and *chokt* is similar, there is a slight decrease for *khutt*. In other words, while values around 1.5 and zero values increase in *pid* and *chokt*, *khutt* displays an X-shaped change, increasing its zero values and decreasing its higher values.

As it was mentioned earlier, the higher the score, the stronger the verb/benefactive-noun connection.

Influence of semantic groups

As it was done for entropy, we measured PMI using semantic groups. The following image displays the results:

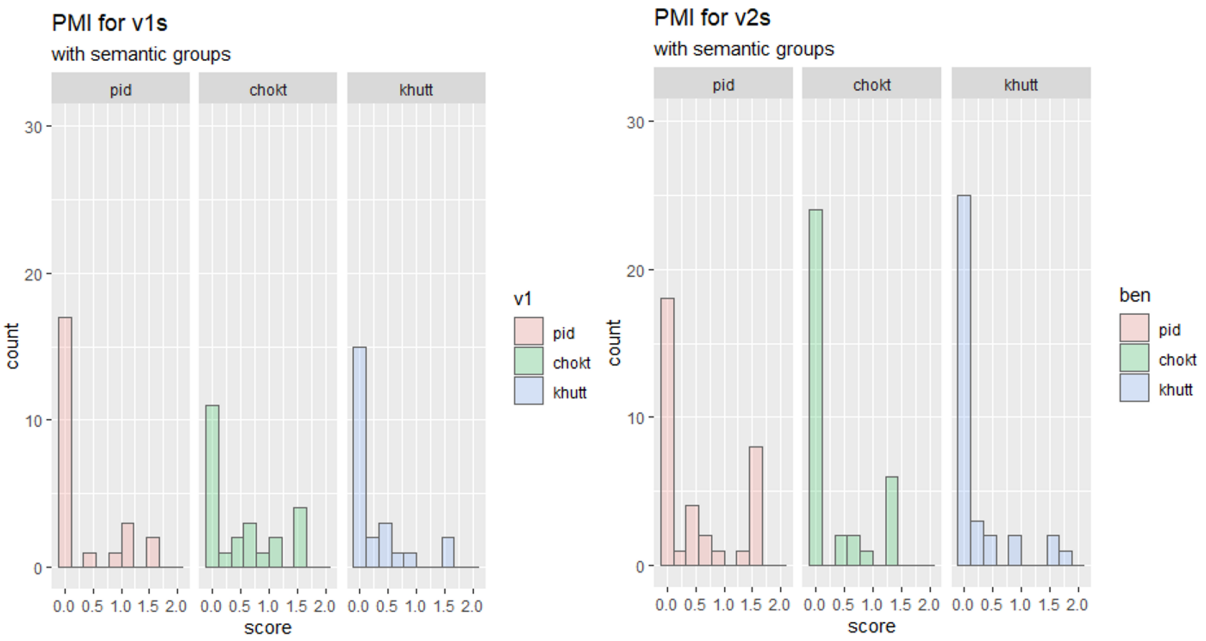

Figure 18. PMI (with semantic groups)

In Figure 18., we observe the results for PMI in each group with the semantic groups already introduced. Compared to Figure 17., without semantic groups, these numbers are smaller (the scale only goes up to 30). This is due to the smaller number of tokes, as many were collapsed for the semantic groups. Furthermore, the only significant increase in PMI values in going from v1 to v2 is experienced by *pid*. The increase in *khutt* and *chokt* is much less pronounced. As for zero values, these have remained like in the former image: an important increase in *khutt* and *chokt* going from v1 to v2. *Pid*, however, does not display an important increase in zero values.

Collostructional analysis

Another powerful tool for measuring the strength between a two groups of words is the analysis of collostructions, as it has been termed by Stefanowitsch & Gries (2003). This method studies constructions (in the parlance of Goldberg), i.e. any linguistic expression that is associated with a particular meaning or function and that is not due to derivation. This wide definition covers anything from single morphemes to larger phrases, like phrasal verbs or idiomatic expressions and also more abstract concepts such as tense or aspect and thus stands as a versatile tool to measure the relational strength of a wide array of elements.

This measurement takes the single and joint frequency of two words a) and b) (or any construction) as well as the frequency of word a) with any other word but word b) and the frequency of all other elements except for the two involved words a) and b). These values are tested with the Fisher exact test. Gries has also developed an R script (2007), which has been used for the present investigation. One of

the possibilities that this script allows, is to compare the distribution, and thus the strength involved, between two sets of words as long as some of them are shared between the two. This would have been most useful in our case, since we want to compare the distribution of verbs and benefactives. This measurement would have decided for each noun how strong was the relation to either the verb or the benefactive. However, due to the insufficient number attestations for most noun/verb-benefactive combinations, the results were not interpretable and this method was not further pursued.

Interplay between entropy and PMI

As it was mentioned in Section 1.3, the results from entropy and PMI are connected in at least two ways. The first one is by taking PMI as a control tool for values in entropy. Since entropy is an average of numbers, PMI can control for unusual values that come out as normal in entropy. In our case, the values of PMI seem similar in all cases, hence comparing entropy is meaningful. The second point is directly comparing PMI with entropy. This is best exemplified in the process of going from v1 to v2, i.e. the coevolution of these measurements. In the plots without semantic groups (Figure 14. and 17.), both high PMI values and entropy increase for *chokt* and *pid*, while these values decrease slightly for *khutt*. The augmentation of zero values in all cases is related to the higher number of noun types: Hapax nouns generates zeros for those verbs/benefactives that do not have that particular hapax noun attested. This, however, does not explain the differential growth of zero values, which will be discussed in the Interpretation. As for the plots with semantic groups (Figure 16. and 18.), we perceive in all cases a growth of both entropy and PMI, albeit to different degrees.

3.6 Interpretation

In this section, we will translate the results presented above into an idiom that can be related to the hypothesis of experiment 2 and also experiment 1. We will start by recapitulating and contextualizing the experiment.

Due to the results from experiment 1, we had a signal (the relation of *khutt* to its etymological origin) whose validity had to be corroborated and its meaning interpreted. From this starting point, the analysis of the results from experiment 2 have been an attempt to relate the same lexeme as a v1 to its v2 by help of empirical data.

In Figure 7. and 8. (first CAs), the spread distribution of benefactives and v1s can be taken as a sign that the obtained nouns can be mostly explained by to these factors (i.e. independent v1s and benefactive v2s) alone. This is central to our argument, since having other factors that interfere decisively in the

larger picture (the distribution of nouns) would make a comparison between two specific variables (v1s and benefactive v2s) difficult. The results from the CA plots suggest that the distribution of nouns can be mostly attributed to these two factors and not to participant and v1s (in the case of complex predicates). This finding, however, could not be proved statistically due to the nature of the data and the overwhelmingly high number of zero values (because of hapax values).

The objective of this investigation was to compare the performance of the *same* lexeme as a v1 and as a v2, taking the described ambivalence of v2s in Chintang as an antecedent of a pattern which might be different to the standard description and prediction of grammaticalization.

The results from Figure 14. (entropy without semantic groups) show *pid* augmenting its entropy from v1 to v2 while *khutt* describes an opposed picture, since entropy goes slightly down from v1 to v2. This can be takes as a signal that while *pid* undergoes the usual process described in the literature, *khutt* does not, for it does not become more abstract as a v2 (measured in relation to nouns). Let us go over this step by step: We want to find out how abstractness differs between the same lexeme as an independent verb and as a grammaticalized benefactive. We thus produced a cloze test to compare the responses elicited by each benefactive and its verbal counterpart (independent v1). We also learned that one way of understanding abstractness is in terms of inclusiveness, i.e. how broad a concept is. This is most easily observed when comparing two objects in terms of which is related or linked to more different contexts/settings/objects. Returning to our study, the hypothesis of Section 3.2 is that the higher the entropy (i.e. the degree of uncertainty), the more abstract a benefactive will be. For this reason, a more abstract benefactive should elicit a more balanced distribution of token/types and also a larger number of types, as it would be associated with a larger group of nouns with greater chances that *any type* may occur (high uncertainty). This can be compared to the measurement of language proficiency (as in Stoll et al., 2012), where a proficient speaker makes use of a wide array of expressions in a balanced way. Conversely, a more concrete benefactive should elicit a less balanced number of tokens with a high concentration on only a few types, i.e. a more predictable distribution.

While the initial distribution showed that *khutt*'s entropy (and abstractness) increases as a v2, the implementation of semantic groups presented a different picture: There is a rise of entropy from v1 to v2 in all cases for all lexemes, albeit to different degrees. This runs counter our initial observation that *khutt* is different from other benefactives in terms of abstraction. Furthermore, it also highlights the importance of semantic groups when analyzing a signal (in our case the entropy of *khutt*). The differential effect of semantic groups reveals that a combination of many similar types in the answers elicited by v1 and non-similar types for v2 normalizes *khutt* with respect to *pid* and *chokt*. A similar picture emerged when we excluded outlier participants from the test: *khutt* as v2 has a score which is approximately 9% higher than its v1. This would suggest that entropy is always augmented when a v1 is used as a v2,

even though the amplitude of the difference varies and *khutt* has a smaller difference, compared to the other lexemes.

As for the results of PMI, in all cases we observe a bimodal distribution with a high amount of zero values and a small amount of values around the score 1.5. The use of semantic groups shrunk the values obtained in PMI due to the smaller number of types (which were collapsed to create the groups, cf. Table 10.). Moreover, this also normalized the divergent values of *khutt* to a certain extent, even though there is no considerable growth in *khutt*'s higher values of PMI as a v2.

Entropy as a measurement for abstraction proved to be a useful tool in theoretical and practical terms. Used in a relative scale, it is also dynamic as it serves to compare different values and thus build the argument of the present investigation. Conversely, its simplicity is also its downside, since the score of entropy is but an average of several values. As such, it may hide unexpected distributions and other information relevant for the researcher. For this reason, we used PMI in order to check for diverging values in each group, as it displayed the strength of each verb/benefactive – noun combination. This was explained in terms of polysemy, the degree of binding force of each lexeme.

A further theoretical insight from the interplay of PMI and entropy is the coevolution of these values when compared v1 to v2: when entropy rises, PMI's higher values and zero values also rise. On the other hand, in the case of *khutt* without semantic groups and before the exclusion of outlier participants, the slight diminishing of entropy was seconded by a) an also slight reduction of PMI 1.5-values and b) a rise in PMI's zero values. The b) latter can be explained by a rise in types, which compared to *chokt* and *pid* is disproportionately low (cf. Table 3. and 4.), while the a) former suggests that *khutt* has less strong verb-noun connections. With the implementation of semantic groups, *khutt*'s small augmentation of entropy is accompanied by an equally meagre augmentation of higher PMI values.

The results, which suggest that abstraction (as measured by entropy) is positively correlated with polysemy (as measured by PMI), run counter other findings from both theoretical ground and empirical data. As an example, Hill et a. found out that abstract words have both more associated concepts than concrete words and that these associations are weaker. They used the USF data, were 6000 participants were asked to fill a cue word with the first (meaningful) word that came to their minds (2014: 163f.). Further research should be able to tell whether this discrepancy is due to methodological or theoretical differences (e.g. the meaning of *abstract words*).

Turning to our hypothesis, we expected to encounter a small degree of difference between v1s and v2s. This goes back to the ambiguous nature of v2s in Chintang that can either take the meaning of a v1 or a v2. In a similar way, we cold expect the distribution of nouns for complex verbs with a benefactive v2 to be similar to that of the one of v1s. However, in the case of *pid* and *chokt*, there seems to be an especially strong cleavage between the associations that trigger v1s and v2s, which would defy our ini-

tial assumption. As for *khutt*, we expected it to have clearly different results that would stick out from the other benefactives. This idea goes back to the results from the first experiment, where translation to Nepali showed some cases of a different construction for *khutt* which included the verb *lyaunu* 'to bring'. While we encountered this situation at first, the use of semantic groups and the exclusion of outlier participants made that initial discovery dubious.

How can we interpret these results? The numbers for entropy, after controlling possible biases, attest lower entropy for a lexeme when it appears as independent v1 and higher entropy when the same lexeme appears as a v2. How is this discovery (predicted in the literature) related to the versatile nature of v2s in Chintang and the especially strong connection of v2 *khutt* to its etymological origin (experiment nr. 1)?

In Chintang, as we have seen, verbs recruited as a v2 do not necessarily comply the expectations of standard grammaticalization theory; while some undergo phonological erosion and semantic bleaching, this does not apply for all v2s. In Table 12. we observe a small typology of complex predicates:

Table 12: Complex predicates in Chintang

	semantic bleaching	phonological erosion
-yakt (IMPERF)	✓	✓
-pid (BEN)	✓	×
-tat (V.AND.BRING)	×	×

In Table 12., we see three different kinds of v2s according to semantic bleaching and phonological reduction. While *-yakt* is a *well-behaved* v2, as it has undergone the two processes described in the literature on grammaticalization, *-tat* seems to be most true to its v1 origin. As for *pid*, it is neither an ambiguous v2 (i.e. semantic bleaching) nor does it display phonological reduction. Some approaches to language change might refer to these examples as varying degrees of grammaticalization within a continuum (cf. Lamiroy & De Mulder, 2011), with *-yakt* being most entrenched in grammar and *-tat* less so.

This table, however, would not be suited for *khutt*, since semantic bleaching has not yet taken place in spite of the grammaticalized meaning as benefaction (differently from *-tat*); both meanings coexist.

We can now return to the measurement of entropy. We know that the grammaticalized meaning is conceived as more abstract than the lexical one. We also know that *khutt* carries much of the semantics of v1 as a benefactive v2 compared to *pid*, where we do not observe such a behaviour. The question is, again, what is the excess of entropy in v2s made of, especially in the case of *khutt*? We are not surprised by the increase of entropy for *pid*, since *pid* as a v2 is not ambiguous; in *khutt*, however, the

increase of entropy, albeit less than *pid*, is due to the lexical meaning 'to bring' as much as it is to the grammaticalized sense of benefaction. In other words, the cloze test cannot tell these two senses apart, at least not in the way our questionnaire was constructed. The excess of entropy in *khutt* as a v2 is thus due to the accumulated effect of both its lexical and its grammaticalized meaning. This renders entropy less useful for ambiguous v2s, as the situation is different from how entropy was used e.g. by Stoll et al. (2012). There, the score obtained for entropy could be measured for each individual speaker. In our case, we cannot tell these two meanings apart and thus the degree by which the lexical and the grammatical meaning have an effect on the entropy of v2s cannot be asserted. This leaves our initial question partly not answered, for the discoveries of this research cannot directly address the ambiguous nature of complex predicates in Chintang.

In sum, the initial signal from experiment nr. 1 about *khutt* could not be confirmed in experiment nr. 2, for entropy (taken as a measurement for abstractness) augments in all three lexemes when they were used as a v2. A higher entropy is thus due to the grammaticalized meaning (benefactive) while the lexical meaning, which we would expect to elicit lower entropy, does not appear in the results or at least cannot be teased apart from the final result.

4 Discussion

In this section, we will go through the main obstacles that this investigation faced, as well as alternative explanation to the findings. Finally, we will mention the broader relevance of our findings.

4.1 Obstacles

Our investigation faced a number of obstacles that should be taken into account after acknowledging the results. Concerning the first experiment with translations from the corpus, as it was mentioned, there are many factors that could have affected the translation. As an example, the translations to Nepali, and this is just an impression, may have been an attempt to follow a word-for-word translation of the Chintang original, rather than rendering it in the most natural way. The most telling fact is the lack of case markers, which are common devices for rendering benefaction in Nepali. All sentences are perfectly grammatical in Nepali, but the absence of these markers indeed raises suspicions.
A future replication of this experiment could be made with free translations from Chintang to Nepali, at best orally. This would be an option to control for the quality of translations.
Concerning the second experiment of this investigation, we implemented a methodology (the cloze test) that has not only been rarely used, but it has also been only applied to well-described languages (e.g. English). We believe that improving this methodology, especially for different cultural settings and diverse languages, will be beneficial for linguistics. Similarly for entropy, this measurement has not being used to calculate abstractness in the available literature and, therefore, a stronger theoretical explanation for its use is required.
Concretely for the cloze test, the first obstacle was the number of tokens. Even though we had a significant number of participants (11), an important number of answers were not valid. This is solely due to the implementation made by the author of this investigation, since the setting of the tests, the questions or the instructions could have been improved.
Concerning the design of the questions, we could have used distracters to improve the quality of answers and hence dissipate doubts about the performance of the participants. Furthermore, the low number of tokens and the high amount of hapax values also made a strong statistical analysis doubtful. For this reason, we decided to refrain from using regression models. One possible way to reduce the large amount of zero values would be to limit possible answers to a given range, i.e. that the participant chooses the answers from a set of possibilities. The downside of such a method is that it would not be a test of free association anymore.
The use of semantic groups has a large potential for improvement, as words were grouped together without a strong theoretical criterion and their influence in entropy seemed important. This means

that more specific groups with fewer tokens each would have had a weaker effect on entropy. For this reason, the implementation of these groups to filter raw answers requires a stronger theoretical basis. The choice of participants could also be a potential issue, as all of them are men, six of them in their 20s, three of them in their 30s and two of them in their 40s. Older speakers of the language would have made a better choice, as they are usually less acquainted than younger speakers to Nepali and have had less exposure to it (due to schooling and new media of communication). In fact, the majority of the participants who were in their 20s left Chintang to seek better education or a difference source on income and returned only sporadically to the village. Furthermore, the engagement of female participants would have provided more a representative sample for the language.

In the case of the provided answers, the high number of repetitions of *saman*, mainly by three participants, has also blurred the results, as this word can readily be translated as 'goods', but also as 'thing', a rather ambiguous and all-encompassing term. This noun could, however, also be read as a sign of misunderstanding of the instructions of the test, as the participants may have felt compelled to say just anything, even if it does not make sense in order to appease the executer of the test. In any case, before applying this method in the future, this issue should be addressed.

4.2 Alternative explanations and speculations

Looking for a possible explanation of the distribution that the two measurements of entropy revealed, an important clue certainly is the inter-speaker variation found in acceptability of combinations of verb+benefactive. Even though all forms were tested by a (young) native speaker, especially older participants of the test rejected some of these combinations. There is, however, no exact number of forms deemed as ungrammatical by some speakers, as these were excluded from the valid answers. A replication of this study should test verb+benefactive combinations with older speakers as well before executing the test.

Language is a system where change is the norm. As for our benefactives, variation of acceptability could be an indication of change in progress. The decisive fact to determine this is whether there is significant variation in age that is not *u*-shaped, i.e. whether younger speakers would ever use these benefactives in the same fashion that older people would (assuming now that age is the variable of change and not something else) (cf. McMahon, 1994: 241f.; Burridge & Bergs, 2016: 18f.). If we take a brief look at the Chintang corpus, we find that the mean of age of people using these benefactives builds two groups: *pid* 25 years; *chokt* 26 years; *dhett* 28 years on one side and *khutt* 38 years on the other. The data collected for the corpus, at least the one having benefactives, was recorded between the years 2004 and 2012, i.e. between 15 and seven years prior to the present investigation. If there is no age grading involved but normal innovation in language, then we should expect to find e.g. *khutt*

nowadays mostly among speakers around 45 and 50. Our oldest participants were, however, 42. Further research, both with the corpus and doing descriptive work, will reveal whether the difference in age is a by-product of the general low attestation of *chokt*, *dhett* and *khutt* or a proof of one benefactive being outdated in modern Chintang. If this be confirmed, we would have good reasons to believe that the difference between certain benefactives is mainly a matter of register/style, as we know that there are no strong grammatical and semantic (here we include abstractness) differences. Furthermore, it would stand out as a case of counter-grammaticalization, as the grammaticalized form fades away while the independent verb prevails.

A further phenomenon that the cloze test revealed was the benefactive-specific nouns that appeared among many speakers. As an example, *ghasa* 'grass' appeared more than 15 times with *dhett* and *saman* 'goods' appeared more than 25 times with *chokt*. This distribution could be explained by a semantic-affinity of certain benefactives. This would explain e.g. why *saman* appeared 27 times with benefactive *chokt* and only 10 times with *pid* and eight times with both *dhett* and *khutt*. This could, however, be related to our point about abstractness, as we expect concrete benefactives to have a small set of contexts where they can appear. The observed circumscription of the number of contexts for the appearance of a concrete benefactive could be the same phenomenon as the possible semantic affinity between a benefactive and a noun or a group of nouns; it depends on the point of view. How exactly this semantic affinity would look like , however, requires further study.

Lastly, the existence of an equivalent to *chokt* in the Chintang dialect of Bantawa raises many questions. It is not clear whether Chintang borrowed this form and thus enriched its own benefactive system or whether they are simply cognates. Furthermore, the use of this benefactive in Bantawa itself is not clear and only more research could bring more conclusive an answer.

4.3 Broader relevance

Concerning v2s as an areas phenomenon, a review of the literature of complex predicates in South Asia reveals that the versatile nature of these constructions (with respect to their verbal origin), is seldom addressed. Especially in grammatical descriptions, the ambivalence of these forms is rarely ever mentioned. The question of how a lexeme changes conceptually once it is recruited as a v2 has only been approached from a theoretical point of view but seldom with empirical data. Furthermore, the prominence of ambiguous v2s in Chintang bespeaks against the common assumption in the literature that the coexistence of the lexical and the grammaticalzed v2 is but a transitory stage, one which will unequivocally end with the former disappearing for good and the latter undergoing some kind of phonological erosion. This reveals a large field of potential research in complex predicates and their relation to their

verbal origins, all framed in a context of areal influence.

4.4 Future work

In the reduced scope of the present investigation, several issues have come up that deserve further exploration. An account of potential betterment for the present research has been mentioned in Section 4.1. For this section, we will suggest some ideas that might continue and enhance the theoretical stance and the practical approach taken in this investigation.

First of all, the use of the Chintang corpus can be used to replicate the results of the cloze test in order to observe the distribution of nouns for v1s and v2s. As the corpus is largely made up of natural speech, it would also overcome the bias that an unnatural situation produces. Furthermore, there is no need to focus, as we did, only on benefactives, since there is a plethora of grammatical categories that are expressed by v2s in Chintang.

A further issue is the inclusion of more languages with complex predicates and perform similar tests in order to find out whether the results obtained for Chintang are tantamount to a cross-linguistic feature of language. Widening the scope and incorporating typologically diverse languages is bound to create a strong case, either for or against the effect of grammaticalization over lexemes, which in turn will help elucidate the relation between lexicon and grammar.

5 Conclusions - Methodological and theoretical contribution

The goal of this study was to conceptualize the difference between v1s and v2s in Chintang, given that many v2s in that language are known to be ambiguous in the sense that they might trigger either the initial, lexical meaning or the grammaticalized one. We narrowed our scope to benefactives, where we find four devices, three of which have a v1 counterpart (their etymological origin).

The first attempt was to use the Chintang corpus and observe how each benefactive was translated to Nepali in a randomly chosen sample. The results show that the translators of the corpus, all natives of Chintang also fluent in Nepali, used a v1+benefactive v2 *di-* (from *dinu* 'to give') construction as the standard for rendering all benefactives, with the exception of *khutt*, which was rendered approximately every third time with a particular construction: v1(non.finite) *lyaunu* 'bring'+benefactive v2. This was taken as a signal that *khutt* retains more of the semantics from its v1 counterpart than the other benefactives.

The second attempt was a cloze test performed with native speakers of Chintang. The aim was to test the difference between the same lexemes as a v1 and as a v2. Before we turn to the results, the one concept behind this test has to be explained: abstractness. One of the many ways to define abstractness is the possibility to be applied to different contexts without hindrance. In this line, a word like *mammal* would be more abstract than *dog*, as it refers to more different animals (i.e. contexts). The literature on grammaticalization anticipates an abstraction in the meaning when a lexeme is recruited as a grammatical device. In the case of complex predicates, an independent verb turns into a verbal suffix and acquires a grammatical function.

We applied this concept to both our benefactive v2s and their independent v1 counterparts in a cloze test of free association, for we hypothesized that a more abstract benefactive and/or verb would elicit more different types (unique noun regardless of its number of repetitions) in a listing exercise whereas a more concrete benefactive and/or verb would elicit a smaller number of types, as it is associated with a small number of specific contexts. We used entropy to compare raw values from different benefactives and verbs and thus measure the *transition* from v1 to v2.

The initial results showed the expected increase in entropy (and thus in abstractness) from v1 to v2 in *pid* and *chokt*, whereas the values were quite similar between the two (v1 and v2) for *khutt*. This was in accordance with the results from the first experiment, whence we expected to encounter little perceived difference between the lexical (v1) and the grammaticalized form (v2). However, further inspection of the data revealed that the deviant values of *khutt* were due to i) the effect of outlier participants and ii) for having considered semantically similar nouns as different types. The former was resolved excluding outlier participants and the latter by grouping similar words into semantic groups. In sum, entropy

seems always to increase from v1 to v2, although in varying degrees. This could only be assessed by controlling for outliers and by using semantic groups.
In terms of methodology, measuring abstractness with entropy and polysemy with PMI came out as promising tools for future studies, even though more research is needed to confirm the extent to which it can be used. Concerning the combination of these two measurements, our results show how higher entropy is correlated to an increase in higher values in PMI. This suggests a direct relation between abstraction and polysemy and inversely, concreteness and vagueness. This forces us to rethink our understanding of these concepts, because we assumed them to be completely independent. In this sense, only further research will tell whether this signal only emerged in our data or whether if carries a fundamental insight about these concepts.
As for the cloze test, it allowed us to render calculable concepts which remain *abstract*. Furthermore, we believe that improving the execution of the test, the number of participants and the questions will be decisive in the quality of this method in future research.

Returning to the ambiguous complex predicates that gave rise to this investigation, the final results seem to confirm the claim that grammaticalization, here in the form of a recruitment as a v2, conveys an abstraction of meaning, insofar as it can be measured in terms of entropy. This also applies for our three benefactives. However, this does not provide an explanation for the coexistence of two meanings, since the measured excess of abstraction may correlate with a grammaticalized, benefactive reading, but it does not say anything about the initial, lexical content of the recruited lexeme. Rather, it can be read as an indication that a better explanation for semantics must be sought after elsewhere, especially since our results suggest that the abstraction cleavage between v1 and v2 is not the same for each lexeme and, specifically, v2 *khutt* came out as i) being more connected to its etymon in translations to Nepali and ii) the margin of difference in entropy (and thus abstraction) was smaller than the other benefactives (albeit not substantively). In other words, using entropy as a measurement of abstraction opens up many possibilities, but its ultimate connection to the semantics of v2s needs further development.
A further discovery that emerged from the results was the positive correlation between entropy and PMI (i.e. abstraction and polysemy). This would imply for our experiment that an inclusive benefactive, with less restrictions to appear in different contexts, might also build stronger connections to these contexts/nouns. Further research will prove the extend to which this holds true and the theoretical implications for our conception of entropy and polysemy.

The conclude, we hope for this investigation to be a contribution to the understanding of complex predicates in general and of Chintang in general. We used an uncommon methodology to explore a grammatical phenomenon using theoretical approaches which are not commonplace in descriptive linguistics: abstractness, entropy and their interplay. The methodological point would be thus that stan-

dard linguistics can benefit from different approaches, especially since larger sets of data may entail patterns which are hidden to the naked eye. Different methods can cover and complement each other. Needless to say that there is potential for improving these methods.

As for the language, our latest report is that Chintang is no longer passed on to the younger generation, a process that started some ten years ago. This is an unequivocal sign of language death in the near future. As several aspects of the grammar remain unclear, we consider it to be a priority for further documentation in this and other languages of the area to be undertaken before both loss of native speakers and language attrition make their inexorable way up to the Kiranti Hills.

References

Aikhenvald, A. Y. (2006). Serial verb constructions in typological perspective. In *Serial verb constructions: A cross-linguistic typology*, pages 1–68. Oxford University Press.

Aikhenvald, A. Y. (2012). 'we can't say it with one word': multiverb constructions. In *The Languages of the Amazon*, pages 304–325. Oxford University Press.

Anderson, G. D. (2006). *Auxiliary verb constructions*. Oxford University Press.

April, M. (1994). *Understanding language change*. Cambridge University Press.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1435):1177–1187.

Barsalou, L. W. and Wiemer-Hastings, K. (2005). Situating abstract concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, pages 129–163.

Bentz, C. (2018). Adaptive languages: An information-theoretic account of linguistic diversity. *Berlin/Boston: Walter de Gruyter*.

Bickel, B. (1996). *Aspect, mood, and time in Belhare: studies in the semantics-pragmatics interface of a Himalayan language*. ASAS, Zürich.

Bickel, B. (1999). Cultural formalism and spatial language in belhara. In Bickel, B. and Gaenszle, M., editors, *Himalayan space: cultural horizons and practices*, pages 75–104. Museum of Ethnography, Zürich.

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

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N. P., Rai, I. P., Rai, M., Rai, N. K., and Stoll, S. (2007). Free prefix ordering in Chintang. *Language*, 83(1):43–73.

Bickel, B. and Gaenszle, M. (2015). First person objects, antipassives, and the political history of the southern kirant. *Journal of South Asian Languages and Linguistics*, 2(1):63–86.

Bickel, B., Rai, M., Paudyal, N., Banjade, G., Bhatta, T. N., Gaenszle, M., Lieven, E., Rai, I. P., Rai, N. K., and Stoll, S. (2010). The syntax of three-argument verbs in Chintang and belhare (southeastern Kiranti). *Studies in ditransitive constructions: a comparative handbook*, pages 382–408.

Bickel, B., Stoll, S., Gaenszle, M., Rai, N. K., Lieven, E., Banjade, G., Bhasa, T. N., Paudyal, N., Pettigrew, J., Rai, I. P., and Rai., M. (2004–2013). Audiovisual corpus of the Chintang language, including a longitudinal corpus of language acquisition by six children.

Bickel, B. and Zúñiga, F. (2017). The'word'in polysynthetic languages: phonological and syntactic challenges. In *The Oxford Handbook of Polysynthesis*, pages 158–185. Oxford University Press.

Bisang, W. (1995). Verb serialization and converbs-differences and similarities. pages 137–88. Berlin and New York: Mouton de Gruyter.

Bisang, W. (2004). Grammaticalization without coevolution of form and meaning: The case of tense-aspect-modality in east and mainland southeast asia. In *What makes grammaticalization? A look from its fringes and its components*, pages 109–138. Berlin and New York: Walter de Gruyter.

Bisang, W. (2008). Grammaticalization and the areal factor. In *Rethinking Grammaticalization: New perspectives*, pages 15–35. Amsterdam/Philadelphia: John Benjamins Publishing.

Boas, H. C. (2010). The syntax–lexicon continuum in construction grammar: A case study of English communication verbs. *Belgian Journal of Linguistics*, 24(1):54–82.

Booij, G. (2012). Construction morphology and the interaction of syntax and word formation. *Los límites de la morfología. Estudios ofrecidos a Soledad Varela Ortega*, pages 105–113.

Borchers, D. (2007). Reasons for language shift: Theories, myths and counterevidence. *Trends in Linguistics - Linguistics of the Himalayas and beyond*, 196:1–22.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, pages 31–40.

Bourdieu, P. (1979). *La distinction. Critique sociale du jugement.* Paris, éd. de Minuit.

Bourdin, P. (2008). On the grammaticalization of 'come' and 'go' into markers of textual connectivity. *Rethinking grammaticalization: New perspectives*, pages 37–59.

Brinton, L. J. (2008). "where grammar and lexis meet": Composite predicates in English. In *Theoretical and Empirical Issues in Grammaticalization*, pages 33–54. Amsterdam and Philadelphia: John Benjamins Publishing.

Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., and Mercer, R. L. (1991). Word-sense disambiguation using statistical methods. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, pages 264–270. Association for Computational Linguistics.

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46(3):904–911.

Bullinaria, J. A. and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods*, 39(3):510–526.

Burgoon, E. M., Henderson, M. D., and Markman, A. B. (2013). There are many ways to see the forest for the trees: A tour guide for abstraction. *Perspectives on Psychological Science*, 8(5):501–520.

Burridge, K. and Bergs, A. (2016). *Understanding language change*. New York: Routledge.

Campbell, L. (1998). *Historical linguistics*. Edinburgh University Press.

Campbell, L. (2000). What's wrong with grammaticalization? *Language sciences*, 23(2-3):113–161.

Campbell, L. and Janda, R. (2000). Introduction: conceptions of grammaticalization and their problems. *Language sciences*, 23(2-3):93–112.

Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.

Colla, D., Mensa, E., Porporato, A., and Radicioni, D. P. (2018). Conceptual abstractness: from nouns to verbs. In *5th Italian Conference on Computational Linguistics, CLiC-it 2018*, pages 70–75. CEUR Workshop Proceedings.

Coupe, A. R. (2008). *A grammar of Mongsen Ao*. Berlin and New York: Walter de Gruyter.

Coupe, A. R. (2018). Grammaticalization processes in the languages of south asia. *Grammaticalization from a typological perspective*, pages 189–218.

Coussé, E., Andersson, P., and Olofsson, J. (2018). Introduction. In *Grammaticalization meets construction grammar*, pages 3–20. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press on Demand.

Croft, W. (2012). *Verbs: Aspect and causal structure*. Oxford University Press.

David, A. B. (2015). *Descriptive grammar of Bangla*. Berlin and New York: Walter de Gruyter.

De Deyne, S. and Storms, G. (2015). Word associations. In *The Oxford handbook of the word*, pages 465–480.

Deane, P. D. (1988). Polysemy and cognition. *Lingua*, 75(4):325–361.

Dehé, N. (2015). Particle verbs in Germanic. In *Word-Formation: an International Handbook of the Languages of Europe*, pages 611–626. Berlin and Boston: Walter de Gruyter.

del Prado Martın, F. M., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94(1):1–18.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, 8(8):1117–1121.

Desagulier, G. (2017). *Corpus Linguistics and Statistics with R*. Heidelberg: Springer.

Diessel, H. (2015). Usage-based construction grammar. *Handbook of cognitive linguistics*, pages 295–321.

Doornenbal, M. (2009). *A grammar of Bantawa: Grammar, paradigm tables, glossary and texts of a Rai language of Eastern Nepal*. PhD thesis, Rijksuniversiteit te Leiden.

Drude, S. (2011). 'derivational verbs' and other multi-verb constructions in aweti and tupi-guarani. In *Multi-verb Constructions*, pages 213–254. Leiden: Brill.

Ebert, K. H. (1994). *The structure of Kiranti languages: comparative grammar and texts*. Seminar für Allgemeine Sprachwissenschaft, Universität Zürich.

Ebert, K. H. (1997a). *Camling/Chamling*. Munich: LINCOM Europe.

Ebert, K. H. (1997b). *A grammar of Athpare*. Munich: LINCOM Europe.

Ebert, K. H. (2008). Converbs in kiranti languages. *From Siberia to Ethiopia—Converbs in a Cross-Linguistic Perspective*, pages 63–89.

Emonds, J. E. (2011). *Lexicon and grammar: The English syntacticon*, volume 50. Berlin and New York: Walter de Gruyter.

Evans, V. (2005). The meaning of time: polysemy, the lexicon and conceptual structure. *Journal of linguistics*, 41(1):33–75.

Frajzyngier, Z. (2008). Grammaticalization, typology and semantics. In *Rethinking Grammaticalization: New perspectives*, pages 61–102. Amsterdam and Philadelphia: John Benjamins Publishing.

Gaenszle, M. (2016). Redefining kiranti religion in contemporary nepal. *Religion, Secularism, and Ethnicity in Contemporary Nepal*, pages 150–191.

Galbraith, R. C. and Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition*, 1(1):56–60.

Garnier, M. and Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, 59:29–44.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. pages 301–334. Cambridge University Press.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., and Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40):10785–10790.

Glynn, D. (2014). Correspondence analysis. *Exploring data and identifying patterns. Teoksessa Glynn, Dylan & Robinson, Justyna (toim.) Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*, pages 443–486.

Gries, S. T. (2007). Coll.analysis. http://www.stgries.info/teaching/groningen/index.html/. R script version 3.2a.

Gries, S. T. (2015). Polysemy. In *Handbook of Cognitive Linguistics*, pages 471–489. Berlin and New York: Walter de Gruyter.

Harder, P. and Boye, K. (2011). Grammaticalization and functional linguistics. In *The Oxford handbook of grammaticalization*, pages 56–69.

Haspelmath, M. (1995). The converb as a cross-linguistically valid category. *Converbs in cross-linguistic perspective: structural and semantic aspects of adverbial verb forms. Berlin: Mouton de Gruyter*, pages 1–55.

Heine, B., Claudi, U., and Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. University of Chicago Press.

Hill, F., Korhonen, A., and Bentz, C. (2014). A quantitative empirical analysis of the abstract/concrete distinction. *Cognitive science*, 38(1):162–177.

Hitchcock, J. T. (1967). A nepalese shamanism and the classic inner asian tradition. *History of Religions*, 7(2):149–158.

Hock, H. H. and Bashir, E., editors (2016). *The Languages and Linguistics of South Asia*. Berlin and New York: Walter de Gruyter.

Hoenigswald, H. M. (1963). Are there universals of linguistic change? *Universals of language*, pages 23–41.

Hook, P. E. (1991). The compound verb in Munda: An areal and typological overview. *Language Sciences*, 13(2):181–195.

Hopper, P. (1991). On some principles of grammaticization. *Approaches to grammaticalization*, 1:17–35.

Hopper, P. J. and Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.

Huettig, F. and McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57(4):460–482.

Iliev, R. and Axelrod, R. (2017). The paradox of abstraction: Precision versus concreteness. *Journal of psycholinguistic research*, 46(3):715–729.

Jacques, G. (2018). Bipartite verbs in Japhug and other Trans-Himalayan languages. *Linguistics of the Tibeto-Burman Area*, 41(2):175–191.

Jacques, G., Lahaussois, A., and Shuya, Z. (in press). Associated motion in Sino-Tibetan/Trans-Himalayan. In *Associated motion*.

King, J. T. (2009). *A Grammar of Dhimal*. Leiden: Brill.

Kinn, T. (2018). Pseudocoordination in Norwegian. *Grammaticalization meets Construction Grammar*, 21:75–106.

Kiparsky, P. (2012). Grammaticalization as optimization. *Grammatical change: Origins, nature, outcomes*, pages 15–51.

Kittilä, S. and Zúñiga, F. (2010). Introduction: Benefaction and malefaction from a cross-linguistic perspective. *Benefactives and malefactives: typological perspectives and case studies*, pages 1–28.

Krawczak, K. and Glynn, D. (2012). *Context and Cognition. A corpus-driven approach to parenthetical uses of mental predicates*, pages 86–99.

Kuteva, T., Heine, B., Hong, B., Long, H., Rhee, S., and Narrog, H. (2019). *World lexicon of grammaticalization*. Cambridge University Press.

Lahaussois, A. (2009). Koyi Rai: An initial grammatical sketch. *Himalayan linguistics*, 4:1–33.

Lamiroy, B. and De Mulder, W. (2011). Degrees of grammaticalization across languages. In *The Oxford Handbook of Grammaticalization*, pages 302–318. Oxford University Press.

Learned-Miller, E. G. (2013). Entropy and mutual information. *Technical report, vol. 4*, pages 1–4.

Lehmann, C. (1995). *Thoughts on grammaticalization*. Munich: LINCOM Europe.

Lehmann, C. (2004). Theory and method in grammaticalization [theorie und methode in der grammatikalisierung]. *Zeitschrift für germanistische Linguistik*, 32(2):152–187.

Lehmann, C., Shin, Y.-M., and Verhoeven, E. (2000). *Direkte und indirekte Partizipation: zur Typologie der sprachlichen Repräsentation konzeptueller Relationen.* Munich: LINCOM Europe.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam and Philadelphia: John Benjamins Publishing Company.

Liljegren, H. (2016). *A grammar of Palula.* Berlin: Language Science Press.

López-Couso, M. J. and Seoane, E. (2008a). Introduction: Further reflections on grammaticalization. In *Theoretical and Empirical Issues in Grammaticalization*, pages 1–14. Amsterdam and Philadelphia: John Benjamins Publishing.

López-Couso, M. J. and Seoane, E. (2008b). Introduction: New perspectives on grammaticalization. In *Rethinking Grammaticalization: New perspectives*, pages 1–14. Amsterdam and Philadelphia: John Benjamins Publishing.

Manning, C. D., Manning, C. D., and Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Masica, C. P. (1993). *The Indo-Aryan languages.* Cambridge University Press.

Meillet, A. (1912 [1921]). L'évolution des formes grammaticales. *Linguistique historique et linguistique générale*, pages 130–148.

Michailovsky, B. (2017). Kiranti languages. In *The Sino-Tibetan Languages*, pages 646–679. New York: Routledge.

Narrog, H. and Heine, B. (2011). Introduction. In *The Oxford Handbook of Grammaticalization*, pages 1–18. Oxford University Press.

Naumann, D., Frassinelli, D., and im Walde, S. S. (2018). Quantitative semantic variation in the contexts of concrete and abstract words. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 76–85.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407.

Opgenort, J. R. (2004). *A grammar of Wambule: grammar, lexicon, texts, and cultural survey of a Kiranti tribe of eastern Nepal.* Leiden: Brill.

Opgenort, J. R. (2005). *A Grammar of Jero: With a Historical Comparative Study of the Kiranti Languages.* Leiden: Brill.

Plevoets, K., Speelman, D., and Geeraerts, D. (2008). The distribution of t/v pronouns in Netherlandic and Belgian Dutch. *Pragmatics and beyond new series*, 178:181–209.

Pokharel, M. P. (1991). Compound verbs in Nepali. *Contributions to Nepalese studies*, 18(2):149–173.

Rai, I., Bickel, B., Lieven, E., Banjade, G., Gaenszle, M., Rai, M., Paudyal, N. P., Stoll, S., and Bhatta, T. N. (2009). Mundum: a case study of Chintang ritual language. *Locality, History, Memory: The Making of the Citizen in South Asia*, pages 20–33.

Rai, M. N. (2017). *A Grammar of Dumi*. Munich: LINCOM GmbH.

Rai, N. K., Rai, M., Paudyal, N. P., Schikowski, R., Bickel, B., Stoll, S., Gaenszle, M., Pettigrew, J., Dirksmeyer, T., Banjade, G., Rai, I. P., Bhatta, T. N., Sauppe, S., Rai, R. M., Rai, J. K., Rai, L. K., Rai, D. B., Rai, G., Rai, D., Rai, A., Rai, C. K., Rai, S. M., and Rai, R. K. (2011). *Chintaṅ śabdakoś tatha vyakaran. [Chintang dictionary and grammar]*. Kathmandu: Chintang Language Research Program.

Rai, V. P. S. (2013). *A Grammar of Chamling*. Universität Bern.

Rousseeuw, P. J. and Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):73–79.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Sakel, J. and Everett, D. L. (2012). *Linguistic fieldwork: A student guide*. Cambridge University Press.

Sapir, E. (2004). *Language: An introduction to the study of speech*. North Chelmsford: Courier Corporation.

Schackow, D. (2015). *A grammar of Yakkha*. Berlin: Language Science Press.

Schackow, D. (2019). Grammatical relations in Yakkha (Kiranti). In *Typological Studies in Language*, pages 469–510. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Schackow, D. and Peterson, J. (2011). Complex predicates in Nepali. http://www.dianaschackow.de/dateien/2011-01-06_ComplPredNep_handout.pdf/.

Schikowski, R. (2013). Nepali verbal morphology. Unpublished.

Schikowski, R. (2018). A grammar of Chintang. Unpublished.

Schikowski, R., Bickel, B., and Paudyal, N. P. (2015). Flexible valency in Chintang. In *Valency Classes in the World's Languages*, pages 669–708. Berlin: Mouton De Gruyter.

Shibatani, M. (2009). On the form of complex predicates: Toward demystifying serial verbs. *Form and Function in Language Research: Papers in Honour of Christian Lehmann*, pages 255–282.

Shibatani, M. and Pardeshi, P. (2002). The causative continuum. *Typological studies in language*, 48:85–126.

Steever, S. B. (1988). *The serial verb formation in the Dravidian languages*. Dehli: Motilal Banarsidass.

Stefanowitsch, A. and Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International journal of corpus linguistics*, 8(2):209–243.

Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., Gaenszle, M., Pettigrew, J., Rai, I. P., and Rai, M. (2012). Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language*, 39(2):284–321.

Stubbs, J. B. and Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *Modern Language Journal 58*, pages 239–241.

Tomasello, M. (2009). *Constructing a language*. Harvard University Press.

Traugott, E. C. (2011). Grammaticalization and mechanisms of change. In *The Oxford Handbook of Grammaticalization*, pages 19–30. Oxford University Press.

Tuggy, D. (2006). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, 4(3):167–184.

Tumbahang, G. B. (2017). *A Grammar of Chhatthare Limbu*. Munich: LINCOM GmbH.

Turin, M. (2007). *Linguistic diversity and the preservation of endangered languages: A case study from Nepal*. International Centre for Integrated Mountain Development (ICIMOD).

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Valenzuela, P. M. (1997). Basic verb types and argument structures in shipibo-conibo. MA thesis.

van Driem, G. (1987). *A Grammar of Limbu*. Berlin and New York: Mouton De Gruyter.

van Driem, G. (1993). *A Grammar of Dumi*. Berlin and New York: Mouton De Gruyter.

van Driem, G. (2002). *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region*. Leiden: Brill.

van Driem, G. (2014). Trans-Himalayan. *Trans-Himalayan Linguistics*, 266:11–40.

van Hout, R. and Vermeer, A. (2007). Comparing measures of lexical richness. pages 93–115. Cambridge University Press.

van Rooij, R. (2011). Vagueness and linguistics. In *Vagueness: A guide*, pages 123–170. Dodrecht: Springer.

Van Valin, R. D. and LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function.* Cambridge University Press.

Viti, C. (2015). On degrammaticalization: controversial points and possible explanations. *Folia Linguistica*, 49(2):381–419.

Wiemer, B. and Bisang, W. (2004). What makes grammaticalization? an appraisal of its components and its fringes. In *What makes grammaticalization? A look from its fringes and its components*, pages 3–20. Berlin and New York: Walter de Gruyter.

Wiemer-Hastings, K., Krug, J., and Xu, X. (2001). Imagery, context availabilty, contextual constraint and abstractness. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 23, pages 1106–1111.

Williams, G. and Millon, C. (2011). Palmer, firth and internet: Drawing together collocational threads. *Corpus Linguistics*, pages 1–34.

Winter, W. and Rai, N. K. (2013). *A Bantawa dictionary.* Berlin and New York: Walter de Gruyter.

Zimmermann, E. (2012). Affix copying in Kiranti. In *Proceedings of ConSOLE XIX*, pages 343–367.

Zimmermann, T. E. and Sternefeld, W. (2013). *Introduction to semantics: An essential guide to the composition of meaning.* Berlin and Boston: Walter de Gruyter.

Zúñiga, F. (1998). Nomina sunt odiosa a critique of the converb as a cross-linguistically valid category. http://zuniga.unibe.ch/down/Zuniga-converbs-1998.pdf/.

A Appendix

In this section, we present the raw data that was used for experiment 2. The following tables have to be read in the following way: the left-most column displays the verb forms that were read aloud to the participants, along with a translation to English. The rest of the columns are the answers provided by the participants, which are numbered ranging from 1 to 11. Blank spaces stand for lack of answers or for answers that were excluded. The order of the verbs is the order in which these were read to the participants.

	1	2	3	4	5
kɨptɨmpidum "we(i) could cut for you"	choyop	tang	saman	tang	
koptuchokta "pick it up for him!"	mobile	sontolong	gilen	saman	
ongsudhettuce "he could peel it of for them"	sontolong	sontolong			chedar
tamsukhuttukum "we(i) pour it for them"	cuwa	cuwa	cuwa		cuwa
maipiceke "he/they give it to us (dual-i)"	sontolong	cuwa	camaca		
choktaŋkhaŋ "you(s) keep passing it to me"	chocolate	sontolong	saman	phekwa	saman
ukhutte "they bring it to him"	phekwa	sontolong	saman	saman	saman
phendubida "you(s) take it off for him"	petti	chocolate	wanam		jacket
osuchokta "you(s) throw it to him"	lungtak	sontolong	saman	saman	saman
khoptudhett "he puts it for him"					saman
khuraŋkhuttahã "you(s) carry it for me"	camaca	sontolong	bhari	saman	bhari
pidukumcɨm "we give it to them"	campol	sontolong	saman	saman	saman
acho'ŋa'a "you(s) pass it to me"			saman	saman	kappara
khuttuce "he brings it to them"	tomato	sontolong	saman	phekwa	saman
akhedabidaŋsehẽ "you(s) have bought it for me"	syau	sontolong	biscuit	upar	phekwa
nasottachokte "he/they moved it for you(s)"	kila		saman	saman	saman
haŋsuchoktukumcɨm "you(ns) send it for them"	mobile	chocolate	saman	phekwa	jhola
naktudhettuce "he asked it for them"		lungtak		saman	saman
napino "he/they give it to you"	phaksa	kok	phekwa	saman	saman
amaichokte "you passed it to us(e)"	phone	kok	saman	saman	saman
akhuttehẽ "you(s) bought it to me"	saman	tei	saman	jhola	saman
tisubidoŋsokha "he has put it to him"	phekwa	aba	saman	phekwa	saman

	1	2	3	4	5
konduchokta "you search it for him!"	chamal	yakkheng	saman	phekwa	
liktudhettakte "he was taking it in for him"		lungtak	cuwa		
loĩsuchoktaŋse "he has taken it out for him"		camace	cuwa	sambok	
khoktubida "he cuts it for him"		sa	sa	sa	sa
akhurachoktaŋsehẽ "you(s) have carried it for me"		bhari	bhari	bhari	bhari
thuktudhette "he cooks it for him"		kok		yakkheng	chula
krukhutta "he rolls it for him"		khi	pira		gari
pidehẽ "I gave (it)"		kok	biscuit	phekwa	phekwa
achoktehẽ "you(s) passed it to me"		cuwa	pira	arkha	saman
ukhuttehẽ "he brought it to me"		phekwa	pira	bier	saman
maipide "he/they gave it to us (p-i)"		kok	pira	chocolate	chamal
amachokte "you passed it to us (p-e)"		yakkheng	pira	samen	saman
nakhuttiki "he/they bring it to you(p)"	sa	makkai	pira	phekwa	biscuit
thubidoko "he drinks it for him"			cuwa	cuwa	cia
thanduchokta "bring it for him!"			cuwa		jhola
khomsakhuttahã "put it on for me!"		football	ghasa		saman
piduhẽ "I give it to them"	phekwa	cuwa	ghasa	saman	
choktuce "he passed it to him"	khana	sambok	cuwa	saman	biscuit
khuttukuŋ "I bring it to him"	jhunar	lungtak		saman	jhola
omabina'a "I throw it for you(s)"		lungtak		shisha	jhola
khoktakhuttoŋse "he has chopped it for him"		sa	sa		
ukhurudhette "they carried it for him"		jhola	ma'mi		doku
thuktukhutta "cook for him!"		kok	kok kok	kok	
apiŋa'a "you(s) give it to me"		phekwa	kok	kok	kok
nachokte "he/they passed it to you(s)"		cuwa	cuwa	kok	yakkheng
ukhuttaŋcihẽ "they(d) brought it to me"		halo	cuwa	jacket	saman
chuptubiduce "he pressed it for them"	dhoka	makkai	achita	gilen	yakkheng
labuŋchoktuŋsuhẽ "I have caught it for him"	mobile	wacile	wa	wasa	bhale
keŋsadhettanuma "hang it for him (you-ns)!"		mala	mala		ghasa
naktukhutta "ask it for him"		sontolong	ghasa	phekwa	biscuit
napide "he/they gave it to you"		kok	phekwa	saman	kok
achoktumhe "you(p) passed it to them"		sontolong	cuwa	saman	kok

	1	2	3	4	5
khuttumhe "we(p-i) brought it to him"		jhola	cuwa	saman	arkha
akɨrabidaŋsehẽ "you(s) have made it roll for him"		sontolong	sing	gari	
lettuchoktukuŋ "I planted it for him"		bakhra		gari	phak
uhektudhette "they cut it for him"		namdhang	sa	ghasa	ghasa
kɨptukhutta "cut it for him!"		coffee	ghasa	kappara	sing
apidehẽ "you gave it to me"		phekwa	phekwa	phekwa	sing
achoktace "you(d) passed it to him"		kok	cuwa		weiwei
maikhuttace "he/they brought it to us (p-i)"		aba	kok	phekwa	kok
atamabidaŋsehẽ "you(s) have poured it for me"		samkhuwa	cuwa		cuwa
phendudhettuce "he took it off for them"		bhari	wanam	ghasa	sing
naosakhutte "he/they threw it for us (p-i)"		lungtak			lungtak
akhedachoktaŋsehẽ "you(s) have bought it for me"		kopi	saman	jhola	biscuit
napiceke "he/they give it to you(d)"		kok	saman	saman	lekar
choknehẽ "I passed it to you(s)"		kursi	saman	jhola	biscuit
khuttoko "he brought it to him"		yakkheng	saman	mobile	sa

	6	7	8	9
kɨptɨmpidum "we(i) could cut for him"			alu	phekwa
koptuchokta "pick it up for him!"	saman	saman	sontolong	phekwa
oŋsudhettuce "He could peel it of for them"				
tamsukhuttukum "we(i) pour it for them"		cuwa		cuwa
maipiceke "he/they give it to us (dual-i)"	cuwa	saman	sa	phekwa
choktaŋkhaŋ "you(s) keep passing it to me"	kok	saman	arkha	saman
ukhutte "they bring it to him"	phekwa	saman	saman	saman
phendubida "you(s) take it off for him"	wanam	saman	jutta	wanam
osuchokta "you(s) throw it to him"			jacket	saman
khoptudhett "he puts it for him"	phaksa			
khuraŋkhuttahã "you(s) carry it for me"	bhari	bhari	saman	sontolong
pidukumcɨm "we give it to them"	saman	phekwa	cuwa	saman
acho’ŋa’a "you(s) pass it to me"		saman	cuwa	
khuttuce "he brings it to them"		sontolong	saman	samance
akhedabidaŋsehẽ "you(s) have bought it for me"		chocolate	saman	teĩ
nasottachokte "he/they moved it for you(s)"	sontolong	biscuit	jag	kok

	6	7	8	9
haŋsuchoktukumcɨm "you(ns) send it for them"	phekwa	saman	saman	saman
naktudhettuce "he asked it for them"		phekwa		
napino "he/they give it to you"		phekwa	saman	sontolong
amaichokte "you passed it to us(e)"	saman	saman	sontolong	saman
akhuttehẽ "you(s) bought it to me"	saman	chocolate	parewa	saman
tisubidoŋsokha "he has put it to him"	saman	phekwa	saman	phekwa
konduchokta "you search it for him!"	saman	ma'mi	phekwa	saman
lɨktudhettakte "he was taking it in for him"		saman		saman
loĩsuchoktaŋse "he has taken it out for him"	phekwa	phekwa	phekwa	saman
khoktubida "he cuts it for him"		saman	sa	phekwa
akhurachoktaŋsehẽ "you(s) have carried it for me"		bhari	saman	saman
thuktudhette "he cooks it for him"	tarkari		sa	gula
kɨrukhutta "he rolls it for him"	cuwa		singran	saman
pidehẽ "I gave it"	mura	saman	saman	phekwa
achoktehẽ "you(s) passed it to me"	mobile	saman	saman	saman
ukhuttehẽ "he brought it to me"		sontolong	sontolong	saman
maipide "he/they gave it to us (p-i)"		phekwa	parewa	saman
amachokte "you passed it to us (p-e)"		phekwa	saman	saman
nakhuttiki "he/they bring it to you(p)"		biscuit	saman	saman
thubidoko "he drinks it for him"		cuwa	arkha	arkha
thanduchokta "bring it for him!"	saman	saman	aba	saman
khomsakhuttahã "put it on for me!"	sing		fruit	sontolong
piduhẽ "I give it to them"	sing	ciso	syau	sontolong
choktuce "he passed it to him"	saman	kok	saman	samance
khuttukuŋ "I bring it to him"	sing		sontolong	saman
omabina'a "I throw it for you(s)"		saman		saman
khoktakhuttoŋse "he has chpped it for him"			sa	sing
ukhurudhette "they carried it for him"	bhari	bhari	saman	saman
thuktukhutta "cook for him!"		yakkheng	yakkheng	kok
apiŋa'a "you(s) give it to me"		kok	sa	phekwa
nachokte "he/they passed it to you(s)"		cuwa	saman	saman

	6	7	8	9
ukhuttaŋcihẽ "they(d) brought it to me"		saman	saman	sontolong
chuptubiduce "he pressed it for them"		gagri	dhoka	bhuja
labuŋchoktuŋsuhẽ "I have caught it for him"	phagcilek	wasa	saman	saman
keŋsadhettanuma "hang it for him (you-ns)"	ghasa	ghasa	ghasa	ghasa
naktukhutta "ask it for him"	dabi	saman	saman	ghasa
napide "he/they gave it to you"	phekwa	saman	sa	phekwa
achoktumhe "you(p) passed it to them"	phekwa	saman		cia
khuttumhe "we(p-i) brought it to him"	saman	sontolong	saman	
akɨrabidaŋsehẽ "you(s) have made it roll for him"	sing	bhari	singran	gari
lettuchoktukuŋ "I planted it for him"	gaice	saman	orange	
uhektudhette "they cut it for him"	ghasa	ghasa	sanwuasa	ghasa
kɨptukhutta "cut it for him!"	sing	saman	bokasa	sing
apidehẽ "you(s) gave it to me"	mobile	kok		cia
achoktace "you(d) passed it to him"	saman	saman	dori	cia
maikhuttace "he/they brought it to us (p-i)"	saman	saman	saman	cia
atamabidaŋsehẽ "you(s) have poured it for me"	bhari	cuwa		cuwa
phendudhettuce "he took it off for them"	ghasa	saman	dori	ghasa
naosakhutte "he/they threw it for us (p-i)"	sing		saman	ghasa
akhedachoktaŋsehẽ "you(s) have bought it for me"	mobile	saman	chawchaw	saman
napiceke "he/they give it to you(d)"	saman	saman	saman	phekwa
choknehẽ "I passed it to you(s)"	mobile	saman	saman	saman
khuttoko "he brought it to him"	saman	sontolong	saman	saman

	10	11
kɨptɨmpidum "we(i) could cut for him"	tang	tang
koptuchokta "pick it up for him!"	tangphekma	lungtak
oŋsudhettuce "He could peel it of for them"		phakce
tamsukhuttukum "we(i) pour it for them"	cuwa	cuwa
maipiceke "he/they give it to us (dual-i)"		phekwa
choktaŋkhaŋ "you(s) keep passing it to me"	sing	lungtak
ukhutte "they bring it to him"	ghasa	sitra
phendubida "you(s) take it off for him"	bakhra	wanam
osuchokta "you(s) throw it to him"	sing	lungtak
khoptudhett "he puts it for him"		phekwa
khuraŋkhuttahã "you(s) carry it for me"	ghasa	cuwa
pidukumcɨm "we give it to them"	cama	phekwa
acho'ŋa'a "you(s) pass it to me"	sing	latti
khuttuce "he brings it to them"	chocolate	cuwa
akhedabidaŋsehẽ "you(s) have bought it for me"	cia	marci
nasottachokte "he/they moved it for you(s)"	sing	
haŋsuchoktukumcɨm "you(ns) send it for them"	phekwa	phekwa
naktudhettuce "he asked it for them"		phekwa
napino "he/they give it to you"	cama	phekwa
amaichokte "you passed it to us(e)"	ghasa	
akhuttehẽ "you(s) bought it to me"	cama	sontholong
tisubidoŋsokha "he has put it to him"	sing	phekwa
konduchokta "you search it for him!"	tangphengma	syau
lɨktudhettakte "he was taking it in for him"		
loĩsuchoktaŋse "he has taken it out for him"	jempel	cuwa
khoktubida "he cuts it for him"	sing	sa
akhurachoktaŋsehẽ "you(s) have carried it for me"		cuwa
thuktudhette "he cooks it for him"	yakkheng	ma'mi
kɨrukhutta "he rolls it for him"	wood	
pidehẽ "I gave it"	cama	phekwa
achoktehẽ "you(s) passed it to me"	sing	lungtak
ukhuttehẽ "he brought it to me"	cama	chocolate

	10	11
maipide "he/they gave it to us (p-i)"	cama	sa
amachokte "you passed it to us (p-e)"	sing	mobile
nakhuttiki "he/they bring it to you(p)"	cama	chocolate
thubidoko "he drinks it for him"	cia	teĩ
thanduchokta "bring it for him!"		bhari
khomsakhuttahã "put it on for me!"		ghasa
piduhẽ "I give it to them"	phekwa	gaice
choktuce "he passed it to him"	ghasa	ghasa
khuttukuŋ "I bring it to him"	biscuit	chocolate
omabina'a "I throw it for you(s)"		mi
khoktakhuttoŋse "he has chpped it for him"	sing	
ukhurudhette "they carried it for him"		murda
thuktukhutta "cook for him!"	yakkheng	sa
apiŋa'a "you(s) give it to me"	cama	
nachokte "he/they passed it to you(s)"	sing	latti
ukhuttaŋcihẽ "they(d) brought it to me"	cama	naglasi
chuptubiduce "he pressed it for them"		marci
labuŋchoktuŋsuhẽ "I have caught it for him"	wacilek	phagcilekca
keŋsadhettanuma "hang it for him (you-ns)"	ghasa	
naktukhutta "ask it for him"	phekwa	dabi
napide "he/they gave it to you"	phekwa	citti
achoktumhe "you(p) passed it to them"	sing	citti
khuttumhe "we(p-i) brought it to him"	phekwa	phekwa
akɨrabidaŋsehẽ "you(s) have made it roll for him"	gari	sing
lettuchoktukuŋ "I planted it for him"	phekwa	pice
uhektudhette "they cut it for him"	ghasa	ghasa
kɨptukhutta "cut it for him!"	sing	tang
apidehẽ "you(s) gave it to me"	phekwa	
achoktace "you(d) passed it to him"	sing	phekwa
maikhuttace "he/they brought it to us (p-i)"	cama	phekwa
atamabidaŋsehẽ "you(s) have poured it for me"		ghasa
phendudhettuce "he took it off for them"	ghasa	phakca

	10	11
naosakhutte "he/they threw it for us (p-i)"	sing	phekwa
akhedachoktaŋsehẽ "you(s) have bought it for me"	biscuit	
napiceke "he/they give it to you(d)"	cuwa	phekwa
choknehẽ "I passed it to you(s)"	sing	citti
khuttoko "he brought it to him"	kopi	citti