
A first take on information structure in Totoli - Reference management and its interrelation with voice selection

Author(s): Sonja Riesberg

Source: Proceedings of the Second International Workshop on Information Structure of Austronesian Languages, 25 December 2015 (ISBN 978-4-86337-212-2 B212), pp. 65-81.

Published by: ILCAA, TUFS

Permanent URL: <http://hdl.handle.net/10108/84510>

Prometheus-Academic Collections is the source of the academic achievements which stores the research and educational achievements of Tokyo University of Foreign Studies and preserves historical materials. It shares such academic achievements on the Internet to ensure the open university. All the copyrights reserved by each author, academic society, association, publisher, and other right owners. The whole or parts of the materials can be used (referred, reproduced, printed) with copyrights acknowledged for the academic and personal use only.

A first take on information structure in Totoli – Reference management and its interrelation with voice selection

Sonja Riesberg
Universität zu Köln

1. Introduction¹

Totoli is a Western Malayo-Polynesian language spoken in the northern part of Central Sulawesi. Like many languages in this group, Totoli is a symmetrical voice language, i.e. it displays more than one transitive constructions – an actor voice and two undergoer voices² – that behave morphologically and syntactically symmetrical. Examples (1)a. and (2)a. below illustrate two dynamic actor voice clauses with their respective undergoer voice constructions in (1)b. and (2)b.

- (1) a. *I Rinto manaip taipang.*
i Rinto **moN**-taip taipang
HON PN AV-peel mango
'Rinto is peeling a mango.'
- b. *Taipang taip i Rinto.*
taipang taip i Rinto
mango peel:UV1 HON PN
'Rinto is peeling a mango.'
- (2) a. *I Winarno mongusut kunji motorna.*
i Winarno **moN**-kusut kunji motor=na
HON PN AV-look.for key scooter=3s.GEN
'Winarno is looking for the keys for his scooter.'
- b. *Kunji itu kusuti i Winarno.*
kunji itu kusut-**i** i Winarno
key DIST look.for-UV2 HON PN
'Winarno is looking for the keys.'

The two morphologically distinct undergoer voices – here glosses as UV1 and UV2 respectively – are lexically determined and unlike in Philippine-type languages not semantically distinct, i.e. in both cases a patient or a theme argument is linked to subject position. In addition to the alternation between actor voice and undergoer voice, there is an obligatory distinction between realis and non-realism mood, as shown in the two undergoer voice examples below. Note that the different mood values are reflected in the English translations by different tenses (past versus future or progressive forms). Table 1 summarises the Totoli voice paradigm for dynamic verbs, including non-realism and realis forms.

¹ I'm in debt to Katharina Haude, Nikolaus Himmelmann, and the participants of the second international workshop on information structure of Austronesian languages 2014 for valuable critique and comments.

² Totoli exhibits a third voice, the locative voice in which a location is linked to subject position. While fully productive, locative constructions have a somewhat special status, as they are syntactically far more restricted than actor voice and undergoer voice constructions. The locative voice is therefore not further considered in this paper. For a detailed description see Himmelmann/Riesberg 2013.

- (3) a. *Niug ana kodoong botak i Jui*
 niug ana ko-doong botak i Jui
 coconut MED POT-want split:UV HON PN
 ‘Jui is splitting a coconut.’
- b. *Niug ana tookamo nibotak i Jui*
 niug ana tooka=mo **ni**-botak i Jui
 coconut MED finish=CPL **RLS**-split:UV HON PN
 ‘Jui split a coconut.’

	NON-REALIS	REALIS
AV	<i>moN</i> ³ <i>mog</i> - <i>mo</i> -	<i>noN</i> - <i>nog</i> - <i>no</i> -
UV1	∅	<i>ni</i> -
UV2	<i>-i</i>	<i>ni</i> - <i>-an</i>

Table 1: Totoli voice paradigm, dynamic verbs

In addition to the paradigm shown in Table 1 there are two more verbal paradigms, the stative paradigm and the potentive paradigm. Stative forms denote qualities or (emotional) states. They are typically intransitive, taking an undergoer subject, but transitive uses are also possible. These then always imply a notion of causativity. Potentive forms, which are formally identical with the stative paradigm, denote events which take place accidentally or actions which are carried out with lack of control on the part of the acting participant. They can also have an ability reading, denoting that something can principally be done or has already been achieved.

The alternations in (1) and (2) two are symmetrical in that all voices are overtly marked⁴ by voice morphology and in that – unlike in an active-passive alternation – the non-subject arguments show the same behavioural properties (e.g. with respect to relativisation, control, raising, word order restrictions etc.). While languages may differ in the degree to which their voice systems are symmetrical, with certain subtle behavioural differences (as recently established in Riesberg 2014), Totoli seems to be a particular prototypical instance of a symmetrical voice language: Totoli shows a nearly 100% symmetry in the behaviour of verbal arguments (one exception being the different realisation of pronouns in actor voice and undergoer voices, see below). In particular, Totoli does not display the same kind of definiteness restrictions known from many other western Austronesian languages, like for

³ The distribution of the three actor voice prefixes is determined mostly by phonological factors: vowel-initial bases, almost all of which are non-derived, take *mog*-, consonant-initial lexical bases take *moN*-, and derived stems mostly take the prefix *mo*-. There is a limited class of consonant-initial verbs which are lexically subcategorized for *mo*-.

⁴ Note that in many Austronesian languages that display symmetrical voice, there is usually one slot in the verbal paradigm that remains morphologically unmarked (here the non-realis form of UV1). However, language inherent evidence as well as cross-linguistic comparison give reason to assume that the non-marked forms are a historical coincident rather than representing the “unmarked” voice (in the sense that the active represents the “unmarked” voice in the active-passive alternation). Thus, language-internally non-marked forms always stand in paradigmatic opposition to marked ones. Cross-linguistically, the non-marked slots do not occur in the same position in the paradigm. For a more detailed discussion on this issue see Riesberg 2014 (especially section 2.2.5).

example Tagalog. In Tagalog definite undergoer arguments usually have to become the subject of the construction and actor voice constructions with definite non-subject undergoer arguments are clearly dispreferred. This is illustrated in the Tagalog example in (4)a., where the undergoer argument *bahay* has to receive an indefinite interpretation (i.e. ‘a house’). If the same state of affairs needs to be expressed with a definite undergoer argument, the speaker has to choose a patient voice construction as in (4)b. (though see Himmelmann 2005: 367 for an discussion of exceptions and counter examples to this rule of thumbs).

- (4) a. *Sumira siya ng bahay.*
 -um-sira siya ng bahay
 -AV-destroy 3s.NOM GEN house
 ‘(s)he destroyed a/*the house.’ (Latrouite 2012: 96)
- b. *Sinira niya ang bahay.*
 -in-sira niya ang bahay
 -PV-destroy 3s.GEN NOM house
 ‘(s)he destroyed the house.’ (Latrouite 2012: 96)

But this is clearly not what we find in Totoli (at least in elicited data); compare the Tagalog data in (4)a. to the Totoli actor voice construction in (2)a., where the undergoer argument is realised as a possessive phrase *kunji motorna* ‘the keys for his scooter’ and thus can/has to be interpreted as definite.

This then brings us to the major research question of this paper: If the two major voice constructions in Totoli are indeed symmetrical and syntactically equal, how do speakers choose which voice to use? Even though we do not find the same strict definiteness restrictions as in other western Austronesian languages, the hypothesis would be that discourse pragmatic factors influence the choice of voice selection made by the speakers during discourse. This paper will therefore investigate reference management in four spoken Totoli narratives and look whether there is a interrelation between the information status of referential expressions and the voice construction. Before looking at the actual numbers and counts from these texts in section 4, section 2 will give an introduction of the different ways to refer to discourse participants in Totoli. Section 3 will introduce and explain the annotation scheme used for the text analyses.

2. Referential expressions in Totoli

As in all languages, there is a wide range of possibilities to refer to discourse participants in Totoli, ranging from zero forms to complex nominal expressions. This section introduces these means of reference without going into detail as to in which information structural contexts they might occur.

2.1 Zero anaphora and bound- and free pronouns

Totoli has two series of personal pronouns; the nominative series consisting of free forms and the genitive series consisting of clitics (cf. Table 2):

	NOMINATIVE	GENITIVE
1SG	<i>aku</i>	= <i>ku</i> ; <i>ku-</i>
2SG	<i>kau</i>	= <i>mu</i> ; = <i>ta</i>
3SG	<i>isia</i>	= <i>na</i>
1PL EXCL	<i>kami</i>	<i>kami</i>
1PL INCL	<i>kita</i>	= <i>ta</i>
2PL	<i>kamu</i>	= <i>ta</i>
3PL	<i>sisia</i>	<i>sisia</i> ; (=na)

Table 2: Totoli pronouns

In actor voice constructions, nominative forms can function either as subjects or as non-subject arguments (cf. (5)a.), in undergoer voice constructions nominative forms usually only occur in subject position, while the non-subject argument is realised by the genitive form, cliticised to the verb. If cliticised to nouns, genitive pronouns mark the possessor in a possessive construction. Note that =*na* is mostly only used for third person singular, whereas the free form *sisia* is used for plural forms. In some instances, however, =*na* can also be found to refer to third person plural actors.

- (5) a. *Aku nongiu' kamu kalangena ia.*
aku noN-iu' **kamu** kalangena ia
1s AV-call **2** a:moment:ago PRX

'I called you this morning.'

[political_meeting.004]

- b. *kalambotimu aku nokulia*
ko-lambot-i=**mu** **aku** no-kulia
POT-remember-UV2=**2s.GEN** **1s** AV.RLS-study

'you remember me study.'

[farming_2.2037]

In spoken discourse, it is common to drop referential expressions if they have been introduced before. This is very common for undergoer voice subjects, as illustrated in the sequence in (6), taken from a narrative. After a first mention of the undergoer subject (*bungo sagin itu* 'the banana fruits'), the following four predicates occur without overt subject expressions. The actor argument, however, is still realised by the third singular pronominal clitic =*na*. This seems to be a common phenomenon in Austronesian languages, see e.g. Himmelmann 1999 on the lack of zero anaphora in undergoer voice constructions in Tagalog. In actor voice constructions, both subject and non-subject argument phrases are often omitted (cf. example (7) where no referring expression is used).

- (6) *Njan nalapitna bungo sagin itu*
njan no-lapit=*na* **bungo** sagin itu
like.that POT.RLS-reach=**3s.GEN** fruit banana DIST
- poopolut niganutna ai nikaanna*
RDP2-polut ni-ganut=*na* ai ni-kaan=*na*
RDP2-penetrate RLS-pull.off:UV1=**3s.GEN** and RLS-eat:UV1=**3s.GEN**

nijjomoona.

ni-RDP1-jomoo=*na*

RLS-RDP1-devour=**3s.GEN**

‘As he reached the banana fruits, he opened (them), ripped off (their peel),
and ate (them). He gorged (them)’ [monkey_turtle.245-249]

- (7) *Ngadaan nousa ana nangkaalamai.*
 nga daan no-usa ana noN-ko-ala=mo=ai
 NEG EXIST ST-long and AV.RLS-ADA-get=CPL=VEN
 ‘It didn’t take long and (they) got (it).’ [monkey_turtle.110-111]

2.2 Demonstratives and demonstrative phrases

Totoli exhibits three demonstrative formatives, roughly marking three levels of distance from the speaker: *ia* signals closeness to the speaker (glossed here as *proximative* = PRX), *ana* signals an intermediate distance from the speaker (glossed as *medial* = MED), and *itu* (glossed as *distal* = DIST) which marks a distance furthest away from the speaker. These demonstratives can function as free demonstrative pronouns, as in (8), as well as determiners in demonstrative phrases, e.g. with nouns ((9)a.), pronouns ((9)b.), or prepositional phrases ((9)c.).

- (8) a. *Ia nollipa nolobaanku Nanong.*
ia no-RDP1-lipa no-loba-an=ku Nanong
PRX ST-RDP1-forget ST-inform-APPL1=1s.GEN PN
 ‘This one has been forgotten, I told Nanong.’ [conversation_4.711]
- b. *Tongaita ana*
 tonga-i=ta **ana**
 ask-UV2=1pi.GEN **MED**
 ‘We ask that.’ [expl_celeb.197]
- (9) a. *Bali aku kode mmake leang sagin ana*
 bali aku kode mon-pake **leang** **sagin** **ana**
 so 1s only AV-use **leaf** **banana** **MED**
 ‘So I just use these banana leafs.’ [red_sugar.393]
- b. *geimo kodoonganta aku ia.*
 geimo ko-doong-an=ta **aku** **ia**
 not ST-like-APPL2=1pi.GEN **1s** **PRX**
 ‘You don’t like me.’ [podok_langgat.186]
- c. *Dei bale itu paapake daster.*
dei **bale** **itu** RDP2-pake daster
LOC **house** **DIST** RDP2-wear house.dress
 ‘In the house, (she) is wearing a house dress.’ [conversation_4.663]

2.3 Nouns and complex noun phrases

As might have become clear from the given examples so far, Totoli has neither definite nor indefinite articles. A bare noun can either receive a definite or an indefinite interpretation, depending on the context. To stress the fact that a specific entity is meant, speakers can use one of the demonstratives discussed in section 2.2 above, but this use is not obligatory for a definite reading. Compare the two instances of the noun *ondan* ‘ladder’ in example (10): The

first instance is the first mention of the ladder in this conversation, and is thus interpreted as indefinite. In the second instance, however, the ladder is already known and thus receives a definite reading.

- (10) *pertama monodokan ondan (...)*
 pertama moN-todok-an **ondan**
 first AV-stand-APPL1 **ladder**
- koopatmo danna limpatan ondan*
 ko-opat=mo daanna limpat-an **ondan**
 ADA-four=CPL then move-APPL1 **ladder**

‘First, you have to put up a ladder (...) There are four (sides to pick).

Then, you have to move the ladder’ [cloves.17&84]

Another strategy for generating a definite reading is by adding the third person genitive pronoun =*na* to the respective noun, which could either mark possession (and therefore single out the entity as specific and definite), or, in some cases mark definiteness without possession. This seems to be a common phenomenon in many Austronesian languages, e.g. also in Indonesian and Balinese, though only little work has been done on this topic (but see e.g. Haiduck 2014 for Balinese). See, for example (11), where the NP *bangunanna* does not mean ‘his building(s)’ or ‘their buildings’ but rather denotes ‘the buildings’ in former times in the village of Bjukan.

- (11) *Tempo ia sampe sekarang Bayugan*
 tempo ia sampe sekarang Bajugan
 time PRX until now PN
- geiga dennia bangunanna*
 geiga dennia bangunan=**na**
 NEG like.this building=**3s.GEN**

‘In these times until now (in) Bajugan,
 they weren’t like this, the buildings.’

[bajugan.44-46]

Noun phrases can furthermore become more complex by being modified by other nouns (cf. e.g. *leang sagin* ‘banana leaf’ in example (9)a.), by stative verbs (e.g. *tampat melea* ‘a large place’), or by relative clauses, as in (12)a. Headless relative clauses can also function as either subjects or non-subject arguments, as in (12)b.

- (12) a. *tau moane anu kodoong kabing*
 tau moane anu ko-doong kabing
 person man REL POT-want marry

‘the man who wants to get married’

[wedding_expl_TTL.026]

- b. *Nokotiing pokotinga i olong.*
 noko-tiing poko-tinga i olong
 POT.AV.RLS-hear POT-say HON monkey

‘(he) heard what the monkey had said.’

[monkey_turtle.277]

2.4 Prepositional phrases

Prepositional phrases in Totoli are usually used to denote peripheral participants, most notably locations and instruments. The preposition *dei* is by far the most frequent one, marking locations, but also goals and recipients (cf. the three examples in (13)). Other prepositions are *uli* ‘from’, *takin* ‘with’ (instrumental), and *lengan* ‘with’ (comitative).

- (13) a. *I Lolio nemea dei tangipa boto*
 i Lolio no-mea dei tangipa boto
 HON PN ST.RLS-live LOC other.side small.lake
 ‘Lolio lived on the other side of the small lake’ [monkey_butterfly.061-062]
- b. *Mallako dei daami.*
 mo-RDP1-lako dei daami
 AV-RDP1-walk LOC abandoned.garden
 ‘Walk to the recently abandoned garden.’ [map_task_2b.407-409]
- c. *kodoong mangambuling dei togu bitu ana*
 ko-doong moN-kambuling dei togu bitu ana
 POT-want AV-return LOC possession bracelet MED
 ‘(it) wants to give the bracelet back to its owner.’ [chicken_eagle.170]

To conclude this section, Table 3 lists all Totoli referential expressions that have been discussed in the previous sub-sections:

PRON	zero	Ø (6)
	free	e.g. <i>aku, kamu</i> (5)a.
	bound	e.g. <i>=ku, =mu, =ta</i> , etc. (5)b.
DEM	PRX	<i>ia</i> (8)a.
	MED	<i>ana</i> (8)b.
	DIST	<i>itu</i>
DP	DPPr	e.g.
	DPM	e.g. <i>leang sagin ana</i> ‘these banana leaves’ (6)
	DPD	e.g. <i>bale itu</i> ‘that house’ (9)c.
N		e.g. <i>ondan</i> ‘a/the ladder’ (10)
N _{poss}	N= <i>ku</i>	e.g. <i>anak=ku</i> ‘my child’
	N= <i>mu</i>	e.g. <i>tangayopan=mu</i> ‘your plants’ (21)
	N= <i>na</i>	e.g. <i>amang=na</i> ‘his father’
	N= <i>kami</i>	e.g. <i>bakele kami</i> ‘our grandmother’
	N= <i>ta</i>	e.g. <i>usat=ta</i> ‘our sibling’
NP	N= <i>sisia</i>	e.g. <i>tinga sisia</i> ‘their language’
		e.g. <i>mangana dedek</i> ‘small child’
PP		e.g. <i>dei daami</i> ‘to the garden’ (13)b.
REL		e.g. <i>tau moane anu kodoong kabing</i> ‘the man who wants to get married’ (12)a.

Table 3: Referential expression in Totoli

3. Information status and the RefLex annotation scheme

This section introduces the annotation scheme used in this study. There are numerous accounts on discourse analysis and on reference management in discourse. By now, it seems to be established that at least three different activation statuses – given, activated, and new – should be distinguished (cf. Chafe 1976, Prince 1981). Other authors have proposed more fine-grained distinctions, such as in the well-known givenness hierarchy established by Gundel, Hedburg and Zacharski 1993. This hierarchy consists of the six statuses given in (14) below. Each of these status is assumed to be “a necessary and sufficient condition for the appropriate use of a different form or forms” (Gundel et al. 1993: 275). As Gundel et al. show in their comparative study on English, Chinese, Japanese, Russian, and Spanish, not all statuses are relevant in all languages. However, for all languages the hierarchy predicts that a given form is inappropriate if the respective status is not met.

(14) in focus > activated > familiar > uniquely identifiable > referential > type identifiable

While this is an very interesting and certainly insightful approach, I found it rather difficult to apply to my own data. When faced with the Totoli texts, I often struggled trying to determine the correct status to a given form. I therefore decided to use the less complex annotation scheme developed by Stefan Baumann and Arndt Riester (cf. Baumann/Riester 2012; 2013). Baumann’s and Riester’s two-dimensional annotation scheme (called RefLex) has been developed to investigate the relationship between information status and prosody. In particular, it claims to enable even non-expert annotators to create consistent annotations and is therefore easier applicable than the six-status approach by Gundel et al. (even though it does not consist of less categories).

The RefLex annotation scheme is two-dimensional in that it annotates the information status of a given referential expression on two levels, i.e. on the referential as well as on the lexical level. Compare the three examples below to see the difference between these two levels (all taken from Baumann/Riester 2013):

- (15) a. *After the holidays, John arrived in a new car, and also Harry had bought a new car.*
- b. *A car was waiting in front of the hotel. I could see a woman in the car.*
- c. *Yesterday, a friend of mine prepared a lasagne for me. I found it hard to enjoy the tasteless stuff.*

In both (15)a. and (15)b. there are two instances of the same referring expression (*a new car* and *a/the car* respectively). In (15)a., the first instance of *a new car* is referentially new, as it is an indefinite expression introducing a new referent. It is also lexically new, as the lexical items have not been used before. The second instance is also referentially new, as it refers to another car than the first mention. However, it is lexically given, as the same lexical material has been used immediately before. In (15)b., again the second instance of *the car* is lexically given, but this time it is also referentially given, as both instances refer to the same referent. Finally, in (15)c. we find the reverse situation in which *the tasteless stuff*, though referentially given as it refers to the same lasagne, is lexically new. The distinction between referential givenness and lexical givenness is important for Baumann and Riester, as it offers, for example, an explanation for the fact that (15)a. and (15)b. receive the same prosodic marking, i.e. deaccentuation of the second mention of *the (new) car*, even though in a. it is referentially new and in b. it is given. For the main research question of this paper – the question whether

there is a relationship between voice on the one hand, and information statuses of referential expression on the other hand – I believe that the distinction of these two levels is not necessary. I will therefore only apply the categories of the referential level proposed in Baumann/Riester (2012; 2013). However, it should be noted that the study presented in this paper is very preliminary. If it turns out that information status on the lexical level *does* play a role for voice selection, it can be easily added in further studies.

For the study presented in this paper, I used a simplified version of RefLex, as proposed in Baumann/Riester (2013). This version contains five labels on the referential level, summarized and briefly explained in Table 4 (for the full, more complex scheme see Baumann/Riester 2012). As mentioned in section 2, definiteness in Totoli is not necessarily overtly marked. Nevertheless, depending on the information status of the respective discourse referent, linguistic expressions receive definite or indefinite interpretations. It therefore seems justified to keep the three-way distinction of definite, definite/indefinite, and indefinite as proposed by Baumann/Riester 2013. In the following sub-sections, I will exemplify the annotation labels, mainly using Totoli data from those texts that have been annotated for the present study.

Definite	
r-given	anaphor corefers with antecedent in previous discourse
r-bridging	anaphor can be resolved to non-coreferring antecedent or within a described scenario
r-unused	discourse-new, non-anaphoric definite expression referring to an item which is generally known or identifiable from its own linguistic description
Definite or Indefinite	
r-generic	abstract or generic item
Indefinite	
r-new	specific or existential indefinite introducing a new referent

Table 4: Labels for the annotation of discourse referents in the RefLex scheme (Baumann/Riester 2013)

3.1 Given versus new

The given versus new distinction has already been briefly illustrated with the English example in (15)a. The Totoli example below consists of the first three intonation units of a story about a monkey and a turtle, a folk story that is well known not only in the Tolitoli area, but also in other parts of Sulawesi and the Philippines. The three bare nouns in the first intonation unit, *bolong* ‘a monkey’, *pomponu* ‘a turtle’, and *guan* ‘a garden’, introduce the most important participants of the story. In intonation units two and three two of them, the monkey and the turtle, are taken up again, this time expressed by a zero form.

(16)	<i>sirita</i>	<i>bolong</i>	<i>ai</i>	<i>pomponu</i>	<i>nogutu</i>	<i>gauan</i>
	sirita	bolong	ai	pomponu	no-gutu	gauan
	story	monkey	and	turtle	AV.RLS-make	garden
		<u>new</u>		<u>new</u>		<u>new</u>

<i>pomoona</i>	<i>molipidoan</i>	Ø
pomoo=na	moli--an pido	Ø
first=3s.GEN	RCP good	Ø

given

<i>njan</i>	<i>pombulina</i>	<i>nolidaisan</i>	Ø
injan	pombuli=na	noli--an dais	Ø
then	later=3s.GEN	RCP.RLS bad	Ø

given

‘(This is) the story about a monkey and a turtle making a garden.

First, they were friends, and then they hated each other’ [monkey_turtle.001-003]

A similar example has already been given in (10), where both, given and new information is expressed by a bare noun (i.e. *ondan* ‘a/the ladder’). Note, however, that distance between the newly introduced referent and its second mention in the two examples discussed differ strongly: While in (16) we are dealing with consecutive intonation units, there are 64 units between the first and the second mention of the ladder in (10). As I believe that the distance between the occurrences of referents may play a role in voice selection (but also on the respective referring expression), I decided to use one more label that is missing from Table 4, but which is part of the more complex scheme in Baumann/Riester 2012, namely the category given-displaced. This label is used for exactly those cases as in (10), i.e. where there is a coreferring antecedent that has been mentioned at some point in the preceding discourse, but not in the immediately preceding utterances (the fixed unit in Baumann/Riester 2012: 138 is five intonation units or clauses, not counting smaller units, such as brief back channelling etc.).

3.2 Accessible

In addition to the three possibilities discussed in the previous section (i.e. given, given-displaced, and new), it is sometimes the case that a referent is neither given, as defined above, nor new, but still activated, or accessible. This is for example the case when an expression denotes a part of an entity that has been mentioned before, as in the example below, again taken from the story of the monkey and the turtle. Here, the stem, being a part of the banana tree (which has been mentioned), is accessible, even though it has not been mentioned before.

(17)	<i>sabab</i>	<i>ana</i>	<i>sagin</i>	<i>mapanggat</i>	<i>batangna</i>
	sabab	ana	sagin	mo-panggat	batang=na
	because	if	banana	ST-high	stem=3s.GEN
			<u>given</u>		<u>bridging</u>

‘because as for the banana (tree), its stem is high’

[monkey_turtle.81-82]

Baumann and Riester use the term *bridging* rather than *accessible* (Chafe 1976) or *inferable* (Prince 1981), because it subsumes not only meronyms (as just illustrated in (17)), but also other context-dependent expressions which do not possess a coreferential antecedent (Baumann/Riester 2013: 22), as, for example, in (18). This example is taken from the end of a story about a chicken and an eagle. The chicken borrows a bracelet from the eagle and then loses it. This, so the story goes, is the reason why chickens always keep scraping, even if they have enough food, and why eagles prey on chicks. The hearer, having heard the whole story, can therefore access the expression *utang* ‘the debt’, though it has not been introduced in the preceding discourse.

1. Which forms can be/are used for which information status?
2. How is the distribution of form and grammatical relation in the different voices?
3. How is the distribution of status and grammatical relation in the different voices?

4.1 Form and information status

Looking at how different referring expressions are used in Totoli narratives, the distribution is not particularly surprising. As one might expect, all three kinds of pronouns (zero, free, and bound forms) as well as demonstrative pronouns are only used when the referent is given, or given-displaced. One exception is the occurrence of a first person singular inclusive bound pronoun that the speaker uses to refer to himself and the listener. Not having been used in the discourse before but referring to the speech act participants, it has been classified as *unused*. Likewise expectable is the distribution of bare nouns. As already mentioned in section 2.3 (cf. e.g. the discussion of example (10)), bare nouns in Totoli can receive both a definite and an indefinite interpretation. This is also reflected in the use of bare nouns in the narratives investigated: While roughly two thirds of the occurrences denote referents that are given or given-displaced, almost one third is used for newly introduced referents. In fact, bare nouns make more than one third of the forms used for introducing new participants (31 of 79). A fairly large amount of bridging instances is also realized by bare nouns, as well as by possessive constructions (together 23 of 33). Especially the latter is, again, not surprising, as most of the instances where bridging occurs consist of part-whole-relations, such as illustrated in example (17). What might be a little less expected is the fact that demonstrative phrases, in addition to their frequent use for denoting given (and given-displaced) referents, are also fairly often used for introducing new participants. One of the most striking examples for this is the following, which is the very first utterance in the story about a chicken and an eagle. It is the first time the chicken is mentioned, yet it is expressed by the demonstrative phrase *manuk ana* ‘that chicken’.

(20)	<i>manuk</i>	<i>ana</i>	<i>masahabatan</i>	<i>ai</i>	<i>kueang</i>	
	manuk	ana	mo--an sahabat	ai	kueang	
	chicken	MED	RCP friend	with	eagle	
	‘the chicken was friends with the/an eagle’					[chicken_eagle.028-32]

Another unexpected instance found in the corpus is the occurrence of a zero form used to introduce a new participant. The example is given in (21); it is part of a story about the seven daughters of the king who run away from home. In the particular example, the undergoer subject of the predicate *aling* ‘to remove’ is left unexpressed, even though the item that is removed has not been mentioned before. In fact, it is neither mentioned in the following discourse, so it seems that the information about what is taken away is not important for the further development of the story. The reason why it is mentioned in the first place is not entirely clear, and as this is the only example of such a use of a zero form, it might also be possible that we are dealing with an error/*lapsus linguae*. Table 5 summarises the distribution of form and their respective information statuses in the four narratives investigated⁵.

⁵ Note that this table does not contain all of the 803 instances of referring expressions mentioned above. In order to keep things simple and more manageable, those cases that were classified as generic or abstract, as well as items that refer to text-external context, are excluded from the summary in Table 5.

(21)	<i>jaam</i>	<i>opat</i>	<i>nangalai</i>	<i>ssaakan</i>	<i>pottuluanna</i>
	jam	opat	-ngo- no-lai	sasaakan	po-RDP1-tuli-an=na
	clock	four	-COLL- AV.RLS-run	all	GER-RDP1-sleep-NR=3s.GEN
	<i>ialingnako</i>		Ø		
	i-aling=na=ko		Ø		
	RLS-remove=3s.GEN=AND		Ø		

‘At four o’clock they all ran from their beds. They took (something) away’
[king_frog.119-123]

	zero	free	bound	dem	DP ⁶	N	Nposs	NP	PP	REL	<i>sum</i>
given	185	109	87	14	31	49	12	3	22	1	513
given-dis	12	6	3	-	29	37	7	5	37	1	137
bridging	1	-	1	-	2	10	13	-	6	-	33
unused	-	-	1	1	4	4	2	-	1	1	14
new	1	1	-	1	8	31	9	6	22	-	79
<i>sum</i>	199	116	92	16	74	131	43	14	88	3	

Table 5: Forms and their statuses

4.2 Form, grammatical relation, and voice

This section will look at potential asymmetries between actor voice constructions and undergoer voice constructions with respect to the linguistic form of different grammatical relations (primarily of subjects and non-subject arguments). One of these asymmetries have been mentioned before, namely the fact that non-subject pronouns in undergoer voice constructions but not in actor voice constructions are cliticized to the verb. However, this kind of asymmetry originates in the morpho-syntactic system of the language. A more interesting question is thus, whether there are other asymmetries in the form of actor voice and undergoer voice subjects and non-subject arguments that are due to information structural factors rather than morpho-syntactic ones. This seems to be indeed the case (cf. Figures 1-4).

Note, first of all, that some of the different forms that have been introduced in section 2 and that have been kept apart in Table 5, have been combined for the purpose of investigating research questions 2 and 3: In the following, I will compare the use of zero forms, pronouns and lexical nominal material in subject and non-subject argument function in the two voice types. The category *pronoun* thus includes both, free forms (personal and demonstrative) and bound forms. The category ‘lexical nominal material’ should be understood as an umbrella term opposed to zero forms and pronouns. It includes bare nouns, determiner phrases, possessive constructions, and modified nouns as described in section 2.3. Furthermore, the label *actor voice* subsumes both actor voice constructions with dynamic verbs, as well as stative and potentive actor voice constructions. Likewise, *undergoer voice*, as used in this section, includes the ‘ordinary’ dynamic undergoer voice constructions, and stative and potentive constructions.

⁶ I use DP for determiner phrases and NP for complex noun phrases, as introduced in sections 2.2 and 2.3.

Looking at subjects first, the distribution of different forms between actor voice and undergoer voice is roughly equal, the difference mainly pertaining to the use of lexical material: Undergoer voice constructions make slightly more use of DPs, NPs and Ns (37% in UV, 26% in AV). The difference between the use of zero forms and pronouns respectively is relatively small (zero: 39% UV vs. 45% AV, pronouns: 24% UV vs. 29% AV).

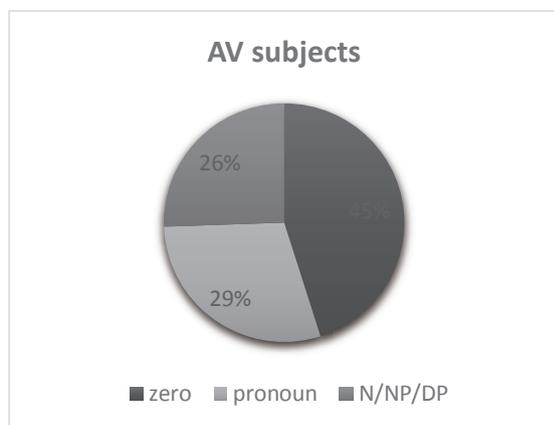


Figure 1: Forms of AV subjects

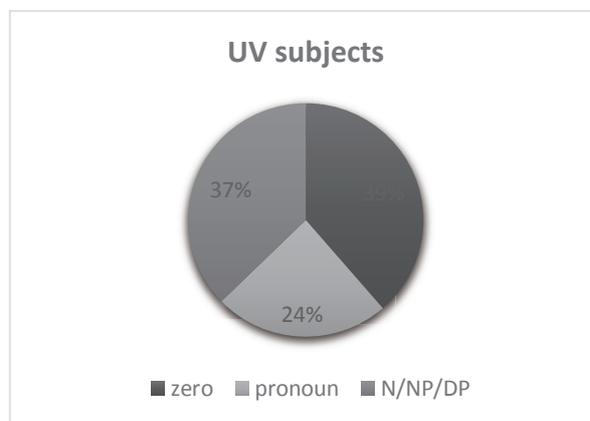


Figure 2: Forms of UV subjects

Looking at the forms of the non-subject argument phrases, the difference between actor voice and undergoer voice is, however, striking. The most noticeable difference pertains to the use of pronouns: In undergoer voice constructions, 66% of all non-subject arguments are realised as (bound) pronouns, whereas the corpus contains only five instances of pronominal non-subjects in an actor voice construction⁷. On the other hand, in actor voice constructions, the vast majority of non-subject arguments (74%) are realised as bare nouns, possessive phrases, complex nouns, or determiner phrases. In undergoer voice constructions, these make only 19% of all instances. The high number of pronominal non-subject arguments in undergoer voice constructions can probably be explained by the high number of so called “rapid action sequences” (Himmelman 1999: 244) typical for Totoli narratives (and Austronesian narratives in general). Typically, these sequences occur in the undergoer voice, where the subject, after being initially mentioned in the first unit, is dropped for the rest of the sequence, the actor, however, keeps being realised as a bound pronoun. An example of such a rapid action sequence of this kind was given in example (6) (cf. also the above mentioned observation by Himmelman 1999 that western Austronesian language have a tendency not to drop actor arguments in undergoer voice constructions). Yet, as Figure 4 shows, zero forms are possible in Totoli undergoer voice constructions (16% in the investigated corpus). Whether these are in fact zero anaphora or whether these actor-less constructions can/must be accounted for by other factors (as Himmelman 1999: 255 does for Tagalog), is left for further research.

⁷ Note, however, that the corpus is still very small. More data is certainly necessary to make more reliable claims, especially with regard to non-subject arguments in actor voice constructions.

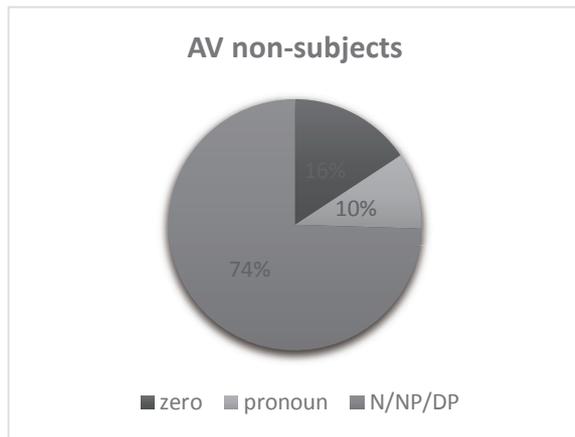


Figure 3: Forms of AV non-subject arguments

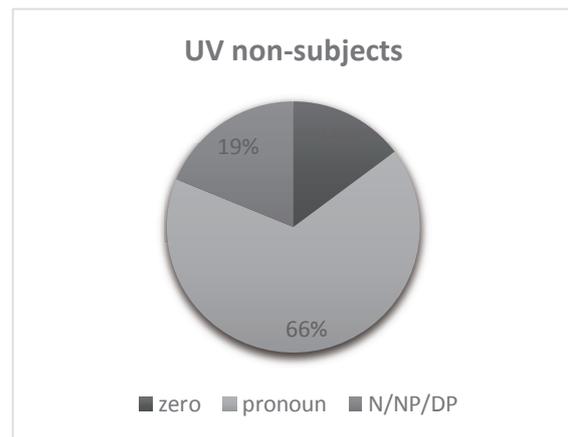


Figure 4: Forms of UV non-subject arguments

4.3 Status, grammatical relation, and voice

The previous section has given a first impression over asymmetries in the realisation of referential expressions in actor voice and undergoer voice constructions, and it has become clear that these asymmetries mainly manifest in the realization of non-subject arguments. This section will now shift the perspective, not looking at the form but at the status of the referring expressions.

Again, starting with the subjects, the difference between actor voice and undergoer voice is only marginal. Lumping together given and given-displaced referents, these two categories constitute the largest group in both voices (96% in AV, 86% in UV). New referents in subject position are slightly more frequent in undergoer voice than in actor voice constructions.

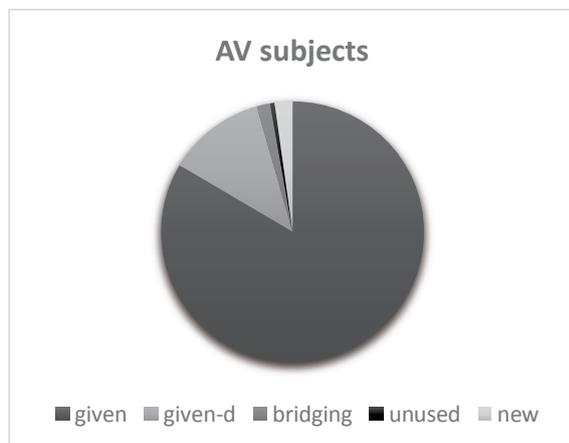


Figure 5: Status of AV subjects

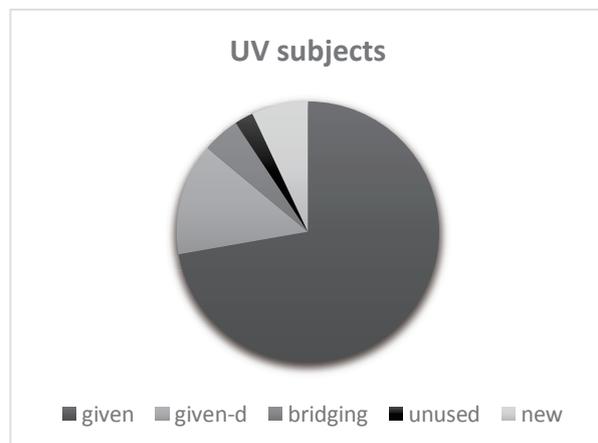


Figure 6: Status of UV subjects

Just as in the previous section, a more striking difference between actor voice and undergoer voice can be seen when looking at the statuses of referents in non-subject function. Similar to what we saw for referents in subject function, the group of given and given-displaced referents in non-subject function is the largest one in both voices. However, the difference between actor voice and undergoer voice is more pronounced when looking at non-subject function than it is for referents in subject function (73% in AV, 96% in UV). The most striking difference between actor voice and undergoer voice with respect to referents in non-

subject function consists in the use of new referents, which is significantly higher in actor voice constructions (21%) than in undergoer voice constructions (1%).

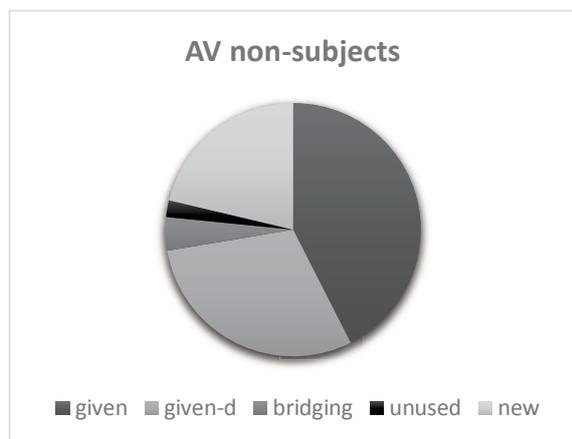


Figure 7: Status of AV non-subject arguments

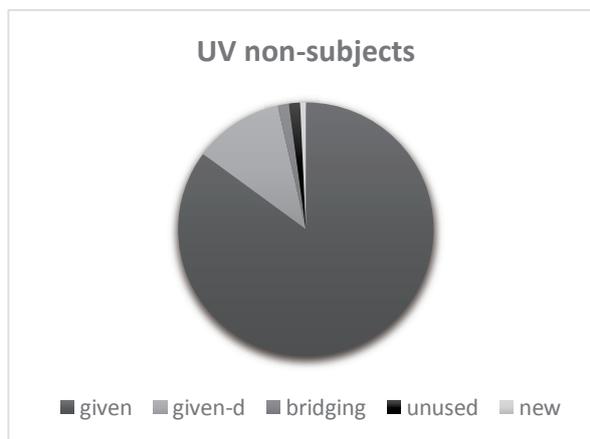


Figure 8: Status of UV non-subject arguments

5. Discussion and summary

This paper has presented some preliminary results of a corpus study of four spoken Totoli narratives. The major aim was to investigate referential expression, as well as the question whether and how information structural factors influence the voice selection in discourse. It should be clear that many important factors that would lead to a more complete picture of reference management and its interaction with voice selection have not been taken into account in this pilot study. These include, among others, distinguishing intransitive and transitive clauses and keeping apart the different verbal paradigms (dynamic, stative, and potentive). Problematic is certainly also the current treatment of zero-forms, where zero anaphors are not separated from other uses of zero forms. Finally, note that due to its small size, and probably also due to the nature of the selected texts, the corpus contains fairly few instances in which new participants are introduced into the discourse (79 in total, out of which the 22 prepositional phrase were not considered in the analyses in sections 4.2 and 4.3). A larger corpus is thus necessary to make more reliable claims about reference management in general and the introduction of new participants in particular.

Despite these deficiencies, some generalisations could be made as to which referential expression are the preferred choice for a given information status, as summarised in Table 5. Furthermore, it was shown that there are asymmetries in the realisation of subjects and non-subject arguments when comparing actor voice constructions with undergoer voice constructions. These are especially pronounced in the use of pronouns, particularly in non-subject argument function. With respect to the status of referents, it was shown that there is a strong tendency to introduce new participants as undergoers: These are most often realized as non-subject arguments of actor voice constructions, but new referents in subject function of undergoer voice constructions are also fairly frequent (and quite more so than in subject function of actor voice constructions or in non-subject function of undergoer voice constructions).

References

- Baumann, S. & Riester, A. 2013. Coreference, Lexical Givenness and Prosody in German. In *Lingua* 136. 16-37.
- Baumann, S. & Riester, A. 2012. Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects. In Elordieta, Gorra & Pilar Prieto (Eds.). *Prosody and Meaning*, 119-162. Berlin, New York: Mouton De Gruyter.
- Chafe, W. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In Charles Li (ed.), *Subject and Topic*, 25-56. New York: Academic Press.
- Gundel, J. K., Hedberg, N. & Zacharski, R. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. In *Language* 69:2. 274-307
- Haiduck, L. 2014. *The Suffix -ne in Balinese*. Köln: Universität zu Köln Master's Thesis, http://www.academia.edu/10241220/The_Suffix_-ne_in_Balinese.
- Himmelman, N. P. 2005. Tagalog. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*, 350-376. London & New York: Routledge.
- Himmelman, N. P. 1999. The Lack of Zero Anaphora and Incipient Person Marking in Tagalog. *Oceanic Linguistics* 38. 231-269.
- Himmelman, N. P. & Riesberg, S. 2013. Symmetrical Voice and Applicative Alternations: Evidence from Totoli. *Oceanic Linguistics* 52. 396-422.
- Latrouite, A. 2012. Differential Object Marking in Tagalog. In: Lauren Eby Clemens, Greg Scontras, and Maria Polinski (eds.), *Proceedings of the 18th Meeting of the Austronesian Formal Linguistics Association*, 94-109. Harvard University, Cambridge.
- Prince, E. F. 1981. Toward a Taxonomy of Given-New Information. In Peter Cole (ed.), *Radical Pragmatics*, 223-256. New York: Academic Press.
- Riesberg, S. 2014. *Symmetrical Voice and Linking in Western Austronesian Languages*. (Pacific Linguistics 646). Boston/Berlin: De Gruyter Mouton.