Proceedings

of

the Second International

Workshop

on

Information Structure

of

Austronesian Languages

Proceedings
of
the Second International Workshop
on
Information Structure
of
Austronesian Languages

The Second International Workshop on Information Structure of Austronesian Languages

Dates: 13-15 February, 2015

Venue：   Room 304, Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies (TUFS)

Workshop Website
http://lingdy.aacore.jp/en/activity/is-austronesian-2.html

(i)

Proceedings of the Second International Workshop on Information Structure of Austronesian Languages
Contents

Analyzing questions under discussion and information structure in a Balinese narrative

Arndt Riester (Institute for Natural Language Processing, University of Stuttgart)

Abstract

I argue against the skepticism recently expressed by Matić and Wedgwood (2013) regarding the possibility of defining a cross-linguistic category of focus. I sketch an interpretation-based and cross-linguistically applicable method of information-structural analysis, which makes use of Questions under Discussion. The method is demonstrated on a Balinese narrative text.

1. Introduction: is focus a universal linguistic category?

In a recent article, Matić and Wedgwood (2013) – henceforth MW – question the possibility of universal categories of information structure; in particular, the focus notion. Their essay is an outspoken criticism of parts of the current practice in the area of comparative information structure, backed up by both empirical and theoretical arguments. In my paper, I will turn against the pessimistic conclusions drawn by MW, and instead argue in the opposite direction, namely that universal notions of information structure are definable in a clear universal and pragmatic sense, and that they can – and, in fact, should – be used in linguistic research in order to ensure an objective means of cross-linguistic comparison. I will demonstrate a meaning-based annotation procedure that arrives at an information-structural analysis without making any language-specific assumptions about focus realization in particular languages. The following two quotes by MW illustrate their critical stance towards a universal notion of focus (emphasis added):

*"**[F]ocus is an inherently problematic category**, which has been used to draw together phenomena in the wrong way: as instances of a single underlying entity, as opposed to potentially independent entities that produce interestingly similar effects." (p.129)*

*"[W]e do not think that any one definition of focus need be basic or universal, and we see no basis for any such assumption [...]; our contention is that, from a linguistic point of view, **the***

term focus may simply not denote a cohesive phenomenon about which to theorise." (p.134)

The main worries articulated by MW pertain to the fact that, often, in the literature, functionally different phenomena have been misanalyzed as being realizations of the same underlying information-structural feature. Moreover, they show that, often, certain morphemes or structural operations in different languages have been prematurely classified in terms of abstract information-structural functions, while counter-evidence has been ignored or explained away. A further issue in MW's criticism relates to the well-known problem of imprecise or contradictory definitions of information-structural concepts found in the literature.

In Section 2, I will refute or at least weaken some of the arguments put forward by MW. In particular, I will address the notorious role of alleged focus markers or syntactic focus positions in comparative research and the conclusions that should be drawn from this situation, as well as the prevailing terminological and conceptual confusion with respect to a precise and cross-linguistically stable definition of focus. In Sections 3 and 4, I turn to the practicalities of analyzing information structure in a meaning-based, language-neutral way. The Appendix contains a comprehensive information-structural analysis of a Balinese narrative.

2.   Linguistic practice and the possibility of a universal focus notion

MW provide examples from various typologically unrelated languages that document how light-mindedly researchers have declared certain particles or syntactic positions to be *focus markers*, while it should have been obvious under more careful observation that the devices in question can mostly also be used with a non-focal meaning, or sometimes with a special meaning on top of what is ordinarily considered to be the core meaning of focus. For instance, elements in the preverbal "focus position" in Hungarian are known to additionally signal *exhaustivity*, and various "focus particles" in the Bantu language Aghem seem to indicate different degrees of *contrast* and *correction*. Other alleged focus markers express linguistic functions that are correlated with focus but nevertheless distinct from it, e.g. markers of *realis mood* in Somali (Afro-Asiatic), or morphemes expressing *direct evidentiality* in Quechua.

I entirely agree that both overly uniformist and excessively detailed focus classifications have often done more harm than good in the history of information-structure theory, and they have, indeed, contributed to the perception that cross-linguistic studies that concern the realization of information structure are vain and fruitless. However, other than MW, I do not

conclude from this that cross-linguistic comparison is indeed futile or impossible but, on the contrary, that we need clearer definitions and better, language-independent, methods of information-structural analysis. Most importantly perhaps, we should give up the idealistic conception that information-structural functions must be hardwired to a single particular morpheme, pitch accent, or syntactic construction. Skopeteas and Fanselow (2010) have shown, in an elicitation study on the identificational vs. non-identificational distinction for subject and for object focus in languages as diverse as Georgian, Hungarian, Québec French and American English, that there is clear "evidence against a cross-linguistic 1:1 mapping between types of focus and structural operations" (Skopeteas and Fanselow, 2010: 194) but there are nevertheless robust statistical differences of focus realization across these languages. This result could not have been achieved without the postulation of interpretive *tertia comparationis*, i.e. abstract notions of information structure. Certain constructions, syntactic operations, morphemes or prosodic features in a specific language have a statistical tendency to be used in the expression of focus or one of its subclasses, but we cannot expect them to do so unanimously, since there will always be other aspects of form and meaning that interfere. To tease all the factors apart is a legitimate and important linguistic goal but, in order to achieve a deeper understanding, it is indispensable to start out from some clearly defined abstract interpretational categories.

The fact that some languages seem to mark certain sub-divisions of focus which other languages ignore, or the fact that prototypical focus constructions in certain languages express meanings that exceed the core meaning of focus should not bother us too much in this regard. The important point, in the first place, is that we have a clear idea of what the core meaning of focus is. To study additional meaning aspects is legitimate and valuable but should be seen as a later step.

As I said, I am fully in line with MW's worries concerning the premature association of certain morphemes or syntactic positions with information-structural functions. While this may often seem tempting, it is, in fact, bad linguistic practice. To illustrate the worries, I simply choose the example of English, which is known for its prosodic marking of focus. Empirical evidence shows that, often, focus constituents carry a high (or falling) nuclear pitch accent. But, of course, this is not a license for calling the H* accent in general a *focus accent*. First of all, it is not too difficult to find corpus examples in which topical, backgrounded, or not-at-issue material is marked by the same type of pitch accent. Second, there are cases of focus constituents, especially complex ones, that are prosodically realized in a more elaborate manner, e.g. by means of a complex, internally structured, prosodic contour that consists of a series of different pitch accents. Third, there are other cases of focus (so-called second-occurrence foci, cf. Beaver and Velleman 2011) that are not marked by any pitch

movement at all. Hence, unfortunately, the conclusion is that linguistic reality is "dirty" and complicated. We cannot expect focus to always have the exact same realization, although statistical effects are clearly present. But should this seriously be considered an argument against the benefit of an abstract concept of *focus*? I do not think so.

But, then what do we mean by an abstract concept of focus? The most common, and by now more or less consensual, definition of focus given in the literature is that focus is what constitutes the answer to an explicit or implicit question (also referred to as the *Question under Discussion (QUD)* or *current question,* cf. Roberts 1996, Beaver and Clark 2008). Finding the QUD and, therefore, identifying the focus has its own issues, which I will address in Section 3. MW do not actually dispute this definition. However, they address the notorious problem of subcategorizing focus, in particular the problem of delimiting a category of *contrastive focus* from a more basic category of *ordinary (information) focus*. (Similar arguments concern the definition of *identificational* – or *exhaustive* – *focus,* and other sub-classes that have been proposed in the literature.) Doubts arise with regard to the benefit of such sub-classifications, which is why I will try to provide some clarifications. It should be noted that there are many accounts in the literature which plainly ignore the existence of a *contrastive* vs. *non-contrastive* divide, e.g. Rooth (1992) or Büring (2008). But even in approaches that do assume a distinct category of *contrastive focus*, opinions diverge of how to define *contrast* and whether it should be treated only as a subcategory of focus or as an optional add-on feature that also combines with topics / themes, cf. Vallduví & Vilkuna (1998).

A major part of the confusion relates to the unclear notion of *alternatives,* whose availability, on some accounts, is taken to be a defining criterion for *contrastive focus* (e.g. Selkirk 2008, Katz and Selkirk 2011) while on other accounts their presence is attributed to all sorts of *focus*, including the non-contrastive kind (novelty focus). The culprit for this unfortunate confusion is probably Rooth (1985, 1992, 1996), whose theory of *Alternative Semantics* contains all ingredients for a comprehensive understanding of focus but is unfortunately presented in a somewhat ambiguous manner. "[E]voking alternatives is the general function of focus." (Rooth 1996: 276). This dictum has often been misinterpreted. In fact, it is necessary to keep apart two notions of alternatives: firstly, sets of alternatives "evoked" by the F(ocus)-feature (so-called *focus-semantic values*) are simply defined in terms of the semantic type of the F-marked expression: every expression that has the same semantic type as the focused expression is automatically an alternative. Alternative sets of this kind are, naturally, quite big, and might be called "anonymous" (Riester and Kamp 2010) or "raw" (Büring 2013) alternative sets. For instance, the focus-semantic value of the noun *tree* consists of all other nouns contained in a speaker's lexicon. It is obvious that such a big

alternative set cannot receive an extensional characterization, i.e. we are simply unable to list all the members of this set, and it is perhaps confusing to think of an F-marked expression as a set at all, and not just, say, as a placeholder for expressions of the same semantic type.

The second type of alternatives could be described as *contextually salient alternatives*. Contextually salient alternatives come much closer to an intuitive and pragmatically relevant notion of alternatives, and they are precisely what we need in defining a notion of *contrast*. It is important to note that, in Rooth's system, contextually salient alternatives are not identified or evoked by focus itself but by means of a special anaphoric operator, written as ~ ("squiggle"), which attaches to a constituent that contains both a focus and some backgrounded material. This constituent is called a *focus domain*. Note that Rooth talks about the ~ operator as "restricting" the original focus-semantic value, but, again, this might be a misleading way of speaking, under the assumption that we usually do not know which elements are contained in the original focus-semantic value in the first place. Instead of "restriction", therefore, I prefer to talk about the "identification" of contextually salient alternatives: a focus domain is an anaphoric expression that wants to identify one or several alternatives in the current context (or, at least, a question antecedent). To be clear: without a ~ operator, there is no discourse interaction at all. The F-feature is not itself anaphoric; a focused constituent which is not embedded in a focus domain does not have any *specific* alternatives that can be named. It is merely an anonymous alternative set (and it can, therefore, only represent *new* but not *contrastive* information). This opens up a possibility to define a notion of *contrastive focus* (and, conversely, a notion of a *non-contrastive focus*): a *contrastive focus* is a focus whose alternatives can be unanimously identified in the discourse context (É. Kiss 1998: 267, Brunetti 2009, Riester and Baumann 2013: 233), while a *non-contrastive focus* is not – as it is often mistakenly put – a focus "without alternatives" but a focus whose alternatives simply remain unidentified and anonymous. In the analysis below, I will not distinguish contrastive from non-contrastive foci because I will only concentrate on a basic focus notion.

An argument that I must reject is the one that "focus is often poorly defined" (MW: 135). While this may hold true for parts of the literature, there is no reason why this woeful situation should persist nowadays. It has become entirely clear that focus must neither be defined in terms of *newness,* nor *contrast,* nor *exhaustivity* (and certainly not in vague terms like *importance, unexpectedness* etc.) All these notions describe focus-related but ultimately distinct phenomena. To define focus in terms of the *availability of alternatives* is not, per se, wrong but, as I have sketched above, typically gives rise to misunderstandings. The only clear definition of focus is that of *being the part of an assertion that answers an explicit or implicit Question under Discussion (QUD)* (e.g. Roberts 1996). Of course, this requires us to explain

what QUDs are in general, and how they can be determined. Monological text typically does not contain any overt QUDs at all. But also dialogue typically contains fewer questions than one would perhaps expect. Note that it may be the case that, occasionally, the implicit QUD in a conversation deviates from an overtly asked question. (Sometimes, people choose to answer a different specific question than the one that was explicitly asked.) To identify a QUD at every point in discourse can be controversial, and it requires a holistic interpretation of the discourse at hand.

Upshot: The discussion above shows, in my opinion, that research on information structure has to internalize two lessons: first, it looks as if there is little hope that we will ever identify a single expression, construction, syntactic position or pitch accent type within one language that unanimously serves as a marker of focus. From this, it follows, second, that the marking of focus, or any other information structural category, will always show a merely statistical distribution. Neither of this, however, justifies the negative conclusion that there can be no abstract interpretive concept of focus, at the outset. On the contrary, the whole situation, to my view, only increases the need for a precise interpretive definition of focus, in combination with a clear meaning-based procedure or recipe how to identify *tokens* of focus constituents in natural language data. Such a recipe is what I am going to try to provide in the next Section. In order to underscore the fact that the described procedure of analysis is indeed language-independent, cross-linguistically applicable, and inherently meaning-based, I will resort to the somewhat unusual experiment of analyzing a narrative text from a language in which I do not have any prior expertise: the Austronesian language Balinese (cf. Arka 2003). It goes without saying that this requires the availability of thorough linguistic glosses and a translation which is close to the original text.[1]

3. Annotating Questions under Discussion and information structure

In the following, I will make the assumption that discourse is not linear but hierarchically organized in the form of a discourse tree. This assumption is commonly found in theories of discourse structure (e.g. Mann and Thompson 1988, Asher and Lascarides 2003) and information structure (Roberts 1996, Büring 2003, Beaver and Clark 2008) but implementations differ. On the one hand, theories of discourse structure usually assume that

---

[1] The Balinese text in the appendix, *Bulan Kuning*, including glosses and translations, was kindly provided to me by Asako Shiohara on the occasion of the International Workshop on Information Structure of Austronesian languages at Tokyo University of Foreign Studies, February 2015. I have slightly changed the translations to bring them a bit closer to the Balinese original, and I

text is built from so-called *elementary discourse units* (roughly: clauses), which themselves represent the nodes of discourse trees. QUD-based theories of information structure typically assume that discourse trees are abstract objects structured by means of increasingly specific questions which are ordered by an entailment relation.

The discourse trees I have in mind combine elements from both discourse structure and theories of Questions under Discussion. The goal is to transform natural discourse into a tree whose non-terminal elements are questions and whose terminal elements are the assertions of the text, in their linear order, as shown in Figure 1.
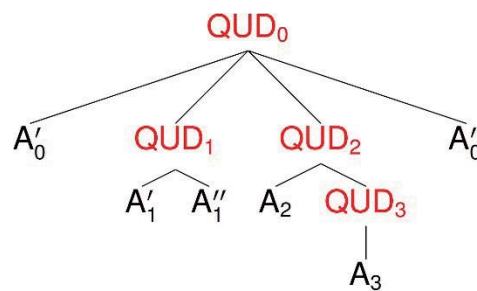
Figure 1 Discourse tree with Questions under Discussion

It is the task of the analyst to reconstruct the QUDs of the text, and, in the course of this, the geometry of the discourse tree. Well-written texts and clearly structured spoken discourses possess an accordingly clear discourse structure. In the following, I am going to sketch the necessary steps for an analysis procedure.

Step 1: Read the entire text carefully and make sure to understand what it is about and whether it makes sense. It is difficult or entirely impossible to analyze text which is incoherent or incomprehensible.

Step 2: Split the text into clauses at sentence-level conjunctions, i.e. isolate single assertions. Do not separate sentential arguments from their embedding verbs.

Step 3: Identify parallel structures, i.e. assertions which provide different partial answers to the same question. The goal is to identify as many parallelisms as possible, thus capturing a maximum of coherence in the text. For instance, in the abstract Figure 1, assertions A_1' and A_1'' have been identified as partial answers to the same question QUD_1. Partial answers to the same question need not be immediately adjacent. For instance, the first (A_0') and the last (A_0'') assertion in Figure 1 are both partial answers to QUD_0, although they are separated by intervening material. This material must elaborate on or

provide some background to A₀'. (1) is an example from our Balinese narrative. A sentence has been split into two clauses at the conjunction *tur*.

(1) a15':   Raksasa dadi   Betara

            Raksasa become   god

            Raksasa became a god.

    a15'':   tur   mawali   buin   ka   suarga.

            and   return   again   to   heaven

            and returned to heaven.

The clauses are parallel to the extent that they have the same subject or topic *Raksasa,* which is left implicit in (a15''). I will adopt the convention to label each clause with an *a* (for *answer* or *assertion*) and a number that matches the number of the respective question. A series of parallel answers to the same question, e.g. (q1), is marked by (a1'), (a1''), (a1''') etc.

Step 5: Formulate the respective QUD. This is not an arbitrary move. Most importantly, the QUD must be such that all the assertions below the question actually are congruent with it (i.e. they must indeed answer the question). In general, this means that the QUD can in principle target any constituent of the assertion. For instance, (a15') could be the answer to any of the following questions: *What happened? Who became a god? What happened to Raksasa? What did Raksasa become? Who became what?* However, the selection of the proper question is restricted as soon as we have several (partial) answers. In this case, the question must contain "the lowest common denominator" of the two partial answers, i.e. the semantically constant element contained in all available answers, while the alternating parts are replaced by a wh-phrase. For the little discourse above, this means that the question can only be (q15), as can be seen in Example (2).

(2) q15: {What happened to Raksasa?}

    > a15':   Raksasa   became a god

                [topic   ] [focus   ]

                [focus domain   ]

    > a15'':   and   (he)   returned to heaven.

                [topic]   [focus   ]

                [focus domain   ]

The constant material in both assertions is called the background. In case the background contains a referring entity, that entity is labeled *(aboutness) topic*, cf. Reinhart (1981), Krifka (2007). In (a15''), the topic is an empty category which can, but need not, be made explicit. (Note that not all backgrounds contain a topic, but all (non-contrastive) aboutness topics are backgrounded.) The constituents that provide the actual answers to the question (q15) – here, the two VPs – are assigned the label *focus*. The background (or topic) in combination with the focus together form the so-called *focus domain* (the phrases Rooth (1992) would mark by use of the ~ operator). The focus domains of the parallel answers must "match" each other as well as the question (Büring 2008), which simply means that they must share the same background. In (2), I have used the > symbols and indentation in order to represent the tree structure in Figure 2.
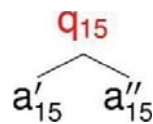
Figure 2 A Question under Discussion with two partial answers

A second constraint that puts a limit on the formulation of any QUD is that QUDs should always make reference to the immediately preceding discourse, i.e. a QUD (except at the beginning of a discourse) must contain given material. (If it doesn't, this means that the text is not very coherent, i.e. that the writer or speaker randomly switches to a completely different topic. Usually, this does not happen in rational discourse.) An example is shown in (3), and its abstract representation in Figure (3).

(3) q0: {What is the way things are?}

> a0: Ada   koné   anak   luh   balu   madan   Mén Bekung.
    exist hearsay person   female   widowed named   mother Bekung
    [focus   **]**
    It is told that there was a widow called Mén Bekung.

> q1: {What about Mén Bekung?}

>> a1: Ia   nongos   di   sisin   alas-é   gedé.
    3   live   at   side   forest-DEF big
    topic   **[focus**   **]**
    [focus domain   **]**
    She lived beside the big forest.

$$q_0$$

$$a_0 \quad q_1$$

$$a_1$$
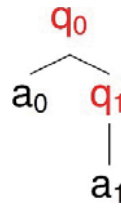
Figure 3 Subquestion (q1) seeking additional information about the previous discourse (a0)

Note that the first question (q0), the initial question of the discourse, is what Roberts (1996) calls the *Big Question*, the most general question that can be asked. The answer to this question is an all-focus sentence (a0), which is not embedded in a focus domain. (It entirely consists of new information.) Question (q1) is a *subquestion* to the pair (q0-a0), since it solicits additional information about what was said in (a0). Note that I deviate from Roberts (1996) to the extent that I do not require that subquestions, e.g. (q1), must be entailed by their super-question, (q0). This means that answers to subquestions can but need not be answers to a super-question. In fact, subquestions often trigger explanations or background information which do not, themselves, answer questions that are higher up in the tree.[2]

Step 6: Identify *not-at-issue* material. The parts of a clause that do not answer the current QUD can be grouped together under the notion of *not-at-issue content*, or *conventional implicatures* (Potts 2005, Simons et al. 2010). In general, this refers to optional information that does not contribute to the truth or falsity of the assertion. The two major types of not-at-issue material are, firstly, *supplements* such as parentheticals, non-restrictive modifiers and other appositive material like the reason clause in Example (4). The second type of not-at-issue content consists of *evidentials* – see Example (5) – and other speaker-oriented material, e.g. sentence-adverbs like *luckily*.

(4) q11: {What did Bulan Kuning learn about the magic stones?}

> a11:   Sawireh   ia   sayangang-a

            because   3   love-3.ACT

            Because she was loved,

            [not-at-issue   ]

---

[2] An issue that I will not address in this paper are the more complicated question-answer structures which arise in connection with *contrastive topics* (Büring 2003). While those are, in general, very important, I skip this issue here, since there seem to be no contrastive topics contained in the very simple narrative structure of our text.

orahin-a	ia	kagunan	manik-é	maketelu	ento
tell-3.ACT	3	use	jewel-DEF	three	that

topic

[focus   **]**

[focus domain   **]**

she was taught the use of those three magic stones.

(5) q2: {What happened to her on a particular day?}

   > a2: ka-kritaang jani ia   maan nuduk   anak   cerik   pusuh.

          PASS-tell   now 3   get   AV-find   child   small   small

[not-at-issue   ] topic   [focus   **]**

                   [focus domain   **]**

       It is now told that *she found a tiny little child.*

The two types of not-at-issue content have in common that they do not contribute to answering the QUD. In principle, it would be possible to change the QUD, thereby turning not-at-issue content into at-issue (focus) content. For instance, the QUD in (4) could be *{Why was she taught the use of the magic stones?}* (answer: *[Because she was loved]focus*), and the QUD in (5) could – perhaps – be formulated as *{How certain is it that she found a tiny little child?}* (answer: *[It is (merely) told]focus*) But, in fact, these QUDs seem very unlikely, since they do not make much sense in the larger context of the story. Table 1 gives an overview of our basic information-structure inventory.

Tag	Definition
Focus	The part of a clause that answers the current QUD
Focus domain	A piece of discourse which contains both a focus and some background, and which recurs elsewhere in the context differing with regard to the focus (possibly in the form of a constituent question)
Background	The non-focal part of a focus domain (that part which is already mentioned in the current QUD)
(Aboutness) topic	Referential entity in the background; ideally an element of a coreference chain
Not-at-issue	The part of a clause which provides optional information from with regard to the current QUD

Table 1 Inventory of information-structure labels

To conclude my brief introduction into QUD analysis, I should point out that narrative text, as the one provided in the appendix, is not the most obvious type of discourse to demonstrate the analysis of information structure since narratives are primarily structured on the temporal dimension. This is not to say that information structure is not important in this genre but its impact on text structure is much bigger in informative discourse such as news, interviews or articles in encyclopedias.

For similar reasons, the QUDs in narratives are often a bit monotonous (*What happened? What happened next?* etc.) The entire text below is divided into several more or less temporally separated sections (smaller discourse trees) rather than forming a single big unit.

4. Two additional levels of information-structure annotation: referential and lexical information status

As a final issue, complementing the QUD and focus analysis of the narrative, I will briefly sketch another aspect of information-structural analysis: information status. Since I (and my co-author) have extensively written about information status (and the *RefLex* annotation scheme) elsewhere (Baumann and Riester 2012, 2013; Riester and Baumann 2013), I will only give a very rough overview here. Information status (Prince 1981, 1992) describes the classification of linguistic expressions according to their degree of cognitive activation. Like in the case of focus, I assume that information status categories are definable as abstract interpretive categories, which can be utilized cross-linguistically. It is possible to distinguish two levels of information status, a referential level (Table 2) that describes the cognitive status of referring expressions, and a lexical level that classifies content words in terms of possible semantic relations to earlier content words (Table 3). As for the referential level, I take it that an important criterion for distinguishing referring expressions in a text is whether they are *uniquely identifiable* with respect to a certain kind of context or whether they are not unique. Unique identifiability is a property associated with *definiteness* already since Frege (1892). (For a comprehensive account and history of definiteness, see Elbourne 2013). However, markers of definiteness suffer the same problem that Matić and Wedgwood (2012) have diagnosed for focus particles: they are language-specific and they are typically not strictly defined in abstract interpretive terms. In Table 3, therefore, no reference is made to definiteness. Instead, referring expressions are distinguished according the context classes with respect to which they are unique. All further details of the *RefLex* scheme can be found in the annotation guidelines (Riester and Baumann, in prep.)

Tag	Description	Uniqueness and Context
r-given-sit	Symbolic deixis	Entity unique in text-external context
r-environment	Gestural deixis	
r-given	Coreference anaphora	Entity unique in previous discourse context
r-given-displaced	Coreference anaphora with remote antecedent (> 5 clauses)	
r-cataphor	Cataphora	Entity unique in upcoming discourse context
r-bridging	Bridging / associative anaphora	Entity unique in previous frame context
r-bridging-contained	Bridging anaphor with embedded antecedent	Entity unique in global context
r-unused-unknown	Hearer-unknown, discourse-new	
r-unused-known	Hearer-known, discourse-new	
r-new	Discourse-new item	Non-unique entity
Optional flags		
+generic	Generic or non-specific item	
+predicate	Used in predicative construction	

Table 2 Referential information status. Annotation units: referring expressions

Tag	Description	Cognitive status
l-given-same	Repetition	active, i.e. salient concepts
l-given-syn	Synonym of previous item	
l-given-super	Hypernym of previous item	
l-given-whole	Holonym of previous item	
l-accessible-sub	Hyponym of previous item	semi-active, i.e. derivable concepts
l-accessible-part	Meronym of previous item	
l-accessible-stem	Part of the word has occured previously	
l-new	Unrelated within last 5 clauses	inactive concepts

Table 3 Lexical information status. Annotation units: content words

Appendix: Analysis of a Balinese narrative

Bulan Kuning

q0: {(Part 1:) What is the way things are?}

> a0:	Ada koné		anak	luh	balu	madan	Mén Bekung.	
	exist hearsay		person	female	widowed	named	mother Bekung	
			l-new	**l-new**	**l-acc-sub**	**l-new**	**[l-new**	**]**
			[r-new					**]**
		[not-at-issue]						
	[focus							**]**

It is told that there was a widow called Mén Bekung.

> q1: {What about Mén Bekung?}

>> a1':	Ia	nongos	di	sisin	alas-é	gedé.	
	3	live	at	side	forest-DEF	big	
		l-new		**l-new**	**l-new**	**l-new**	
	r-given				**[r-unused-known]**		
			[r-bridging-contained			**]**	
	topic	**[focus**				**]**	
	[focus domain					**]**	

She lived beside the big forest.

>> a1'':	Gegina-né	sai-sai	ngalih	saang	ka	alas-é.
	job-3.POSS	usually	AV-look for	firewood	to	forest-DEF
	l-new		**l-new**	**l-acc-part**		**l-given-same**
	(r-given)					
	r-bridging-cont			**r-new+generic**	**[r-given**	**]**
	(topic)					
	[focus					**]**
	[focus domain					**]**

Her work was to go to the forest looking for firewood.

> q2: {What happened to her on a particular day?}

>> a2:

ka-critaang	jani	ia	maan	nuduk	anak	cerik	pusuh.
PASS-tell	now	3	get	AV-find	child	small	small
l-new			l-new	l-new	l-acc-sub	l-new	l-given-syn
	r-given				[r-new		]
[not-at-issue]	topic		[focus				]
[focus domain							]

It is now told that she found a tiny little child.

>> q3: {What happened to the child?}

>>> a3:

Sawireh	kulit-né		putih
because	skin-3.POSS		white
l-acc-part	l-new		l-acc-part
	(r-given)		
	r-bridging-contained		
[not-at-issue ...			

tur	mua-né		bunter	buka	bulan-é
and	face-3.POSS		round	like	moon-DEF
	l-new		l-new		l-new
	(r-given)				
	r-bridging-contained		[r-unused-known]		
			... not at-issue]		

lantas	ka-adanin	Ni	Bulan Kuning.
then	PASS-name	Ms.	yellow moon *(idiom)*
	l-new	l-new	[l-new]
		[r-unused-unknown+predicate]	
[focus			]

Because her skin was white and her face was round like the moon, she was given the name
Bulan Kuning.

q4: {(Part 2:) What happened when Bulan Kuning was older?}

> a4:	Sedek	dina	anu	Ni Bulan Kuning	ajaka	bareng
	at	day	unknown	Ms. Bulan Kuning	with	together
		l-new	l-new	[l-given-same]		l-new
	[r-new		]	[r-given]		
				[topic]		
	[background			] [focus ...		
	[focus domain ...					

	ngalih	saang	ka	tengah	alas-é.
	AV-look for	firewood	to	center	wood-DEF
	l-given-same	l-given-same		l-new	l-given-same
					r-given-same
		r-given+generic	[r-bridging-contained		]
					... focus]
					... focus domain]

One day, Ni Bulan Kuning joined in searching for firewood in the middle of the forest.

> q5: {What happened to her in the middle of the forest?}

> > a5':	Di	tengah	alas-é	Ni Bulan Kuning	paling
	at	center	forest-DEF	Ms. Bulan Kuning	get lost
		l-given-same	l-given-same	[l-given-same]	[l-new]
			r-given		
	[r-given		]	[r-given]	
				[topic]	
	[background			] [focus]	
	[focus domain				]

In the middle of the forest Bulan Kuning got lost

> > a5'':	tur	palas	ngajak	meme-n-né.
	and	separate	with	mother-INS-3
		l-new		l-new
			(r-given)	
		[r-given-displaced		]
		[focus		]
		[focus domain		]

and was separated from her mother.

> > q6: {How did she feel about being lost?}

> > > a6: Ni Bulan Kuning   jejeh   pesan   atin-né.

  Ms. Bulang Kuning   scared   much   heart-3 *(idiom)*

  [l-given-same   **] l-new**   **l-acc-part**

                                                         (r-given)

  [r-given   **]**   **[r-bridging-contained]**

  [topic   **] [focus**   **]**

  [focus domain   **]**

  Bulan Kuning was very scared.

> > a5''': Mara   ia   inget

  when   3   become concious

                         [l-new   **]**

              r-given

  [not-at-issue   **]**

  dapetang-a,   ibane   suba   saup-a   tekén   I Raksasa

  find-3.ACT   oneself   already   take-3.ACT by   Mr. Giant

  l-new   **l-new**   **[l-new**   **]**

              r-given   **[r-unused-unknown]**

              topic

  [focus   **]**

  [focus domain   **]**

  When she became conscious again she found herself held by Raksasa (the Giant).

> > q7: {How did she feel about this?}

> > > a7: Ia   ngetor   baan   jejeh-ne

  3   tremble   because of   fear-3.POSS

              l-new   **l-acc-stem**

                                             (r-given)

  r-given   **[r-bridging-contained+generic**   **]**

  [topic]   **[focus**   **]**

  [focus domain   **]**

  She trembled because she was scared.

q8: {(Part 3:) How did her relationship with Raksasa continue?}

> a8':	Ia	sayangang-a	pesan	baan	I Raksasa.
	3	love-3ACT	much	by	Mr. Giant
	l-new				**[l-given-same]**
	r-given			**[r-given**	**]**
	topic	**[focus**			**]**
	[focus domain				**]**

She was very much loved by Raksasa.

> q9: {What about Raksasa?}

> > a9:	I Raksasa	ngelah	manik	sakti	telung	besik
	Mr. Giant	AV-have	jewel	magic	three	item
	[l-given-same]		**l-new**	**l-new**		**l-given-super**
	[r-given	**]**	**[r-new**			**]**
	[topic	**]**	**[focus**			**]**
	[focus domain					**]**

Raksasa had three manik (jewels, magic stones),

> > q10: {What kind of stones?}

> > > a10':	luire:	manik	api,
	that is:	jewel	fire
		l-given-same	**l-new**
		[r-new	**]**
			focus
		[focus domain	**]**

that is, a fire jewel,

> > > a10'':	manik	yeh,
	jewel	water
	l-given-same	**l-new**
	[r-new	**]**
		focus
	[focus domain	**]**

a water jewel,

\> \> \> a10''':   manik   angin.

jewel   wind

l-given-same   **l-new**

[r-new   **]**

focus

[focus domain   **]**

and a wind jewel.

\> \> q11: {What did Bulan Kuning learn about the magic stones?}

\> \> \> a11: Sawireh   ia   sayangang-a,

because   3   love-3.ACT

l-given-same

r-given

[not-at-issue   **]**

orahin-a   ia   kagunan   manik-é   maketetelu   ento.

tell-3.ACT 3   use   jewel-DEF   three   that

l-new   **l-new**   **l-give-same**

r-given   **[r-given**   **]**

[r-bridging-contained   **]**

topic

[focus   **]**

[focus domain   **]**

Because she was loved, she was taught the use of those three magic stones.

q12: {(Part 4: What happened with the stones one day when Raksasa was away?}

> a12:

Katuju	I Raksasa	luas,
when	Mr. Giant	go out
	[l-given-same]	l-new
	[r-given]	

[background   ]
[focus domain ...

lantas	plaibang-a	manik-é	makejang	tekén	I Bulan Kuning.
then	run with-3.ACT	jewel-DEF	all	by	Ms. Bulan Kuning
l-new	l-given-same	[l-given-same]			
		[r-given] [r-given]			
		topic			

[focus   ]
                                        ... focus domain]

When Raksasa happened to be out, then all the magic stones were taken away by Bulan Kuning.

> q13: {How did Raksasa react?}

> > a13':

Saget	teka	I Raksasa
unexpectedly	come	Mr. Giant
	l-new	[l-given-same]
		[r-given]

[not-at-issue]   focus   [topic   ]
                 [focus domain   ]

Unexpectedly, Raksasa returned.

> > a13'':

sahasa	nguber	Ni Bulan Kuning.
fiercely	AV-chase	Ms. Bulan Kuning
l-new	l-new	[l-given-same]
		[r-given]

[focus   ]
[focus domain   ]

He fiercely chased Bulan Kuning.

> > a13''':   Makir-é   Ni Bulan Kuning   bakatanga   tekén   I Raksasa

when   Ms.Bulan Kuning   catch-3.ACT   by   Mr. Giant

[l-given-same   ]   l-new   [l-given-same]

[r-given   ]   [r-given   ]

[not-at-issue   ]

lantas   sabat-a   baan   manik-é   ento

then   throw-3.ACT   with   stone-def   that

l-new   l-giv-same

[r-given   ]

[focus ...

[focus domain ...

kanti   I Raksasa   mati

until   Mr. Giant   die

[l-given-same]   l-new

[r-given   ]

... focus]

... focus domain]

When Bulan Kuning was nearly caught by Raksasa, he got these stones thrown at himself until Raksasa died.

> q14: {What happened to Bulan Kuning?}

> > a14:   Ni Bulan Kuning   tengkejut   ningeh   sabda

Ms. Bulan Kuning   surprised   hear   voice

[l-given-same   ]   l-new   l-new   l-new

[r-given   ]   r-new

[not-at-issue   ]

sawiréh   Ni Bulan Kuning   nyupat   I Raksasa.

because   Ms. Bulan Kuning   AV-purify   Mr. Giant

[l-given-same   ]   l-new   [l-given-same]

[r-given   ]   [r-given   ]

[topic   ]   [focus   ]

[focus domain   ]

Bulan Kuning was surprised to hear a voice saying that Bulan Kuning had purified Raksasa.

> q15: {What happened to Raksasa?}

>> a15':

Raksasa	dadi	Betara
Raksana	become	god
l-given-same	**l-new**	**l-new**
r-given		**r-new+predicate**
[topic]	**[focus**	**]**
[focus domain		**]**

Raksasa became a god

>> a15'':

tur	mawali	buin	ka	suarga.
and	return	again	to	heaven
	l-new			**l-new**
			[r-unused-known]	
	[focus			**]**
	[focus domain			**]**

and returned to heaven.

q16: {(Part 5:) What happened to Bulan Kuning, in the end?}

> a16':

Ni Bulan Kuning	kaicén	kasaktian	maubad-ubadan.
Ms. Bulan Kuning	PASS-give	power	treat
[l-given-same]	**l-new**	**l-new**	**l-new**
[r-given]		**[r-unused-unknown+generic]**	
[topic]	**[focus**		**]**
[focus domain			**]**

Bulan Kuning was given the power to treat people.

> a16'':

Jani	Ni BulanKuning	mulih
now	Ms. Bulan Kuning	go home
	[l-given-same]	**[l-new]**
	[r-given]	
	[topic]	**[focus]**
	[focus domain	**]**

Then she went home

> a16''':   lantas   ngalih   Mémén Bekung.

         then   AV-look for   mother Bekung

         l-new   **[l-new   ]**

                        [r-given-displaced]

         [focus   ]

         [focus domain   ]

         and saw Mémen Bekung.

3: 3rd person, ACT: actor, AV: actor voice, DEF: definite, PASS: passive voice, POSS: possessor

References

Arka, I. Wayan (2003). *Balinese Morphosyntax: A Lexical-Functional Approach.* Canberra: Pacific Linguistics.

Asher, Nicholas and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Baumann, Stefan and Arndt Riester (2012). Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In: G. Elordieta & P. Prieto (eds.) *Prosody and Meaning.* Berlin: Mouton de Gruyter, pp. 119-161.

Baumann, Stefan and Arndt Riester (2013). Coreference, lexical givenness and prosody in German. *Lingua* 136: 16-37.

Beaver, David and Brady Clark (2008). *Sense and Sensitivity: How Focus Determines Meaning.* (Explorations in Semantics 5). Oxford: Wiley-Blackwell.

Beaver, David and Dan Velleman (2011). The communicative significance of primary and secondary accents. *Lingua* 121: 1671-1692.

Brunetti, Lisa (2009). On links and tails in Italian. *Lingua* 119: 756-781.

Büring, Daniel (2003). On D-Trees, beans, and B-Accents. *Linguistics & Philosophy* 26(5): 511-545.

Büring, Daniel (2008). What's new (and what's given) in the theory of focus? In: *Proceedings of the 34th Annual Meeting of the Berkeley Linguistics Society*, pp. 403-424.

Büring, Daniel (2013). Givenness, contrast, and topic. Talk presented at *DGfS*, Potsdam, Germany.

Elbourne, Paul (2013). *Definite descriptions* (Oxford Studies in Semantics and Pragmatics 1). Oxford University Press.

É. Kiss, Katalin (1998). Identificational focus vs. information focus. *Language* 74: 245-273.

Frege, Gottlob (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100:25-50.

Katz, Jonah and Elisabeth Selkirk (2011). Contrastive focus vs. Discourse-new: Evidence from Phonetic Prominence in English. *Language* 87(4): 771-816.

Krifka, Manfred (2007). Basic notions of information structure. In C. Féry and M. Krifka (eds.) *Interdisciplinary Studies on Information Structure* 6. Potsdam, pp. 13-56.

Mann, William and Sandra Thompson (1988). Rhetorical Structure Theory: Towards a functional theory of text organization. *Text* 8: 243-281.

Matić, Dejan and Daniel Wedgwood (2013). The meanings of focus: the significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(1): 17-163.

Potts, Christopher (2005). *The Logic of Conventional Implicatures*. Oxford University Press.

Prince, Ellen (1981). Towards a taxonomy of given-new information. In: P. Cole (ed.) *Radical Pragmatics* (Syntax and Semantics 14). New York: Academic Press, pp. 223-256.

Prince, Ellen (1992). The ZPG letter: Subjects, definiteness and information status. In W. Mann and S. Thompson (eds.) *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. Amsterdam: Benjamins, pp. 295-325.

Reinhart, Tanya (1981). Pramatics and linguistics: An analysis of sentence topics in pragmatics and philosophy I. *Philosophica anc Studia Philosophica Gandensia Gent* 27(1):53-94.

Riester, Arndt and Stefan Baumann (2013). Focus triggers and focus types from a corpus perspective. *Dialogue and discourse* 4(2): 215-248.

Riester, Arndt and Stefan Baumann (in prep.) RefLex scheme – Annotation guidelines. URL: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/arndt/

Riester, Arndt and Hans Kamp. Squiggly issues: Alternative sets, complex DPs and intensionality. M. Aloni et al. (eds.) *Logic, Language and Meaning. Revised Selected Papers from the 17th Amsterdam Colloquium*. Berlin: Springer, pp. 374-383.

Roberts, Craige (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics* 49. The Ohio State University.

Rooth, Mats (1985). Association with Focus. Doctoral Dissertation, University of Massachusetts, Amherst.

Rooth, Mats (1992). A theory of focus interpretation. *Natural Language Semantics* 1(1): 75-116.

Rooth, Mats (1996). Focus. In: S. Lappin (ed.) *The Handbook of Contemporary Semantic Theory*. London: Blackwell, pp. 271-297.

Selkirk, Elisabeth (2008). Contrastive focus, givenness and the unmarked status of 'discourse-new'. *Acta Linguistics Hungarica* 55: 331-346.

Simons, Mandy, Judith Tonhauser, David Beaver and Craige Roberts (2010). What projects and why. In: *Proceedings of SALT 20*, pp. 309-327. Vancouver.

Skopeteas, Stavros and Gisbert Fanselow (2010). Focus types and argument asymmetries: A cross-linguistic study in language production. In: C. Breul and E. Göbbel (eds.) *Comparative and Contrastive Studies of Information Structure* (Linguistics Today 165).

Amsterdam: Benjamins, pp. 169-197.

Vallduví, Enric and Maria Vilkuna (1998). On rheme and kontrast. In: P. Culicover and L. McNally (eds.) *The limits of syntax* (Syntax and Semantics 29). New York: Academic Press, pp. 19-106.

Discourse and information structure in Kadorih

Kazuya INAGAKI

Visiting Research Fellow
(Kobe University of Foreign Studies)

kazuyainagaki@gmail.com

1. Overview

In Inagaki (2014), I introduced prosodic features and constructions signaling information structure in Kadorih such as topic-comment, entity-introducing, event-reporting and identificational (or cleft) constructions. In this paper, I will point out that the description of prosody such as pause, pitch, length and loudness is important for analyzing information structure of Kadorih. Among these prosodic features, pause and length have not been an issue drawing sufficient attention in previous studies on information structure. This paper will describe the role these prosodic features play in discourse and demonstrate their importance for interpreting information flow, boundary or structure. I will show information structure of Kadorih within the framework proposed by Chafe (1976), Prince (1981), Lambrecht (1994) and others.

As Chafe (1979) pointed out, pause and other prosodic features function as delimiters in a discourse. In addition, as described in numerous studies, including Inagaki (2014), prosodic features crucially relate to understanding of a sentence in terms of information structure. These aspects in speech tend to be missed when analyzing a spoken text which has already been transcribed by a linguist. This might be because we believe it is more important to see distinctive formal or segmental units which associate with semantic meanings. However, non-distinctive properties, or even non-verbal properties may well be clues for an addressee to interpret the addressor's speech in terms of information structure. For example, eye movement directed to some entity in the deictic world may signal a topic of the following statement.

In the following sections two issues will be discussed. First, in section 2, distribution and behavior of prosodic features in discourse will be considered. Pause, length, and pitch will be analyzed from the viewpoint of acoustic phonetics. Subsequently, prosodically supported structures will also be analyzed. Second, section 3 will illustrate the role of these prosodic features in signaling information structure.

2. Discourse analysis

In this section, I will focus on prosody in Kadorih discourse. In the following subsections, pause and its surrounding prosodic features, length and pitch will be discussed.

2.1 Pause in discourse

2.1.1 Overview

The duration and number of pauses can vary considerably among specific discourse instances. This can be clearly seen if we compare the details of pauses in different texts. Table

1 summarizes the number of pauses per minute and the minimum, maximum and average pause duration in four discourse samples given by four Kadorih speakers.

Table 1: The number and duration of pauses in different texts

text	speaker	number/min.	duration(avg.)	duration(min.)	duration(max.)
A	female	25	816ms	115ms	3174ms
B	female	20	688ms	125ms	1155ms
C	male	19	728ms	157ms	2140ms
D	male	18	1703ms	134ms	8844ms

The texts selected here are all narrated stories belonging to the genres of folktale, daily happening, and procedure explanation.

The number of pauses per minute seems to vary from text to text, but the figures are not significantly different from each other. These figures (18 to 25 per minute) can be recalculated as 3 to 4 pauses per 10 seconds. And I believe this pausing rate is true for many fluent Kadorih speakers. If we count the number of pauses in a text uttered by a less fluent speaker, for example, we will find more than 6 pauses per 10 seconds, in other words, more than 36 pauses per minute. This is an unlikely situation for Kadorih native speakers. By contrast, we will have very few opportunities to find an extreme case of a speaker who can keep almost nonstop talking for many minutes at the rate of only 1 or 2 pauses per 10 seconds.

What is common to the pause duration in the four texts here is the minimum value. The speakers tend to use a short pause of 100 to 200 milliseconds. If a pause is produced for 70ms, it is too short to perceive. Generally speaking, silent intervals over 250ms are reliably pauses, and those under 50 or 60ms are not considered as pauses (Robb et. al. 2004, cf. Kendall 2009).

On the other hand, it should be noticed here that pauses ranging from 300ms to 700ms are most commonly used by all the above speakers. However, as can be seen in Table 1, the values of maximum duration of pauses differ from text to text, or speaker to speaker. That is why the average pause duration varies and ranges from 688ms to 1703ms here.

2.1.2 Pre-pause elements

This subsection introduces what kinds of elements occur before pause. More or less all segments, syllables, prefixes (probably with the exception of the infix <*an*>), words, phrases, clauses, or sentences can be followed by a pause. The list in (1) contains elements frequently occurring before pause, other than clause- or sentence-final elements.

(1) a. Fillers: *anu(i)* 'um' , *e* 'er', *he* 'yeah', *nah* 'well', *nng* 'hm', *o* 'er'
   b. Interjections: *(a)yu* 'Come on!', *ei* 'Hey!', *i* 'Hey!'
   c. Conjunctions: *jadi* 'so', *inonko* 'and then', *iyo* 'then', *rimai* 'the story goes that', *sehingga* 'so that', *tahpi* 'but', *turus* 'and then'
   d. Discourse connective phrases: *(amun) jadi orih* 'after having been done, insomuch that', *umbot orih* 'after having finished that'
   e. Approximation adverb: *kira-kira* 'roughly', *kurang-labih* 'more or less', *mungkin* 'possibly', *sekihtar* 'approximately'

Some other function words such as prepositions, classifier-like abstract nouns and the relativizer *(i)jo* are also frequently followed by a pause. The relativizer *(i)jo* precedes a pause

+ clause, prepositions precede a pause + word/phrase, and abstract nouns precede a pause + word/phrase as in (2)–(4). In this paper, pauses are transcribed by three dots (…).

(2)   Pre-pause element: relativizer

　　a. *tohtok orih tahkan dasar sarupih jo … pakakisung ah nai rih*
　　　'The cutting [of the stem post] from the bottom of a side plank <u>which</u> … is the highest part [of a boat when we turn the boat downside up]'

　　b. *orih naing gunai jo … palapah pisang naing arai tuh rih*
　　　'That's the function <u>which</u> … of this so-called "palapah pisang"'

(3)   Pre-pause element: preposition

　　a. *harun to masang aang anui rih … aang … tohun sarupih jo pakadiang ah rih*
　　　'Only after that we install [palapah pisang] on um … <u>in</u> … the midst of that top side plank'

　　b. *kira-kira himat diang isut umba … sambungan doro jo sarupih numur duo rih*
　　　'Roughly, it must be installed at little higher place <u>than</u> … the seam [of the No.3 side plank] with the No.2 side plank'

　　c. *amun io nokuh … booi salut oh …he jelas io anak Rungan*
　　　'If it (=river) goes <u>to</u> … the flow goes downriver … yeah it is for sure the tributary of Rungan river.'

(4)   Pre-pause element: classifier-like abstract noun

　　a. *kira-kira kakahpai sekihtar karo … ohpat senti ka ih*
　　　'Roughly, its (=beam installed on the bottom plank) thickness is about (as <u>much as</u>) … four centimeters as well, the thickness.'

　　b. *"peda borai to kumai" hion … Tempun Tajawun*
　　　' "We are sick of eating" <u>said</u> … Tempun Tajawun'

　　c. *holang … uut Jehoi umba uut … Rungan*
　　　'The <u>border [area]</u> between … a headstream of Jehoi and a <u>headstream</u> … of Rungan'

　　d. *amun io nyalut nokuh hila … booi*
　　　'If it (=river) flows to the <u>direction</u> … downriver'

2.2 Lengthening in discourse

This section describes a phonetic phenomenon of vowel/nasal lengthening, which is common in Kadorih discourse. It should be kept in mind that the duration of vowel/nasal depends on the speech rate, and that it can also vary among discourse samples.

The following examples in (5) are two extracts from the same story. In all examples below, lengthened segments are transcribed by capital letters (e.g. AA).

(5) a. *ihtOO anui hinOO nguting ah koi-koIIk sekihtAAr karo duo sentII kahajon bindang ah …*
　　'<u>We</u>, um, <u>again</u>, cut it (=galvanized iron) small (into pieces) with scissors, the size of a piece is <u>about</u> (as <u>much as</u>) two centimeters.'

　　b. *kira-kira kakahpai sekihtar karOO … ohpat senti ka ih … kakahpai rih …*
　　'Roughly, its thickness is <u>about</u> (as <u>much as</u>) four centimeters as well, the thickness.'

As can be seen in (5a), many word-final vowels may undergo lengthening at the same time in a single intonation unit. (5a) includes five lengthened vowels: *o* of *ihto* 'we (inclusive)', *o* of *hino* 'again', *i* of *koi-koik* 'small', *a* of *sekihtar* 'approximately', and *i* of *senti* 'centimeter'. The sequence *sekihtar + karo* 'approximately + quantity' in (5a) is a very useful expression for talking about quantities, and is also used in other places in the text. When we compare the same sequence used in other places, we will find both *sekihtAAr karo* as in (5a) and *sekihtar karOO* as in (5b).

The underlined elements in (5), *sekihtar*, *karo*, *ihto* and *hino* are frequently in texts explaining various procedures. We can easily draw comparisons of vowel length between them. The table given in (6) shows the result of the comparisons.

(6)   Vowel length

word	sample	in (5ab)	average	minimum	maximum
sekiht AA r	12	372ms	372ms	234ms	602ms
kar OO	11	358ms	469ms	208ms	729ms
iht OO	8	506ms	387ms	252ms	506ms
hin OO	16	421ms	317ms	194ms	469ms
sekiht a r	15	85ms	79ms	60ms	105ms
kar o	8	60ms	61ms	49ms	89ms
iht o	27		106ms	57ms	167ms
hin o	7		106ms	77ms	144ms

The data of lengthened vowels is shown in the top four rows under the header, and that of normal short vowels can be found in the bottom four rows. It is obvious that the lengthened and normal short vowels are categorically different in duration from each other. While the final short vowels are usually produced with the duration ranging from 60ms to 140ms, the lengthened vowels are usually within the range of 200ms to 600ms.

Even though not frequently, prefixes can undergo lengthening. Some examples are given in (7) below.

(7) a. *ohpat sentII ka.. ... e kAAmihpih ah anui nai ...*
        'four centimeters as well, … er, its thinness …'

    b. *io rih eam io ahkaNN karo sapating lah amun tAA ... pingas atawa iOO ... EE hakAAsatuk umba injah ah gi jorIIh hapa nahai amun jo ...*
        'It (=boat) won't get loose anymore, even if it is stomped on, or even if it, er, comes into contact with each other (=with another boat) when controlling, or even if …'

In (7a), the vowel of a prefix *ka-* is lengthened. This prefix morphologically derives an abstract noun by attaching to an adjective such as *mihpih* 'thin'. Similarly, prefixes *ta-* and *haka-* undergo lengthening in (7b). In these cases, lengthening probably applies at the first foot-final syllable of each word: (*ka*)(*mih.pih*) consists of two feet (*ka*) and (*mih.pih*), and the first foot-final syllable is *ka*), (*ta*)(*pi.ngas*) has (*ta*) + (*pi.ngas*) and the relevant syllable is *ta*), and (*ha.ka*)(*sa.tuk*) has (*ha.ka*) + (*sa.tuk*) and the relevant syllable is *ka*).

2.3 Roles of pause and lengthening

This section describes the roles played by pauses and lengthening in discourse. An example in (8) includes a chain of clauses and pauses extracted from the text of procedure explanation.

(8) a. *mahkUU nah* ...
   b. *amun io jadi uras ndai pahku AAh uras segAAh* ...
   c. *sarupih jo mapan tawang tuh rih nyamah tatukup nai uras umbot pahku AAh* ...
   d. *harun to anUU* ...
   e. *EE ngurah* ...
   f. *nguhca nyahtOOng* ...
   g. *ahkan hapa to malisan AAh* ...
   h. *lisan tuh rih* ...
   i. *sanyapur umba minyAAk galapung nyahtong rih nyapur umba minyAAk harun io manjadi lisAAn* ...

Translations of the respective extracts and durations of each pause are given in (9).

(9) a. to nail, well …(1390ms)
   b. when the nailing is completely done and all [of them] are stable …(407ms)
   c. [when] the nailing is completed for the side planks [which were jointed with] the bottom plank all the way up to the stem ...(1001ms)
   d. only after that we, um …(917ms)
   e. er look for …(856ms)
   f. pound resin …(1170ms)
   g. for us to use for caulking it (=boat) …(1046ms)
   h. that caulking glue …(1084ms)
   i. mixed with oil, the resin powder [we] mix it with oil, only after that it becomes caulking glue …(659ms)

There are two kinds of pauses, namely hesitation and juncture pause (Laver 1994: 537–539). The lines in (8a) and (8d) end, respectively, with fillers *nah* and *anUU* followed by a long pause respectively (1390ms and 1001ms). This kind of pause can be seen as hesitation pause produced after a hesitation filler. In addition, hesitation is also expressed by final lengthening on the hesitation fillers such as *anUU* in (8d) and *EE* in (8e).

On the other hand, the final pause in (8b), which is realized after a subordinate clause, can be seen as a juncture pause. This kind of pause is useful for a narrator to show a boundary between sentences, clauses, as well as subject–predicate, or topic–comment pairs. In fact, the relatively short pause (407ms) between (8b) and (8c) does not show a boundary between a subordinate and main clause. The clause in (8c) is only juxtaposed within the large subordinate clause which extends through (8b) and (8c). The clause in (8c) supplements a condition which is expressed in the previous clause in (8b), that is, 'when the nailing is done'. Interestingly, this supplementation by juxtaposition in Kadorih discourse often co-occurs with final lengthening. The shared topic argument *pahku AAh* 'the nailing' which is dislocated in (8b) and appended in (8c) plays a 'pivot' role for juxtaposing, and lengthened syllables demarcate the clauses obtained by means of juxtaposition. Of course, the intervening supplemental clause *uras segAAh* 'all are stable' is also a juxtaposed clause, and the boundary is also indicated by final lengthening.

This kind of clause combining by demarcative lengthening can be seen in (8i) too — the juxtaposition boundary is indicated by the final lengthening in *minyAAk* 'oil'. In (8h), the

pre-pause noun phrase *lisan tuh rih* 'that caulking glue' is the shared topic argument for the following statement, and the post-pause clauses *sanyapur umba minyAAk* 'mixed with oil' and *galapung nyahtong rih nyapur umba minyAAk* 'the resin powder [we] mix it with oil' in (8i) are the juxtaposed comments for that topic. In other words, the two clauses which convey new information are juxtaposed in (8i). On the other hand, in (8b–c), the clauses which convey information that is given (or accessible to the hearer) are juxtaposed. Therefore, lengthening in Kadorih is commonly observed in the cases involving hesitation and demarcation and it can be applied independently of information status of the units involved.

The final pause in (8h) can be seen as a juncture pause which demarcates the topic argument in (8h) from the comment parts in (8i). However, a certain hesitation may also be observed in the final pause in (8h) because of the semantically heavy content conveyed by the following comment parts. Generally speaking, hesitation pauses tend to be produced before a semantically heavy unit (Maclay and Osgood 1959). Thus, the functions of hesitation and juncture are not mutually exclusive, and a pause or lengthened segment may serve these two functions simultaneously.

2.4 Prosodic features and information flow in narratives

This subsection briefly considers the information flow by comparing prosodic features in folktale and non-folktale narratives.

In all kinds of narrative texts, pause and lengthening tend to be used more frequently in the earlier part of a text. These prosodic features which may be roughly associated with silence can signal that the speakers are relaxed and control their speech production, but they can also indicate that they are under pressure and feeling bewildered.

Fillers and hesitation pauses tend to be used more often in non-folktale texts. A folktale has fixed storyline and constructions. Therefore, what the narrator must do is to recall and produce those fixed structures, without any need to create new structures. If narrator's memory and production skills are strong enough, fillers and hesitation pauses will not be much used. On the other hand, other kinds of stories usually require creating new sentences and gathering threads of a story. Non-folktale storytellers will need more time for these mental processes, so that they tend to use fillers and hesitation pause many times.

Interestingly, sentence-internal boundary markings realized by lengthening and pitch contour alone, namely without juncture pause, are observed more frequently in non-folktale texts. When we listen to a non-folktale text in Kadorih, we will feel that many sentences are structured in a somewhat random fashion. It can be said that non-folktale storytellers tend to be less concerned with aesthetic differences between demarcations obtained by means of pause, lengthening and pitch. On the other hand, folktale narrators tend to be more concerned with this difference and they may more consciously know some sort of predominance of pause over lengthening and pitch. This can be another reason why information flows in folktales are usually smoother than those in non-folktales.

2.5 The importance of pitch

Non-folktale storytellers sometimes do not use either pause or lengthening to mark a boundary. When we listen to such non-paused and non-lengthened sequences, we will be a little bit confused. The extract shown in (10b) is a representative example.
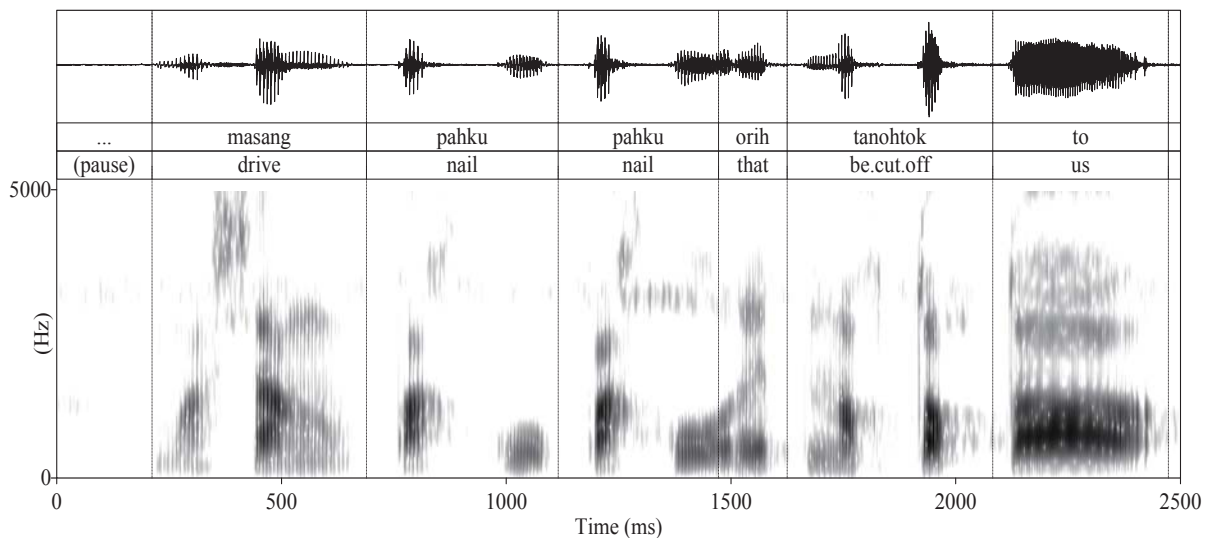
(10) a. *harun to masang AAh ...*
      b. *masang pahku <u>pahku orih tanohtok tOO</u> ...*
      c. *tanohtok ngindou to kuhung ah rih ...*
          'Only after that we set it (=No.2 side plank) up …
           drive nails <u>that nails were cut off by us</u> …
           their heads were cut off and thrown away by us …'

   In (10b), there are two clauses juxtaposed without any intervening pause or lengthened segment, *masang pahku* and *pahku orih tanohtok to*. In (10a), the storyteller states the preceding operation of boat-making, and begins to move on to the next topic, that is, the next operation in (10b), but in the middle, he suddenly jumps on to another relevant topic, that is, cutting off the nail heads. After lengthening the last vowel and pausing, he supplements the information about the nails in (10c).
   The annotated waveform and spectrogram of (10b) is shown in (11).

(11)   Annotated waveform and spectrogram of (10b)

	masang	pahku	pahku	orih	tanohtok	to
...	drive	nail	nail	that	be.cut.off	us
(pause)						

(Hz)

5000

0

0   500   1000   1500   2000   2500
                              Time (ms)

   As can be seen in the figure in (11), there is no salient pause or lengthened segment between the first *pahku* part (around 1000ms) and the second *pahku* part. From the waveform or spectrogram, we can see a short gap between the end of the first *pahku* and the beginning of the second *pahku*. However, this gap is just the silent period for the occlusion of *p* in the second *pahku*. And it should be noted that the gap between *masang* and *pahku* (105ms) and that between *pahku* and *pahku* (100ms) are almost identical in duration. That is why we can perceive the sequence of words *"masang | pahku | pahku"* as pronounced at the same interval of time.
   When we listen to the whole sequence in (10b), we can be strangely aware of the existence of the boundary between the first *pahku* and the second *pahku*. The acoustic cues to this perception is the pitch, specifically, falling pitch, rising pitch, and pitch uptrend (cf. Ladd et. al. 1985). The figure in (12) shows the acoustic correlates of these pitch contours.

(12)　　　Annotated spectrogram with the F0 contour superimposed
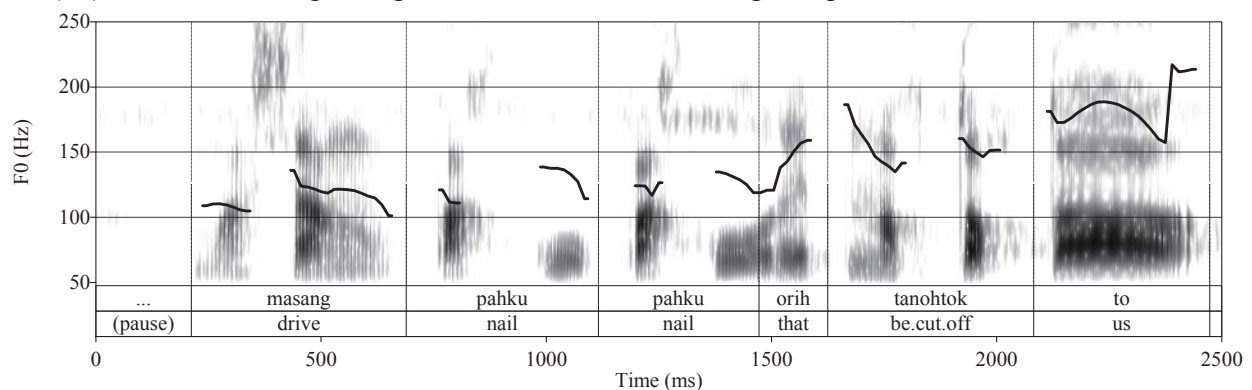
The figure in (12) has the same range of the sequence as the one given in (11), and the fundamental frequency of the sequence is superimposed on the spectrogram. The first *pahku* is pronounced with a slight falling pitch at the end. The correlated F0 contour slightly drops from 138Hz to 114Hz. And, the following topic argument *pahku orih* is pronounced with a rising pitch at the end, which correlates to the sharp F0 movement observed on *orih*, rising from 120Hz to 186Hz. Moreover, the latter part of the sequence, *pahku orih tanohtok to* is produced with a gradual rise which can be compared with the former part. These pitch contours can contribute to our perception of the boundary between *pahku* and *pahku*. Both falling and rising pitch are the normal indicators of a clause boundary in Kadorih, and pitch uptrend signals that the uptrend part is different enough from the rest of the sequence.

When a listener keeps up with a relatively faster pace of information flow in Kadorih, the more important acoustic cue may be the fundamental frequency (pitch) although gap (pause) and duration (length) are important as well.

3. Information structure and prosody

This section describes some useful prosodic features for understanding information structure of Kadorih, focusing mainly on pause.
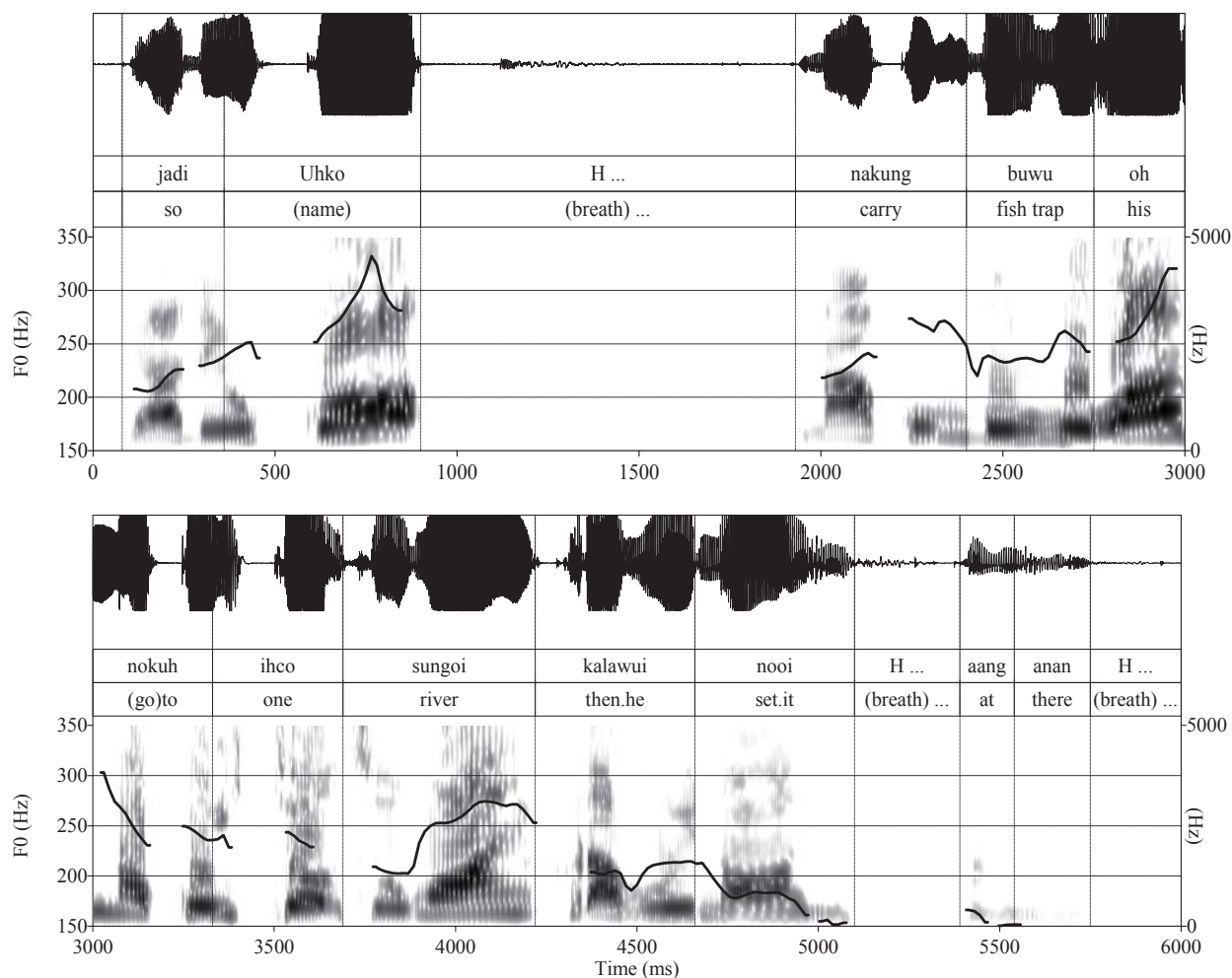
3.1 Breathy pause

Relatively long pause beginning with a breath can be observed for all speakers of Kadorih. This pause can be called "breathy pause", as opposed to silent pause which does not involve a breath. In Kadorih, it is a prototypical pause after a discourse connective or conjunction, a topic argument, or a whole/quoted sentence. (13a) and (13b) show such breathy pause after a whole sentence, and (13c) shows a breathy pause after a conjunction. In (13d), the first pause occurs after the quoted sentence, and the second one occurs after the whole utterance. A breathy pause is indicated by a capital *H* plus three dots in each extract.

(13) a. *tahpII kihtai ohcin naang aro H...*
 'But he saw there were many birds (breath) … (1077ms)'
 b. *iyo UhkOO noon buwu ah aang taruk kacu H...*
 'Then, Uhko set his fish trap on a tree (breath) …(2438ms)'
 c. *inonko H... (800ms) ihco ondou atuh Uhko ngindoi ah ...*
 'And then, (breath) … Uhko waited for it all day long'
 d. *"dinun ohcin naang aro poh Uhko" H... hion Mulau H...*
 ' "Uhko! you got so many birds" (breath) …(759ms) said Mulau (breath) … (1916ms)'

As shown in (13), a breathy pause contributes to the demarcation between cohesive information which is manifested in sentences. On the other hand, a breathy pause often occurs after a topic argument in a sentence. In such a case, this kind of pause usually serves as post-topic marker. The whole extract in (14) can be grammatically analyzed as a sentence, which is preceded by a conjunction. (14a) shows a conjunction and a topic argument *Uhko*, which are followed by a breathy long pause. In (14b), there are three clauses, that is, (a) 'carried his fish trap', (b) 'went to a river', and (c) '(then) he set it'. None of them is separated by a pause. The prepositional phrase in (14c) supplements the information where Uhko set his fish trap, that is, in the river. Only between (14b) and (14c), a breathy short pause is used.

(14) a. *jadi UhkOO H...*
     b. *nakung buwu oh nokuh ihco sungoi kalawui nooi H...*
     c. *aang anan H...*
     (One day, Uhko went setting his fish trap)
     'So, Uhko (breath) … (982ms)
      carried his fish trap went to a river and then he set it (breath) … (270ms)
     there (breath) … (753ms)'

The sentence in (14) is a topic-comment construction which has multiple comments at the same time. In other words, the topic *Uhko* is shared by the three comments expressed by the clauses in (14b). The information structure of (14) can be represented schematically as in (15) with respect to its presupposition and assertion.

(15)   Sentence: [[$_{TOP}$ *Uhko*] [$_{CMT}$ *nakung buwu oh*]
                                   [$_{CMT}$ *nokuh ihco sungoi*]
                                   [$_{CMT}$ *kalawui nooi aang anan*]]
        Pragmatic presupposition: Uhko is a topic for comment x
        Pragmatic assertion 1:   x = carried Uhko's fish trap
        Pragmatic assertion 2:   x = went to a river
        Pragmatic assertion 3:   x = set Uhko's fish trap there (=in the river)

The extract in (14) was uttered after *inonko Uhko ondou atuh tulak, nyalan noon buwu ah* 'One day, Uhko went setting his fish trap'. At the time of the utterance, hearers already know that (a) there is Uhko's fish trap, (b) Uhko goes somewhere, and (c) Uhko has a purpose to set his fish trap. Thus, going and setting actions were already given information. However, both actions can be new information because there are newly established relations: the relation between the going action and the exact place where Uhko went, and the relation between setting action and the exact place where Uhko set his fish trap. In addition, at the time of utterance, either fish trap, going action, or setting action is not limited as a single issue. Accordingly, none of them has been available as a sufficient and salient presupposition for the following assertion. Therefore, none of the clauses can be regarded as an identificational construction like 'it was a river where Uhko went' or 'it was a river where Uhko set his fish trap'.

The three comment clauses in (14b) do not seem to be demarcated. There is no intervening pause or prominent lengthening. However, as we already observed in 2.5, perceptible pitch plays a role for dividing (14b) into three comment clauses. Its acoustic analysis is shown in (16).

(16)   Annotated waveform and spectrogram of (14b)

We can find the first sharp rise in fundamental frequency, up to around 330Hz, in the column *Uhko*. The second, equally sharp rise can be found in the *oh* column at the very end of the upper figure. These two sharp rises have about 80Hz increase in F0. The last rise with such a large increase appears in the last syllable of *sungoi*, at around 4000ms. The first sharp rise is used for the topic argument *Uhko*, and the remaining two are main acoustic cues for demarcating the three comment clauses. In other words, the subsequent rising pitch may enable comment clauses to be easily linked to the foregoing topic even when the clauses are simply juxtaposed without any boundary pausing or lengthening.

From a grammatical point of view, the first two comments of the three, *nakung buwu oh* 'carry his fish trap' and *nokuh ihco sungoi* 'to a river' may be combined into a single clause through regarding *nokuh* as a preposition ('Uhko brought his fish trap *to* a river'). However, the verb *nakung* lexically means nothing more than 'carry something over one's shoulder' or just 'shoulder', so that it is inherently not a legitimate motion verb like *ngomin* 'bring'. Moreover, a translation of (14) into Indonesian given by a (bilingual) native speaker contained three independent clauses demarcated by two commas as in (17). Thus, it is more appropriate not to combine these two comment clauses.

(17) a. *Jadi Uhko memikul bubunya,*   'So Uhko carried his fish trap,
     b. *berjalan menuju sebuah sungai,*   walked toward a river,
     c. *lalu memasang bubu di situ.*   and set the fish trap there.'

The reason why the storyteller unnecessarily separates (14b) and (14c) by a pause may be as follows. This story will revolve around Uhko's strange behavior, setting a fish trap on a tree. Perhaps, the storyteller wanted to emphasize the key point where Uhko set his fish trap or emphasize that Uhko once did so in the right place, namely in a river.

3.2 Constrictive pause

A short pause may be a useful signal of information structure as well. (18) shows a relatively short pause beginning with a glottal constriction (glottal stop or creaky voice). This kind of pause can be called "constrictive pause", and it can serve as a pre-focus marker. The whole extract in (18) contains two such pauses in a stretch of clauses.

(18) a. *iyo Uhko buli ʔ...*
      b. *ngomin ohcin naang aro ʔ...*
      c. *ahkan Mulau H...*
          'Then, Uhko went back home (constriction) …(221ms)
           brought many birds (constriction) … (189ms)
           for Mulau (breath) … (1347ms)'

The whole example in (18) can be seen as a topic-comment construction which has two comments at the same time. The shared topic argument is again *Uhko*. The information structure of (18) is represented in (19) and the acoustic analysis of (18) is shown in (20).

(19)   Sentence: [[TOP *Uhko*] [CMT *buli*]
                                       [CMT *ngomin ohcin naang aro ahkan Mulau*]]
        Pragmatic presupposition: Uhko is a topic for comment x
        Pragmatic assertion 1:   x = went back (home)
        Pragmatic assertion 2:   x = brought many birds for Mulau

(20)   Annotated waveform and spectrogram of (18)
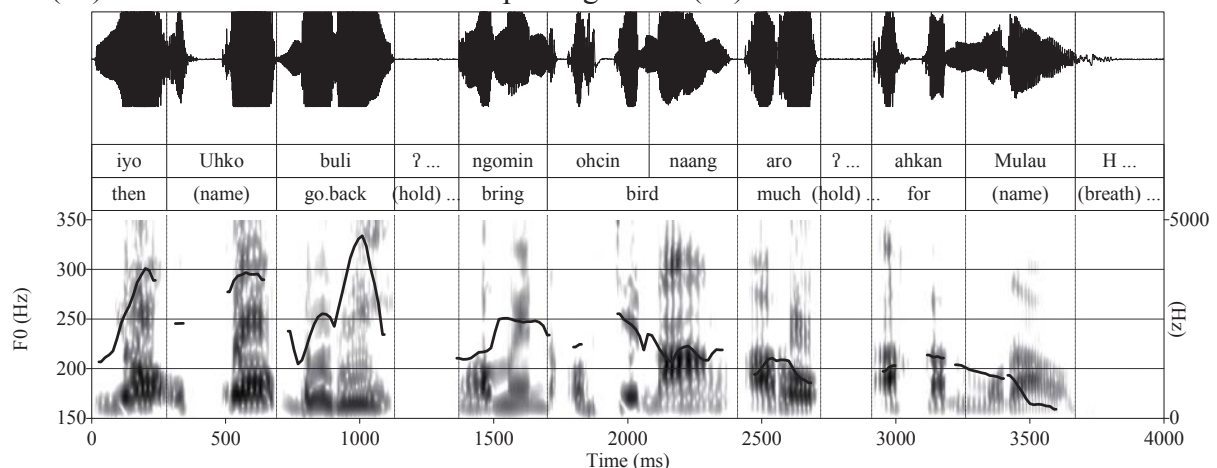
It can be said that a Kadorih storyteller uses a constrictive pause for a particular emphasis. After emphasizing the unlikely event of *(Uhko) ngomin ohcin naang aro* 'bringing many birds' in (18b) (although *Uhko* went setting his fish trap), the story progresses in a new direction. On the other hand, a new entity is also introduced by the prepositional phrase *ahkan Mulau*, who is Uhko's wife. The emphasis obtained by using constrictive pause helps addressees to figure out the new direction and new entity introduced here.

In addition to the constrictive pause, pitch and loudness may function as a means of emphasis. As can be seen from the sharp F0 movement of the first three words in (20), the frequency range of the first three words is greater than that of the remaining part, and the waveform and spectrogram of the first three words shows higher amplitude and darker or stronger intensity than that of the remaining part.

An extract in (21) shows another kind of emphasis. The quoted statements in (21) are uttered by Mulau although storyteller made slip of the tongue.

(21) a. *"[...] eam puji ulun noon buwu aang taruk kacu" hion Uhko ʔ...*
  b. *"aang danum" ʔ...*
  c. *aang sungoi" ʔ...*
   ' "[…] people have never set a fish trap on a tree" said [Mulau] (constriction) … (304ms)
    "in water" (constriction) … (220ms)
     in a river" (constriction) … (746ms)

In (21a), Mulau shares general knowledge about setting a fish trap with the bird-brained Uhko. At this point, the proposition 'people set a fish trap not on a tree' becomes available for shaping a pragmatic presupposition. The information structure of (21bc) can be represented as in (22).

(22) Sentence: *aang danum, aang sungoi*
    Pragmatic presupposition: People set a fish trap (not on a tree but) in x
    Pragmatic assertion 1: x = water
    Pragmatic assertion 2: x = river

The pragmatic assertions expressed by (21b) and (21c) are identificational ones, that is, (21bc) identifies where people set a fish trap. The addressee, Uhko, once set his fish trap appropriately in a river as shown in (14c). Thus, we can guess that Uhko knows where to set a fish trap and that Mulau can guess so. However, Mulau emphasizes the place where a fish trap must be set in order to warn Uhko not to act crazy.

(23) Annotated waveform and spectrogram of (21)
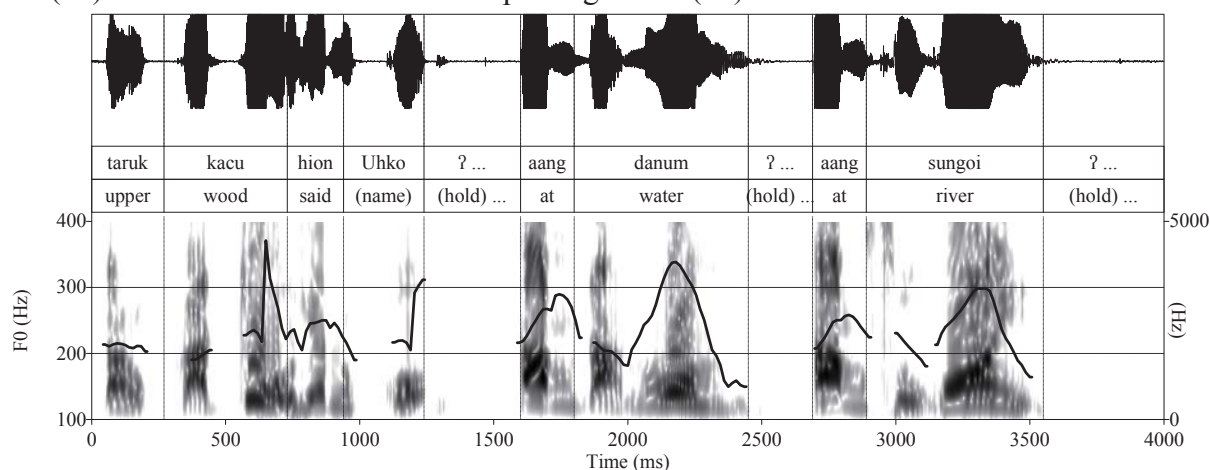
In (23) as well, it can be recognized that pitch (and possibly loudness) play a role to signall the focus. Fundamental frequency sharply rises and falls in the columns of *danum* 'water' and *sungoi* 'river', which are places focused by Mulau.

4. Conclusion

In this paper, the distribution and behavior in discourse of pauses and lengthened segments has been analyzed from the viewpoint of acoustic phonetics. Pausing and lengthening in Kadorih express hesitation and demarcation. Basically, they are used independently of information status of the respective units. Additionally, I considered the difference between prosodic features in folktale and non-folktale narratives with respect to the information flow. In particular, relatively faster pace of information flow in non-folktales utterances usually requires the addressee to be more sensitive to the fundamental frequency (pitch) than to gap (pause) or duration (length). In section 3, I described two conspicuous pauses in Kadorih, namely breathy and constrictive pauses. A breathy pause can be used as a marker indicating the boundary between cohesive information or as a post-topic marker, and a constrictive pause can be used as a marker emphasizing the turning point of information flow in a discourse or as a pre-focus marker. In addition to these pauses, other prosodic features, such as pitch or loudness may play instrumental role signaling information flow, boundary or structure.

References

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In *Subject and Topic*, ed. by Charles N. Li, chapter 2, 25–55. New York: Academic Press.

————. 1979. The flow of thought and the flow of language. In *Discourse and Syntax,* ed. by Givón, Talmy, volume 12 of *Syntax and Semantics*, 159–81. New York: Academic Press.

Kendall, Tyler S. 2009. Speech rate, pause, and sociolinguistic variation: An examination through the Sociolinguistic Archive and Analysis Project. Ph.D dissertation, Duke University.

Inagaki, Kazuya. 2014. Information structure in Kadorih. In *Proceedings of the International Workshop on Information Structure of Austronesian Languages*, 229–44. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.

Lambrecht, Kund. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*, volume 71 of *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.

Maclay, Howard and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous speech. *Word* 15: 19–44.

Ladd, D. Robert, Silverman, Kim E. A., Tolkmitt, Frank, Bergmann, Günther, Scherer, Klaus R. 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. *Journal of the Acoustical Society of America* 78(2): 435–444.

Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Prince, Ellen. 1981. Toward a taxonomy of given–new information. In *Radical Pragmatics*, ed. by Peter Cole, 223–255. New York: Academic Press.

Robb, Michael P., Margaret A. Maclagan, and Yang Chen. 2004. Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics and Phonetics* 18: 1–15.

Acknowledgement

Topic-Marking constructions in Bantik

Atsuko Utsumi

Meisei University

Abstract

This paper aims to show how a topic noun phrase (NP) is marked syntactically in the Bantik language. Constructions that introduce a discourse-new entity, cleft-sentence construction, and left-dislocation will be investigated. All the above constructions can mark an NP that is referred to or activated in the immediately preceding discourse. A cleft construction in Bantik places a contrastive topic NP in sentence-initial position, which is followed by the linker *nu*, which introduces the main clause. Left-dislocation in Bantik is defined as a construction where the left-dislocated NP is referred to again by a pronoun in the main clause. A new entity is often introduced by a topic-introducing construction that employs the existential marker *pai*.

1. Introduction

The Bantik Language[1] is an Austronesian language spoken in North Sulawesi, Indonesia. It is said to belong to the Sangiric subgroup within the Philippine group, Western Malayo-Polynesian (cf. Noorduyn 1991, Sneddon 1984 among others). It is said to be spoken by around 10,000 people in nine villages in the vicinity of Manado, a provincial city of the North Sulawesi, and two more villages around 100 kilometers away from Manado (cf. Noorduyn 1991). Every speaker of Bantik also speaks Manado dialect of Indonesian. As people born after 1970 mostly use the Manado dialect and people born after 1980 basically do not use Bantik, it is clearly in danger of extinction.

The syntactic features of Bantik are similar to those of other Philippine languages, and so are the topic-introducing constructions in the following sections, but the comparison between them is beyond the focus of this paper.

The example sentences are taken from elicitation, naturalistic data, and translation of folk tales in Indonesian to Bantik. Naturalistic data consist of folk stories "*Biou ni-timpunuu bo i-boheng* (The tale of a turtle and a monkey)" and "*Batu Madengkei* (Mandengkei stone)", and naturally occurring conversation "*Memperbaiki Rumah* (conversation about reconstucting a house)", and "*Hidupan* (life)".

Translated stories are: "*Biou ni-toadaʔ bo i-lummuutu* (The tale of Toada and Lumimuutu)*, "Kokokuk* (The tale of a kokokuk bird)", and "*Burung Taoun dan Burun Ngulngul* (The tale of Taon Bird and Ngulngul bird)".

---

[1] Bantik has five vowels /i, e, a, o, u/ and fourteen consonants /p, b, t, d, k, g, s, h, j, ɾ, m, n, ŋ, ʔ/. The glottal stop occurs only base-finally with a few exceptions. A word consists of a base, or a base with one or more affixes. The basic word order is SVO while VOS word order frequently occurs when the verb is in an Undergoer Voice. Like many other Philippine type languages, Bantik has more than one Undergoer Voices (at least two) in addition to an Active Voice. For a detailed description of Bantik, see Bawole 1993 and Utsumi 2005.

In the following discussion, several technical terms are used.   "Discourse topic" is used to refer to an entity that is important for the text and referred to repeatedly.

2. Referential statuses and linguistic expressions in Bantik

In this section, expressions that denote referential statuses of an NP will be briefly explained.   I will employ the framework of the referential hierarchy that is proposed in Gundel et al 1993, Gundel 2003, and Hedberg 2013, among others.   The description in (1) is an illustration an illustration of the givenness hierarchy given in Hedberg 2013, modified by the author, along with English expressions and the meaning of each cognitive status.   A linguistic item in the list can refer to an entity lower in the cognitive hierarchy, so *it* in English can also be used to refer to an activated entity.

(1) Cognitive status

Cognitive status	Meaning	English example
in focus	associate representation in focus of attention	*it*
activated	associate representation in working memory	*this/that/this NP*
familiar	associate representation in memory	*that NP*
uniquely identifiable	associate unique representation in DP	*the NP*
referential	associate unique representation	indefinite *this NP*
type identifiable	associate type representation	*a NP*

This universal givenness hierarchy can be applied to Bantik expressions as well.   Below (2) is a tentative table of Bantik expressions that relates to the givenness or referentiality that described in Utsumi 2014b.   *In focus* entities are normally expressed by zero-forms or connective pronominal forms.   If an entity is activated, the proximal or distal demonstratives follow the NP in question.   When it is familiar, the medial demonstrative follows the NP.   As for uniquely identifiable NPs, one of the following linguistic entities precedes them: an NP introducing entities *tou*/*side*, or the mirative demonstrative *ite/ete*. *Tou* is used for a singular NP, and *side* for a plural NP.   *Ite* is used to denote a discourse-new entity that is already present in the non-linguistic context.   A discourse-new entity that is not known to the addressee is sometimes introduced by a bare NP, but it is also introduced by *pai*, an existential marker, in other instances.

(2) Bantik expressions and referential statuses

in focus >	**activated** >	**familiar** >	**uniquely identifiable** >	**referential** >	**type identifiable**
Ø	*ie* (proximal)		*tou/side* NP		ØNP
pronouns		*ene* (medial)		*ite* (mirative proximal)	*pai* NP
(connective forms)	*eʔe* (distal)	full pronouns		*ete* (mirative distal)	

Both the existential marker *pai* and *tou/side* introduce a new entity into the discourse, but the

NPs they introduce differ in their information status.   In Section 3, usages of the existential marker *pai* will be described in comparison with usages of *tou/side,* in order to clarify their functions.

3. Expressions for introducing a new entity: *pai* and *tou*

A discourse-new entity can be entirely new to the addressee, but it can also be present in the non-linguistic context, or can be assumed from a certain linguistic entity in the preceding discourse.   If an entity is entirely new and cannot be assumed from the preceding discourse, the existential marker *pai* will be used to introduce it.   If, on the other hand, it can be uniquely identified from the non-linguistic context or assumed from the preceding discourse, *tou/side* will occur before it.   In this section, their similar but clearly different functions will be described.   *Tou/side* will be glossed as PRO, because they behave as a free variable pronoun.

3.1 Usages of *pai*

Bantik has an existential marker *pai*.   This marker has three main usages: first, it is used to denote that something exists, and in this case, it most often co-occurs with a PP that shows location, as in example (3).   Second, it is used to show possession as in example (4).   Third, it is used to denote a resultative state of the following clause, as in examples (5) and (6).

(3) *su   pandihiʔ   nu-ɾaodoʔ   su   bukidiʔ   wulur   maatus*

    LOC   near   LK-sea   LOC   hill   Wulur   Maatus

    pai   batu   kaɾaŋ.

    EXIST   stone   Karang

    'Near the sea, at the Wulur Maatus Hill, there was a stone (named) Karang.'   (*Biou ni Toadaʔ bo i-Lumimuutu*)

(4) *iaʔ   pai   sinage   bua   nu   Bali.*

    SUBJ.1sg   EXIST   friend   from   LK   Bali

    'I have a friend from Bali'. (Elicitation)

(5) *pai   [i-deki   na-idaoʔ=te   su   buha   e]*

    EXIST   SUBJ-Deki   NA-reach=COMP   LOC   Buha   DP

    'Deki has arrived at Buha.'

(6) *pai   i-stefi   ma-mokou   ɾakuʔ=ku.*

    EXIST SUBJ-Stevy   AV.NPST-wash   clothes=LK.1sg

    'Stevy has already (started) washing my clothes'

The topic introducing function of *pai* derives from the first usage.   A sentence with *pai* is often found in the first sentence of the text, as in examples (7) and (8).   The *pai* in example (3) can also be analyzed as having a topic-introducing function, which is used in the first sentence of the folk story.

(7) *pona　e　　**pai**　　i-opoʔ,　　　　aden=ne　　　i-opoʔ　　　　　roŋkoro.*

　　 before　DP　　exist　　SUBJ-old.man　　name=LK.3sg　SUBJ-old.man　　Rongkoro

　　 'Once upon a time, there was an old man named Grandpa Rongkoro.' (*Batu Madengkei*)

(8) *su　　rou　　pona　　**pai**　　side　　　　　dua　　siŋka-tuhaŋ.*

　　 LOC　day　　before　　EXIST　SUBJ.3pl　　　two　　one-sibling

　　 'Days and days ago, there were two brothers.' (*Kokokuk*)

3.2 *Tou/Side*: the marker that indicates an activated entity

　　　　We saw in the previous section that an NP that denotes an entirely new entity often follows the existential marker *pai*.　In contrast, when an entity has some connection with the previously emerged entity, *tou*[2] and *side*[3], which show the information status of "referential[4]" or "uniquely identifiable[5]", are attached.　*Tou* co-occurs with an NP that denotes a singular entity, both human and non-human, whereas *side* co-occurs with an NP that denotes plural entities.　Descriptions of *tou* and *side* are presented in Utsumi 2014.　*Tou/side* cannot mark an entity which cannot be assumed from the previous context, but neither can it be used to mark an entity that is fully "activated" or "in focus" in the cognitive hierarchy. An entity "in focus" or "activated" is referred to by a pronoun, the proximal demonstrative, or the distal demonstrative.

　　　　Tou/side can be regarded as a pronominal entity that precedes nouns, adjectives, and verbs, to form a nominal phrase.　Especially, a noun that denotes a human being cannot stand alone but should take *tou/side* as shown in (9).　In example (10), *tou* precedes an adjective, whereas in (11), *side* precedes a verb.　*Tou/side* in the two examples forms nominal phrases.

(9) *i-tou　　　　mahuanei　　　ene　　　　　　ma-turau　　si　　sie.*

　　 SUBJ-PRO　male　　　　dem.medial　　AV.NPST-live　LOC　here

　　　 'That man lives here' (Elicitation)

(10) *isie　　　　ma-ruan　　si-tou　　　ma-pedekeʔ?*

　　 SUBJ.3sg　AV.NPST-buy　OBJ-Prou　ADJVZ-short

　　　 'S/he will buy a short one' (Elicitation)

(11) *su　seʔe　　　　ma-ka-serei　　si-side　　　mam-bere*

　　 Loc　there.distal　AV.NPST-POT-see　OBJ-PRO　　AV.NPST-work

---

[2]　*Tou* is supposed to derive from *toumata* meaning "human being" in Bantik. The reconstruction of "human being" in Proto-Sangiric languages (to which Bantik belongs) is **tau* (Sneddon 1984).

[3]　*Side* is originally a third person plulral pronoun.

[4]　The term "referential" is used here according to the definition in Hedberg 2013. A referential satisfies one of the following two conditions: (i) "It is mentioned subsequently in the discourse," and (ii) "it is evident from the context that the speaker intends to refer to some specific entity."

[5]　The definition of the term "uniquely identifiable" here is taken from Hedberg 2013. Something that is uniquely identifiable satisfies both of the following conditions: (i) "the referent form contains adequate descriptive/conceptual content to create a unique referent," and (ii) "a unique referent can be created via a "bridging inference" by associating with an already activated referent."

'There, (we) can see workers (*lit.* those who work)' (Elicitation)

It is to be noted that *tou/side* is optional for non-human NPs as shown in (12)a and (12)b.

(12)a. *manuʔ*   *ene*   *ma-ituŋ*

     that.medial   ADJVZ-black

  'That chicken is black.'

   b. *i-tou*   *manuʔ*   *ene*   *ma-ituŋ*

     SUBJ-PRO chicken   that.medial   ADJVZ-black

  'That chicken is black.'

Although a non-human entity that is new to the discourse can occur without *tou* or *pai*, this is not always the case.   Example (13) shows that a non-human entity that is salient in the non-linguistic context is introduced by *tou*.   The speaker comments on the usage of *tou* in example (13) that if the addressee is aware of the presence of the coconut tree (*pun m-baŋo*), then *tou* is preferred.   It seems a non-human entity that is new to the discourse but is known immediately by the addressee preferably co-occurs with *tou/side*.

(13) *i-**tou***   *pun m-baŋo*   *ma-raŋkasaʔ*   *apadeʔ=ku*

    SUBJ-PRO   tree LK-coconut ADJVZ-tall   belong=NI.1sg

  'The tall coconut tree belongs to me' (Elicitation)

In the naturalistic data, a discourse new entity may or may not co-occur with *tou/side*.   In example (14), *korano* "king" appears for the first time but is not accompanied by a marker: the existential marker *pai* does not appear either.   In contrast, when it appears for the second time as shown in example (15), the referential marker *tou* appears before it.   It seems that the presence of *tou/side* is not obligatory but preferred when the addressee is assumed to be able to identify the NP.   Example (16) shows the usage of *side* that is attached to an NP that denotes an activated entity in the immediately preceding sentence.

(14) *dadiŋihiʔ=te* *i-korano*   *ni-boheŋ*.

   listen=COMP   SUBJ-king   LK-monkey

  'Monke's king listend.'   (*I-timpunu bo i-boheng*: line 54)

(15) *na-maro=te*   *i-**tou***   *korano* *nu-boheŋ*   *kasiʔ*

    AV.PST-announce=COMP   SUBJ-PRO   king   LK-monkey   INT

    pa-dadiŋihiʔ, *'kiteʔ*   *siŋka-maya-n*   *ie,*   *ka-kanioʔ*   *bo*   *bagai.'*

    CAUS-listen   SUBJ.1pl.EXC   one-all-AN   this   RED-small   and   big

  'The king of monkeys announced and let (them) llisten (to him), "We are one people, including small ones and big ones."'   (*I-timpunu bo i-boheng*: line 62)

(16) *manuʔ*   *si-yopi*   *ni-tekos-an.*

chicken   LOC-Yoppy   PST-steal-GV

isie   *na-moagaʔ*   *si-side*   *ma-na-nekosoʔ*

SUBJ.3sg   AV.PST-beat   OBJ-PRO   MA-RED-steal

'A chicken at Yopi's house was stolen.   He beat the thieves.'   (Elicitation session)

To conclude, both the existential marker *pai* and the pronominal *tou/side* can mark an entity that occurs for the first time in the discourse.   The existential marker *pai* is used to introduce a brand new entity in the discourse whereas *tou/side* marks identifiable entities.   To describe them in the cognitive hierarchy terms, *pai* marks "type identifiable" entities and *tou/side* marks "referential" or "uniquely identifiable" entities.

4. Cleft constructions

In this section, the cleft construction and its function will be discussed.   The cleft construction in Bantik places an NP in sentence-initial position, which is followed by the linker *nu* that introduces the main clause.   Cleft constructions in the below examples are placed in brackets, as shown in example (17).

A clefted NP denotes an entity that is contrasted with another entity, and it is already a given entity in the discourse.   In the cognitive hierarchy, a clefted NP mostly denotes "activated" or "familiar" entities that are contrasted with another entity.   *Ene* "that" in example (17) indicates the hot season that is described in the immediately preceding clause.   Here, *ene* is covertly contrasted with the other season in the tropical area, i.e., the rainy season.   This is an instance of a contrastive topic[6] NP in a cleft position.

This cleft construction, as often is the case with other languages in the Philippines and Indonesia, is also used in content questions as in example (18).   Example (19) has the same construction but the linker *nu* introduces a relative clause.   ""

(17) *duŋkuɾu*   *nu-ene*   *maŋoɾou*   *bo*   *suaya*   *nu-ɾou,*

before   LK-that   hot.season   and   light   LK-day

[ene=te   *nu*   *paŋ-uɾiʔ ma-ihaʔ]*

that=COMP   LK   APP-say ADJVZ-hot

'Back then, (it was in a) hot season, and the sunshine, that is the one that could be said to be hot'.
(*Biou ni Toadaʔ bo i-Lumimuutu*)

(18) *[isai nu*   *na-mihei*   *doitiʔ*   *si-kau]*

who LK   AV.PST-give   money   OBJ-2sg

'Who gave you the money?' (Elicitation)

(19) *babaeheʔ*   *su-beo,*   *isie*   *ma-ka-deaʔ*   *[apa*   *nu*

6   I use the term "contrastive topic" as it appears in Lambrecht (1994 : 291).   Topics, including "contrastive topics" are outside the scope of negation.

reward	LOC-wild.pig	SUBJ.3pl	AV.NPST-POT-find	what	LK

ni-kan   n-toumata]

PST-eat   LK-people

'As the reward to the wild pig, he (=wild pig) can find what was eaten by people'

When more than two entities are compared and one of them is picked up, the cleft construction is used to mark the "contrastive focus".   A contrastive focus excludes the possibility of the other option. The fronted NP, *anaʔ* "child" in (20), is previously mentioned, and was picked up in those sentences again to be contrasted with *iaʔ* "I" in the previous sentence.   In this sentence, *anaʔ* are the ones who support the speaker's life, since *iaʔ*, the speaker herself, is not able to do so anymore.   The sentences in example (21) are taken from a free conversation between four people.   They have been neighbors for a long time and share knowledge of the village people.   When they talk about a possible candidate for a folk-story teller, several elderly people come to mind.   The people who are compared become the contrastive foci of the last two sentences.   Words in angled brackets in the below examples are loanwords from Indonesian.

(20) *<ʒadi>   iaʔ   ie   gagudaŋ=te,*

    then   SUBJ.1sg   now   old=COMP

    <ʒadi> <hidop>=ku   ie,   doŋka   [anaʔ   nu   ma-miahaʔ   e]

    then   life=LINK.1sg   this   then   child   LK   AV.NPST-feed   DP

    'Then I got old, so my life is (like) this, and now it is my children that support (me).'   (*Hidupan,* monologue)

(21)L: *ada   isie   man-deaʔ   ma-biou   e,   i-maʔ*

    if   SUBJ.3sg   AV.NPST-find   MA-story   DP   SUBJ-mother

    ma-tihoʔ   ma-biou   e.

    AV.NPST-know   AV.NPST-story   DP

    'If she is looking for storytelling, (your) mother knows how to tell stories.'

    T: *i-maʔ   aya=te   ma-kuari.*

    SUBJ-mother   not=COMP   AV.NPST-able

    L: *[i-maʔ   bun   nu   <musti> ma-tihoʔ   e].*

    SUBJ-mother   Bun   LK   musti   AV.NPST-know   DP

    'Mother Bun should know (telling the stories).'

    E: *[i-maʔ   len   nu   <maksud>=nu]*

    SUBJ-mother   Len   LK   goal=LK.2sg

    'Mother Len is who you mean'.

In example (22), the timing of a marriage is a local discourse topic.   A man and a woman finally got married one day when the condition for the marriage was met, and the NP that denotes that day (*rou ene* "that day") becomes the contrastive focus.   It is the new piece of information that the storyteller wants to

convey to the hearer, and it cancels out the other possibilities.

(22) *i-toadaʔ*   *bo*   *i-ɾumimuʔutu*   *siŋka-pahi-sabu*   *side*   *dua*

   SUB-Toada   and   SUBJ-Lumimuutu   once-RCP-meet   SUBJ.3pl   two

na-pahi-sasuka   *nu*   *tikin*   *n-side.*

AV.PST-RCP-measure   NU   stick   LK-3pl

'When Toada and Lumimuutu met, they compared their sticks.'

s-in-eɾe-an   *mai*   *n-side*   *tikin*   *aya=te*   *pada.*

PST-see-GV   already   LK-3sg   stick   not=COMP   same

'It seemed their sticks were not the same (length) anymore.'

yo   *[ɾou*   *ene=te*   *nu*   *side*   *dua*   *na-kabiŋ]*

then   day   that=COMP   LK   SUBJ.3sg   two   AV.PST-marry

su   *timbou*   *nu-kantaŋ=ne.*

LOC   top   LK-mountain=LK.3sg

'So that was the day when the two of them got married at the top of the mountain.'

5. Left-dislocation sentences

    Left-dislocation in Bantik is defined as a construction where the left-dislocated NP is referred to again by a pronoun in the main clause.   A fronted NP functions as a sentential topic or a contrastive topic in this construction, which is very commonly found in the naturalistic data.   It is predominantly an entity "in focus" that becomes a referent of the fronted NP.

    For example, *i-tuadiʔ=ne* in example (24), which follows example (23), is referred to again in the main clause by *=ne* (third person pronoun, connected form[7]). In the previous context, the two brothers were introduced as discourse topics and the elder brother has been explained, so *i- tuadiʔ=ne* in this sentence is clearly a contrastive topic.   The left dislocated items are shown in the brackets [   ], and repeated NPs are bold-faced in the following examples.

(23) *su*   *ɾou*   *pona*   *nu*   *ie*   *pai*   *side*   *dua*   *siŋka-tuhaŋ*

   LOC   day   before   LK   this   EXIST   SUBJ.3pl   two   one-sibling

   'Days and days ago, there were two brothers.' (*Kokokuk*, folk story)

(24) *[i-tuadiʔ=ne]*   *aden=**ne***   *i-gimon*

   [SUBJ-younger.sibling=LINK=3sg]   name=LINK.3sg   SUBJ-Gimon

   'The younger brother, his name was Gimon'   (*Kokokuk*, folk story)

This construction is also used to introduce a discourse topic as in example (25).   Here, the topic *manuʔ taonan* "taonan bird" is referred to again by a possessive pronoun *=ne* (third person pronoun, connected

---

[7] A connected form of a pronoun is used to denote the possessor when it follows an NP, and the actor when it follows a verb in undergoer voice.

form).   A similar construction is shown in examples (26) and (27).

(25) *su*   *huaŋ*   *nu*   *<dunia>,*   *ni-ka-deaʔ-en*   *manuʔ*   *taonan*

   LOC   in   LK   world   PST-POT-find-GV   bird   Taon

   bo   *manuʔ*   *bayan*

   and   bird   Bayan.

   'In the (ancient) world, Taonan bird and Bayan bird were found'.

   [manuʔ taonan]   *aoʔ=**ne***   *bagai*   *bo*   *bombuɾu=ne*   *ma-ituŋ*

   [bird   Taon]   body=LK.3sg   big   and   feather=LK.3sg   ADJVZ-black

   'Taonan bird, its body was big and its feathers were black' (*Burung Taoun dan Burung Ngulngul*)

(26) *[isie]*   *duŋkuɾu.nu.ene,*   *<mayat>=**ne***   *ni-diŋan* *<kulurahan>*   *siŋkil.*

   SUBJ.3sg   at.the.same.time   body=LK.3sg   PST-take region   Singkil

   'As for him, at that time, his body was taken to Singkil'. (*Memperbaiki Rumah*)

(27) *bo*   *[sapi]*   *anaʔ=**ne***   *e*   *na-i-pahuʔ*   *su*   *p-in-a-ŋikiʔ-an*

   and   cow   child=LK.3sg   DP   AV.PST-NVlt-coil LOC   PST-APP-tie-GV

   'And the cow, its child was unintentionally coiled to the pole (to which it was tied).'

Example (19), shown below as (28), also has a left-dislocated NP.   An NP that is repeated can be placed at the end of the clause as in example (29).

(28) *[babaeheʔ*   *su-beo],*   ***isie***   *ma-ka-deaʔ*   *apa*   *nu*

   reward   LOC-wild.pig   SUBJ.3pl   AV.NPST-POT-find what   LK

   ni-kan   *n-toumata*

   PST-eat   LK-people

   'As the reward to the wild pig, he (=wild pig) can find what was eaten by people'

(29) *[anaʔ]*   *kadua=ne,*   *i-fɾaŋki*   *kumaunaŋ,*   *ma-tuɾau*

   child   second=LK.3sg   SUBJ-Franky   Kumaunang   AV.NPST-live

   su   *buha,*   ***isie***

   LOC   Buha   SUBJ.3sg

   'The second child, Franky Kumaunang, he lives in Buha'.

5. Conclusion

     In this paper, markers for discourse new entities and two constructions that show information structure in Bantik were described.   We looked at the two markers for NPs that show information status, which are *pai* and *tou/side*.   The existential marker *pai* is used for an entity that is entirely new to the discourse, or a "type identifiable" entity.   For "uniquely identifiable" or "referential" entities which can be assumed from the previous context or can be inferred from the non-linguistic context, *tou/side* is used to mark the NP.

   Contrastive topics and contrastive foci are shown by cleft construction, whereas discourse topics are marked by left-dislocation.   A cleft construction requires an NP that is "activated" or "familiar" to be fronted.   It is always the case that a clefted NP denotes an entity that is contrasted overtly or covertly with the other entity/entities in the immediately preceding discourse.   Although it is possible to find clefted pronouns in the elicitation data, they are infrequently fronted in the naturalistic data.

   Left dislocation is used when an NP denotes an entity "in focus".   It is a salient local topic, and is referred to again by a pronoun in the main clause that follows it.

   The correlation between the information status of NPs and these constructions should be studied in more detail in the future.

Abbreviations

1sg	first person singular
1pl.EXC	first person plural exclusive
1pl.INC	first person plural inclusive
2sg	second person singular
2pl	second person plural
3sg	third person singular
3pl	third person plural
-AN	suffix -*an* which has a function of nominalization, or of forming derivational verbs
AV.NPST-	prefix attached to verb base, indicating non-past tense and Actor Voice
AV.PST-	prefix attached to verb base, indicating past tense and Actor Voice
CONT	enclitic =*te* that indicates continuative aspect
COMP	enclitic =*ken* that indicates completive aspect
DP	discourse particle
-GV	suffix attached to verb bases, which indicates goal voice
INT	interjection
LK-	noun marker *ni-/nu-* that denotes genitive or actor in undergoer voice sentences, or linker that connects two NPs
POT-	potentive prefix *ka-* which attaches to verb bases
PRO	pronoun *tou/side* that forms an NP with a noun, an adjective, or a verb
REL	relativiser *nu*
SUBJ	nominative case marker attached to subject nominals

References

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski.   1993.   Cognitive Status and the Form of Referring Expressions in Discourse.   Language 69(2).   274-307.

Hedberg, Nancy. 2013. Applying the Givenness Hierarchy Framework: Methodological Issues. In this volume.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form.* Cambridge: Cambridge University Press.

Lyons, Christopher. 1999. *Definiteness.* Cambridge: Cambridge University Press.

Noorduyn, J. 1991. *A critical Survey of Studies on the Languages of Sulawesi.* Leiden: KITLV Press.

Sneddon, James N. 1984. *Proto-Sangiric and the Sangiric languages.* [Pacific Linguistics Series B, No.91] . Canberra: The Australian National University.

Utsumi, Atsuko. 2014a. 'Daimeisi *tou* no youhou (Usages of the pronoun *tou*)'. Journal of the Department of Japanese Culture, Meisei University. Vol.20.

Utsumi, Atsuko. 2014b. 'Newly introduced NPs and given NPs in the Bantik discourse'. *Proceedings of the International Workshop on Information Structure of Austronesian Languages.* p.183-192

Focus and argument indexing in Makasar

Anthony JUKES

Centre for Research on Language Diversity
La Trobe University

Makasar and other languages of South Sulawesi share a grammatical pattern in which (in basic examples) an NP can be 'fronted', and the fronted NP is then not indexed with a pronominal clitic, unlike most core arguments. This pre-predicate position is analysed as focus, and its interaction with the indexing system serves several of the functions typically fulfilled by a voice system in other West Austronesian languages. However this 'basic' characterisation, especially with regard to focus, misses subtleties and irregularities in complex sentences which also need to be accounted for.

1. Introduction

Makasar (also referred to as Makassar, Makassarese or Macassarese — the endonym is *basa Mangkásara'*) is one of the larger regional languages of eastern Indonesia, spoken by the Makasar people in and around the city of Makassar in the province of South Sulawesi. The number of speakers is estimated at about two million (Jukes 2006), making Makasar the second largest ethnic group in Sulawesi — the largest being Bugis with an estimated 3,600,000 (Pelras 1996:1). The language is still widely spoken, though there has been a significant shift away from it in Makassar city itself.
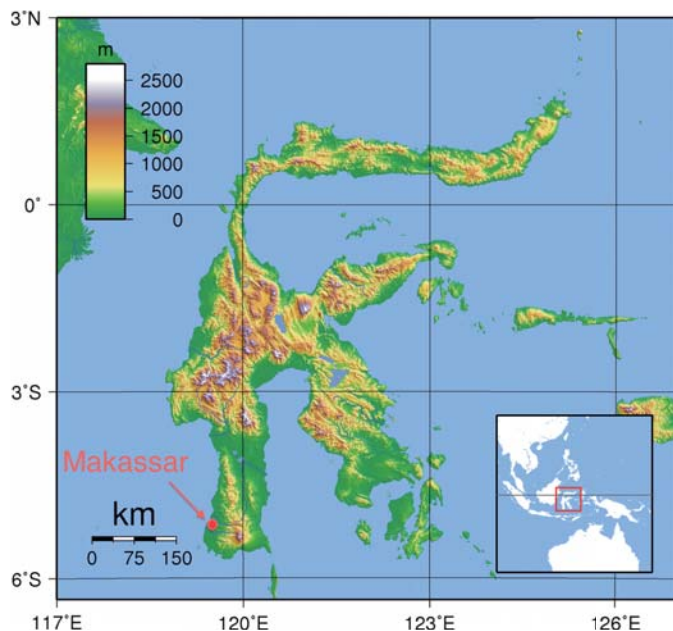
Figure 1: Sulawesi and Makassar

Makasar is a member of the South Sulawesi language subgroup, within the (Western) Malayo-Polynesian branch of the Austronesian language family (Blust 2009). Its closest relatives are the nearby languages Konjo and Selayarese, sometimes thought of as dialects of Makassar. More distantly related are the other languages of South Sulawesi such as Bugis, Mandar, and Sa'dan Toraja. Adelaar (1994, 2005) has also shown the subgrouping relationship between South Sulawesi languages and the Tamanic languages in Borneo.

2. Basic clause structure

Makasar is head-marking and morphologically ergative, with grammatical relations being primarily signified by pronominal clitics on the predicate ('argument indexes' to use Haspelmath's (2013) terminology). The pronominal clitic system is shown in Table 1, along with the associated free pronouns and possessive suffixes.[1]

	Free Pronoun	Proclitic (ERG)	Enclitic (ABS)	Possessive suffix (POSS)
1s	*inakke*	*ku=*	*=a'*	*-ku*
2 fam	*ikau*	*nu=*	*=ko*	*-nu*
2 pol/1pl inc.	*ikatte*	*ki=*	*=ki'*	*-ta*
1 pl exc.[2]	*ikambe*		*=kang*	*-mang*
3	*ia*	*na=*	*=i*	*-na*

Table 1: Pronominal elements

2.1 Intransitive clauses

In intransitive clauses there will be an absolutive enclitic (=ABS) indexing the sole argument S, if S is definite or otherwise salient in the discourse, and not in focus (§5.2). The ABS enclitic tends to attach to the first constituent and is thus a second-position or 'Wackernagel' clitic.

Intransitive verbs are typically marked with a verb prefix, usually *aC–* as in (1), but a small set of basic verbs such as *tinro* 'sleep' (2) does not require these.

(1)   *A'jappai Balandayya*

    aC–   jappa =i   balanda -a
    INTR– walk =3ABS Dutch   -DEF

    The Dutchman is walking

(2)   *Tinroi iAli*

    tinro =i   i   Ali
    sleep =3ABS PERS Ali

    Ali is sleeping

Many other types of phrase may head intransitive clauses, for example adjectives (3), nominals (4) including pronouns (5), and prepositional phrases (6):

(3)   *Bambangi alloa*

    bambang =i   allo -a
    hot   =3ABS day -DEF

    The day is hot

---

[1] The distinction between affixes and clitics can be drawn partly on phonological grounds — affixes are counted as part of the word when stress is assigned, while clitics are not. However this phonological diagnostic is only useful for enclitics, because stress is counted back from the right edge of the word.

[2] The 1st person plural exclusive category lacks a proclitic form and is considered archaic.

(4)   *Jaranga'*

jarang =a'
horse =1ABS

I am a horse

(5)   *Inakkeji*

inakke =ja   =i
1PRO   = LIM =3ABS

It's only me

(6)   *Ri balla'nai*

ri   balla' -na   =i
PREP house -3.POSS =3ABS

He's at home

2.2 Transitive clauses

In transitive clauses both proclitic (A) and enclitic (P) are canonically indexed on the verb, and there is no verb prefix.

(7)   *Nakokkoka' miongku*

Na=   kokko' =a'   miong -ku
3ERG= bite   =1ABS cat   -1.POSS

My cat bit me

(8)   *Lakuarengko Daeng Nakku'*

La=   ku=   areng =ko Daeng nakku'
FUT=   1ERG=   name =2   (title) yearning

I'll call you 'Daeng Nakku''

When both arguments are third person it can sometimes be unclear which clitic pronoun indexes which argument, and the order of free NPs does not help to clarify this, as can be seen in (9). In these situations context or pragmatics must resolve the ambiguity.

(9)   *Naciniki tedongku i Ali*

Na=   cini' =i   tedong -ku   i   Ali
3ERG= see   =3ABS buffalo -1.POSS PERS Ali

Ali sees my buffalo / my buffalo sees Ali

Exceptions to the normal transitive pattern occur for three main reasons:

   (1) either A or P may be in focus position (§5.2);

   (2) the clitics may appear on separate words if there is some preverbal element (due to second-position or 'Wackernagel' constraints); or

   (3) the clause may have an indefinite Undergoer argument. Examination of this type of clause — labeled 'semi-transitive' — is the topic of the remainder of this paper.

2.3 Semi-transitive clauses

The term *semi-transitive* refers to clauses which, although clearly describing events involving two participants, only include a clitic pronoun indexing one of those participants — the Actor, as seen in (10) and (11). The clitic is from the absolute set (S/P).

(10) *ammallia' ballo'*

aN(N)–	balli	=a'	ballo'
TR–	buy	=1ABS	palm.wine

I buy palm wine

(11) *angnganrea' unti*

aN(N)–	kanre	=a'	unti
TR–	eat	=1ABS	banana

I eat bananas

Thus, semi-transitive clauses contain verbs which are generally bivalent lexically, but the Undergoer appears as a full NP and is not cross-indexed. The verb is marked with a verb prefix, usually the nasal-substituting *aN(N)–* (see §3). The general rule is that Undergoers must be definite to be cross–indexed — in other words referred to by name or title, otherwise pragmatically salient such as first and second person, or marked with the determiner *–a* or a possessive suffix. Compare the fully transitive parallel to (11):

(12) *kukanrei untia*

ku=	kanre	=i	unti	-a
1ERG=	eat	=3ABS	banana	-DEF

I eat the bananas

In most instances semi-transitive clauses such as (10) and (11) require an overt Undergoer NP and there is no possible intransitive interpretation, (cf **ammallia'* 'I buy'). With a few verbs, for example *kanre* 'eat' and *inung* 'drink', omission of the Undergoer is allowed and results in an intransitive clause which is quite well-formed, though obviously it differs in meaning. This is because these verbs are ambitransitive, equally allowing intransitive and transitive readings.[3]

(13) *angnganrea' taipa*

aN(N)–	kanre	=a'	taipa
TR–	eat	=1ABS	mango

I eat a mango/mangoes

(14) *angnganrea'*

aN(N)–	kanre	=a'
TR–	eat	=1ABS

I eat, I'm eating

The term *semi-transitive* for clauses with indefinite Undergoers was chosen because it captures the fact that these clauses exhibit properties that fall in between those of normal intransitive and transitive clauses. They differ from intransitive clauses because of the obvious fact that they contain Undergoers, both in their logical structure and in their syntax. They differ from fully transitive clauses in that the Undergoer is not marked with a clitic — signalling that it is not like an ordinary P, if it is a P at all.

Other labels which have been or could be used are **actor focus**, **actor voice**, **antipassive**, **extended intransitive**, or simply **intransitive**.

In the following sections I discuss overt marking of focus and topic, which are each associated with particular syntactic positions. The basic facts are not unlike those

---

[3] An alternative analysis gives these verbs an inherent Undergoer, e.g. 'eat (rice)'.

described for Tukang Besi (South-East Sulawesi) by Donohue (2002), and are also similar to those described for Mayan languages by Aissen (1992), which is that there is a clause-initial focus slot, and a clause-external (ie. left-dislocated) topic slot.[4] However the 'basic' characterisation, especially with regard to focus, misses subtleties and irregularities in complex sentences which also need to be accounted for.

3. Focus

In its most basic manifestation, focus involves an NP referring to a core argument being placed in pre-predicate position. There is a prefix *aN–* which explicitly marks Actor focus (appearing in the place of the ERG= proclitic), whereas Undergoer focus is marked by the absence of an =ABS enclitic. (I use the macrorole labels here because both P and P^{INDEF} may be focused).

Thus, arguments which occur as full NPs directly preceding the predicate are not cross-indexed — for example, compare 15 and 16:

(15)   *Tinroi i Ali*

   tinro =i   i   Ali
   sleep =3   PERS Ali

   Ali is asleep

(16)   *I Ali tinro*

   i   Ali tinro
   PERS Ali sleep

   Ali is asleep

This pre–predicate slot is a focus position,[5] which performs a variety of pragmatic functions such as disambiguating, emphasizing, adding certainty or uncertainty. So while 15 is just a statement of fact, 16 with S in focus can express such meanings as: 'Are you sure it's Ali who is asleep?', 'I tell you that Ali is asleep', 'I've heard that Ali is asleep'. It is also the answer to the question *inai tinro?* 'who is asleep?' (interrogative pronouns are typically focused). Another example of how focus conveys extended meanings is the following:

(17)   *Ballakku kicini'*

   balla' ≡ku   ki=   cini'
   house ≡1.POSS   2p= see

   You see **my house**

This could be given as an answer to the question: what can you give as a guarantee for a loan? (The unmarked way of saying 'you see my house' is *kiciniki ballakku* <ki=cini'=i balla'≡ku | 2f=see=3 house≡1.POSS>).

In transitive clauses either A or P can be in focus. The following two sentences show A focus and P focus respectively where both arguments are definite:

---

[4] See also Finer's work on A' positions in Selayarese (Finer 1994).

[5] Specifically, it is a slot for marked argument focus (Van Valin 1999). As for the configuration, Finer (1994) has analysed the focus position (for Selayarese) as Spec of IP.

(18) *Kongkonga ambunoi mionga*

kongkong ≡a   aN– buno =i miong ≡a
dog   ≡def AF– kill =3 cat   ≡DEF

The **dog** killed the cat

(19) *Mionga nabuno kongkonga*

miong ≡a   na= buno kongkong ≡a
cat   ≡DEF 3=   kill dog   ≡DEF

The dog killed the **cat**

Thus, in 18 there is no proclitic indexing *kongkonga* (A), while in 19 *mionga* (P) lacks a corresponding enclitic.[6] Also note that in 18 the verb is marked with the Actor Focus prefix *aN–* (found in clauses where A is in focus and P is definite.

If P is indefinite (ie. if the corresponding non-focused clause is semi-transitive) either argument may still be focused, so 20 shows A focus , while 21 shows P$^{\text{INDEF}}$ focus:

(20) *Inakke angnganre juku'*

inakke aN(N)– kanre juku'
1PRO BV–   eat   fish

I'm eating fish

(21) *Juku' kukanre*

juku' ku= kanre
fish   1=   eat

I'm eating **fish**

Note that in 20 the verb is marked as semi-transitive with the prefix *aN(N)–* (the missing clitic pronoun being 1$^{\text{st}}$ person *=a'*), but in 21 the verb hosts a proclitic, identical to clauses with focused definite P such as 19 above. This suggests that focus promotes P$^{\text{INDEF}}$ to P (ie. promotes it from a non-core to a core argument), with concomitant promotion of S$^{\text{A}}$ to A.[7]

Sentences with indefinite A are marginal as a general rule, and examples 22 and 23 are no exception.

(22) *?Miong ammuno kongkong*

miong aN(N)– buno kongkong
cat   BV–   kill dog

A **cat** killed a dog / **cats** kill dogs

(23) *?Kongkong nabuno miong*

kongkong na= buno miong
dog   3=   kill cat

A cat killed a **dog** / cats kill **dogs**

Note however, that to make it even marginally acceptable in 23 *miong* (A) has been cross-indexed with *na=* even though it is indefinite and indefinite arguments are not

---

[6] When A is in Focus this has obvious similarities with the phenomenon of 'ergative extraction' as described for Mayan languages (Aissen 1992)— except that there is a parallel 'absolutive extraction' when O is in Focus.

[7] Basri & Finer (1987) have a different analysis, in which it is the trace (left behind when P$^{\text{INDEF}}$ is moved) that is definite and which triggers the ERG= marking of S$^{\text{A}}$. I prefer an analysis in which focus itself promotes an argument to core status.

usually cross-indexed. This could again suggest that focusing P^{INDEF} promotes it to P, which further promotes A^{INDEF} to A.

Complex sentences show focus phenomena which differ somewhat from simple examples. For example, NPs may be be in standard (postverbal) position in one clause, and simultaneously occupy focus position (as can be seen by the use of the Actor focus prefix *aN–*) in a subsequent clause. For example, 24 shows the S NP from one clause serving as focused A in the following clause, and then as A in a third clause though the NP is not present in the clause:

(24)   *battu– tommi kongkonga ampasire'bokangi, angkanrei.*

battu	tong	=mo	=i	kongkong	≡a	aN–	pa–	si–	re'bo'	–ang	=i
come	also	=PFV	=3	dog	≡DEF	AF–	CAUS–	MUT–	squabble	–BEN	=3

aN–	kanre	=i
AF–	eat	=3

the dogs came, fought over it, ate it (bembe:100)

Example 25 from the same story shows three clauses with typical focus morphology, but only one in which an NP (*bembea*) actually occupies the focus slot. In the second clause the 1st person (represented by the preposed clitic pronoun on the initial adverbial modifier *dikki'–dikki'*) is marked as focused A by the prefix *aN–* on *ambuangi*, after which the unfocused P of the second clause becomes the focused (but ellipsed) P of the third clause:

(25)   *Bembea mange a'je'ne', kudikki'–dikki' mange ambuangi karungkunna naung ri buttaya, napasire'bokang kongkong.*

bembe	≡a	mange	aC–	je'ne'	ku=	dikki'–	dikki'	mange	aN–	buang	=i	karungkung
goat	≡DEF	go	MV–	water	1=	RDP–	creep	go	AF–	fall	=3	disguise

≡na	naung	ri	butta	≡a	na=	pa–	si–	re'bo'	–ang	kongkong
≡3.POSS	go.down	PREP	land	≡DEF	3=	CAUS–	MUT–	squabble	–BEN	dog

The goat went to bathe, I crept to throw her disguise down to the ground, it was torn apart by dogs (bembe:111)

In the preceding examples, although focus can be identified according to the structural principles as noted for simple clauses, it is unclear what the pragmatic effects are. This requires further investigation not only of focus but of clause integration phenomena.

Finally, 26 is a proverb with two parallel clauses.

(26)   *Tedong lompo mate i rawa ri sirinna na tena naciniki, sama–sama mate ri sirinna taua na nacini'*

tedong	lompo	mate	i	rawa	ri	siring	≡na	na	tena	na=	cini'	=i
buffalo	big	death	PREP	beneath	PREP	cellar	≡3.POSS	and	NEG	3=	see	=3

sama–	sama	mate	ri	siring	≡na	tau	≡a	na	na=	cini'
RDP–	louse	death	PREP	cellar	≡3.POSS	person	≡DEF	and	3=	see

A big dead buffalo in his cellar and he doesn't notice it, a dead louse in someone else's cellar and that, he notices

This example is somewhat confusing because *na* has 3 separate functions: ERG, POSS and the conjunction 'and'. But it is clear that in the first part of the proverb the buffalo is indexed with an ABS enclitic, and in the second the louse is not indexed, though the constructions are otherwise exactly parallel. The difference is that the louse is receiving contrastive focus (represented in English with the cleft construction). Which suggests

that focus is marked not only by pre-predicate position, but also by lack of indexing, reminiscent of Nikolaeva's (1999) analysis of Northern Ostyak:

> The object that does not trigger agreement bears the focus function, and systematically corresponds to the focus position. (Nikolaeva 1999:331).

The extent to which lack of indexing marks focus requires future investigation.

3. Topicalisation

There is a further possibility for preposing elements in a clause, which is left-dislocation. In this (unlike with focus) a clear prosodic break occurs between the preposed element and the remainder of the clause, and if the preposed element is a core argument, cross–indexing does occur (again, unlike focus). This can be seen in both 27 and 28 — in the former A is topicalised and both arguments are cross–indexed, in the latter A is topicalised, P is focused and thus only A is cross–indexed with a proclitic:

(27) *kongkonga, nabunoi mionga*

kongkong	≡a	na=	buno	=i	miong	≡a
dog	≡DEF	3=	kill	=3	cat	≡DEF

the dog, it killed the cat

(28) *kongkonga, mionga nabuno*

kongkong	≡a	miong	≡a	na=	buno
dog	≡DEF	cat	≡DEF	3=	kill

as for the dog, it was the cat that it killed

Example 29 has two clauses illustrating the structural contrast between topic and focus — in the first clause P is topicalised and thus is cross-indexed with an enclitic, while in the second P is in focus and is not cross-indexed:

(29) *Anjo bainea, nalantiki Karaeng ri Massere'; anjo bura'nea nalanti' Karaeng ri Roong*

anjo	baine	≡a	na=	lanti'	=i	karaeng	ri	Massere'	anjo	bura'ne	≡a
that	female	≡DEF	3=	inaugurate	=3	karaeng	PREP	Massere'	that	man	≡DEF

na=	lanti'	karaeng	ri	Roong
3=	inaugurate	karaeng	PREP	Roong

That girl, he made her Karaeng of Massere', that boy he made Karaeng of Roong. (bembe:003)

Topicalisation differs functionally from focus as one would expect. Whereas marked focus is generally used in a contrastive function, topicalisation is most often used when setting a topic either for a whole text (as was the case in 29 as the story is basically about Karaeng Massere'), or for switching between alternative topics. It also clearly differs syntactically. Whereas a focused argument is an argument within the phrase (as indicated by omission of its corresponding clitic pronoun), a topicalised NP is external to the phrase (as indicated by the presence of the clitic pronoun).

Abbreviations

ABS	absolutive	AF	actor focus
DEF	definite	ERG	ergative
FUT	future	INTR	intransitive
LIM	limitative	NEG	negative
PERS	personal prefix	POSS	possessive
PREP	preposition	PRO	pronoun
PROH	prohibitive	SBJV	subjunctive
STV	stative	TR	transitive

References

Adelaar, K.A. 1994. The classification of the Tamanic languages (West Kalimantan). In Dutton, T.E. & Tryon, D.T. (eds) *Language contact and change in the Austronesian world.* Berlin: Mouton de Gruyter. 1–41.

Adelaar, A. 2005. A historical perspective. In Adelaar, A. & Himmelmann, N. (eds) *The Austronesian Languages of Asia and Madagascar.* London: Routledge.

Aissen, J. 1992. Topic and Focus in Mayan. *Language* **68**:43–80.

Arka, I W. 2009. *The core-oblique distinction and core index in some Austronesian languages of Indonesia.* Paper first presented at International ALT VI (Association of Linguistic Typology conference, Padang, Indonesia, July 2005).

Blust, R.A. 2003. Three Notes on Early Austronesian Morphology. *Oceanic Linguistics* **42/2**:438–478.

Blust, R.A. 2009. *The Austronesian Languages.* Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Ceria, V. 1993. Verb morphology and valence change in Selayarese. *University of Pittsburgh Working Papers in Linguistics*, **II**:76–185.

Cummings, W. 2000. Reading the histories of a Maros chronicle. *Bijdragen tot de Taal–, Land– en Volkenkunde* **156**:1–31.

Dixon, R.M.W. 2012. *Basic Linguistic Theory* Vol.3. Oxford: Oxford University Press.

Dryer, M. 2007. Clause Types. In Shopen, T. (ed.) *Clause Structure. Language Typology and Syntactic Description: Volume 1.* Cambridge: Cambridge University Press.

Friberg, B. 1988. Ergativity, focus and verb morphology in several South Sulawesi languages. In Harlow, R & Clark, R. (eds) *VICAL 2: Western Austronesian and contact languages: papers from the Fifth International Conference on Austronesian Linguistics.* Auckland: Linguistic Society of New Zealand. 103–130.

Friberg, B. 1996. Konjo's peripatetic person markers. In Steinhauer, H. (ed.) *Papers in Austronesian Linguistics No. 3.* Canberra: Pacific Linguistics. 137–171.

Friberg, B. 2002. Grammatical construction differences between Makasar and Konjo. Handout from 9ICAL (the Ninth International Conference on Austronesian Linguistics).

Hanson, C. 2001. A description of basic clause structure in Bugis. *La Trobe Papers in Linguistics*, **11**:143–160.

Hanson, C. 2003. *A grammar of Bugis based on the Soppeng dialect*. Ph.D. thesis, La Trobe University.

Hasan Basri. 1998. Number marking in Selayarese. In Pearson, M. (ed.) *Recent Papers in Austronesian Linguistics: Third and Fourth meetings of the Austronesian Formal Linguistics Association*. University of California Los Angeles. 45–59.

Hasan Basri. 1999. *Phonological and syntactic reflections of the morphological structure of Selayarese*. Ph.D. thesis, SUNY Stony Brook.

Hasan Basri, Broselow, E., Finer, D. 1999. Clitics and crisp edges in Makasar. *Toronto Working Papers in Linguistics: Proceedings of AFLA VI (The sixth meeting of the Austronesian Formal Linguistics Association)*.

Hasan Basri & Chen, Y. 1999. When harmony meets reduplication in Selayarese. *Toronto Working Papers in Linguistics: Proceedings of AFLA VI (The sixth meeting of the Austronesian Formal Linguistics Association)*.

Haspelmath, M. 2013. Argument indexing: a conceptual framework for the syntax of bound person forms. In: Bakker, Dik & Haspelmath, Martin (eds.) *Languages across boundaries: Studies in memory of Anna Siewierska*, 197-226. Berlin: De Gruyter Mouton.

Himmelmann, N. 1996. Person marking and grammatical relations in Sulawesi. In Steinhauer, H. (ed.) *Papers in Austronesian Linguistics No. 3*. Canberra: Pacific Linguistics. 115–136.

Himmelmann, N. 2002. Voice in western Austronesian: an update. In Wouk, F. & Ross, M. eds) *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics. 7–16.

Himmelmann, N. 2005. Typological characteristics. In Adelaar, A. & Himmelmann, N. (eds) *The Austronesian Languages of Asia and Madagascar*. London: Routledge.

Jukes, A. 1998. *The phonology and verbal morphology of Makassar*. M.A. thesis, University of Melbourne.

Jukes, A. 2005. Makassar. In Adelaar, A. & Himmelmann, N. (eds) *The Austronesian Languages of Asia and Madagascar*. London: Routledge.

Jukes, A. 2006. *Makasar*. PhD Thesis, University of Melbourne.

Kaufman. D. 2008. South Sulawesi pronominal clitics: form, function and position. *Studies in Philippine Languages and Cultures* **17:**13-65.

Lee, J.K.L. 2008. Transitivity, Valence and Voice in Mandar. *Studies in Philippine Languages and Cultures* **19**:55-66.

Manyambeang, A.K., Mulya, A.K., & Nasruddin. 1996. *Tatabahasa Makassar*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Mead, D. 2003. Evidence for a Celebic supergroup. In Lynch, J. (ed.) *Issues in Austronesian phonology*. Canberra: Pacific Linguistics.

Mead, D. 2005. Mori Bawah. In Adelaar, A. & Himmelmann, N. (eds) *The Austronesian Languages of Asia and Madagascar*. London: Routledge.

Mithun, M. & Basri, H. 1987. The phonology of Selayarese. *Oceanic Linguistics* **25**: 210–254.

Nikolaeva, Irina. 1999. Object Agreement, Grammatical Relations, and Information Structure. *Studies in Language* 23/2: 331-376.

Pelras, C. 1996. *The Bugis*. Oxford: Blackwell.

Reid, L. and Liao, Hsiu-chuan. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* **5/2**:433-490.

Salombe', Cornelius. 1982. *Bahasa Toraja Saqdan: Proses morfemis kata kerja*. [Toraja Saqdan language: Verbal morphology]. Jakarta: Penerbit Djambatan.

Siewierska, A. 2004. *Person*. Cambridge: CUP.

Van Valin, R.D. 1999. A Typology of the Interaction of Focus Structure and Syntax. In Raxilina, E. & Testelec, J. (eds) *Typology and the Theory of Language: From Description to Explanation*. Moscow: Languages of Russian Culture. 511–524.

Van Valin, R.D. and La Polla, R. 1997. *Syntax*. Cambridge: Cambridge University Press.

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

Sonja Riesberg
Universität zu Köln

1. Introduction[1]

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 voices[2] – 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-realis 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-realis 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 in 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	*moN-*[3]	*noN-*
	mog-	*nog-*
	mo-	*no-*
UV1	Ø	*ni-*
UV2	*-i*	*ni- -an*

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 marked[4] 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	*aku*	*=ku; ku-*
2SG	*kau*	*=mu; =ta*
3SG	*isia*	*=na*
1PL EXCL	*kami*	*kami*
1PL INCL	*kita*	*=ta*
2PL	*kamu*	*=ta*
3PL	*sisia*	*sisia*; (*=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 is 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 Bjugan.

(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 discusses in the previous sub-sections:

PRON	zero	Ø (6)
	free	e.g. *aku, kamu* (5)a.
	bound	e.g. *=ku, =mu, =ta*, etc. (5)b.
DEM	PRX	*ia* (8)a.
	MED	*ana* (8)b.
	DIST	*itu*
DP	DPPr	e.g.
	DPM	e.g. *leang sagin ana* 'these banana leafs' (6)
	DPD	e.g. *bale itu* 'that house' (9)c.
N		e.g. *ondan* 'a/the ladder' (10)
Nposs	N=*ku*	e.g. *anak=ku* 'my child'
	N=*mu*	e.g. *tangayopan=mu* 'your plants' (21)
	N=*na*	e.g. *amang=na* 'his father'
	N *kami*	e.g. *bakele kami* 'our grandmother'
	N=*ta*	e.g. *usat=ta* 'our sibling'
	N *sisia*	e.g. *tinga sisia* 'their language'
NP		e.g. *mangana dedek* 'small child'
PP		e.g. *dei daami* 'to the garden' (13)b.
REL		e.g. *tau moane anu kodoong kabing* '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 <u>a new car</u>, and also Harry had bought <u>a new car</u>.*

b.   <u>*A car*</u> *was waiting in front of the hotel. I could see a woman in <u>the car</u>.*

c.   *Yesterday, a friend of mine prepared <u>a lasagne</u> for me. I found it hard to enjoy <u>the tasteless stuff</u>.*

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 Bauman/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)	*sirita*	*bolong*	*ai*	*pomponu*	*nogutu*	*gauan*
	sirita	bolong	ai	pomponu	no-gutu	gauan
	story	monkey	and	turtle	AV.RLS-make	garden
		new		**new**		**new**

pomoona	*molipidoan*	Ø
pomoo=na	moli--an pido	Ø
first=3s.GEN	RCP good	Ø
		given

njan	*pombulina*	*nolidaisan*	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 form 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)	*sabab*	*ana*	*sagin*	*mapanggat*	*batangna*
	sabab	ana	sagin	mo-panggat	batang=na
	because	if	banana	ST-high	stem=3s.GEN
			given		**bridging**

'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.

(18) *pogitanapo*   *turuus*   *ana*
      pog-ita=na=po   turuus   ana
      GER-search=3s.GEN=INCPL always   MED
      given   **given**

      Ø   *tuju*   *kueang*   *menagi*   *utang*
      Ø   tuju   kueang   menagi   utang
      Ø   bewitch   eagle   AV:demand   debt
      given   **given**   **bridging**

      'It (the chicken) is still searching for it (the bracelet). It is bewitched
      by the eagle, who demands his debt'   [chicken_eagle.211-214]

3.3 Discourse-new versus hearer-new

Finally, one more distinction needs to be discussed in more detail, namely the difference between *new* and *unused* (see Table 4). This distinction pertains to the fact that certain referents are known to the hearer, even though they have not been introduced into the discourse before, i.e. even though they are discourse-new. Well known and often cited examples are *the president of the United States*, or *the moon*, which are uniquely identifiable and therefore in English are used with the definite article, even if they have not been mentioned before. A similar case is illustrated by the Totoli example in (19). Both proper nouns, *Palu* and *KPUD*, have not been mentioned in the discourse before. However, they are known to the hearer: Palu is the capital of the province Sulawasi Tengah (Central Sulawesi), the KPUD (= *Komisi Pemilihan Umum Daerah*) is the local commission for the general elections.

(19) *tau*   *tukka*   *dei*   *Palu*   *tempona*   *ana*
      tau   tukka   dei   Palu   tempo=na   ana
      person   older.sibling   LOC   PN   time=3s.GEN   MED
      given   **unused**

      nobali   *anggota*   *KPUD*
      no-bali   anggota   KPUD
      AV.RLS-become   member   PN
                            new   **unused**

      'The oldest brother, who is in Palu now, became a member of the KPUD'
      [Nahre's_life.093-096]

4. Preliminary results

 For this very preliminary study discussed in this paper, 4 spoken Totoli narratives were annotated (together 31:25 minutes of spoken speech). The texts were coded for four variables: First, the grammatical relation of the respective referring expression (i.e. whether it occurred in subject, non-subject, or oblique function); second, the voice of the constructions, as introduced in section 1 (i.e. whether we are dealing with an actor voice, or an undergoer voice construction (either dynamic, stative, or potentive)); third, the form of the referring expression, as discussed in section 2 and summarized in Table 3; fourth, the information status of the referent denoted by the referring expression, as discussed in section 3 and summarized in Table 4. Altogether, 803 referring expressions where coded.

The annotations were used to address the following three research questions, which will be discussed consecutively in the next three sub-sections:

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) | *manuk* | *ana* | *masahabatan* | *ai* | *kueang* |
      | 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 linguea*. Table 5 summarises the distribution of form and their respective information statuses in the four narratives investigated[5].

---

[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) *jaam*   *opat*   *nangalai*   *ssaakan*   *pottuluanna*
      jam   opat   -ngo- no-lai   sasaakan   po-RDP1-tuli-an=na
      clock   four   -COLL- AV.RLS-run   all   GER-RDP1-sleep-NR=3s.GEN

      ialingnako   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	*sum*
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*
sum	*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 phases, 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).
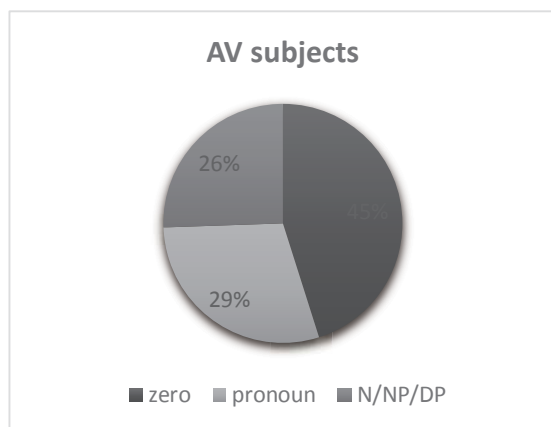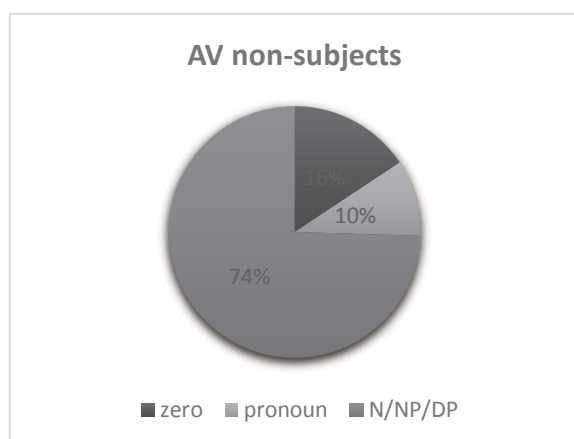

Figure 1: Forms of AV subjects
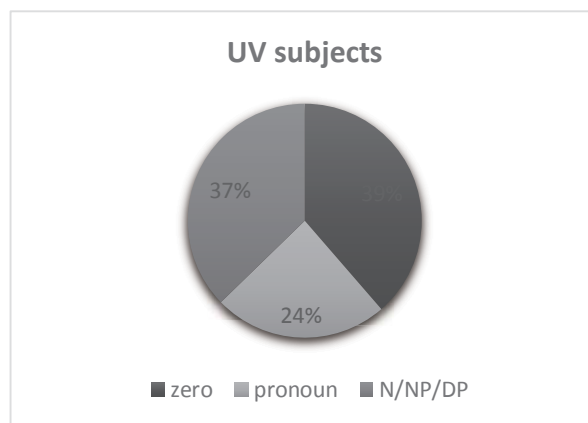
Figure 2: Forms of UV subjects

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 construction[7]. 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" (Himmelmann 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 realized 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 Himmelmann 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 Himmelmann 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
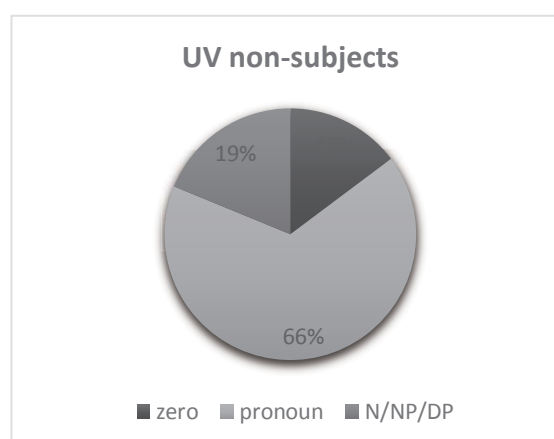
**Figure 4:** Forms of UV non-subject arguments

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.
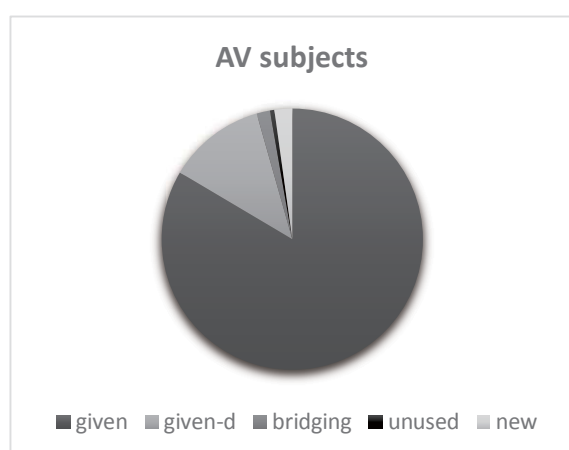
Figure 5: Status of AV subjects
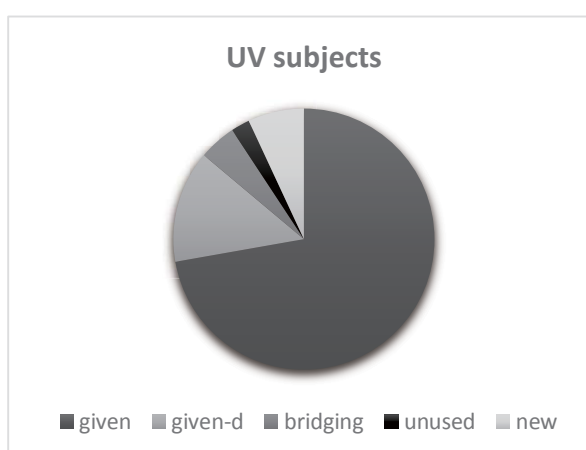
Figure 6: Status of UV subjects

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%).
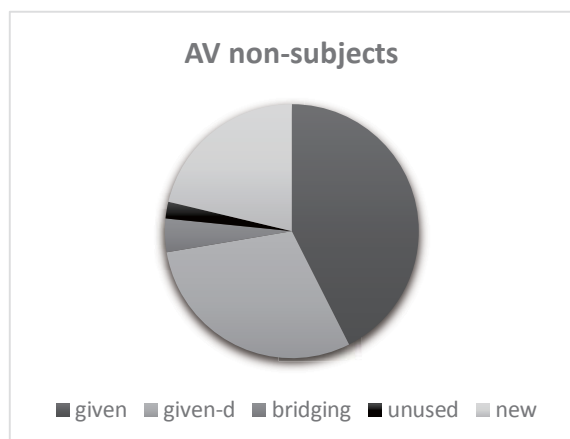
Figure 7: Status of AV non-subject arguments
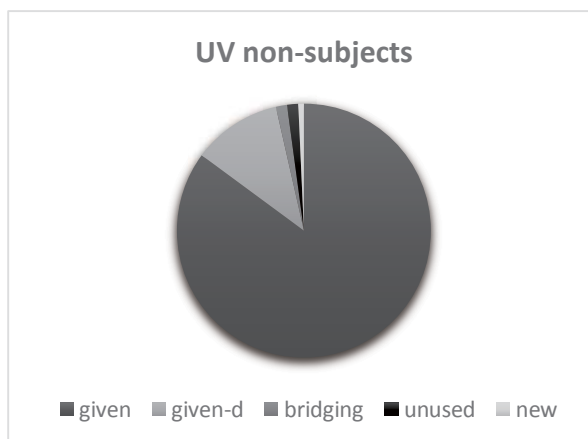
Figure 8: Status of UV non-subject arguments

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, Gorka & 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 Topiv*, 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.

Himmelmann, N. P. 2005. Tagalog. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*, 350-376. London & New York: Routledge.

Himmelmann, N. P. 1999. The Lack of Zero Anaphora and Incipient Person Marking in Tagalog. *Oceanic Linguistics* 38. 231-269.

Himmelmann, 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 18t 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.

Person restriction on passive agents in Malay and givenness[*]

Hiroki Nomoto
Tokyo University of Foreign Studies

1 Introduction

The agent of *di-* passives in Malay appears to be restricted to third person. In Nomoto and Kartini (2014), we analysed this restriction as resulting from the influence of the givenness of the eventuality described by the passive verb phrase on that of the agent. Specifically, the low givenness/salience (= high newness) of the former forces the latter to be also low. Since first and second person agents are speech act participants (i.e. speakers and hearers) and highly given, they are not suitable as a *di-* passive agent.

This paper elaborates on our previous analysis, with particular focus on the following two theoretical issues: the givenness of implicit passive agents (analysed as *pro*) and givenness of eventualities. The notion of givenness is usually discussed of individual-denoting referential noun phrases (e.g. Chafe 1976; Prince 1992; Gundel et al. 1993). However, the notion is also relevant to other constituent types such as verb phrases, and plays a role in information structure-related linguistic phenomena (Schwarzschild 1999; Riester 2008). In discussing these theoretical issues, this paper also makes a few modifications to our previous analysis of Malay passives in Nomoto and Kartini (2014).

This paper is structured as follows. Section 2 presents the voice system of Malay assumed in this paper, including different passive subtypes in Malay and how their agents are expressed. Section 3 introduces the person restriction on the agent in *di-* passives and the essence of Nomoto and Kartini's (2014) analysis of it in terms of information structure, particularly givenness. The section thus contains a brief review of the notion of givenness. Sections 4 and 5 discuss issues concerning givenness that arise from our analysis: the status of the implicit passive agent (section 4) and givenness of eventualities (section 5). Section 6 is the conclusion.

2 Passives in Malay

Malay has two types of passive(-like) constructions: morphological passives with the prefix *di-* (1a) and bare passives (1b).[1] They are so called based on their surface morphological characteristics. The verb bears the overt passive voice marker *di-* in the former whereas it bears no overt voice marker in the latter.[2] Besides this morphological difference, the two passives also differ

[1]Non-standard abbreviations used (those not included in the Leipzig Glossing Rules): ACT: active; FAM: familiar; PART: particle; POL: polite.

[2]Bare passives are referred to by various names in the literature: 'object-preposing construction' (Chung 1976; Willett 1993), 'Passive Type 2' (Dardjowidjojo 1978; Sneddon et al. 2010), 'pasif semu' [pseudo-passive] (Asmah

in the status of the agent. The agent in *di-* passives appears to be optional whereas that in bare passives is obligatory and immediately precedes the verb. Note that I use the term 'passive' to refer to a construction type in which the theme argument does not occur as a direct object but as a subject, regardless of the syntactic status of the agent (see below for further details).

(1)   a.   *Di-* passive

Dokumen itu   sudah   *di*-semak   oleh mereka.
document that already   PASS-check by   them
'The document has already been checked by them.'

b.   Bare passive

Dokumen itu   sudah   *(mereka) semak.
document that already   they   check
'They have already checked the document.'

Di- passive agents are encoded in three ways, as shown in (2). In the '*pro* type' (2a), no overt agent occurs, though the presence of an agent is entailed. In the '*oleh* type' (2b), the agent is introduced by the preposition *oleh* 'by'. Finally, in the 'DP type' (2c), the agent immediately follows the verb, with no preposition.

(2)   a.   *Pro* type

Surat itu   sudah   di-poskan *pro*.
letter that already   PASS-post

b.   *Oleh* type

Surat itu   sudah   di-poskan *oleh kerani*.
letter that already   PASS-post by   clerk

c.   DP type

Surat itu   sudah   di-poskan *kerani*.
letter that already   PASS-post clerk
'The letter was already posted (by the clerk).'

I propose the structures in (3) for the three *di-* passive subtypes and the bare passive. In *di-* passives, the verb moves from V to v to Voice, to supply the prefix *di-* with a verbal host to attach to.[3] Nomoto and Kartini (2014) analyse an implicit agent as a null unspecified pronoun (*pro*) rather than being absent altogether from the structure.[4] This ensures that the presence of an agent is entailed even if it is not explicitly expressed. The meaning of *pro* can be left unspecified, but it can also be specified either overtly by an *oleh* 'by' phrase or covertly by the context outside of the passive clause. In other words, *pro* is involved in the *oleh* type as well as the *pro* type. By contrast, the DP type *di-* passive and bare passive must have an overt agent DP.[5]
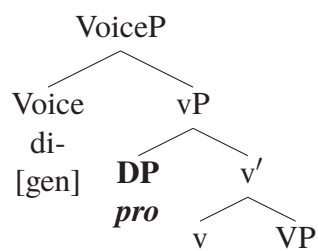

2009), 'object(ive) voice' (Arka and Manning 1998; Cole, Hermon, and Yanti 2008), and so forth. See Nomoto (2006) for a summary of various existing terms.

[3]I revised the structures proposed in Nomoto and Kartini (2014). In the latter paper, we posited the voice markers *di-* and Ø in v. The verb movement in *di-* passives lacked a clear motivation in this analysis, unlike the current one. Cole et al. (2008) also posit the voice-related prefixes *di-* and *meN-* in the Voice head distinct from v.
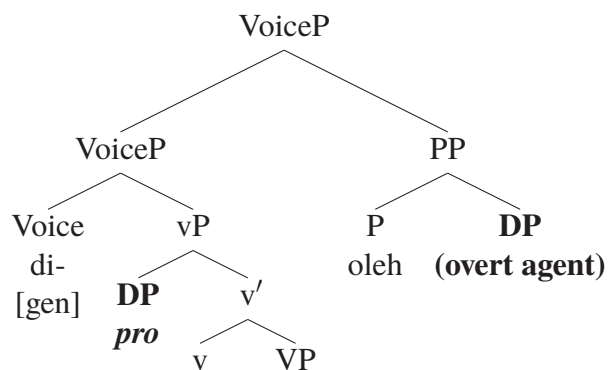
[4]Alternatively, the agent argument can be existentially closed (cf. Legate 2010, 2012, 2014; Kartini and Nomoto 2012).

[5]One known problem with positing the agent of the DP type *di-* passive in Spec,vP is that it cannot bind a reflexive in the subject position, unless it is a pronoun, as in (ia) (Arka and Manning 1998; Cole et al. 2008; Kroeger 2014). Cole et al. (2008) thus posit the agent DP below the theme position as a V′ adjunct, as in (ib).
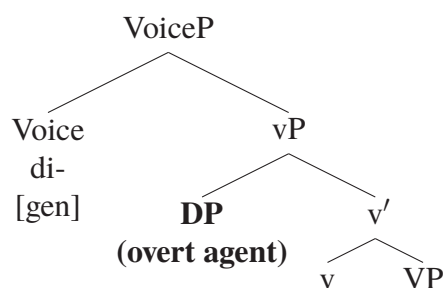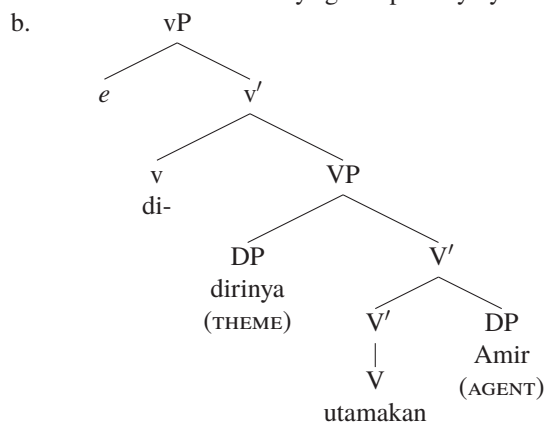
(3)   a.   *Pro* type *di-* passive

```
            VoiceP
           /      \
      Voice        vP
       di-        /  \
      [gen]    DP      v′
               pro    /  \
                     v    VP
```

b.   *Oleh* type *di-* passive

```
                    VoiceP
                  /        \
            VoiceP           PP
           /      \         /   \
      Voice        vP      P      DP
       di-        /  \    oleh   (overt agent)
      [gen]    DP      v′
               pro    /  \
                     v    VP
```

c.   DP type *di-* passive

```
            VoiceP
           /      \
      Voice        vP
       di-        /  \
      [gen]    DP      v′
            (overt agent)  /  \
                          v    VP
```
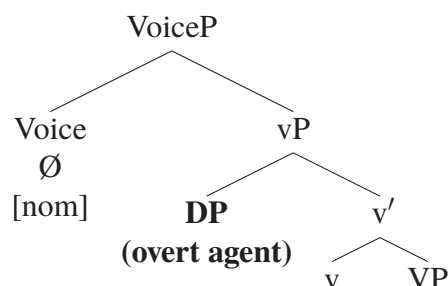

(i)   a.   Diri-nya (sendiri) selalu di-utamakan   -nya/*Amir.
           self-3   own   always pass-prioritize -3/Amir
           'Himself was always given priority by him/*Amir.'

      b.
```
           vP
          /  \
         e    v′
             /  \
            v    VP
           di-  /  \
              DP      V′
           dirinya   /  \
          (THEME)  V′    DP
                   |    Amir
                   V   (AGENT)
               utamakan
```

d.   Bare passive

```
                    VoiceP
                   /      \
               Voice       vP
                 Ø        /   \
               [nom]    DP      v′
                     (overt agent)  / \
                              v    VP
```

I assume that Voice licenses the agent DP introduced in Spec,vP through Case assignment. In Malay, abstract Case is thought to be reflected on the type of clitics: *di-* with [gen(itive)] licenses enclitics whereas Ø with [nom(inative)] licenses proclitics.

At this point, I should clarify the relation between *voice categories* such as active and passive, and *voice markers* that I assume. My definition of voice categories is based solely on whether and how an argument is licensed. The active-passive distinction (in my definition) is concerned with how the functional head v licenses an internal argument, which is introduced by a lexical verb (V).[6] In short, the active-passive distinction is a property of v. The active v licenses an internal argument by assigning accusative case to it whereas the passive v lacks this ability and cannot license it. The v head has another role; it introduces an external argument. The external argument thus introduced needs to be licensed, and it is Voice that does this job. Crucially, in my definition, how an external argument is licensed is a different voice distinction than the active-passive distinction and is orthogonal to the active-passive distinction. It enables one to subclassify actives/passives.[7]

However, Voice is not totally irrelevant to the active-passive distinction. This is because it has a selectional restriction on the type of vP it takes to its complement, though it does not determine the type. *Di-* selects for a passive vP, that is, a vP headed by the passive v. It is in this sense that *di-* is a passive voice marker. Similarly, the prefix *meN-* in the *meN-* morphological active (4a) is an active voice marker in the sense that it selects for an active vP. The null Voice head involved in the bare active (4b) and bare passive has no selectional restriction, that is, it is compatible with either an active or a passive vP.[8] Table 1 summarizes voice categories and voice markers in Malay.

(4)   a.   *MeN-* active

        Mereka sudah   meny-[s]emak dokumen itu.
        they   already ACT-check   document that

---

[6]From a typological perspective, it is more adequate to refer to what I call 'passive' as the undergoer voice. This is because the term 'passive' is normally used to refer to a construction in which the external rather than internal argument is suppressed (Keenan and Dryer 2007). My choice of the term 'passive' here is based on the norm in Malay linguistics, whereby *di-* clauses are referred to as 'passives'. It goes without saying that what is important is not the name but the properties of the construction at issue.

[7] While bare passives are a subtype of the passive in my definition of voice categories, they are seen as a third kind of voice, the so-called 'object(ive) voice', in a definition that conflates the two independent factors of Voice and v (cf. symmetric(al) voice hypothesis). See also Table 1.

[8]Many researchers make little of the bare active despite its frequent use in daily speech. It is either simply ignored or seen as the *meN-* active whose *meN-* is omitted/deleted. In the present analysis, the bare active involves the unmarked voice marker and no such omission/deletion takes place.

Table 1: Voice markers and voice categories in Malay

Construction	Voice (marker)	vP selection	Case by Voice	Case by v	Voice category
meN- active	*meN*-	active	nominative	accusative	active
di- passive	*di*-	passive	genitive	none	passive
bare passive	Ø	either	nominative	none	passive
bare active				accusative	active

b.   Bare active

Mereka sudah   semak dokumen itu.
they   already check document that
'They have already checked the document.'

3   Person restriction on *di*- passive agents and givenness

Prescriptive grammars of Malay (and Indonesian) state that the agent in *di*- passives should be third person and prohibit first and second person agents. Researchers are not unanimous as to whether this statement is descriptively accurate. In order to resolve this empirical unclarity, Nomoto and Kartini (2014) examined various texts in Formal and Colloquial Malay, and showed that the restriction exists as a strong tendency rather than an absolute syntactic rule. No similar person restriction exists for the agent in bare passives.

How can we explain these facts? Since the restriction is not an absolute syntactic rule, a plain syntactic account will be too restrictive and face an undergeneration problem. Nomoto and Kartini (2014) thus propose an account in terms of information structure, particularly givenness.

Before introducing our account more specifically, I would like to briefly overview the notion of givenness. Givenness has to do with the speaker's assessment of the addressee's consciousness/attention state and knowledge with regard to a referent (e.g. Chafe 1976; Prince 1992; Gundel et al. 1993; Lambrecht 1994). Initially, the notion was conceived as a binary distinction between 'given' and 'new', where a given referent is already activated in the speaker's consciousness at the time of utterance whereas a new referent is not and newly activated by the relevant utterance. However, it is nowadays common to identify multiple statuses with different degrees of givenness.

One of the popular theories of givenness is the Givenness Hierarchy of Gundel et al. (1993), which has been adopted in studies of many languages including Austronesian languages such as Bantik (North Sulawesi, Indonesia; Utsumi 2014) and Kalanguya (Northern Phillipines; Santiago 2014) (see Hedberg 2014 for a list of other languages). The Givenness Hierarchy and English examples that represent each status are given in (5).

(5)   The Givenness Hierarchy

in focus	>	activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
it		*that* *this* *this* NP[9]		*that* NP		*the* NP		indefinite *this* NP		*a* NP

What distinguishes the Givenness Hierarchy in (5) from other similar theories of givenness is that the status categories form a hierarchy in such a way that a status entails all the statuses to its right. Thus, if a referent is 'in focus', it is also 'activated', 'familiar', 'uniquely identifiable', 'referential' and 'type identifiable'. This feature elegantly captures the fact that one form can be employed for multiple adjacent statuses. For example, '*the* NP', categorized as 'uniquely identifiable', can be used to refer to referents of higher statuses such as 'familiar' and 'in focus' as well. Furthermore, consisting of a single dimension, the Givenness Hierarchy also enables an easy comparison between different statuses with respect to degrees of givenness. The feature is crucial for the account of the person restriction on passive agents in Malay proposed by Nomoto and Kartini (2014).

Although studies of givenness usually centres around individual-denoting noun phrases, the notion is not exclusively for individuals, but it also applies to other semantic types. Thus, the Coding Protocol for Statuses on the Givenness Hierarchy (Gundel et al. 2006) takes into consideration eventualities and propositions when discussing the usage of the italicized nominals in (6) and (7).

(6)   John fell off his bike. *This/it* happened yesterday.

(7)   A: John fell off his bike.
       B: *That*'s not true.

This/it in (6) refer to the event of John's falling off his bike that is introduced into the discourse and consequently the addressee's consciousness by the first sentence. *That* in (7) refers to the proposition associated with that same event.[10] While Gundel et al.'s concern lies in the italicized nominal expressions, an adequate description of them presumes that their non-individual antecedents have givenness statuses. For example, one should be able to say things like "the event described by sentence S is 'in focus', 'activated', etc."[11]

With this background, let us return to the issue of the person restriction on passive agents. To recapitulate, there is a strong tendency that the agent in *di-* passives is third person whereas no such tendency exists for the agent in bare passives. According to Nomoto and Kartini (2014), first and second person referents are rare in *di-* passives because *di-* passive agents are supposed to be low in givenness/salience. First and second person agents are speech act participants and highly given. Hence, they are not so suitable for *di-* passive agents. The lack of a similar person restriction in bare passives means that bare passives impose no givenness specification. Note

---

[9]The DP hypothesis is assumed here, whereby the traditional noun phrases are analysed as determiner phrases with a determiner head and an NP complement ([$_{DP}$ D NP]).

[10]Similarly, German *da* can refer not only to individuals but also to eventualities and propositions associated with them (p.c. Arndt Riester). (i) is an example taken from *Grosses Deutsch-Japanisches Wörterbuch* (second edition, Shogakukaku, 2000).

(i)   Er schenkte mir eine Brosche, und ich freute   mich   sehr *dar*-über.
       he presented me a   broach   and I   pleased myself very DA-about
       'He presented me a broach, and I was very pleased about *it*.' [*it*: (i) his presenting me a broach, (ii) the broach]

[11]It may sound worthless to think about different givenness statuses for eventualities/verb phrases because they almost always introduce new events in typical narratives, as claimed by Nikolaus Himmelmann during the workshop. The latter fact definitely makes it much more difficult to find examples of different givenness statuses for eventualities/verb phrases (if any) compared to individuals/noun phrases. However, I am uncertain whether it serves to completely reject the initial hypothesis that the same theory of givenness applies to the two types.

that even though the restriction is directly relevant to the agent DP, one cannot just focus on the agent DP. This is because the restriction is not on the agent DP in general, but only on that in *di-* passives.

One may wonder if the restriction is present only in *di-* passives but not in bare passives because the agent is suppressed in *di-* passives. However, such an argument does not go through, because an overt agent is obligatory in the DP type of *di-* passives, as seen in section 2. Moreover, Nomoto and Kartini analyse the *pro* and *oleh* types as containing a null unspecified pronoun *pro*, that is to say, the agent is not suppressed in all types of *di-* passives. I will postpone discussing *pro* to the next section, as its givenness status needs special attention.

Nomoto and Kartini do not state how low the relevant givenness status is. However, a sentence like (8) shows that the relevant givenness status is the lowest one, 'type identifiable'. The noun phrase *dua hingga lima ekor burung betina* 'two to five female birds' is only type identifiable.

(8)   [Burung unta makan secara berkumpulan yang terdiri daripada enam hingga 10 ekor.]
        Se-ekor burung jantan akan di-temani   oleh *dua hingga lima ekor burung*
        one-CLF bird   male   will PASS-accompany by   two to   five CLF bird
        betina.
        female
        '[Ostriches eat in a group of six to ten.] A male bird will be accompanied by *two to three female birds*.'   (DBP Corpus)

Since a form associated with a status can be employed for higher statuses on the Givenness Hierarchy, if *di-* passive agents must be only 'type identifiable', any form can occur as *di-* passive agents. What is crucial, however, is that the Givenness Hierarchy predicts that not all forms are equally frequent; forms with a distant status becomes infrequent. This explains the fact that first and second person agents do occur but much less frequently than the third person pronouns (cf. (1a)). Even though personal pronouns are generally very high in givenness (probably 'in focus'), first and second pronouns are more given than third person pronouns.[12]

Furthermore, if a status associated with a high status is expressed by a form associated with a lower (entailed) status, Gricean (1975) maxim of quantity gives rise to the implicature that the high status does not hold. In the present context, this means that if a first or second person referent is chosen as a *di-* passive agent, it is presented as if the speech act participant at issue were a third person referent, making an otherwise subjective description objective. The following example cited by Nomoto and Kartini (2014) as a "marked case" is a case in point. The agent in this example involves coordination and refers to a first person (exclusive) plural referent.

(9)   Malaysia di-wakili   oleh *saya dan tiga   orang lagi   rakan*.
        Malaysia PASS-represent by   me   and three CLF   more colleague
        'Malaysia is represented by *me and three other colleagues*.'   (DBP Corpus)

Nomoto and Kartini's account predicts that if a referent can be referred to by either a third person pronoun or an alternative form of a lower status, the latter is more likely to occur as a *di-* passive agent. In order to verify this prediction, it is necessary to identify the givenness statuses for major referential expressions in Malay. This task is beyond the scope of this study, and I leave it for future research.

Now, what determines the low givenness status of *di-* passive agents? Since the property is

---

[12]I thank Novi Djenar for bringing my attention to this point.

concerned only with *di-* passives and is not shared by other voice types, it should be attributed to an element that only *di-* passives have. The most likely source of the low givenness status, of course, is the prefix *di-*. This is also a logical extension of the situation in the nominal domain to the verbal domain. The givenness properties of noun phrases are usually regarded as lexically specified. For example, determiners such as *this*, *that*, *the* and *a* in English encode as part of their meanings different degrees of givenness associated with the DP headed by them, as specified in the Givenness Hierarchy. Although I know of no serious study that applies the Givenness Hierarchy to Malay, demonstratives such as *itu* 'that' and *ini* 'this' are thought to encode particular degrees of givenness in addition to their deictic meanings. As seen above, the notion of givenness is not limited to noun phrases/individuals but also relevant to verb phrases/eventualities. If so, there should be morphemes that encode givenness in verb phrases too. Moreover, a parallelism between the nominal and verbal domains suggests that such morphemes should encode givenness on top of other meanings or functions. *Di-* is thought to be one such morpheme. Specifically, it encodes a low level of givenness, besides its syntactic function as a passive voice marker.

Two possibilities are conceivable as to how *di-* encodes a low givenness status. First, the prefix *di-* encodes the givenness status of the agent directly. This option may sound reasonable, but it is in fact not so straightforward. This is because under the current syntactic analysis (cf. section 2), the prefix *di-* takes a passive vP and licenses the agent argument in Spec,vP by assigning it a genitive case. In other words, the only way for *di-* to access the agent directly is by means of case assignment. One is thus forced to claim that genitive case is associated with low givenness.

Another possibility is that the low givenness of a *di-* passive agent is a consequence of the property of the passive clause (vP) of which it is part. That is to say, a property that was initially thought of as one of an argument (DP) is in fact a property of an eventuality (vP). This option may sound counterintuitive at first, but it is more compatible with the current syntactic analysis. Given that *di-* takes a vP, it is possible that it specifies the property of the vP it takes (including the kind of v, i.e. passive v). Nomoto and Kartini (2014) adopt this second option, and claim that the prefix *di-* selects for a vP describing an eventuality that is low in givenness and the low givenness of the eventuality makes its agent also low in givenness. The givenness property of the passive prefix *di-* can be formulated as a selectional restriction, as in (10).

(10)   The prefix *di-* in Voice selects for a vP describing an eventuality that is low in givenness.
        [$_{\text{VoiceP}}$ *di-* [$_{\text{vP[low givenness]}}$]]

An alternative formulation would let *di-* mark the givenness of the phrase it heads, i.e. VoiceP. Given that *di-* passives and bare passives share the same kind of vP (i.e. passive vP, cf. Table 1) and no person restriction exists on the agent in bare passives, the givenness of passive vPs can be either high or low. Thus, the formulation as a selectional restriction in (10) rejects a vP describing a highly given eventuality whereas the alternative formulation alters the givenness of such a vP to fit its requirement.

It is difficult to decide on which formulation is adequate based on empirical data. I opt for the formulation in terms of a selectional restriction, because it operates in other areas of Malay grammar (Nomoto 2013b). We have seen above that overt voice markers such as *meN-* and *di-* select for a vP of an appropriate type (see Table 1). Overt number marking by means of classifiers and reduplication restrict an otherwise uinrestricted (i.e. number-neutral) noun denotation to singularities and pluralities respectively (Nomoto 2013a). Soh and Nomoto (2011, 2015) propose that the active prefix *meN-* selects for an eventuality with stages in the sense of Landman (1992, 2008) to capture the aspectual contrast between sentences with and without *meN-*, as in

(11). This selectional restriction can be formulated in a parallel fashion to that of the passive marker *di-* above, as in (12).

(11)   a.   Harga minyak turun selama/dalam tiga   hari.
              price oil   fall   for/in   three day
              'The oil price fell for/in three days.'
        b.   Harga minyak men-[t]urun selama/*dalam tiga   hari.
              price oil   ACT-fall   for/in   three day
              'The oil price was falling for three days.'   (Soh and Nomoto 2015:151–152)

(12)   The prefix *meN-* in Voice selects for a vP describing an eventuality with stages.
        *[$_{\text{VoiceP}}$ *meN-* [$_{\text{vP}[-\text{stages}]}$]][13]

Soh and Nomoto's finding about *meN-*'s aspectual meaning indirectly supports Nomoto and Kartini's (2014) claim that *di-* encodes givenness. Since the active voice marker is more than a purely syntactic formative, it is not surprising if the passive marker also has a semantic/pragmatic function. However, Nomoto and Kartini's claim needs empirical support based on concrete data. That said, it is not very easy to prove the low givenness of *di-* passive clauses for a few reasons. I will discuss these issues in section 5.

4   Givenness of implicit passive agents

In the previous section, I put aside the implicit agent *pro* involved in the *pro* type *di-* passive with an implicit agent. The *pro* type is most frequently used amongst the three types of *di-* passives. At first brush, the prevalence of the *pro* type appears to run counter to our information-based analysis of *di-* passive agents. This is because it is generally agreed upon in the literature of information structure that the level of givenness inversely correlates with the amount of overt material, i.e. the more given a denotation is, the less phonetic material the linguistic expression associated with it contains. Gundel et al. (1993) thus identify "Ø (zero) NPs" as the form with the highest givenness status "in focus" in Mandarin Chinese, Japanese, Russian and Spanish. If what is represented as *pro* above is the same entity as their "Ø NPs," *pro* should be more given than the overt first and second person pronouns. Under Nomoto and Kartini's analysis, whereby the agent of *di-* passive is low in givenness, one would expect the *pro* type *di-* passive to be at least as infrequent as *di-* passives with first and second person agents, quite contrary to actual fact.

   Nomoto and Kartini argue that *pro* is not a kind of "Ø NP," and that *pro* is low in givenness due to its unspecified nature. If so, the prevalence of the *pro* type *di-* passive makes perfect sense. Indeed, the interpretation of *pro* is not always clear. In many cases, it seems most appropriate to analyse *pro* as "unspecified," though its referent is obvious in some cases. Moreover, according to Nomoto (to appear), the same null unspecified pronoun *pro* is employed in the following anaphoric expressions: as a possessor argument of *diri* '(physical) self' and with the intensifier *sendiri* 'own, alone', as shown in (13a) and (13b) respectively.

(13)   a.   diri *pro*   'oneself'   b.   kereta *pro* sendiri   'one's own car'
              diri-ku   'myself'   kereta-ku sendiri   'my own car'
              diri-mu   'yourself'   kereta-mu sendiri   'your own car'
              diri-nya   'himself/herself'   kereta-nya sendiri   'his/her own car'

---

[13]Soh and Nomoto assume that *meN-* occupies v rather than Voice. Hence, their original formulation differs slightly from the one presented here.

Notice that *pro* occurs in the same position as enclitics do. The agent position of *di-* passives also allows enclitics. Both positions are associated with genitive case. The relevant position is not associated with accusative case, given that *pro* cannot occur as a preposition object position (e.g. *di-semak oleh pro [PASS-check by]). Incidentally, Nomoto and Kartini claim that the agent in bare passive is obligatory because the agent position in bare passives is not a genitive case position allowed for *pro*.[14]

While the unspecified nature of *pro* is sufficiently reasonable, considering it as distinct from Ø causes proliferation of covert forms.[15] I thus argue that, insofar as Malay is concerned, *pro* and Ø capture different stages of the same entity: *pro*/Ø is inherently low in givenness (in the lexicon, before interpretation) but can be understood as referring to highly given referents through contextual restriction (after interpretation).

The situation is comparable to the interpretation of pronouns. Suppose that John is talking with Mary about their mutual friend Ali. In this situation, *I* refers to John, *you* to Mary and *he* to Ali. However, these are by no means the lexical meanings of *I*, *you* and *he*. The observed meanings are the results of interpretation with respect to a particular context. For example, the lexical meaning of *he* is a singular individual who is not a speech act participant and is male. The context restricts the set of its possible referents further until the set becomes a singleton consisting only of Ali.

Pro in *di-* passives can be restricted in the same way. Although the referent of *pro* can be left unspecified, when it undergoes contextual restriction, it can sometimes refer to a highly given referent such as first and second person referents. In (14), the originally unspecified referent of *pro* is restricted by the context to a first person referent, i.e. the writer of the article in question or 'the media' including the writer. The person who met *beliau* 'him' cannot be the reader (second person) or a third party excluding the writer/reader (third person). Likewise, in (15), the originally unspecified referent of *pro* is restricted to first person referents, this time, overtly by the agentive phrase *oleh kita* 'by us'.

(14)   Beliau di-temui   *pro* selepas merasmikan Seminar Pengurusan Sukan Institusi
        he   PASS-meet   after officiate   seminar management sport institution
        Pengajian Tinggi (IPT) 2010 di UiTM kampus Khazanah Alam Bandar Jengka di
        study   high   2010 at UiTM campus Khazanah Alam Bandar Jengka at
        sini.
        here
        'He was met (by *pro*) after he had officiated the 2010 Higher Academic Institution Sports Management Seminar at UiTM, Khazanah Alam Bandar Jengka campus here.'
        (*Utusan Malaysia*, 01/01/2011)

(15)   Usia tidak mengampunkan segala dosa yang di-buat *pro* oleh kita.
        age not   forgive   all   sin REL PASS-do   by   us
        'Age does not forgive all the sins that were committed by (*pro* =) us.'   (DBP Corpus)

Classical Malay provides a case where an overt pronominal passive agent is restricted by an *oleh* 'by' phrase. *Di-* passives in Classical Malay have an additional subtype that is no longer available in Modern Malay. I refer to this type as the 'hybrid type', as the agent is expressed simultaneously by an *oleh* phrase as well as the third person enclitic *-nya*. It is situated between

---

[14]This means that obligatory agent expression is not a defining property of bare passives. Rather, it is a matter of licensing condition of *pro*. If so, it is predicted that the agent could be optional (i.e. allow *pro*) in comparable constructions in other languages. Kroeger (2014) reports two candidates for such languages: Pangutaran and Mualang.

[15]I put aside instances of Ø that arise from ellipsis.

the *oleh* type and the DP type.[16] An example is given in (16). A direct translation of the clause that reflects the compositional interpretation would be '… by a third person referent who is him'.

(16)   maka lalu di-baca-*nya oleh* baginda surat itu.
        and   then PASS-read-3 by   him   letter that
        'and then the letter was read by him.'   (*Hikayat Maharaja Marakarma*: 139b)[17]

Nomoto and Kartini (2014) report that first and second person agents are found least frequently in the DP type. With the revised syntactic structures in (3), Nomoto and Kartini's analysis offers a possible explanation for this fact. Unlike the *pro* and *oleh* types, the agent in Spec,vP is not further restricted by the context or an *oleh* phrase in the DP type. This suggests that the givenness of the DP in Spec,vP is more important as a determinant of the well-formedness of *di-* passives than that of the final referent after contextual restriction. If the ultimate source of the pressure against first and second person agents is *di-* in Voice, its effect applies to its c-command domain, i.e. vP. It is the agent DP in Spec,vP that is directly affected by the givenness constraint of *di-*. In the *pro* and *oleh* types, the agent comes to refer to a first and second person referent not because of *pro* in Spec,vP but because of the *oleh* phrase or context, which are added outside the scope of *di-*. By contrast, the DP type cannot have a first and second person referent unless the DP in Spec,vP itself is first or second person.

5   Givenness of eventualities

In this section, I discuss issues related to Nomoto and Kartini's (2014) claim that the eventuality described by a *di-* passive verb phrase is low in givenness. Although this claim offers a way to capture the low givenness of the *di-* passive agent without stipulation, it is not so easy to prove its correctness. There are two main reasons for this. First, theories of givenness have developed through studies of noun phrases, which typically denote individuals rather than eventualities. Second, in Modern Malay, most passive clauses have a preverbal subject, unlike earlier stages of the language and some regional Malay varieties. Consequently, the informational status of the verb phrase gets obscured by that of the subject. It is known that the clause-initial noun phrase (i.e. subject) in Malay is highly topical, that is, it often sets the topic on which the rest of the clause make comments (Alsagoff 1992; Nomoto 2009). Before going into details, I should qualify that due to these and other reasons, the rest of this section is rather preliminary in nature. Its aim is more to share problems and their possible solutions than to solve them persuasively.

---

[16]This construction is reminiscent of Legate's (2012) analysis of Acehnese passives, whereby verbal prefixes in v restrict the agent, as shown in (i). Note that in terms of their semantic function, the verbal prefixes in Acehnese correspond to the enclitic *-nya* in Malay rather than the passive prefix *di-*; Acehnese does not have an overt morpheme corresponding to *di-* in Malay. This supports the current analysis, where (Malay) *di-* occupies a head higher than v.

(i)   a.   Aneuk miet nyan *di*-kap   (*lé* uleue nyan).
             child   small that 3FAM-bite by snake that
             'The child was bitten (by the snake).'
       b.   Aneuk miet nyan *meu-/   neu-/   geu*-tingkue *lé* kamoe/ droeneuh/ gopnyan.
             child   small that 1EXCL- 2POL- 3POL-carry   by us   you   him/her
             'The child is carried by us/you/him/her.'   (Legate 2012:497)

[17]Data obtained from the Malay Concordance Project of the Australian National University (`http://mcp.anu.edu.au`).

5.1   Previous studies on the givenness of non-individuals

To my knowledge, it is Schwarzschild (1999) who first provided an explicit definition of givenness for non-individuals. His basic idea is that while givenness of individuals is based on coreference, that of non-individuals is determined by entailment. Consider the example in (17). Uppercase letters here and elsewhere indicate pitch accents.

(17)   NObody murdered JOHN although BOB WANted to kill him.

(adapted from (22) in Riester 2008:79)

Given the first verb phrase *murdered John*, the event denoted by the second verb phrase *kill him* counts as given in some sense. It is possible that the events described by the two verb phrases are identical. That is to say, the killing in question is achieved by means of murder. But even if not, a murdering event necessarily involves a killing event. So, the event of killing John is not completely new. Importantly, in this second case, givenness is not based on coreference but a lexical relation; specifically *murder* is a hyponym of *kill*. As Baumann and Riester (2012) point out, hyponymy like this is verified by entailment, as shown in (18). Schwarzschild developed a procedure to make such verification possible, including existential type-shifting, which turns verb phrase meanings into propositions, as found in (18).

(18) $\exists x.\mathrm{murder}(x, j)$ *entails* $\exists x.\mathrm{kill}(x, j)$
         (someone murdering John)   (someone killing John)

Baumann and Riester (2012) push this idea a step further and propose to differentiate two kinds of givenness, i.e. referential givenness and lexical givenness. These two kinds of givenness are respectively based on coreference and lexical relations such as identity, synonymy and hyponymy. The new notion of lexical givenness captures the fact that sometimes an expression is treated as given due to the presence of a related expression rather than a coreferential entity. For example, in (19), *a big German Shepherd* and *Anna's dog* are not corerential. However, the word *dog*, which heads the latter noun phrase, cannot be accented, which means that it is treated as given. This is because *a big German Shepherd* is a hyponym of *dog*. Similarly, in (20), the two occurrence of *Italian* are not coreferential, with the first one denoting a language and the second one a nationality. However, the second occurrence of *Italian* cannot be accented, and hence is treated as given, because the language name *Italian* is closely related to *Italian* as nationality.

(19)   On my way home, a big German Shepherd barked at me. It reminded me of ANna's
       dog.   (Baumann and Riester 2012:133)

(20)   (Why do you study Italian?) I'm MARried to an Italian.   (Büring 2007)

While Schwarzschild distinguishes between individual-denoting (type *e*) and non-individual denoting (non-type *e*) expressions in his definition of a single notion of givenness, Baumann and Riester associate referential and lexical givenness with referential and non-referential expressions respectively. The borderlines coincide in the case of nominal expressions, but not in non-nominal expressions. For instance, in Schwarzschild's definition, a referential event is given if it is entailed by its antecedent. In Baumann and Riester's theory, on the other hand, a referential event is considered (referentially) given if it has a coreferential antecedent. Unfortunately, Baumann and Riester focus on nominal expressions and do not discuss non-nominal expressions. Nevertheless, Baumann and Riester's theory has a conceptual advantage over Schwarzschild's. It is not obvious in Schwarzschild's theory why only individual-denoting referential expressions

invoke coreference. See Appendix for further details on Baumann and Riester's referential and lexical givenness, and their relation to the Givenness Hierarchy of Gundel et al. (1993).

5.2   Challenges

While the criteria developed for individuals can be extended to eventualities, an actual (referential) givenness identification task is not so straightforward. The difficulty is due to different natures of individuals and eventualities. First, while it is common that an individual is repeatedly referred to in discourse, an eventuality is usually not repeated. Hence, criteria for givenness statuses based on coreference are not helpful in many cases. In theory, an event is high in givenness if the same action involving the same participants has occurred in immediate discourse. Thus, among the four continuations to (21) in (22) (temporal locations put aside), the event described by sentence (22a) is higher in givenness than those described by the other three. In actual discourse, however, such a repetition situation is rare.

(21)   [The speaker and addressee are talking about their roommate John.]
        John$_1$ was eating your bread this morning.

(22)   a.   He$_1$ was eating your bread again.   (same action, same agent, same theme)
        b.   He$_1$ was eating your eggs too.   (same action, same agent, different theme)
        c.   Mary was eating your bread too.   (same action, different agent, same theme)
        d.   Mary was watching TV.   (different action, different agent, different theme)

Second, eventualities typically involve more than one individual. The same action can be conducted by the same agent on a different theme, as in (22b); it can also be conducted on the same theme by a different agent, as in (22c). Is there a difference in givenness between these two cases? If so, which event is higher in givenness?

Relating to the second difference, assuming that the event argument of a verb is existentially closed at vP (or VoiceP), a sentence involves at least two levels of referential givenness for eventualities, corresponding to different syntactic nodes denoting eventualities: vP (or VoiceP) and TP.[18] Both vP and TP eventualities can serve as an antecedent for nominal expressions indicating particular degrees of givenness such as *(do) it*, *this* and *that* (see (6) for an example of a TP eventuality referred to by *this/it*). Although (22a) and (22c) have different degrees of givenness at the TP level, they do not differ at the vP level, as shown in (23).

(23)   a.   [$_{TP}$ He$_1$ was [$_{vP}$ eating your bread] (again)] (= (22a))
        b.   [$_{TP}$ Mary was [$_{vP}$ eating your bread] (too)] (= (22c))

Moreover, at the vP level, voice alternation affects givenness. Consider the passive counterpart of (23a).

(24)   [$_{TP}$ Your bread was [$_{vP}$ eaten by him$_1$] (again)]

While the sentence as a whole describes the same event as (23a), i.e. John's eating the addressee's bread, the events described at the vP level are different between (23a) and (24), i.e. someone's eating the addressee's bread and John's eating something. The givenness statuses associated with (23a) and (24) should also differ accordingly.

---

[18]I assume that the lexical verb (V) and its projection VP are assigned lexical givenness, in line with Baumann and Riester's (2012) treatment of the lexical noun (N) and its projection NP.

5.3   "Foregrounding" in discourse as a low givenness indicator

Nomoto and Kartini (2014) do not address the issues pointed out above when they claim that *di-* passive verb phrases are low in givenness. Instead, they reinterpret Hopper's (1983) notion of "foregrounding" as indicating low givenness.

   Hopper studies discourse functions of three clause types in the Early Modern Malay text *Hikayat Abdullah*: 'active', 'passive' and 'ergative'. Hopper's active construction is the same as that in the present study. He argues that what is collectively refrerred to as the passive in the present study in fact consists of two distinct voices, though they share the same morphology (i.e. *di-* and proclitics) and are sometimes indistinguishable. The two constructions are distinguished by functional and formal criteria, of which he states the former is primary. In Hopper's definition, the "passive" has the discourse function of backgrounding, and the theme precedes the verb, as in (25). By contrast, the "ergative" foregrounds events, and the theme follows the verb, as in (26a). Clauses with a preverbal theme are not "passive" but "ergative" if (i) the theme is followed by the particle *pun* or *semua-nya* 'all of them', as in (26b), or (ii) it has a foregrounding function (e.g. part of an event sequence), as in the first clause of (26c).

(25)   Hopper's "passive"

   maka dua puncha kiri kanan itu di-matikan
   then   two ends   left right   the PASS-knot
   'and the two ends to the right and left are knotted'   (Hopper 1983:71)

(26)   Hopper's "ergative"

   a.   di-champakkan-nya puntong cherutu itu   ka-dalam kapal
        PASS-throw.away-3 stub   cheroot that into   ship
        'and they threw away the stubs into the boat'

   b.   Maka segala pengana itu pun di-bahagikan-lah   ka-pada segala
        and   all   cakes   the PUN PASS-distribute-PART to   all
        budak-budak
        boy.PL
        'Then all the cakes were passed around to all the boys'

   c.   maka duit   itu di-ambil oleh ibu-bapa-nya, di-belikan-nya   penganan
        then   money the PASS-take by   parents-his   PASS-use.to.buy-3 cakes
        atau barang-barang makanan, di-makan-nya
        or   things   eating   PASS-eat-3
        'Then his parents take the money and use it to buy cakes or other things to eat,
        and they eat them.'   (Hopper 1983:72–73)

He demonstrates the foregrounding and background difference between the two constructions by examining the Transitivity index of each of the 100 clauses (= 50 "ergative" + 50 "passive" clauses). Each clause is inspected as to whether it exhibits a positive (i.e. more Transitive) or negative value for the ten Transitivity parameters proposed by Hopper and Thompson (1980). The results show that that "ergative" clauses are more Transitive than "passive" clauses with respect to all Transitivity parameters, especially kinesis (action involving movement), punctuality (no discernible duration), aspect (telic) and agent potency (animate). Since high Transitivity reflects foregrounding in discourse in Hopper and Thompson's Transitivity Theory, the high Transitivity of the "ergative" justifies Hopper's definition of the construction as a foregrounding construction.

   It must be noted here that Hopper treated all clauses with a preverbal theme as "passive" to guarantee the objectivity of the examination. Hence, Hopper's "ergative" and "passive" data

roughly correspond to verb-initial and theme-initial passives in the term of the present study. In short, verb-initial passives are more Transitive than theme-initial passives, and hence have a foregrounding function. I think that this statement about Early Modern Malay is also valid in Modern Malay, though it is not as evident as in Early Modern Malay due to the general scarcity of verb-initial passives.[19] Although I do not find Hopper's functional definitions and identification of voice categories very useful,[20] his characterizations of verb-initial passives are worth quoting. He writes that a verb-initial passive clause "focuses purely on the event—the change—itself" and "narrates sequenced events which pertain to the main line of the discourse" (Hopper 1983:84). Verb-initial passive clauses are used in the same way in Modern Malay, though they are limited to the literary genre and certain subordinate contexts.

Hopper states that the foregrounding function of verb-initial passive clauses is obliterated by the positioning of a noun phrase before the verb, which he analyses as "a device for arresting the flow of the discourse and holding up the action by momentarily focusing attention away from ACTIONS to PARTICIPANTS" (87). This quote indicates that by "foreground" Hopper means "require or draw attention of the addressee." In terms of givenness, it is a denotation which is not already given enough in the address's consciousness that requires his/her special attention. Hence, in verb-initial passive clauses, the verb is low in givenness. The verb is arguably low in givenness in theme-initial passive clauses too, but the low givenness of the verb does not come to the fore because of the fronted theme.

If the reasoning above is justified, givenness of eventualities correlates with (or possibly *is*) Transitivity. The second last sentence of Hopper and Thompson's Transitivity article is suggestive of this connection:

> While we claim that the discourse distinction between foregrounding and back-grounding provides the key to understanding the grammatical and semantic facts we have been discussing, we also explicitly recognize that grounding itself reflects a deeper set of principles—relating to decisions which speakers make, on the basis of their assessment of their hearers' situation, about how to present what they have to say. (Hopper and Thompson 1980:295)

It seems to me that the relevant "deeper set of principles" is in fact givenness.[21] Givenness has to do with the speaker's assessment of the addressee's consciousness/attention state and knowledge with regard to a referent, which can be eventualities as well as individuals. Based on their assessment, speakers choose a form that encodes the most appropriate givenness status, e.g. *it* over *that* (English DPs), *di-* over Ø (Malay VoicePs).

6   Conclusion

This paper has reviewed and elaborated on Nomoto and Kartini's (2014) analysis of the person restriction on the agent of *di-* passives in Malay. In doing so, I have made the following two main claims. First, the implicit agent *pro* involved in *di-* passives is not distinct from "Ø NPs" in the Givenness Hierarchy: *pro*/Ø is lexically low in givenness due to its unspecified nature and often interpreted as referring to a highly given referent, including speech act participants.

---

[19]See Cumming (1991) and Djenar (2015) for descriptions of verb-initial passives in Modern Indonesian and how they differ from Classical Malay.

[20]See Kroeger (2014) for a critical review of functional definitions of voice categories in Malay/Indonesian.

[21]Reinhart (1984) expresses a different view. She proposes that grounding in discourse is "a temporal extension of the principle governing the spatial organization of the visual field into figure and ground."

Second, morphemes exist that encode givenness not only for noun phrases/individuals but also for verb phrases/eventualities. The paper has also discussed issues concerning givenness of eventualities. The discussion is still premature. Especially, more empirical work is necessary, to demonstrate the low givenness status of *di-* passive verb phrases in Modern Malay and the connection between givenness and Transitivity cross-linguistically.

Appendix. Referential and lexical givenness statuses, and the Givenness Hierarchy

Baumann and Reister's (2012) referential givenness statuses (cf. Table 2) basically correspond to one of the coding criteria for the Givenness Hierarchy (Gundel et al. 2006), and hence match nicely with the Givenness Hierarchy, as shown in Figure 1 (Baumann and Riester 2012:143).[22] This means that one can compare different statuses with respect to the degree of givenness: one status is more given than another.

Table 2: Referential givenness (Baumann and Riester 2012:138)

r-given	anaphor corefers with antecedent in previous discourse
r-given-sit	referent is immediately present in text-external context (in particular discourse participants)
r-unused-known	discourse-new item which is generally known
r-given-displaced	coreferring antecedent does not occur in previous 5 intonation phrases or clauses
r-bridging	non-coreferring anaphor, dependent on previously introduced scenario
r-environment	refers to item in text-external context (conversational environment)
r-bridging-contained	bridging anaphor which is anchored to an embedded phrase
r-unused-unknown	discourse-new item which is identifiable from its linguistic description but not generally known
r-new	specific or existential indefinite introducing a new referent
r-cataphor	item whose referent is established later on in the text

activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
		r-unused-known		r-bridging				
r-given				r-bridging-contained				
r-given-sit		r-given-displaced		r-environment		r-cataphor		r-new
				r-unused-unknown				

Figure 1: Referential givenness and the Givenness Hierarchy

Baumann and Reister's lexical givenness statuses are summarized in Table 3.[23] They discuss neither the relative degrees of givenness for these lexical givenness statuses nor their cognitive statuses, i.e. positions on the Givenness Hierarchy. However, it would be possible to rank these

---

[22]I omitted 'r-generic'. Baumann and Riester align their 'r-new' with 'referential' on the Givenness Hierarchy. This is because they only discuss the middle four statuses available in the latter.

[23]I substituted "expression" for "noun" in their article.

categories to each other, because I see parallelisms between the categories of referential and lexical givenness. To begin, 'l-given-same' is obviously the lexical counterpart of 'r-given', and 'l-new' is that of 'r-new'. The abstract relationship among referents underlying bridging inference involved in 'r-bridging(-contained)' is arguably identical to that holding between linguistic expressions ('l-given-supr', 'l-accessible-sub', 'l-accessible-other'). Furthermore, 'l-given-syn' is comparable to 'r-given-displaced' in that both are pretty close to the highest givennnes but do not quite reach it because they do not share the same phonetic form (sound) or attention span (time) with the antecedent. The preliminary ranking resulting from these parallelisms is: l-given-same > l-given-syn > l-given-supr, l-accessible-sub, l-accessible-other > l-new.

Table 3: Lexical givenness (Baumann and Riester 2012:144)

l-given-same	recurrence of same expression
l-given-syn	relation between expressions at the same hierarchical level (synonyms)
l-given-supr	expression is lexically superordinate to previous noun
l-accessible-sub	expression is lexically subordinate to previous noun
l-accessible-other	two related expressions, whose hierarchical lexical relation cannot be clearly determined
l-new	expression not related to another expression within last 5 intonation phrases or clauses

References

Alsagoff, Lubna. 1992. Topic in Malay: The Other Subject. Doctoral Dissertation, Stanford University.

Arka, I Wayan, and Christopher D. Manning. 1998. Voice and grammatical relations in Indonesian: A new perspective. In *Proceedings of the LFG98 Conference*, ed. Mirriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications.

Asmah Hj. Omar. 2009. *Nahu Melayu Mutakhir*. Kuala Lumpur: Dewan Bahasa dan Pustaka, 5th edition.

Baumann, Stefan, and Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In *Prosody and Meaning*, ed. Gorka Elordieta and Pilar Prieto, 119–162. Berlin: Mouton de Gruyter.

Büring, Daniel. 2007. Semantics, intonation and information structure. In *The Oxford Handbook of Linguistic Interfaces*, ed. Gillan Ramchand and Charles Reiss, 445–474. Oxford: Oxford University Press.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In *Subject and Topic*, ed. Charles N. Li, 25–55. New York: Academic Press.

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In *Subject and Topic*, ed. Charles N. Li, 57–99. New York: Academic Press.

Cole, Peter, Gabriella Hermon, and Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118:1500–1553.

Cumming, Susanna. 1991. *Functional Change: The Case of Malay Constituent Order*. Berlin: Mouton de Gruyter.

Dardjowidjojo, Soenjono. 1978. *Sentence Patterns of Indonesian*. Honolulu: University of Hawaii Press.

Djenar, Dwi Noverini. 2015. Active/passive alternation and switch reference in Indonesian narrative discourse. In *Proceedings of the Second International Workshop on Information Structure in Austronesian Languages*.

Grice, H. Paul. 1975. Logic and conversation. In *Syntax and Semantics: Speech Acts*, ed. Peter Cole and Jerry L. Morgan, volume 3, 41–58. New York: Academic Press.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274–307.

Gundel, Jeanette K., Nancy Hedberg, Ron Zacharski, Ann Mulkern, Tonya Custis, Bonnie Swierzbin, Amel Khalfoui, Linda Humnick, Bryan Gordon, Mamadou Bassene, and Shana Watters. 2006. Cording protocol for statuses on the Givenness Hierarchy. URL `http://www.sfu.ca/~hedberg/Coding_for_Cognitive_Status.pdf`.

Hedberg, Nancy. 2014. Applying the Givenness Hierarchy framework: Methodological issues. In *Proceedings of the International Workshop on Information Structure of Austronesian Languages*, 1–18.

Hopper, Paul, and Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56:251–299.

Hopper, Paul J. 1983. Ergative, passive, and active in Malay narrative. In *Discourse Perspectives on Syntax*, ed. Folora Klein-Andreu, 67–88. New York: Academic Press.

Kartini Abd. Wahab, and Hiroki Nomoto. 2012. Passives without 'by' in Malay. Paper presented at the 19th Meeting of the Austronesian Formal Linguistics Association (AFLA).

Keenan, Edward L., and Matthew S. Dryer. 2007. Passive in the world's languages. In Timothy Shopen (ed.), *Language typology and syntactic description: Volume I, clause structure*, 325–361. Cambridge: Cambridge University Press.

Kroeger, Paul. 2014. Passive agents in malay: The binding properties and discourse functions of agentive =*nya*. In *Current Trends in Malay Linguistics*, ed. Siaw-Fong Chung and Hiroki Nomoto, volume 57 of *NUSA*, 5–29. Jakarta and Tokyo: Universitas Katolik Indonesia Atma Jaya and Tokyo University of Foreign Studies.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Landman, Fred. 1992. The progressive. *Natural Language Semantics* 1:1–32.

Landman, Fred. 2008. 1066: On the differences between the tense-perspective-aspect systems of English and Dutch. In *Theoretical and Crosslinguistic Approaches to the Semantics of Aspect*, ed. Susan Rothstein, 107–166. Amsterdam: John Benjamins.

Legate, Julie Anne. 2010. The structure of implicit agents in passives. Paper presented at NELS 41.

Legate, Julie Anne. 2012. Subjects in Acehnese and the nature of the passive. *Language* 88:495–525.

Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese*. Cambridge, MA: MIT Press.

Nomoto, Hiroki. 2006. Voice in Colloquial Malay relatives. In *Gengojouhougaku Kenkyuuhoukoku 12*, ed. Yoichiro Tsuruga, Nobuyuki Miyake, and Yuji Kawaguchi, 97–116. Tokyo University of Foreign Studies.

Nomoto, Hiroki. 2009. On indefinite subjects of pivot verbs in Malay. *International Journal of Corpus Linguistics* 14:221–254.

Nomoto, Hiroki. 2013a. Number in Classifier Languages. Doctoral Dissertation, University of Minnesota.

Nomoto, Hiroki. 2013b. On the optionality of grammatical markers: A case study of voice marking in Malay/Indonesian. In *Voice Variation in Austronesian Languages of Indonesia*, ed. Alexander Adelaar, volume 54 of *NUSA*, 121–143. Jakarta and Tokyo: Universitas Katolik Indonesia Atma Jaya and Tokyo University of Foreign Studies.

Nomoto, Hiroki. to appear. Decomposing Malay anaphoric expressions. In *Proceedings of the Twenty-First Meeting of the Austronesian Formal Linguistics Association (AFLA)*.

Nomoto, Hiroki, and Kartini Abd. Wahab. 2014. Person restriction on passive agents in Malay: Information structure and syntax. In *Current Trends in Malay Linguistics*, ed. Siaw-Fong Chung and Hiroki Nomoto, volume 57 of *NUSA*, 31–50. Jakarta and Tokyo: Universitas Katolik Indonesia Atma Jaya and Tokyo University of Foreign Studies.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information-status. In *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, ed. William C. Mann and Sandra A. Thompson, 295–325. Amsterdam: John Benjamins.

Reinhart, Tanya. 1984. Principles of gestalt perception in the temporal organization. *Linguistics* 22:779–809.

Riester, Arndt. 2008. The Components of Focus and Their Use in Annotating Information Structure. Doctoral Dissertation, Universität Stuttgart. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS). Vol. 14(2).

Santiago, Paul Julian. 2014. Information structure, voice alternation and text genre in Kalanguya. Paper presented at Workshop on 'Cross-linguistic Perspective on the Information Structure of Austronesian Languages'.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7:141–177.

Sneddon, James Neil, Alexander K. Adelaar, Dwi N. Djenar, and Michael Ewing. 2010. *Indonesian: A Comprehensive Grammar*. London: Routledge, 2nd edition.

Soh, Hooi Ling, and Hiroki Nomoto. 2011. The Malay verbal prefix *meN-* and the unergative/unaccusative distinction. *Journal of East Asian Linguistics* 20:77–106.

Soh, Hooi Ling, and Hiroki Nomoto. 2015. Degree achievements, telicity and the verbal prefix *meN-* in Malay. *Journal of Linguistics* 51:147–183.

Utsumi, Atsuko. 2014. Newly introduced NPs and given NPs in Bantik discourse. In *Proceedings of the International Workshop on Information Structure of Austronesian Languages*, 183–192.

Willett, Marie Louise. 1993. Object preposed constructions in Malay. Master's thesis, Memorial University of Newfoundland.

Pragmatic particles and information structure
in colloquial Indonesian dialogue

Yoshimi Miyake

Akita University

Abstract

In this paper, four pragmatic particles from Jakarta Indonesian, i.e. *sih, kok, lho, dong*, will be examined. I will address these particles by examining them from the perspectives of new topic vs old topic, new information vs old information, certainty vs uncertainty, positive evaluation vs negative evaluation, and finally, strong command vs soft command. The data comes from dialogues in two contemporary Indonesian films depicting life in Jakarta.

1. Introduction

Linguists of Indonesian language have claimed that Indonesian pragmatic particles have no significant referential semantic function, but rather they express speaker mood or are used for "emphasis" (Sneddon 2006, Errington 1985). However, these explanations do not answer an important question, namely 'what are the differences between particles when the general meanings of them are the same?'

Indonesian native speakers often equate these particles with full lexical items. For example the particle *deh* denotes *jadi* 'to be realized', or *sudah* 'already completed.' (Atmosumarto 1994). The question then arises, why do the speakers choose to substitute *deh* for *jadi* or *sudah*?

Based on the assumption that they reflect the speaker and hearers' access to the information and knowledge that the speaker wishes to convey, this paper will attempt to consider the functions of discourse particles as they are used in conversations depicted in film. I will also argue that in addition to differences in access to information the particles reflect the emotional condition of the interlocutors. That is, I hypothesize that in informal or casual conversation, the particles play significant roles in linking the knowledge and information accessible to speakers and hearers to their emotional states. I also argue that since many of the particles are polysemous, they require context dependent analysis.

Bataone (n.d.) compares the particles to Indonesian verbs, nouns, adjectives, and adverbs. He explains *sih* as a particle indicating inference or uncertainty, *kok* for surprise, accusation, or negative feeling, *deh* for softening imperatives or leading to finishing up certain tasks. He explains *lho* as a particle which expresses surprise.

Based on the hypothesis that the Indonesian cultural value of solidarity is

marked by the use of particles Wouk (2001) tried to place them along a scale of solidarity. According to her, the tags *kan* and *ya*, for example, denote a certain level of familiarity or informal relationship between the speaker and hearer .

Faizah (2007) conducted extensive studies of *ya* and other particles, focusing on contour variations. She classified the functions of the particles along three axes: Coherence; focusing on speech act types such as statement, command and warning; Topic, focusing on whether it is a new topic or old topic; and finally, politeness, constructing and maintaining a desired level of solidarity.

In this study, I will mainly analyze the particles *sih*, *lho*, *kok*, and *dong*, although brief discussion of another particle *dong* will be added when needed for making comparisons.

1. *sih*

Sih is a phrase final or sentence final particle. (It can also be used in an independent intonation unit.). In the dialogues of three films, *Ada apa dengan Cinta?* 'What's up with Cinta "love"?'. *Arisan* 'Gathering' and *Arisan 2* 'Gathering 2', the most frequently appearing particle is *sih*. *Sih* has the following functions:

1.1. Ironical question: Interrogative + *sih*

Immediately following interrogative markers, *sih* shows a more ironical attitude of the speaker as shown in (1) to (4):

(1) (Observing the rude behavior of the addressee)

kamu itu apa, sih?

you DEM what, *sih*

'What are you, *sih*?'

(2) (Referring to an instance of arrogant behavior by a man)

(dia) laki-laki gimana, sih,   ya?

He   man   how   *sih*   yeah

'What kind of man is he, *sih*,   yeah?'

(3) (noticing the suitcase is very heavy),

isinya   apa   aja   sih?

Content-DEF.   what   just   *sih*

'What's inside, *sih*?'

(4)　(Commenting on a strange hair-do)

　　rambut kamu ngapain, sih?

　　hair　　you　do what　sih

　　What did you do with that hair, *sih*?

Sentence (5a) with *sih* shows that the questioner child is already tired, and he is not happy that somebody else may be coming.　A lack of eagerness, fatigue, or irritation is denoted by *sih*.

(5a) (A little child asks)

　　om, siapa　　lagi,　　sih,　　yang　　datang?

　　uncle who　　else　　sih　　RE.Pro. come

　　'Uncle, who else *sih* is coming?'

(5b) without *sih* is a pure question sentence:

(5a) *om,　　　siapa　lagi　　yang　　dating?*

　　uncle　　who　　else　　RE.Pro. come

　　'Uncle, who else is coming?'

1.2.　Complaining

　Sih added to a descriptive statement connotes a negative feeling or complaining.

(6)　*primitive banget, sih,　　airportnya.*

　　simple　　very　　sih　　airport

　'The airport is so primitive, *sih*.'

(7) (A child is crying)

　　kamu, sih,　　buat　nangis.

　　you　sih,　　make　cry

　　'You *sih* are the one who made him/her cry.'

1.3.　Topic marker, comparing and contrasting

　Sih can be used as a topic marker for comparing or contrasting.　In (8) and (9) each of the speakers compares the current situation with past.

(8) *tahun lalu, sih, mamanya minta tas Birkin special edition.*

   year   past sih mom-DEF. ask   bag   Birkin special edition

  '(Compared with this year) last year, *sih*, my mom asked for a special edition Birkin bag.'

(9) *dulunya   sih,   datang ke   sini   cuma   main   saja.*

   before   *sih*,   come   to   here   only   play   just

  '(Compared with now) I used to come here only to have fun.'

   In (10), the speaker compares what he wants and the reality.

(10)   *maunya sih, gitu,*

       desire   sih, that way

    tapi ntar aku   disangka   orang   gila   lagi

    but later I   be considered   person   crazy   again

    sama   orang-orang   Jakarta.

    with/by   people PL.   Jakarta

    '(Differently from the reality), my desire is like that, but later I will be considered a crazy person again by Jakartan people.'

In (11) the speaker is comparing her own condition to the addressee:

(11)   (Hearing that the woman has cancer)

    buat   saya,   sih,   berkah.

    for   me,   *sih*,   blessing

    'For me, it is a blessing

  saya juga kanker….rahim.

  me   also cancer …..uterus

  'I also have cancer, uterine cancer.'

1.4.   Reasoning/justification

   Sih connotes reasoning or justifying.

(12)   (I don't buy that)

   soalnya mahal, sih

   because expensive *sih*

   'Because it is expensive, *sih*.'

1.5.   Idiomatic expression: *masa sih?*

Masa sih meaning 'how come' is an idiomatic expression.

(13)   (being given a compliment for looking slimmer)

masa sih,		*ini*	*kayaknya*	*efek*	*baju,*	*lho.*
how come, *sih*,		this	apparently	effect	outfit	*lho*

'How come, *sih*, it seems an effect of the outfit, *lho.*'

(14)   (Hearing that Sakti was afraid of telling Mei that he is a gay.)

Sakti, masa sih		*gue*	*pikiran*	*secepit*	*gitu?*
Sakti, how come *sih*,	I	think	that-narrow	like that

'Sakti, how come *sih* you thought that I am that narrow-minded?'

2. *Lho*

Lho can be located at the sentence initial position as an exclamation for surprise, which I will discuss later.   *Lho* as a phrase-final or sentence final-particle can be compared or contrasted with *sih* discussed in Section I. I will start my discussion by comparing them in sentence-final position.

2.1.   Information and knowledge: *sih* and *lho*

When used at the end of a phrase/sentence, both particles *lho*, and *sih* emphasize the speaker's emotion (Bataone n.d.).   More importantly, *lho* provides new information while *sih* connotes a rather negative statement.   Compare (15a) and (15b).

(15a) *itu (Sakti)*	*orangnya*	*baik*	*banget, lho.*
(15b) itu	orangnya	baik	banget, sih.
that	personality-DEF. good	very	lho/sih

'He is such a nice guy, *lho/sih.*'

The difference between (15a) and (15b) is only the use of *lho* rather than *sih.* In (15a), the speaker introduces a character called Sakti as a very nice person, on the other hand in (15b), the speaker introduces Sakti as a very nice person but suggest that there may be some problems in him. (15a) is new information, wheares (15b) could be a response sentence to a certain statement such as "Sakti helped me".

In (16a), a child speaker uses *lho* to provide new information about himself.   Then the

uncle confirms the information, praises him, by using *sih* in (16b),

(16a)   *uncle, aku sekarang   udah   bisa   nulis   namaku   lho,   uncle.*
        uncle, I   now   already   can   write   name-my   lho,   uncle.
        'Uncle, I can already write my name, *lho*, uncle.'

(16b)   *eh,   pinter   banget, sih.*
        EXC.   smart   very,   sih
        'Wow, (you are) so smart, *sih*.'

2.2.   Definiteness and emphasis

Lho connotes definiteness. In (17) the speaker boasts that she has joined five social meetings, by adding *lho*.

(17) *aku sebulan   ikut   arisan   lima   kali,   lho.*
   I   every month   join   arisan   five   times,   lho
   'I join *arisan* (socialization meeting) five times a month, lho.'

2.3.   Warning

Lho warns or gives strong advice. *Lho* added to an imperative form connotes a bad result if the hearer does not follow the command.

(18) *tunggu, lho.*
   wait IMP, lho.
   'Wait, (otherwise…) *lho*.'

   (18a) without *lho* is a simple command form.
   Cf. (18a)   Tunggu,
                  'Wait.'

(19) *kamu makan dulu,   nanti   lapar,   lho.*
   you   eat   before, later   hungry lho.
   'You eat now, otherwise you will get hungry, *lho*.'

(20) *cepetan   bikin   appointment keburu,   penuh lho,   nanti.*
   quickly   make   appointment hurry   full   lho   later
   'Hurry up and make an appointment, otherwise it will get full *lho*, later.'

3. *Dong*

Above an imperative form which was added the particle *lho* was discussed. For comparison with *lho*, two other end of sentence particles, namely *deh*, and *dong* will be discussed.

3.1.   Soft imperative

After the imperative form of a verb, *dong* is added to soften the imperative as in (21).   Adding *dong* is often used for flirting purpose, especially when it is used by a woman as in (22). In both instances, the vowel for [o] is lengthened.

(21) *tenang, dong.*

　　calm, dong

　　'Calm down, *dong.*'

(22) *sayang, jangan pulang,   dong,   ya?*

　　Darling, don't   go home   dong   yeah

　　'Darling, don't go home, *dong*, yeah?'

3.2.   Pleading/ interrogative

This *dong* is similar to 3.1. above, but it does not co-occur with an imperative.

(23) (Hearing that people know he is gay)

　　　　kalau   ga   ada   yang   bilang   trus   siapa yang   bilang, dong?

　　　　If   NEG   exist   that   say   then   who that   say

　　　　 'If there is nobody who said (so), then who is the one?'

(24) a.　*kata   dokter, aku   nggak   bisa   punya   anak.*

　　　　say   doctor I   NEG   can   have   child

　　　　'The doctor said that I cannot have children.'

　　 b.　*sayang, kamu   omong   sesuatu,   dong.*

　　　　darling you   say   something,   dong.

　　　　'Darling, you say something, dong.'

3.3. Definiteness

(25) a.　*ya, kamu   dong   yang   cukongin.*

　　　　yeah, you   *dong*   that   contribute

'You *dong* are the one who should pay.'

b. *masa* *tamu* *disuruh* *bayar, sih*?

   how come guest forced to pay, sih

'How come that guests are forced to pay, sih?'

3.4. Cheering up

(26)a. *kok* *bisa,* *dokter*?

   kok, can doctor

'*Kok*, is it possible?'

b. *bisa* *aja,* *dong*.

   can plain *dong*

'You can do that, *dong*'

4. *Kok*

Kok is also a highly polysemous particle. *Kok* can be located sentence initially, a oint that I will discuss later. A sentence final particle *kok* connotes that the uttered sentence may well be against the expectation of the hearer.

4.1. Unexpected statement

(27)a. *omong-omong,* *Nino* *ga* *datang*?

   talking.about Nino NEG. come

'Talking about Nino, isn't he coming?'

b. *Nino* *datang* *kok*.

  Nino come *kok*

'Nino is coming, *kok*.'

(28)a. *lo* *nggak* *capek, nih*?

  you NEG tired, *nih*

'Aren't you tired?'

b. *enggak,* *gue* *enggak cape, kok*.

  no I NEG. tired *kok*

'No, I am not tired, *kok*.'

(29)a. *ini* *kenal* *dulu,* *Andien*.

  this get.to.know before Andien

'I am introducing (her) first, Andien.'

b. *udah*   *kenal,*   *kok.*

already   know   *kok*

'We are already acquainted, *kok*.'

4.2.   Statement of fact

Kok is added to a truthful statement to show the speaker's assertion versus presumed skepticism of the hearer.

(30) *sebenarnya aku juga masih mau tinggal disini, kok.*

actually   I   also   still   want stay here   kok

'To tell the truth, I also want to still stay here, *kok*.'

4.3.   Mitigating

In both (31) and (32), each of the speakers tries to show his/her caring attitude toward the addressee.

(31)   (Being asked if Bu Mei is tired)

Bu Mei   *cuma*   *ngantuk, kok.*

Bu Mei   only   sleepy   kok

'I am just getting sleepy, *kok*.'

(32) *ikut juga tidak apa.apa, kok.*

join also NEG   what kok

'We don't mind you coming with us, *kok*.'

4.4.   Sentence initial particles *kok* and *lho*

Kok and *lho* are the only particles which can be located in the sentence initial position. The difference between *lho* and *kok* seems subtle. There is a time delay in (34) where *lho* is used. Also, *lho* seems stronger. For example, (34) was uttered by a psychiatrist from whom a patient asked for a few more sleeping pills.

(33) *kok, bisa*   *gitu*?

kok can   like that

'*Kok*, can it be like that?'

(34) *lho, saya*   *sudah*   *kasih.*

lho, I   already give

'*Lho*, I already gave (that) to you.'

5. Discussion

This paper examined the four particles *sih, dong, lho,* and *kok.*   There are many other particles, such as *nih, deh, tuh, kan, ya, gak,* etc., whose use I have not yet studied. The four particles that are discussed here fit my hypothesis that they connote the speaker's proximity or distance vis-à-vis the information they s/he provides. Even if Indonesian speakers cannot articulate the similarities and differences among these particles in abstract terms, they are aware of the similarities and differences among them.   The functions of the particles are shown in Table 1.

Interrogation	Command / Request	Description/Evaluation	Inference	Contrast / Comparison
sih		*sih*	*sih*	*sih*
dong	*dong*	*dong*		
	(?) lho	*lho*		
		kok	*kok*	*kok*
	(deh)			

Table 1.   The functions of particles

The four most frequent particles, *sih, dong, lho*, and *kok,* can be either descriptive or evaluative. The emphatic particle *dong* has three of the five functions: questioning, command or request, and description or evaluation.

The most frequent particle *sih* has a descriptive/evaluative function, as well as an (ironically) questioning function. Unlike *dong, lho,* or *kok*, particle *sih* is used more for expressing uncertainty or for soliciting listeners' agreement. *Sih* is ubiquitous: it can be attached to a noun, adjective, verb, or adverb. It should also be noted that '*sih*' can be used in monologs.

Moreover, the high frequency of *sih* reflects the fact that a number of Indonesian colloquialisms are accompanied by tags, with which the interlocutors try to confirm their knowledge, information, or even their feelings about each other. Furthermore, the high frequency of *sih* furthermore shows that speakers tend to try to solicit agreement from hearers by adding tags, although often the topic supplemented with *sih* may connote rather negative sentiments.

Wouk 1998, 2001 and Faizah 2007 tried to analyze the pragmatic particles from a politeness point of view as well by having recourse to the idea of 'solidarity'. The particle *dong* can be discussed as an indicator of solidarity, but it should be noted that other particles such as *sih* connote uncertainty or doubt, and may even express an impolite attitude, while *lho* can connote assertiveness - linguistic behaviors far from 'solidarity' or 'politeness'.

Because of the polysemy of the particles, the study of them should be context-dependent. If we try to analyze particles in isolated sentences, we will encounter apparently contradictory functions. Connotations of the particles can be properly understood only by considering them in conversation sequences or pairs.

Finally, comparative or contrastive particles can be lodged only within the same column as shown in Table 1. Contrasts of new vs. old information, certainty vs. uncertainty, positive evaluation vs. negative evaluation, and stronger command vs. softer command, can be described as shown in Table 2.

1. <u>New information</u>   <   >   <u>Old information</u>
   lho   *sih*
2. <u>Certainty</u>   <   >   <u>Uncertainty</u>
   lho   *sih*
3. <u>Positive evaluation</u>   <   > <u>Negative evaluation</u>
   dong   *sih*
4. <u>Strong command</u>   <   > <u>Soft command</u>
   0   *(deh)* ???

Table 2.   The functions of particles in contrastive pairs

References

Faizah S. (2007)   Pragmatic particles: A cross-linguistic discourse analysis of interaction. Unpublished Ph.D. Dissertation, University of Alabama.

Miyake, Y. (2014a) Mapping hedges and tags in colloquial Indonesian. Paper presented at ISMIL, Procida, Italy, June.

Miyake, Y. (2014b) Reconsidering discourse particles in colloquial Indonesian, Paper presented at Annual Meeting, Japan Society of Indonesian Studies. Kanda International University. Chiba, November.

Sneddon, J. (2003). *The Indonesian language: Its history and role in modern society.* Sydney: University of New South Wales Press.

Wolff, J., Oetomo, D., & Fietkiewicz, D. (1992). *Beginning Indonesian through*

self-instruction. Ithaca: Cornell University Press.

Wouk, F. (1998). Solidarity in Indonesian conversation: The discourse marker *kan*. *Multilingua*, *17*, 379-406.

Wouk, F. (1999a). Dialect contact and koineization in Jakarta, Indonesia. *Language Sciences*, *21*, 61-86.

Wouk, F. (1999b). Gender and the use of pragmatic particles in Indonesian. *Journal of Sociolinguistics*, *3*(2), 194-219.

Wouk, F. (2001). Solidarity in indonesian conversation: The discourse marker *ya*. *Journal of Pragmatics*, *33*, 171-191.

http://www.indonesia.co.jp/bataone/ruangbahasatop.html

http://www.bahasakita.com/emotive-particles-sih-deh-kok-dong-combination/

http://en.wikipedia.org/wiki/Indonesian_slang

Information structure and discourse markers in Tok Pisin: differences in genres[1]

Masahiko NOSE (Shiga University)
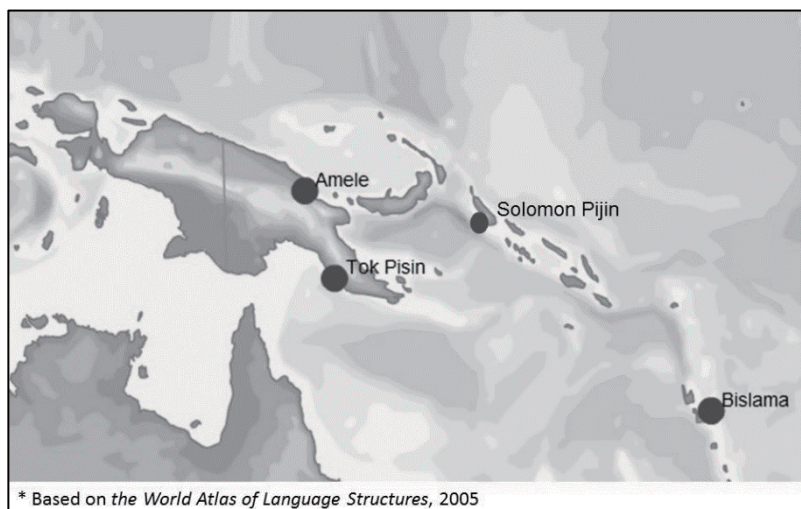mnosema@gmail.com

Abstract:

This study describes the information structure of Tok Pisin, an English-based creole language, spoken in Papua New Guinea. It has a rigid word order SVO, and does not have a grammatical voice. This study illustrates several cases of grammatical behavior related to the information structure and clarifies grammatical means of Tok Pisin to deal with information flow. This study utilizes the written texts of the New Testament and translated folk tales, as well as spoken data from radio news and chats in the field. Topic and comment in Tok Pisin are affected by emphasizing, topicalization (or fronting), and by using several discourse markers. Finally, I claim that Tok Pisin has an extremely limited means of dealing with information structure in its grammar, although we observed slight differences between the spoken and written data.

Keywords: information structure, topic, focus, discourse marker, Tok Pisin

1. Introduction

There are approximately 1000 native languages in the Melanesia area. They are classified mainly as Austronesian and New-Guinea (or Papuan) languages. Moreover, some lingua franca[2] languages are spoken for communicating with each other (see Figure 1). The three main languages (Tok Pisin, Solomon Pijin, and Bislama) are English-based creole languages and their lexifier is English.

This study describes the information structure in Tok Pisin and will pay special attention to both spoken and written data. Crowley (2004) described the grammar of Bislama and wrote about some aspects of its information flow. For this study, I conducted fieldwork in the Amele-speaking area in Madang Province (Nose 2014) and gathered Tok Pisin data (Amele in Figure 1).

Section 2 presents basic information about Tok Pisin, and explores several previous studies on creole languages. Section 3 presents examples of spoken and written data of Tok Pisin; in particular, we observe changing word orders and usage of discourse markers. Section 4 presents a discussion on managing information flow and the differences in genres, and Section 5 is the conclusion of the study .

* Based on *the World Atlas of Language Structures*, 2005

Figure 1: Amele and the three Melanesian Pidgin

2.   The grammar of Tok Pisin and preliminary studies on information structure

This section introduces basic information about Tok Pisin and, subsequently, indicates how previous studies have described the information structure of creole languages. This generally concerns creole languages (McWhorter 2011, Veenstra & Besten 1994), Melanesian pidgins (Crowley 2004, Dutton 1985, Mihalic 1971), and information structure(Andersen 1983, Foley 2007).

2.1.   Introduction of Tok Pisin grammar

Tok Pisin is an English-based creole language and one of the official languages of Papua New Guinea. Grammatically, Tok Pisin has an isolating tendency and little inflection on nouns and verbs. It has a rigid SVO order and an adjective-noun and demonstrative-noun order.

(1) Mi   bin   rid-im dispela nicepela buk long   haus   bilong   mi.
    1sg[3]   past read   this   nice   book prep   house   prep   1sg
    "I read this nice book in my house"

In (1), the subject *mi* appears at the beginning of the sentence and, in contrast, the object *buk* appears later, while the verb *ridim* occupies the middle position between the subject and the object. The form *bin* indicates the past tense marker, and the *-im* of the verb *ridim* is a transitive marker. There are two prepositions in Tok Pisin; one is the multifunctional preposition *long*, and the other is the possessive preposition *bilong* (Mihalic 1971: 38). Possessive formation is expressed in "object *bilong* owner," as *haus bilong mi* "my house" in (1). Tok Pisin has only one active voice and, therefore, the sentence (1) cannot be passivized.

2.2.   Discourse markers of Tok Pisin

Mihalic (1971:40-41) described the usage of conjunctions in Tok Pisin as fulfilling discourse markers. The most frequent marker is *na*, which means "and" and "nor." Other significant markers are *olsem* "in order to," "and," "then" and "as," and *tasol* "but," "however." Moreover, we found several kinds of discourse markers in (2) in the process of our translation (Nose and Tamo (2015 forthcming)).

(2) Representative discourse markers in Tok Pisin:
  na "then"; *tasol* "but"; *mi tok olsem* "in fact, that is to say"; bikos "because, that's why";
  long mi "for me, as for me"; *kain olsem* "for example"; *long sampla kain tok* "in other words";
  nambawan taim "firstly"; *bihain orait* "next, secondly"; *bihain pasten* "finally";
  kain samting olsem "how could I say, something like"

In (2), the forms *long, olsem,* and *bihain* are preferable and they function as connectives, in terms of text organization and speaker-hearer interaction, and they influence the information structure, as well (Schiffrin 1987, Heine 2013). In section 3 and 4 we discuss how several discourse markers influence the information structure.

2.2 Some previous studies

This section describes the characteristics of creole languages in terms of several previous studies (Crowley 2004, McWhorter 2011), and the definition and properties of information structure, based on Kiss (2002), Veenstra & Besten (1994) and Sankoff (1983). As McWhorter (2011) argued, creole languages tend to have simple grammar. According to Dutton (1985) and Mihalic (1971), Tok Pisin does not have the complicated characteristics of grammar, and has no means to specify topic and focus. However, Sankoff (1993) presented another perspective.

2.2.1 Creole has simple grammar

McWhorter (2011) indicated that creole languages are generally simple in phonology, morphology, and syntax,and he explained the reasons for creole languages being at the infancy stage and why they have not experienced much grammaticalization. Actually, the grammar of Tok Pisin is simpler than that of the other native New-Guinea and Austronesian languages, and therefore, Tok Pisin is easier to learn. Thus, it spread as a lingua franca throughout Papua New Guinea. For example, verbs in Tok Pisin have few inflections and the words are rather isolated; moreover, word orders are fixed in SVO. Crowley (2004) described the grammar of Bislama, but he did not point out whether there is a topic/focus system in Bislama, although he enumerated several discourse markers relevant to the information structure. Moreover, Sankoff (1993) discussed topic and focus in Tok Pisin, and Meyerhoff (2011) described some discourse markers in Bislama.

2.2.2 Information structure from a typological perspective

Word order and information structure are related to each other, and they have been a popular topic for studies since the Prague school and classical studies in word order typology (cf. Andersen 1983). This study reviews the information structure and word order of Tok Pisin using Hungarian sentence examples (Kiss 2002).

Hungarian is a language with free word order, but its orders are influenced by topic and focus. In (3), the transitive sentence in Hungarian can be changed in (3a-d), in terms of information structure.

(3) Hungarian:
    a. Péter el-olvassa   a köny-et   a   bolt-ban.
      Peter prev-read.3sg.pres   the book-acc the store-in
      "Peter read through the book in the store."
    b. A könyvet Péter elolvassa a boltban.
      "As for the book, Peter reads it through in the store."
    c. Péter a könyv-et olvassa el a boltban.
      "It is the book that Peter reads through in the store."
    d. A boltban a könyvet olvassa el Péter.
      "In the store, it is the book that Peter reads it through."

In (3), (3a) has a neutral order, SVO order, and the first element *Péter* is the subject and also occupies the topic position. In (3b), the sentence-initial topic becomes the direct object *a könyv*, and in (3c), the topic is *Péter*, but *a könyv* occupies the focus at the preverbal position. Finally, in (3d), *a boltban* occupies the topic and *a könyv* occupies the focus, respectively[4]. Generally, topic comes at the beginning of the sentence, and the focus position is near the verb, as shown in Hungarian (Andersen 1983:70-71, 74-75). However, unlike Hungarian, creole languages do not have a developed information structure in their grammar.

Veenstra & Besten (1994:303) suggested the concept of fronting instead of the topic and focus system similar to Hungarian. Veenstra & Besten claimed that fronting is functioning as topicalization, left-dislocation, and focusing. Fronting indicates moving an element to the initial position in the sentence. Fronting occurs in the following three conditions: in declarative contexts, interrogative contexts, and relative clauses. Topicalization is applied in declarative contexts and this movement also includes cleft, and focusing. Wh-movement is applied in interrogative contexts, and the third one in relative clauses.

(4) Tok Pisin:
a. **Long ples**   mi   lainim tok ples   Amele.
   prep   village 1sg   learn   native tongue Amele
   "At village, I will learn a native language Amele."
b. **Husait**   em i   nicepela strait?
   who   3sg cop   nice   very
   "Who is the fairest of us all?" (FT)
c. Em kamap   **gutpela   hap**   we   i   gat haus   i   stap.
   3sg   come up   good   place   where   cop get house   cop   be located
   She came to a clearing where there was a small cottage. (FT)

In (4a), the word *long ples* takes the initial position in the sentence. This fronting is topicalization. Next (4b) is an example of wh-movement, and the interrogative pronoun *husait* (who) occupies the initial position. The relative clause in (4c) is also a fronting movement. However, focusing in Tok Pisin has another view (Sankoff 1993). Therefore, we observe another creole, Saramaccan (Veenstra & Besten 1994:306), as shown in (5).

(5) Saramaccan:
   Di buku   **we**   mi bi   lesi.
   the book   focus 1sg   past   read
   "I read THE book." we: focus marker

There is a focus marker, *we*, in Saramaccan. Sankoff (1993) claimed that there are postponed focus particles in Tok Pisin; *yet, tru, moa, tasol,* and *wanpela*[5]. In the following paragraphs, this study provides spoken and written data of Tok Pisin, and we examine how information structure works and

we consider whether there is focus marker in Tok Pisin.

3.   Information structure in different genres

This section investigates the information structure in Tok Pisin by observing several different kinds of texts. Mainly, we examine fronting and other means of topicalization/focusing; moreover, we observe the usages of discourse markers. This study classifies the texts of Table 1 according to genres; namely, spoken/written and city/ rural texts[6].

	City Tok Pisin	Rural Tok Pisin
Spoken discourses	YM	FTC, Bel
Written texts	RA, NT	FT

Table 1. Spoken and written Tok Pisin data in this study

3.1.   Spoken data

The spoken data are YM, FTC, and Bel, and there are several differences between city and rural genres. Lexically speaking, YM has frequent usages of English words, and in contrast, FTC and Bel maintain Tok Pisin vocabulary. Grammatically, not all discourses keep a strict word order, and we can observe many discourse markers that are relevant to the information structure.

First, in (6), the fronting element *puttim tan long* is emphasized by stress. Therefore, adding stress phonologically is a means of focusing in spoken Tok Pisin.

(6) FTC/ rural
puttim tan   long,. maus bilong em….   em em yet… em   pinis a? em i   go … long brik haus
put   tongue prep   mouth prep 3sg   3sg 3sg yet   3sg   perf intj 3sg cop go prep block house
"put tongue on his mouth, (he licked his lips, as he went to the brick house)"
(7) Bel/ rural
wankain olsem
same   like (DM)
"It is the same like this."

Another means of focusing is to say only the necessary word, as in (7). In (7), the fragment words *wankain olsem* are foregrounded in the discourse and other information is not specified, but they are supplemented by the context. In (8), it is an example of fronting, but the word "Snow White" is repeated in the following sentence and the element is emphasized by this repetition. Example (9) is peculiar in that the 3rd person singular pronoun *he* (instead of *em*) is used. The speaker (newscaster on the radio) accidentally used the English word *he*. Needless to say, the specific name (HK) of the governor is operated by fronting.

(8) FTC/ rural
Snow White …. **Snow white** em   pilim tait,
Snow White   Snow White 3sg   feel   tired

"Snow White feels tired"

(9) YM/ city

Gabanaa Haveila Kaabo, he   givim bikpela tenkyu bilong em i   go long travel foundation
Governor   3sg give big   thanks prep   3sg cop go prep travel foundation
"Governor HK gives big thanks to the travel foundation."

(10) FTC/ rural

i   no ken writim pinis   mi   askim yu, yu   writim pinis oo
cop no can write   perf   1sg ask   2sg 2sg write   perf intj
"You cannot write it up, and I am asking you, have you already written it?"

In (10), the sentence does not have a subject and the verb comes first. It seems to be fragment information, but this type is frequently observed in both spoken and written texts. In this case, focus is considered to be on the negative action or verb phrase.

(11) YM/ city

na ting blong ol **na** i   tok **olsem**
DM think prep 3pl DM cop talk like (DM)
"And their idea, they say like this"

In (11), the discourse marker *na* is used. This *na* means "and" and "then," and this marker is apt to appear at the beginning of the sentence. The discourse marker *na* can connect the previous and present sentences; moreover, speakers can give additional information through the second *na* in (11). In (11) and (12), the form *olsem* is used, which means "like, that, in this way." This is also a discourse marker and it connects the discourse or omits background information[7].

(12) FTC/ rural

em mas   meri mi   ting   wanpla tok **olsem**
3sg must   girl   1sg think one   talk like (DM)
"She must be a girl, and I think one says like this."

(13) FTC/ rural

Nau   mipla go long   42, 41 pinis
now (DM) 1pl   go prep   page 42, 41 finish
"Now we go to the page 42, the page 41 is finished."

Finally, in (13), the form *nau* is naturally derived from the English "now". It is also a discourse marker (as pointed out by Sankoff 1993) and this marker can introduce a new topic at the beginning of the sentence.

3.2.   Written data

In this section, we examine written texts of Tok Pisin and examples of their behavior are observed through fronting, discourse markers, and other means. Moreover, we try to find differences

in genres. RA and NT are regarded as city Tok Pisin, and FT is a translation work of Nose and Tamo (2015 forthcoming).

(14)RA/ city
Long Madang Provins blong PNG, ol   keis   blong measles i   wok   long go daun.
prep Madang Province prep   PNG   3pl case   prep measles cop work prep go down
"In Madang Province, PNG, all cases of measles are going down."

In (14), the locative phrase with preposition *long* is moved to the initial position by fronting. The fronting of locative phrases is frequently observed in Tok Pisin.

(15) NT/ city
Na   em i   salim sampela man i   go long Betlehem **na** ol   ples   klostu long en,
DM 3sg cop send   some   man cop go prep Bethlehem DM 3pl   place   near   prep it
na ol i   kilim i   dai olgeta pikinini man i   no winim   tupela yia   yet.
DM 3pl cop kill   cop die all   child   man cop no suppress two   year yet
"and he sent forth and put to death all the male children who were in Bethlehem and in all its districts, from two years old and under."

The discourse marker *na* is used frequently in written texts as well as in spoken texts. In particular, in (15), the three *na* markers connect discourses, each meaning "and."

(16)FT 5/ rural
SnowWhite i   no save   **olsem** em   haus bilong ol   sevenpla   dwarf.
Snow White cop no know   DM   3sg house prep 3pl seven   dwarf
"What SnowWhite did not know was that the cottage belonged to seven dwarfs."
(17)FT 7/ rural
Olsem na   em i   go   long tower na   wokim apul   long   outsait
DM   DM 3sg cop go   prep tower DM   work   apple   prep   outside
luk   olsem   apul   tasol em pulap   poison.
look   like (DM) apple   only 3sg full   poison
"So she went to her tower where she made an apple."

The discourse marker *olsem* is frequently observed, as shown in (16) and (17). The form *olsem* has several usages. First, in (16), this marker functions as a complementizer or for quoting (Meyerhoff 2011: 256). In contrast, in (17), *olsem* and *na* appear in sequence and their meaning is "so, then," functioning as a conjunction.

(18)FT 4/ rural
so   em i   sleep insait   long wanpla bilong ol   dispela bed.
DM 3sg cop sleep inside prep one   prep 3pl this   bed

"and so she yawned and stretched and lay down to sleep in one of the beds."

On the other hand, in (18), the English word "so" is used as a conjunction. The meaning of *so* is almost the same as *olsem* or *na*.

(19) NT/ city

Nau Herot i   save,   ol   saveman bilong hap   sankamap
now Heod cop know 3pl   wise man prep   side   of sun rise
ol i   giamanim em, **orait** bel   bilong em i   hat   moa.
3pl cop deceive   3sg   DM   belly prep 3sg cop hard   more
"Then, Herod, when he saw that he was deceived by the wise men, was exceedingly angry"

In (19), there are two discourse markers, *nau* and *orait*. *Nau* functions by presenting a new topic or story, and it is used preferably in spoken discourse, like in (13). In contrast, *orait* means "so" in English. Both *nau* and *orait* play a role in expressing the speaker's (or storyteller's) confirmation (cf. Schiffrin 1987: 230, "speaker progression").

4. Discussion

This section summarizes the observations of the previous sections and discusses several usages. The present study examined spoken and written texts and tried to find their differences and to summarize several sentence types related to their information structure. Overall, the observed grammatical means are summarized in (20). There are limited means expressing topic and focus in Tok Pisin; moreover, only a few discourse markers are preferred for usage. Additionally, this section discusses the differences in genres that were observed.

(20) Grammatical options regarding information structure in Tok Pisin
  a. Fronting: topic and focus: repeating, cleft[8]
  b. Emphasizing (focusing)[9]: necessary word only, emphatic words (*wanpla, tru, tasol:* Sankoff (1993)'s focus particles)
  c. Using discourse markers: connecting discourse, introducing, omission
    c-1: conjunction: *na, tasol, so, yet, olsem (like this, thus)*
    c-2: demonstrative: *ya, olsem (introducing quotation),*
    c-3: confirmation: *nau, orait;* confirmation, new topic
  d. others (verb-initial, *yu tok*): verb-focus, turn-taking[10]

Mainly, Tok Pisin utilizes emphasizing in spoken discourse, and prefers fronting in both genres. Although we did not find a proper example in this study, cleft construction (see footnote 8) is also possible as a fronting movement. Discourse markers have a function of introducing topics, connecting sentences, and implying speaker's confirmations, but these kinds are limited and only *na* and *olsem* are commonly observed[11]. Other usages are considered, but this study does not discuss them.

Next, we try to clarify the differences in genres. Spoken discourse can depend on phonological clues (stress and pause); moreover, the speaker can choose to relay the important information only. Instead, written data prefers using discourse markers; in particular, the discourse markers *na* and *olsem* are frequently used for coherence. Next, we consider the differences between city and rural areas, but there is no significant difference in the information structure between them. Fronting is observed in both genres, and some English words, *so, he*, etc. are preferred in the spoken/city texts.

   Moreover, we discuss that discourse markers can become topic/focus markers. For example, McWhorter underscored that the discourse marker *nɔ'ɔ*, "then" in Saramaccan, has been grammaticalized to a new information marker, as shown in (21).

   (21) Saramaccan (McWhorter 2011:127-128)
   A bu'nu. **nɔ'ɔ** mi o'   ta'   ha'ika i.
   it good   DM   I   future imperfective listen   you
   "Good. So I'll be listening for you (waiting for your answer)"

Sankoff (1993) indicated that Tok Pisin has the focus marker *yet,* originally meaning "yet, still." Sankoff claimed that the form *yet* in (22) has been grammaticalized from intensifier to focus marker . However, this study did not find such a usage, and the usage of *yet* is extremely limited in texts. This study considers that *yet* is not a fully grammaticalized marker in Tok Pisin, and that it only has fronting or emphasizing functions for focusing (the usage *yet* is included in (20b)).

   (22) Tok Pisin: Sankoff (1993: 131)
   Tok "Orait   yu   **yet**   kilim pikinini bilong mi."
   say   alright 2sg focus kill   child   prep   1sg
   "(She) said, "Alright, **you're the one** who killed my child""

   We subsequently discuss the usages of discourse markers. The discourse marker *na* is frequently observed in connecting sentences. In (23), these sentences are taken from FT and the parallel texts in Tok Pisin, Amele, and English. Foley (2000:387) claimed that the New Guinea languages have common discourse characteristics. First, given and presupposed information is normally omitted, and second, only one piece of new information is introduced per clause. By contrasting the parallel texts in (23), these characteristics are applicable to Tok Pisin and also visible through language contact. That is, the discourse markers *na, olsem* are used to form consistent discourse structures, as observed in New Guinea languages.

   (23) Tok Pisin/ Amele: FT (cf. Nose 2014)
   Na ol   i   kolim em   Snow White.
   DM they cop call   3sg Snow White
   Odocob   uqa ayan   snow-white boin.
   And then (DM) 3s   name-poss snow white call-3s.past
   "and then she called snow white."

Finally, this paper considers answers to why Tok Pisin has a limited number of managing information structures, and explains the decisive differences between spoken and written texts. First, Tok Pisin is a creole language and it still has the characteristics of a simple grammar in its construction, which cannot yet fully grammaticalize topic/focus markers. As a result, it depends on fronting, emphasizing, and several discourse markers, as shown in (20). In spoken discourse, speakers utilize stress, pause, emphasizing, and repeating, while, in contrast, discourse markers are preferred in written texts. This study claims that discourse markers such as *nau, orait, na,* and *olsem* function to introduce new information or focus to sentences (cf. Foley 2000: 386-387, Nose 2014).

5. Conclusion

We summarize the characteristics of topic and focus in Tok Pisin that this study has found.

First, fronting is the most frequent method of indicating topic/focus. Tok Pisin has fronting movement through the indication of topicalization and partly through focusing; but focusing is realized through other options, namely adding stress, emphasizing, and cleft constructions. In particular, spoken discourses prefer adding stress, using emphasis, and repeating, while, in contrast, several discourse markers are frequently observed in both genres, and they are effectively used to organize foreground and background information in written texts. However, these discourse markers have not yet been grammaticalized into topic or focus markers.

Notes:

1. I would like to thank Neret Tamo, Nelau Lagia and the villagers in Sein, Madang Province, Papua New Guinea for their data and kindness. I claim sole responsibility for any errors.

2. The lingua franca in Melanesia are Indonesian in the Western part of New Guinea Island, Hiri Motu in the Western Province and around Port Moresby in Papua New Guinea, Tok Pisin in the other area of Papua New Guinea, Solomon Pijin in Solomon Island, and Bislama in Vanuatu. Tok Pisin, Solomon Pijin and Bislama are called Melanesian Pidgin.

3. Abbreviations: acc, accusative; cop, copula; DM, discourse marker; future, future tense; intj, interjection; sg, singular; past, past tense; perf, perfective; pl, plural; prep, preposition; prev, verbal prefix, pres, present tense; 1, 2, 3, first person, second person, third person respectively.

4. Thus, word order position and information structure in Hungarian are summarized in (i).

    (i)   (Topic) X (Focus)-VERB X

    Topic: Sentence initial position, Focus: Preverbal position and X: neutral position or supplementary element

5. There is no description of information structure in Tok Pisin and Bislama, as far as we checked Dutton (1985), Crowley (2004), and Mihalic (1971).

6. Tok Pisin is mainly a spoken language and there are few materials of written texts (mainly Christian texts). This study collected both spoken and written discourses and they are shown in Texts section underneath.

7. Crowley (2004), Meyerhoff (2011) and Meyerhoff & Niedzielski (1998) discussed the usages of DM *olsem*. Meyerhoff & Niedzielski (1998) claimed that the form *olsem* functions as complementizer, clausal anaphor (thus, like this), attention shifter (rather anyway) and hedge.

8. Tok Pisin has cleft constructions, although the texts that this study used were not found. These examples (ii) and (iii) are from my recording, September 6. 2014. Cleft construction has a focusing effect and it is included in fronting movement.

   (ii)   food asde   mipla kaikai

          food yesterday 1pl   eat

          "The food we ate yesterday"

   (iii)   dispela   haus hap   youngpela boi bin wokin

          this   housethere young   boy past build

          "This house there young boy built"

9. There are emphatic words for expressing importance in information: *tru* (very, really), *tasol* (only, just)

10. Nose (2014) claimed that Tok Pisin has a kind of turn-taking expression, *yu tok/yu toktok* "you say" and *mi tok* "I tell you."

11. The options shown in (20) are realized in (iv).

   (iv)   Mi kaikai yam. "I eat yam"

          a.   Fronting: ***Yam***, mi kaikai/   Yam mi kaikai yam.

          b.   Emphasizing: Yam/ Mi kaikai yam tasol.

          c.   Using Discourse markers:

               1. Na mi kakai yam/ Olsem mi kaikai yam.

               2. Mi kaikai yam ya/ Mi kaikai (yam) olsem.

               3. Nau mi kaikai yam/ Mi kaikai yam orait.

          d. Others: Kaikai yam/ yu tok mi kaikai yam.

Texts:

1. Bel: Bel speaker interview, elicitation of Bel grammar, spoken in Tok Pisin (Recorded on September 1, 2012)

2. FT: Nose Masahiko & Neret, Tamo. 2015 forthcoming. *Folk Tales from Around the World: Amele-Tok Pisin-English multilingual book*. Ms.

3. FTC: Recoded discourse in translating FT (Recorded on August 25, 29, 2014)

4. NT: Matthew 2:16-17 (*Herot i tok na ol i kilim planti pikinini i dai*) In: *Nupela Testamen: bilong bikpela Jisas Kraist*. Canberra/Port Moresby: The British and Foreign Bible Society in Australia, 1969.

5. RA: Web news in Radio Australia; *Ol toktok long sik measles long Solomon Islands na PNG* (23, October, 2014), Accessed on October 27, 2014 (http://shar.es/1HOaJ6)

6. YM: Yumi FM radio news (Recoded on September 3, 2013)

References

Andersen, Paul. K. 1983. *Word order typology and comparative constructions*. Amsterdam/Philadelphia: John Benjamins.

Crowley, Terry. 2004. *Bislama reference grammar.* Honolulu, Oceanic Linguistics Special Publication.

Dutton, Tom. 1985. *A new course in Tok Pisin (New Guinea Pidgin)*, Pacific Linguistics D-67, Canberra: Australian National University.

Foley, William. A. 2000. The languages of New Guinea. *Annual review of anthropology*, 357-404.

Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics*, 51(6), 1205-1247.

Kiss, Katalin. É. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press.

McWhorter, John. H. 2011. *Linguistic Simplicity and Complexity: Why Do Languages Undress?* Berlin: Walter de Gruyter.

Meyerhoff, Miriam. 2011. *Introducing Sociolinguistics (2nd edition)*. London: Routledge.

Meyerhoff, Miriam, & Nancy. Niedzielski, 1998. The syntax and semantics of olsem in Bislama. Recent Papers in Austronesian Linguistics. *UCLA Occasional Papers in Linguistics,* 20, 235-243.

Mihalic, Francis. 1971. *The Jacaranda dictionary and grammar of Melanesian Pidgin*. Hong Kong: Jacaranda.

Nose, Masahiko. 2014. Information structure in Amele, Papua New Guinea. Proceedings of the International Workshop on Information Structure of Austronesian Languages. Research Institute for Languages and Cultures of Asia and Africa: 105-112

Sankoff, Gilian. 1993. Focus in Tok Pisin. In: Burne, Francis, & Donald Winford(eds.). *Focus and grammatical relations in creole languages.* Amsterdam/Philadelphia: John Benjamins: 117-140.

Schiffrin, D. 1987. *Discourse markers (No. 5).* Cambridge: Cambridge University Press.

Veenstra, Tonjes, & Hans den Besten. 1994. Fronting. In: Arends, Jacques, Muysken, Pieter, & Smith, Norval(eds.). *Pidgins and creoles: An introduction*. Amsterdam/Philadelphia: John Benjamins: 303-315.

The role of Bunun deixis in information structure: An initial assessment

Rik De Busser
National Chengchi University

1. Introduction

This paper will discuss the role of deixis on information structure in the Takivatan dialect of Bunun. For the sake of simplicity, the discussion will be restricted to topicality. We will distinguish two types of topicality:

(1) **Clausal topicality**: the expression of the topic in a clause, i.e. the grammatical or functional identification of the pragmatically most salient participant of that clause.
(2) **Discourse topicality**: the expression of the discursive text topic, i.e. the establishment of discursive themes through the creation of strings of cohesion (cf. Halliday & Hasan 1976). A single text or discourse can have more than one discourse topic.

We will also restrict the discussion to deictic paradigms or elements whose primary function is the expression of either spatial deixis (*this*, *that*) or anaphoric deixis (*the aforementioned*).

1.1. Takivatan Bunun

Takivatan is one of the five dialects of the Bunun language (ISO 639-3: bnn), an Austronesian language mainly spoken in the central regions of Taiwan, mainly in villages in relatively remote mountainous locations. There are no official figures on the actual number of speakers; currently 56,004 people are officially registered as Bunun, but the actual number of fluent speakers is certainly considerably lower (at most 60% of that number). The upper limit for Takivatan speakers is certainly not higher than 1700 individuals.

The Takivatan dialect is largely agglutinating with a very strongly developed verbal morphology. It has a Philippine-type argument alignment system (see De Busser 2011), with a basic contrast between actor, undergoer and locative alignment marked by suffixes on the verb, and has a very productive of valency-changing verbal affixes. Only non-third-person pronouns make a case distinction.

1.2. Cross-linguistic functions of spatial deixis

As said before, the discussion here restricts itself to paradigms whose primary use is the expression of spatial (and often by extension temporal) deixis. The discussion will exclude person deixis, i.e. personal pronouns and other words used for indicating person contrasts, and phenomena like TAM systems, which are dedicated to the grammaticalized expression of temporal contrasts. We will discuss anaphoric markers and expressions of manner, because they both have an anaphoric deictic function.

For the sake of simplicity, we will start start the discussion from the point-of-view of what could be called 'canonical' deictic markers, demonstrative pronouns. Himmelmann (1996:218ff) divides the cross-linguistic functions of demonstratives in four major categories:

- ▪ **Situational use**: "reference to an entity present in the utterance situation" (Himmelmann 1996:219)
- ▪ **Discourse deixis**: a metaphorical extension of spatial deixis that refers to the distance of a referent in the current text or conversation.
- ▪ **Tracking use**: the use of demonstratives for keeping track of textual referents. In other words, the use of deictics for creating textual cohesion in the sense of Halliday & Hasan (1976)
- ▪ **Recognitional use**: "[…] the intended referent is to be identified via specific, shared knowledge rather than through situational clues or reference to preceding segments of the ongoing discourse" (Himmelmann 1996:230).

What is quite striking about this classification is that only the first of these categories is directly involved in the expression of spatio-temporal deixis. The primary function of the other three categories appears to be related to organizing information in texts and conversations. This is especially the case for the tracking use of demonstratives, in which demonstrative reference is used for creating cohesive chains of reference in a text, thus allowing the discourse participants to keep track throughout a text of the different entities mentioned in that text.

It is important to realize that when Himmelmann talks about these different demonstrative functions, he generally seems to assume that they are *primary* functions of demonstratives. Thus, when he discusses the tracking use of Tagalog *ito* in example, the primary function of this form is not spatial deictic reference, but establishing an anaphoric link to the referent *isang manlalakbay* in the previous clause.

(1)   may kasaysayan sa isang manlalakbay
        may   ka-saysay-an   sa   isa-ng   maN-CV-lakbay
        EXIST   ?-statement-LOC   LOC   one-LK   IRR.ACT-RED-travelling
        '(One incident) is told about a traveler;'

        ang manlalakbay na ito ay si Pepito
        ang   maN-CV-lakbay   na   **ito**   ay   si   Pepito
        SPEC   IRR.ACT-RED-travelling   LK   **PROX** PM   PN   P.
        '**This** traveller (his name) was Pepito.' (from Himmelmann 1996:229)

Finally Himmelmann (1996:210) clearly distinguishes between definiteness markers and demonstratives (although he mentions on p. 243 that the extensional use of demonstratives for discourse reference and tracking is a step in the grammaticalization pathway leading to the creation of definiteness markers or third person pronouns). This is a distinction that I will not make such a clear distinction for Takivatan Bunun. The only grammatical forms that conceivably could be analysed as markers of definiteness markers – the bound suffixes discussed in 2.1.1 – do also encode a distance contrast. These bound markers clearly establish a referent or an event (see below) as definite or identifiable, while their spatial deictic function is often more difficult to determine. They can be used for establishing what Himmelmann (1996:210) calls "associative-anaphoric" links.

2. Bunun deixis and information structure

2.1. General overview

Most deictic paradigms discussed in this paper make a basic three-fold distinction between the morphs -*i* 'proximal', -*un* 'medial' and -*a* 'distal' (see De Busser 2009, Chapter

9). This is the case for the bound definiteness markers, third person personal pronouns, demonstrative paradigms, and the place words *ʔiti/ʔitun/ʔita* (see 9.2, 9.3, and 9.4 respectively). Exceptions are the anaphoric marker *sia* and the manner word *(m)aupa* 'thus'; both express anaphoric deixis but express no distance contrast.

The prototypical function of deictic elements is marking distance contrasts, as in illustrated in the example below for definiteness markers.

(2)   Asa pisihalun itu Kaliŋkuti, pisihalunti, na asa tunhan Nantuta
      asa   pi-sihal-un   itu   Kaliŋku-**ti**
      have.to CAUS.STAT-good-UF   this.here K.-**DEF.REF.PROX**
      'You have to put everything in order **here** in Kaliŋku, …'

      pi-sihal-un-ti
      CAUS.STAT-good-UF-DEF.REF.PROX
      '… and when it is in order here,…'

      na   asa   tun-han   Nantu-**ta**
      CONS have.to   PERL-go   N.-**DEF.REF.DIST**
      'you have to go over **there** to Nantou.' (TVN-012-002:49)

Apart from spatial deixis, Takivatan Bunun deictic markers have developed a number of meaning extensions: temporal deixis, empathy, animacy, and endorsement. In fact, certain deictic forms are primarily used for expressing non-spatial meanings. For instance, the bound medial definiteness marker *-kun*, when it appears on nominal forms, almost exclusively marks that the speaker has an emotional connection to the noun, usually a person, that is marked by *-kun*; this is called empathy in De Busser (2009:422–425).

(3)   Muska [ma]limadia minsumina Linikun.
      muska   mali-madia   min-suma-in-a   Lini-**kun**
      but   SUPERL-many INCH-return-PRV-LNK L.-**DEF.SIT.MED**
      'But then, after a long time (my dear friend) Lini came back.' (TVN-008-002:179)

Interestingly, while all the above-mentioned functions of deixis are discussed in De Busser (2009), no reference is made to the role of deictic markers in establishing contrasts in information structure, except for the dedicated anaphoric function of the marker *sia*. Curiously, this seems to suggest that deictic paradigms have no significant function in the realization of information structure. The sections below will explore to what extent this is a valid assumption. I will then investigate how different deictic paradigms are involved in the realization of sentence-level and discursive topics.

The following deictic paradigms relevant to the present discussion:

- bound definiteness markers (2.1.1)
- third person pronouns (2.1.2)
- free demonstratives (2.1.3)
- dedicated place and manner words (2.1.4)
- the anaphoric marker *sia* (2.1.5)

2.1.1. *Definiteness markers*

Takivatan Bunun has a set of six bound markers that express a three-fold contrast in distance (proximal, medial, distal) and a two-fold contrast in what could be called ontological status (referential vs. situational).

Table 1. Takivatan Bunun definiteness markers

	Referential	Situational
Proximal	*-ti*	*-ki*
Medial	*-tun*	*-kun*
Distal	*-ta*	*-ka*

In a number of ways, these bound definiteness markers are unusual. First of all, calling them definiteness markers is somewhat controversial, given that (a) they encode a distance contrast and (b) while the attachment of a definiteness marker causes the host referent to be definite, referents can be definite without one of these markers being present (in other words, they are optional).

Secondly, these bound markers can occur on words in many word classes, including nouns, verbs (!), the anaphoric marker *sia*, and the manner word *maupa*.[1] Third, definiteness markers distinguish between what has been called in De Busser (2009) referential and situational forms. The former pertain to the material properties of a referent of an event, while the latter put more emphasis on the spatial and/or temporal properties of a referent or event. This analysis has been elaborately supported by evidence in De Busser (2009:426–440).

Below is an example of a distal situational marker on a verbal host. The use of a situational marker indicates that emphasis is placed on the distal *location* of the event, rather than on what actually happened.

(4)   Mukvaikuka vaŋlað.
        mukvaiku-**ka**   vaŋlað
        bend-**DEF.SIT.DIST**   riverside
        'The river makes a bend **over there**.' (TVN-xx2-001:3)

Example (5) contains a proximal situational marker and a distal referential marker. The distal referential *-ta* indicates that the referential properties, in this case the physical identity, of the person marked are important.

(5)   … laʔadusduki Qusunsubali sia Maiata tama.
        la-adus-du-**ki**   Qusunsubali
        COVER-carry-EMO-**DEF.SIT.PROX**   Q.
        sia   Maia-**ta**   tama
        ANAPH   M.-**DEF.REF.DIST**   father
        '[…] **from here** we went together to (**that**) Qusunsubali,
        to the father of Maia.' (TVN-008-002:69)

Definiteness markers are very common in narrative discourse; their frequency varies in informal spoken language, but is generally lower there.

2.1.2. Third person pronouns

Personal pronouns do generally not express a distance contrast, with the exception of the paradigms for the third person singular and plural.[2]

---

[1] But not on third person pronouns or demonstrative pronouns.

[2] The main reason for not analyzing the forms in Table 2 as demonstrative forms is that they appear to be historically related to the root *-is*, which in Isbukun Bunun has been analysed as a bound third person pronoun (e.g. in Zeitoun 2000) and occasionally occurs in Takivatan Bunun, mainly in fixed constructions.

Table 2. Third person personal pronouns

	Singular	Plural
Proximal	*isti*	*inti*
Medial	*istun*	*intun*
Distal	*ista*	*inta*

Third person pronouns usually refer to human or other higher animate referents.

(6)   Hanʔak daiða maluskun inta.
       han-ʔak   daiða   ma-luskun   **inta**
       be.at-1S.TOP   over.there   DYN-together   **3P.DIST**
       'I am there with them together.' (TVN-008-vxxx:1)

Unlike other pronominal numbers, the third person forms do not have distinguish between different grammatical roles (agent, undergoer, location) and – in Takivatan Bunun – have no bound equivalent forms. Third person pronouns are relatively uncommon in comparison to first and second person forms and free demonstrative forms.

2.1.3. Free demonstrative paradigms

Takivatan Bunun has a complex demonstrative paradigm, which encodes: (a) a two-way visibility distinction; (b) a four-way plurality distinction; (c) and a three-way distance dimension:

Table 3. Demonstrative forms

Visibility	ROOT	Plurality	Distance
Ø- 'VIS'	*ai-*	*-p-* 'singular'	*-i* 'PROX'
n- 'NVIS'		*-ŋk-* 'vague plural'	*-un* 'MED'
		-nt- 'paucal'	*-a* 'DIST'
		-t- 'inclusive generic'	-Ø 'USPEC'

Not all combinations of morphs have been attested in naturalistic language and there is great variety in the frequency of use. For instance, in the paucal paradigm only the distal forms *ainta* 'DEM.VIS.PAUC.DIST' and *nainta* 'DEM.NVIS.PAUC.DIST' have been attested. Underspecific forms (forms without a distance marker) in general are relatively rare.

The most commonly occurring demonstratives by a large margin are singular and vague plural forms. Below is an example of the visibility contrast expressed by the singular distal demonstratives *(n)aipa*.

(7)   Na, ukin aipa ʔita namudanin, musbai naipa maqmut
       na   uka-in   **aipa**   ʔita
       INTER   NEG.have-PRV   **DEM.S.DIST.VIS**   there.DIST
       '**It [the deer, visible]** wasn't there anymore, …'
       na-muda-in
       IRR-walk-PRV
       '… it had gone, …'
       musbai   **naipa**   maqmut
       run.away   **DEM.S.DIST.NVIS**   night
       '… **it [non-visible]** had run away during the night.' (TVN-008-002:135)

Inclusive generic forms refer to an indeterminate number of referents which always includes the speaker.

(8) Haiða aitun ludun tikisuna, [...]
haiða **aitun**   ludun   tikis-un-a
have   **DEM.IG.MED.VIS** mountain   small-EMPH-SUBORD
'**Our people** had a small mountain, [where in the old days they would go hunting.]' (TVN-012-002:162)

2.1.4. Multi-categorial place and manner deictics

A dedicated set of words is used for expressing space, time and manner. These forms can occur in adverbial and in verbal slots. The forms *ʔiti/ʔitun/ʔita* typically express spatial and, somewhat less commonly, temporal distance. They make the typical three-fold distinction between proximal, medial and distal.

Table 4. Manner words

		Spatial	**Temporal**
PROX	*ʔiti*	'here'	'at this moment'
MED	*ʔitun*	'there (medial)'	'at that moment (medial)'
DIST	*ʔita*	'there (distal)'	'at that moment (distal)'

An example of the proximal place word used as a verb:

(9) Iʔitiʔak.
i-**ʔiti**-ʔak
STAT-**here**-1S.F
'I am **here**.' (BNN-N-002:52)

In the example below, the distal form occurs both in a verbal form and clause-finally in an adverbial slot.

(10) Munʔita madas pudaku atikisunaŋ ʔita
mun-ʔita
ALL-there.DIST
'[The shaman] has to **go there …**'
madas pu-daku   tikis-un-aŋ   **ʔita**
take   place-ritual.object   little-EMPH-PROG   **there.DIST**
'… and put a little bit of the ritual token **over there**.' (TVN-012-001:44)

The form *(m)aupa* expresses manner and has a similar syntactic distribution as the place words above. It can be translated as 'thus' or 'in this/that manner' and in expressing such a meaning often refers back anaphorically to a previous event in the text or discourse. As the example below illustrates, it can be – and often is – modified by a bound definiteness marker, most commonly the referential distal form *-ta*.

(11) Maupata madaiŋʔað tu m baðbaði Diqanin tu masihalaŋ kakaunun
maupa-ta   ma-daiŋʔað tu   baðbað-i   Diqanin   tu
thus-DEF.REF.DIST   STAT-old COMPL have.conversation-PRT Heaven   COMPL
ma-sihal-aŋ   ka-kaun-un
STAT-good-PROG   things.to.eat
'And **like that**, the elders talked to Heaven in order to keep producing good crops.' (TVN-012-001:46)

2.1.5. *The anaphoric marker*

An extremely common marker that is exclusively used for anaphoric (or exophoric) reference is *sia*. It refers to referents or events that were previously mentioned in a discourse or are assumed to be commonly known to all discourse participants. *Sia* combines with bound definiteness markers and is used both in nominal (12) and verbal slots (13).

(12)   Ma, samantukandu siatun [...].
         ma   samantuk-an-du   **sia**-tun
         INTER   spy.on-LF-EMOT   **ANAPH**-DEF.REF.MED
         '[The deer…] I kept a close watch on **it** [in order to shoot it]' (TVN-008-002:184)

(13)   Siata.
         sia-ta
         ANAPH-DEF.REF.DIST
         '[I will now explain how we Bunun in former days were, how our elders said: if you want to grow up, you have to live attentively, if you see a one-eyed man, if there is a cripple, you cannot laugh, it is a taboo, you cannot make jokes about them.] It was **like that**.' (TVN-013-001:4)

Example (13) illustrates a common usage of *sia* in narrative prose as an end-of-story marker, in which case its antecedent is an entire text rather than a single referent or event.

2.2. Austronesian definiteness and information structure

In the study of various Austronesian languages, a case has been made that topics (or topical subjects, or the like) must be definite. For instance, Schachter (1976:494) says of Tagalog:

> "Formally, the topic is marked either by the use of a topic pronoun form or by a prenominal topic marker. Notionally, the topic is always interpreted as definite."

In the same volume, Keenan (1976:252) states:

> "Surface subjects of Malagasy simplex Ss [sic] are necessarily definite. Semantically this means there are always objects which the subject phrase refers to, and further this referentiality is not lost when the sentence is negated or questioned."

Keenan examples that this requires that Malagasy subjects "either be proper names, definite pronouns, or common nouns with demonstrative adjectives or definite articles." (Keenan 1976:253).

The Takivatan corpus does not corroborate this necessary link between definiteness and topicality. For instance, it is possible for the clausal topic of a sentence to be indefinite and non-specific. The example below is the elicited answer to the question *Did you plant many yams*.

(14)   Sauðunin ðaku.
         suað-un-in   ðaku
         sow-UF-PRV   1S.N
         'Many were planted by me.' (TVN-xx2-003:39)

Since this is an undergoer construction (as indicated by the suffix *-un*), the topic of this sentence must be the implied subject 'many' and this undergoer topic is indefinite in the given context. In addition, all deictic elements involved in establishing definite referents can occur in topic and non-topic positions and some, such as the definiteness markers and the

anaphoric marker *sia,* can even mark predicates rather than arguments (see e.g. (13)). This means that there is no hard requirement for Takivatan subjects to be definite.

However, this does not mean that there is no correlation between definiteness and topicality in a more general sense. On a conceptual level, it does make sense that pragmatically salient elements in a clause or discourse are more commonly realized as definite entities. In fact, the Animacy Hierarchy makes this connection explicit in that it "arranges entities in the order of their INTRINSIC TOPICALITY, i.e. the degree to which they are likely to be definite and referential" (Hopper & Thompson 1980:286). It is just that this correlation is not absolute.

All words in Takivatan that are associated with the explicit expression of definiteness have a deictic function. It is therefore safe to assume that there will be a strong correlation, whatever its nature, between Takivatan deixis and topicality, or more generally the degree of information salience.

3. The role of Takivatan deictics in information structure

This appears at odds with the following statement:

> "Many studies on spatial deixis put great stress on the use of deictic markers for anaphoric reference and discourse deixis […]. In Takivatan, the distance dimension in any of the deictic paradigms is rarely used unambiguously for anaphoric reference, most likely because of the existence of the anaphoric marker *sia* […]." (De Busser 2009:425)

What does this mean? In 1.2, we saw that a number functions that have been commonly associated with deixis (or more narrowly, demonstration) is related to the organization of information structure. For instance, in many languages demonstratives have developed an anaphoric function and as such are important grammatical tools in establishing textual cohesion. Among Takivatan deictic that have a tripartite distance distinction, I have so far found not a single example where spatial deictic contrast has developed an unambiguously anaphoric meaning extension. For instance, there are no instances in the corpus where the proximal definiteness marker *-ti* means 'the one just mentioned' and *-ta* 'the one mentioned longer ago'.

The absence of such metaphorical extensions of the spatial into the discursive domain in Takivatan should not surprise us, because the language has a dedicated anaphoric marker *sia* which is fulfils what Himmelmann calls a 'tracking function'. Another word that has an obvious textual function is *maupa* 'thus', which has a clear discourse-anaphoric function.

However, this does not mean that spatial deictic words and morphemes have no function in the realization of Takivatan information structure. Although we established above that there is no absolute correlation between deixis and *clausal* topicality in Takivatan, deictic forms are, by the virtue of being definite and referential, involved in the realization of topical progression, i.e. they have a function in maintaining discursive topics.

To illustrate this, we will now look at a narrative sequence from a hunting story. A group of hunters, which includes the speaker (VT) in his younger days, have gone into the mountains to hunt for deer. One of the men has gone on a reconnaissance trip and has just arrived back in the hunters' temporary camp.

(15) [A] Aupa tuða... niaŋ tu nanu sanavan minsumina ... Tiaŋ, minabaʔav tupa naip tu:

[A1] aupa   tuða ni-aŋ   tu   nanu   sanavan min-suma-in-a   Tiaŋ
thus   real NEG-PROG COMPL really evening INCH-return-PRV-LNK T.

[A2] mina-baʔav   tupa naip   tu
ABL-high.location say   DEM.S.NVIS COMPL
'But, when it wasn't really evening yet, Tiang had returned, he had come back from the
mountain and told us:'

[B] Na, maqtu laqbiŋina, naʔasa dusa ta matiskun, maluʔumi han baʔav daiðaki, pinkaunun
isian baʔavta, ŋabul.

[B1] na   maqtu   laqbiŋin-a   na-asa   dusa ta   ma-tiskun
well be.possible tomorrow-LNK IRR-have.to two   COMPL DYN-in.a.group

[B2] maluʔum-i   han baʔav   daiða-ki
disperse-PRT   be.at high.location there-DEF.SIT.PROX

[B3] pinkaun-un   i-sia-an   baʔav-ta   ŋabul
go.up-NR.INSTR LOC-ANAPH-LF   high.location-DEF.REF.DIST deer
'Well, tomorrow is possible, two of us will have to go together, and disperse when we
get to this place, and we will climb upwards to the deer that is in that place above.'

[C] A, namaqaisaq dauka, saqnutai du sia ʔukai laqaiban.

[C1] a   na-ma-qaisaq   dau-ka
INTER   IRR-DYN-in.that.direction   EMO-DEF.SIT.DIST

[C2] saqnut-ai-du   sia   ʔuka-i   laqaiban
get.stuck-PRT-EMO ANAPH NEG.have-PRT   route
'A, if he will go in that direction, he will get stuck there, without a way out.'

[D] Ansaisaŋa Atul Daiŋ tu "nis, matiŋmutin tamudana madav."

[D1] ansais-aŋ-a   Atul daiŋ   tu
forbid-PROG-ENUM   A.   large   COMPL

[D2] ni-is   ma-tiŋmut-in   ta-mu-dan-a   maðʔav
NEG-3S.F   STAT-morning-PRV   ?-ALL-road-LNK   embarrassed
'But Big Atul forbade us: "no, when it has become morning, we will leave, it is
embarrassing.'

[E] Na... s… ʔukin aipa ʔita namudanin, musbai naipa maqmut.

[E1] na   ʔuka-in   aipa   ʔita   na-mu-dan-in
well NEG.have-PRV   DEM.S.DIST.VIS there.DIST   IRR-ALL-go-PRV

[E2] mu-isbai   naipa   maqmut
ALL-cause.to.move DEM.S.DIST.NVIS   night.time
'Well, it will not be there anymore, it will be gone, it will have run away during the
night.' (TVN-008-002:130-134)

Figure 1 provides a visual representation of the anaphoric and exophoric links that establish
discourse cohesion through the use of deictic words in this narrative segment. Discourse
participants are marked by a square; anaphoric or exophoric links established by deictics are
represented as arrows. Note that only explicit elements in the text are encoded; non-expressed
arguments, even those that might be signalled by verbal morphology, are not taken into
account.[3]

---

[3] Ellipsis (or better non-expression) is an important indicator of discursive prominence in many Austronesian
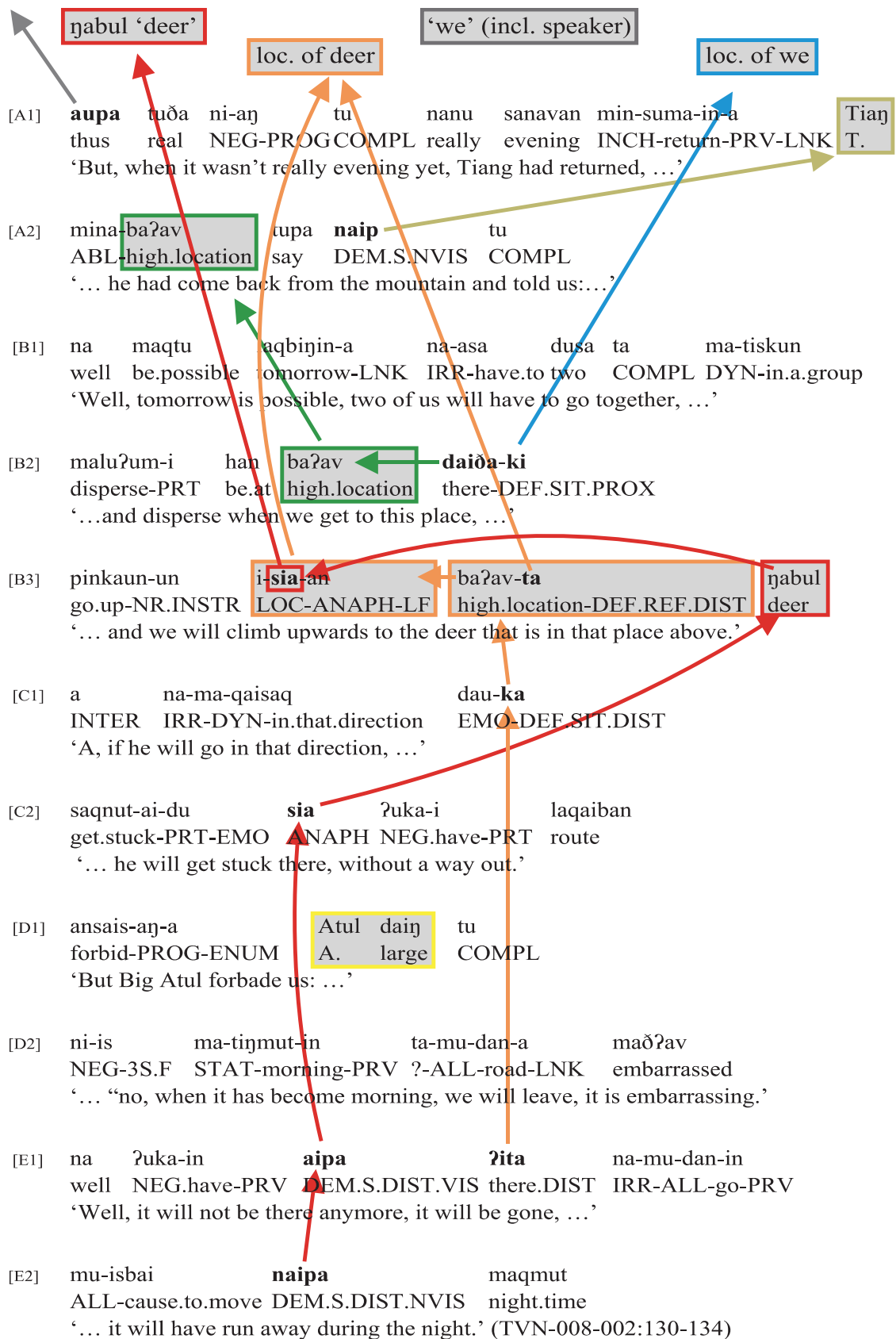languages, and in Takivatan Bunun argument ellipsis is extremely common.

ŋabul 'deer' loc. of deer 'we' (incl. speaker) loc. of we

[A1] **aupa** tuða ni-aŋ tu nanu sanavan min-suma-in-a Tiaŋ
     thus real NEG-PROG COMPL really evening INCH-return-PRV-LNK T.
     'But, when it wasn't really evening yet, Tiang had returned, …'

[A2] mina-baʔav tupa **naip** tu
     ABL-high.location say DEM.S.NVIS COMPL
     '… he had come back from the mountain and told us:…'

[B1] na maqtu aqbiŋin-a na-asa dusa ta ma-tiskun
     well be.possible tomorrow-LNK IRR-have.to two COMPL DYN-in.a.group
     'Well, tomorrow is possible, two of us will have to go together, …'

[B2] maluʔum-i han baʔav **daiða-ki**
     disperse-PRT be.at high.location there-DEF.SIT.PROX
     '…and disperse when we get to this place, …'

[B3] pinkaun-un i-**sia**-an baʔav-**ta** ŋabul
     go.up-NR.INSTR LOC-ANAPH-LF high.location-DEF.REF.DIST deer
     '… and we will climb upwards to the deer that is in that place above.'

[C1] a na-ma-qaisaq dau-**ka**
     INTER IRR-DYN-in.that.direction EMO-DEF.SIT.DIST
     'A, if he will go in that direction, …'

[C2] saqnut-ai-du **sia** ʔuka-i laqaiban
     get.stuck-PRT-EMO ANAPH NEG.have-PRT route
      '… he will get stuck there, without a way out.'

[D1] ansais-aŋ-a Atul daiŋ tu
     forbid-PROG-ENUM A. large COMPL
     'But Big Atul forbade us: …'

[D2] ni-is ma-tiŋmut-in ta-mu-dan-a maðʔav
     NEG-3S.F STAT-morning-PRV ?-ALL-road-LNK embarrassed
     '… "no, when it has become morning, we will leave, it is embarrassing.'

[E1] na ʔuka-in **aipa** **ʔita** na-mu-dan-in
     well NEG.have-PRV DEM.S.DIST.VIS there.DIST IRR-ALL-go-PRV
     'Well, it will not be there anymore, it will be gone, …'

[E2] mu-isbai **naipa** maqmut
     ALL-cause.to.move DEM.S.DIST.NVIS night.time
     '… it will have run away during the night.' (TVN-008-002:130-134)

Figure 1. Active topic chains in example (15)

The result is a schema that gives a rough impression of the topical chains established in the textual sequence above, with different colours representing different chains. These topical chains weave this sequence together into a coherently interpretable whole. From Figure 1 we can deduce the following about textual coherence in Takivatan Bunun:

- New discursive topics in a text are often established by common or proper nouns in the case of entities, or by locative nouns or verbs in the case of locations.
- Once established, entitities and locations are maintained by deictic elements, which often have deictic reference as their primary function in discourse, i.e. they are not primarily discourse-deictic markers. The relation between a referent and its anaphoric antecedent is often one that Halliday & Hasan (1976:314) call CO-INTERPRETATION: deictic elements link back to a previous reference in the text, but the anaphor and the anaphoric target do not necessarily refer to identical semiotic denotata. For instance, *aipa* in (15) E1) and naipa in (15) E2) both refer to the same deer (*ŋabul* in B3), but in (E1) it is a visible deer, and in (E2) a non-visible deer that has already run away. Similarly, in example (16) below, (B1) *Ɂita* 'there.DIST' refers back to the root *quma* 'field' in (16) A1), but the semantic target of both words is not identical: *Ɂita* refers to a location, whereas *quma* is part of a verb (*namuqumaka*) referring back to an event.
- Important discursive topics can be reinforced by an occasional repetition of nominal forms, e.g. *ŋabul* 'deer' in (B3). This is what Halliday & Hasan (1976:279) refer to as REITERATION.
- Highly salient topics do not need to be expressed; the topical arguments are simply ellipted in subsequent sentences and not marked by any deictic element.

Below is a longer narrative segment by another narrator, followed by its analysis. In it, the narrator (TM) explains how in traditional Bunun society the most important work on the field, in this case the harvest, could only be undertaken after consulting prophetic dreams.

(16)　[A] Maqai maqabasi tupa tu madaiŋɁaði namuqumaka taŋusaŋ matibahi.

[A1] maqai　　ma-qabas-i　　　　　　tupa　tu
　　　if　　　　DYN-in.former.times-PRT　say　COMPL
　　　ma-daiŋɁað-i　　　na-mu-quma-ka
　　　STAT-old-PRT　　　IRR-ALL-field-DEF.SIT.DIST

[A2] taŋus-aŋ　　mati-bahi
　　　first　　　　PROG-have.prophetic.dream
　　　'If in the old days the elders said they wanted to work on the land, they interpreted a prophetic dream beforehand.'

[B] Namaqun Ɂita maqai masihala bahia, tudip, na, sintupadu tu maqai Ɂitun asa namasihal kakaunun.

[B1] na-maqun　　　Ɂita
　　　IRR-cut.off　　there.DIST

[B2] maqai　ma-sihal-a　　　　　　bahi-a　　　　　　　　　tudip
　　　if　　　STAT-good-SUBORD　prophetic.dream-SUBORD　that.time

[B3] na　　　sin-tupa-du　　　　tu　　　　maqai　Ɂitun
　　　well　　RES.OBJ-say-EMO　COMPL　if　　　there.MED

[B4] asa　　　　na-ma-sihal　　　　ka-kaun-un
　　　be.able　　IRR-STAT-good　　CV-eat-UF
　　　'And when they wanted to go there to harvest (lit: when they wanted to cut off things in that place), if the dream was good, that meant in those days that if you were there, you could eat very well.'

[C] A maqai dipi madiqla bahia tupa tu asa ni ʔituni nalauq, nitu na … masihala kakauna
sanasia maqai, amin tu maqai ʔitun namuqða kuðaki madiqla bahi, na haiða matað.

[C1] a   maqai   dip-i   ma-diqla   bahi-a
INTER   if   then-PRT   STAT-bad   prophetic.dream-LNK

[C2] tupa   tu   asa   ni   ʔitun-i
say   COMPL   have.to   NEG   there.MED-PRT

[C3] nalauq   ni   tu   na   ma-sihal-a   ka-kaun-a
otherwise NEG   COMPL   well   STAT-good-LNK   CV-eat-LNK

[C4] sana-sia   maqai
ACCORDING.TO-ANAPH   if

[C5] amin   tu   maqai ʔitun   na-muqða   kuða-ki
all   COMPL   if   there.MED   IRR-again   work-DEF.SIT.PROX

[C6] ma-diqla   bahi
STAT-bad   prophetic.dream

[C7] na   haiða   matað
well   have   die

'And if the dream was bad, then they said that you must not go there, because otherwise
you would not eat well, if you followed the rule, but if anyone at all went back to that
place to work, and there was a bad dream, people would die.'

[D] A, maqai mataisaq … matataisaq a madadaiŋʔað tu, … maqai munʔitaʔa mavia mataisaq
tu saduʔuki siatu, sinsusuað bunuað masmamua mavisqai, mavilasa tupaka madadaiŋʔað
tu na maqtu munquma istaʔai nakasihalain kakaunun namasihala bunun.

[D1] a   maqai   ma-taisaq
INTER   if   DYN-dream

[D2] ma-ta-taisaq   a   madadaiŋʔað   tu
DYN-CV-dream   INTER   elder   COMPL

[D3] maqai   mun-ʔita   a   ma-via   ma-taisaq   tu
if   ALL-there.DIST   HESIT   DYN-why   DYN-dream   COMPL

[D4] saduʔu-ki   sia   tu
see-DEF.SIT.PROX   ANAPH   COMP

[D5] sin-su-suað   bunuað   mas-ma-muav   ma-visqa-i
RES.OBJ-CV-sow   plum   BE-CV-excessive   STAT-abundant.with.fruit-PRT

[D6] mavilas-a
have.large.fruits-LNK

[D7] tupa-ka   ma-da-daiŋ-ʔað   tu
tell-DEF.SIT.DIST   elder   COMPL

[D8] na   maqtu   mun-quma   ista-ai
well   be.possible.to ALL-field   3S.DIST-PRT

[D9] na-ka-sihal-in   ka-kaun-un
IRR-ASSOC.DYN-good-PRV   CV-eat-UN

[D10] na-ma-sihal-a   bunun
IRR-STAT-good-LNK   people

'And if they dreamt… if the elders dreamt that, if they went over there, they suddenly
dreamt that they saw that the plum tree had grown so that it was full of fruits and had
large fruits, then the elders would say that it was permitted for them to the land to work,
and they would produce good fruits, and the people would also be fine.' (TVN-012-
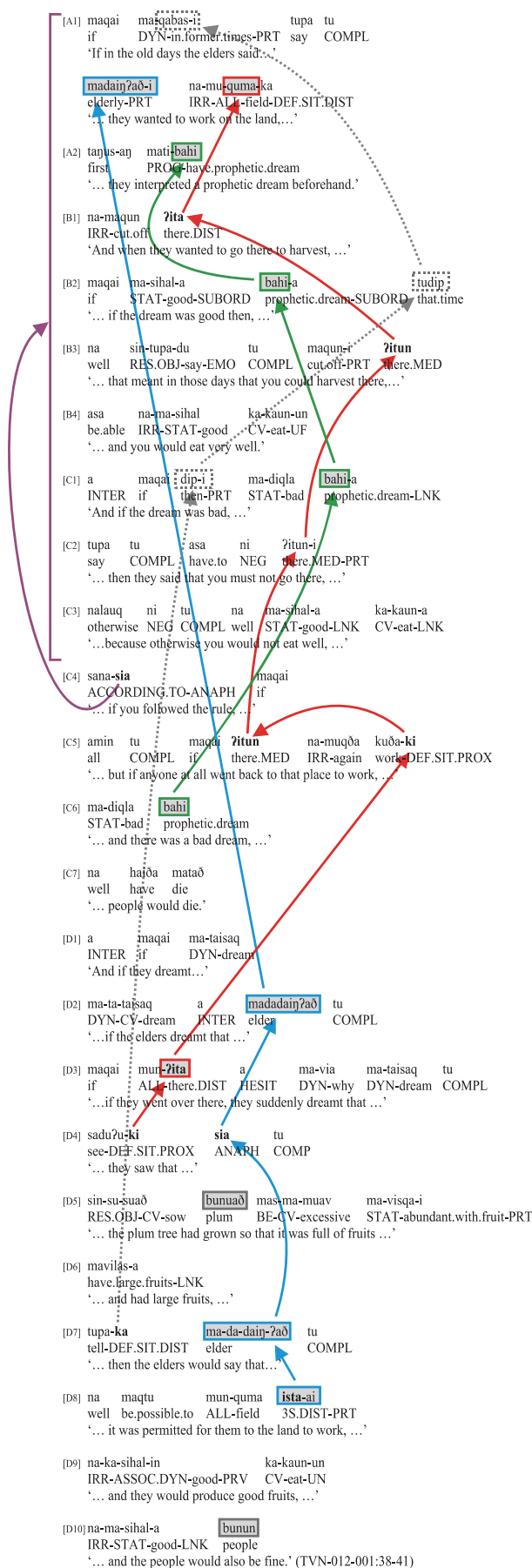001:38-41)

[A1] maqai maⁿqabas-i tupa tu
if DYN-in.former.times-PRT say COMPL
'If in the old days the elders said...'

madaiŋʔað-i na-mu-quma-ka
elderly-PRT IRR-ALL-field-DEF.SIT.DIST
'… they wanted to work on the land,…'

[A2] taŋus-aŋ mati-bahi
first PROG-have.prophetic.dream
'… they interpreted a prophetic dream beforehand.'

[B1] na-maqun ʔita
IRR-cut.off there.DIST
'And when they wanted to go there to harvest, …'

[B2] maqai ma-sihal-a bahi-a tudip
if STAT-good-SUBORD prophetic.dream-SUBORD that.time
'… if the dream was good then, …'

[B3] na sin-tupa-du tu maqun-i ʔitun
well RES.OBJ-say-EMO COMPL cut.off-PRT there.MED
'… that meant in those days that you could harvest there,…'

[B4] asa na-ma-sihal ka-kaun-un
be.able IRR-STAT-good CV-eat-UF
'… and you would eat very well.'

[C1] a maqai dip-i ma-diqla bahi-a
INTER if then-PRT STAT-bad prophetic.dream-LNK
'And if the dream was bad, …'

[C2] tupa tu asa ni ʔitun-i
say COMPL have.to NEG there.MED-PRT
'… then they said that you must not go there, …'

[C3] nalauq ni tu na ma-sihal-a ka-kaun-a
otherwise NEG COMPL well STAT-good-LNK CV-eat-LNK
'…because otherwise you would not eat well, …'

[C4] sana-sia maqai
ACCORDING.TO-ANAPH if
'… if you followed the rule, …'

[C5] amin tu maqai ʔitun na-muqða kuða-ki
all COMPL if there.MED IRR-again work-DEF.SIT.PROX
'… but if anyone at all went back to that place to work, …'

[C6] ma-diqla bahi
STAT-bad prophetic.dream
'… and there was a bad dream, …'

[C7] na haiða matað
well have die
'… people would die.'

[D1] a maqai ma-taisaq
INTER if DYN-dream
'And if they dreamt…'

[D2] ma-ta-taisaq a madadaiŋʔað tu
DYN-CV-dream INTER elder COMPL
'…if the elders dreamt that …'

[D3] maqai mun-ʔita a ma-via ma-taisaq tu
if ALL-there.DIST HESIT DYN-why DYN-dream COMPL
'…if they went over there, they suddenly dreamt that …'

[D4] saduʔu-ki sia tu
see-DEF.SIT.PROX ANAPH COMP
'… they saw that …'

[D5] sin-su-suað bunuað mas-ma-muav ma-visqa-i
RES.OBJ-CV-sow plum BE-CV-excessive STAT-abundant.with.fruit-PRT
'… the plum tree had grown so that it was full of fruits …'

[D6] mavilas-a
have.large.fruits-LNK
'… and had large fruits, …'

[D7] tupa-ka ma-da-dain-ʔað tu
tell-DEF.SIT.DIST elder COMPL
'… then the elders would say that…'

[D8] na maqtu mun-quma ista-ai
well be.possible.to ALL-field 3S.DIST-PRT
'… it was permitted for them to the land to work, …'

[D9] na-ka-sihal-in ka-kaun-un
IRR-ASSOC.DYN-good-PRV CV-eat-UN
'… and they would produce good fruits, …'

[D10] na-ma-sihal-a bunun
IRR-STAT-good-LNK people
'… and the people would also be fine.' (TVN-012-001:38-41)

Figure 2. Active topic chains in example (16)

Despite the fact that there is a relatively large difference in the deictic forms used in (15) and (16), it is clear that deictic expressions function very much in the same way: one the one hand, they have a clear spatial deictic function; on the other they are used to maintain discursive topics throughout the narrative segment.

What is interesting in (16) is that a repetition of nominal forms (e.g. *bahi* in A2, B2, C1, etc.) appears to be used for indicating contrast, while deictic forms are employed to establish a consistent, stable theme (e.g. the string maintaining the salience of *quma* 'land' throughout the segment). Further research will indicate whether this is a peculiarity of this particular narrator, or a general strategy in Takivatan Bunun.

4. Conclusion

The two examples above indicate that it is correct, as De Busser (2009:425) asserted, that in actual text deictic forms that make a distance distinction are primarily involved in the expression of spatial (or temporal) deixis. There is also no clear correlation between spatial deictic forms and intra-sentential topicality: (1) deictics occur on both topical and non-topical arguments, and on predicates and adverbials; and (2) topical arguments do not need to be marked by deictics.

However, in narrative discourse deictic elements are important tools in the creation of textual cohesion. Typically, discursive topics (as opposed to clausal topics) are established by full nominal reference and are then subsequently maintained by a combination of deictic markers and the ellipsis of topical arguments. The primary function of these deictic markers is in all instances above still spatial deictic reference.

It is also clear from the examples that there are complex interactions between different deictic paradigms. How these interactions exactly work will be the subject of future research.

Bibliography

Busser, Rik De. 2009. Towards a Grammar of Takivatan Bunun: Selected Topics. Melbourrne: Research Centre for Linguistic Typology, La Trobe University PhD.

Busser, Rik De. 2011. Towards an analysis of argument alignment in Takivatan Bunun. *Studies in Language* 35(3). 523–555. doi:10.1075/sl.35.3.02deb.

Halliday, M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. (English Language Series). London: Longman.

Himmelmann, Nikolaus P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In Barbara Fox (ed.), *Studies in Anaphora*, 203–252. Amsterdam: John Benjamins.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.

Keenan, Edward Louis. 1976. Remarkable subjects in Malagasy. In Charles N. Li (ed.), *Subject and Topic*, 247–302. New York: Academic Press.

Schachter, Paul. 1976. The subject in Philippine language: Topic, actor, actor-topic, or none of the above. In Charles N. Li (ed.), *Subject and Topic*, 491–518. New York: Academic Press.

Zeitoun, Elizabeth. 2000. 布農語參考語法. (5). Taipei: YLib / 遠流.

The definite marker in Balinese

Asako Shiohara (ILCAA, Tokyo University of Foreign Studies)

Ketut Artawa (Udayana University)

1 Introduction

The presence of the definite suffix -*é* (-*né* after vowel) is a distinctive feature that Balinese exhibits among the languages in the Western Malayo-Polynesian languages spoken in Indonesia. Most of the languages do not have a special marker for indicating definiteness, while some languages, such as colloquial Malay and Javanese, employs the third person genitive suffix (-*nya* in Malay, -(*n*)*é* in Javanese) as a definite marker, as a result of semantic extension. Sentence (1) is an example of colloquial Malay. The third person genitive form *nya* in The NP *garam-nya* 'salt-3GEN' may be interpreted as some third person that can be identified by the speaker and the hearer or the definite marker, indicating the salt is identifiable from the linguistic or non-linguistic context, for example, being present at the place of utterance.

(1)   *ambilkan   garam-nya*

     take   salt-3GEN

      "Take his salt." or "Take the salt (e.g. on the table)."

Sentence (2) and (3) are Balinese examples corresponding to example (1) above. In sentence (2) below, the NP with -*é* refers to the entity that is linguistically or non-linguistically identifiable. Unlike Malay and Javanese, the form is distinguished from the third person genitive suffix -*né,* which occurs in example (3).

(2) *jemakang   uyah-é*!

   take salt-**É**   'Take **the** salt (e.g., on the table).'

(3) *jemakang   uyah-né*!

    take   salt-**3GEN**   'Take **his/ her/ its/ their** salt.'

The suffix -*é* may occur when the head noun is modified by other constituent that shows the referent is definite, such as a demonstrative, the first and second person pronoun, or a relative

clause.

(3) *anak-é ento* 'that person' (a demonstrative modifier)

(4) *pianak tiang-é* (a pronominal modifier)

(5) *jelema-n-é   ané   maling   dompet*

    person-INS-É   REL   steal   wallet

    'The man who stole a wallet' (relative clause)

The only exception is when the head noun is attached by the third person pronominal suffix –*né* of low register, with which the suffix -*é* cannot co-occur.

*(6) **pianak-né-n-é*

Balinese has several other third person pronoun distinguished according to the social status of the hearer or the referent. (See Arka (2005: 174)) The other three third person forms are realized as an independent pronoun, which occurs with the suffix -*é*.

(7) *pianak ipun-é*

(8) *oka-n-ida-n-é*

This study aims to describe the semantic conditions in which the suffix -*é* occurs. Before looking in to the point, the morpho-syntactic property of the suffix –*é* will be briefly seen with in section 2. In section 3, we will give an attempt to describe semantic range that the suffix -*é* denotes by the elicited data. In section 4, we will examine the actual occurrence in narrative text.

2 Morphophonological and morpho-syntactic property of -*é*

Balinese has a morphophonological rule that a sound *n* is inserted between a vowel final stem and a vowel initial suffix (e.g., *beli-n-ang* (⬅*beli* 'buy'+ -*ang* (applicative suffix)). The suffix –é also follow this rule, as shown in *jelema-n-é* (*jelema* 'person'+the suffix –*é*, buku-n-é 'book' + the surrix –*é*).

The suffix -*é* is normally attached to the head noun.

(9) *anak-é luh* 'the woman' (an adjective modifier)

(10) *anak-é ento* 'that woman' (a demonstrative modifier)

(11)   *marga-n-é di   Bali*

road-ins-é in   Bali   'The roads in Bali' (a PP modifier)

(12)   *jelema-n-éané*   *maling*   *dompet*
person-ins-é   rel   steal   wallet
'The man who stole a wallet' (relative clause)

The only exception is an NP in which the head noun is modified by a noun; in this structure, the suffix *-é* is attached to the modified noun, as in *montor jepang-é* 'the Japanese car', *sebun kedis-é* 'the bird's nest'; a personal pronoun exhibits features similar to a noun in this environment, as seen in *pianak tiang-é* 'my child (child 1SG-É)', while a demonstrative pronoun does not, as seen in *anak-é ento* 'that woman'.

3 Semantic range the suffix *-é* covers

Like all the other grammatical categories, the semantic range 'definite marker' varies cross-linguistically. To see the semantic features that 'definite markers' in many languages share, we will start with introducing the use of the English definite article *the*, based on the discussion of Lyons (1999: 1-15) and Quirk et al. (1985: 265-268). Roughly speaking the definite article indicates that the referent of the NP is identifiable to the addressee linguistically or non-linguistically. Three main three categories of the use are shown below.

I. Situational use
The reference of the NP is identified by the extralinguistic situation that the speaker and the hearer share in examples (13)-(16).

(13) Just give **the shelf** a quick wipe will you, before I put this vase on it.

In example (13) the referent of *the shelf* is identifiable from the physical situation in which the speaker and hearer are located.

(14) I hear **the prime minister** behaved outrageously again today.

In (14) the relevant situation is wider; it is most plausibly a country in which the speaker and hearer are located, and the referent of **the prime minister** is identifiable as the prime minister of

the country.

(15) **The moon** was very bright last night.
(16) **The president of Ghana** is visiting tomorrow.

Example (15) and (16) belong to the extreme type of situational use; the situation is the whole world here, in which 'the larger situation is scarcely distinguishable from general knowledge' (Quirk et al. 1985: 266). In (15) and (16), the referent of the moon and the president of Ghana is identifiable from the wider situation of the whole world, or general knowledge.

II. Anaphoric use
The referent of the NP with the definite article is identifiable from the linguistic information given earlier in the discourse.

(17) **An elegant, dark-haired woman**, a well dressed man with dark glasses, and **two children** entered the compartment. I immediately recognized **the woman**. **The children** also looked vaguely familiar.

III. Associative use
This use can be thought of as a combination of the anaphoric and general knowledge types. The referent of the driver in (18) is identifiable because it can be associable from the referent of *a car*, which is mentioned in the previous sentence.
(18) I had to get a taxi from the station. On the way **the driver** told me there was a bus strike.
(19) They've just got in from New York. **The plane** was five hours late.

I made an elicitation research as to how these situations expressed in (13)-(19) are expressed in Balinese. Balinese speakers judged the suffix -*é* covers all the semantic types in the list above, except (16) in the situational use. Balinese sentences corresponding (13)-(19) above will be shown below.

Situational Use
Example (20)-(23) are examples of situational use, which roughly corresponds to the English example (13)-(16). In (20)-(22) the Balinese NP corresponding to the English definite NP

occurs with the suffix –é, while in (23) it is not.

(20) *Sap-sapin gen* **rak-é** *bedik,*
      wipe   just   shelf-**É**   a.little

      setonden tiang   ngejang pas-é   ené   duur-né
      before   1sg   ao.put   vase   this   above-3GEN

   "Just give the shelf a quick wipe will you, before I put this vase on it."

(21) *Icang   ningeh   buin*   **presiden-é**
      I   hear   again   president-**É**

      melaksana   sing   lung   jani
      behave   not   good   today

      "I heard the president behaves impolitely again today."

(22) **Bulan-é** *galang   magladaran   ibi   peteng*
      moon-**É**   bright   very   yesterday night

      "The moon was very bright last night."

(23) **Presiden Ghana-∅** *lakar   teka   mani*
      president Ghana   will   come   tomorrow

      "The president of Ghana is visiting tomorrow".

Anaphoric use

Examples (24) is an example of anaphoric use, which corresponds to example (16).

(24) *Anak   luh   jegeng   lan   anak   cenik   ajak   dadua*
      person   woman   elegant and   person small   with   two

      macelep ke kamar-é.
      enter   to   room- know

      Prajani   icing   nawang   anak-e   luh   ento.

immediately	1sg	know	person-É	woman	that

Anak-é	*cenik*	*ento*	*masi*	*cang*	*mase*	*nawang.*
person-É	small	that	still	I	vaguely	know

"An elegant woman and two children entered the room. I immediately recognized the woman. The children also looked vaguely familiar".

Associative Use

Examples (25) and (26) are examples of associative use.

(25)
Icang	*musti*	*numpang*	*taksi*	*uli*	*penambangan.*
1sg	need	ride	taxi	from	station

Di jalan	*sopir-é*	*nyambat*	*ada*	*kecelakaan*	*tunian*	*suba.*
at way	drive-E	tell	exist	accident	a while ago	already

  "I had to get a taxi from the station. On the way the driver told me there was an accident a few hours ago".

(26)
Ia ajak	*tetelu*	*mara*	*san*	*teka*	*uli*	*New York.*
3 with	three	newly	just	come	from	New York

Kapal-é	lambat	limang	jam
plane-	late	five	hour

"The three people have just got in from New York. The plane was five hours late".

From what we have seen so far, we could say that the NP marked with suffix *-é* in Balinese covers the semantic range that is quite similar to that of English definite NP, except the case in which the referent is clearly unfamiliar between the speaker and the hearer, as in (23).

Another difference between definite markers in Balinese and English is caused by the presence of the third person genitive suffix *-né* in Balinese, which is a counterpart of *-nya* in Malay. This pronominal suffix occurs instead of *-é* when the referent can be related to an already mentioned entity, and is therefore definite. For instance, the pronominal suffix *-né* cannot be replaced by the definite suffix *-é* in example (27).

(27) *Umah   icang-é   resem.   Kakus-**né** (*-é)   uwug,   raab-**né** (*-é)   bolong*

     house 1SG-É   shabby. toilet-3GEN   broken, roof-3GEN   have a hole

     'My house is shabby. The (lit. its) toilet is broken and the (lit. its) roof has a hole.'

4. Condition in which the suffix -é occurs in narrative text

This study aims to describe the semantic conditions in which the suffix *-é* occurs. It plays almost the same with the English definite article, in that it indicates that the referent is identifiable from the addressee.

4.1 Data

  We examined sentences included in short stories that appear in Balinese school textbooks for elementary school students in order to examine how the sufix *-é* actually occurs in discourse. Table 4 shows the list of the stories examined.

Balinese has been taught for a long time in Bali province at least 50 years ago. I employed the three types of textbook below as a source.

Series Title	published year
Sari	1972
Titi Basa Bali	1981
Kusmasari	1995

	Title	token/ type	Occurrence of -é	genre	Series title
1	Masekolah 'going to school'	207/123	6	essay	Sari
2	Nyakan 'cooking rice'	232/ 126	18	essay	Sari
3	Katak teken sampi 'the frogs and cow'	255/134	9	folktale	Sari
4	Gajah Mada (A prime minister of the Majapahit Empire)	146/89	9	history	Sari
5	Galungan (one of Balinese holidays of celebration)	199/113	19	essay	Sari
6	Puputan Margarana (A battle between Indonesia and Holland)	317/151	31	history	Sari
7	Kakap Emas (the golden axe, or a honest wood cutter)	685/293	40	folktale	Sari
8	Koperasi Sekolah (School Co-op)	224/117	18	essay	Titi Basa Bali
9	Lampu Kuning (the yellow light)	336/161	24	essay	Titi Basa Bali
10	Blabar (flood)	267/154	14	essay	Titi Basa Bali
11	Nguwangun Bale Banjar (Building a meeting place of the village)	58/36	5	essay	Kusmasari
12	Bulan Kuning (Yellow moon, a folktale)	209/107	8	folktale	Kusmasari

The number of the total words included are 3038. The suffix *-é* occurs in 201 NPs. We can see the following points as to the formal property of the –é marked NP.

(1) *-é* always occur when an NP includes other modifying constituents that makes the referent identifiable, such as personal pronoun except the third person pronoun *-né*, a proper noun, a demonstrative in an NP. (59 examples)

(2) *-é* often occur when NP includes a relative clause[1]. (17 examples)

(3) *-é* always occur when the head NP denotes a positional relation of a entity, irrespective of whether the entity is linguistically or non-linguistically indefinable or not. (13 examples)

> e.g.   *duur batu lémpéh-é* 'on the step stone'
>
> *selangan punyan jagung-é* 'between the corn tree'

(4) Proper nouns exhibit various behavior as to the co-occurrence with *-é*.

(i) Personal names and kin terms are not marked by –é, but marked by the article *I*, which precedes a male personal name (e.g. *I Wayan* 'Mr. Wayan'), kin term (e.g. *I Mémé* 'Mother') and *Ni*, which precedes a female (e.g. *Ni Sari* 'Ms.Sari').

(ii) Some compound nouns indicating a place name do not occur with the suffix *-é*, while some do. It depends on the head noun.

(a) Place names occurring without the suffix *-é*;

> *Désa Marga* 'the Marga village'
>
> *Karajaan Majapait* 'the Majapahit Kingdome'

(b) Place names occurring with the suffix *-é*;

> *tukad Gangga-n-é* 'the Ganges River'
>
> *kota Singaraja-n-é* 'Singaraja City'

As for the *-é* marked NPs that do not have the formal features shown above, most of the *-é*

---

[1]A relative clause does not necessarily compatible with the suffix *-é*, if the referent of the whole NP is not identifiable from the preceding discourse.

(27) …*lantas   masuah   tur   nganggo   baju   ané   kedas.*
      then   comb   and   use   clothes REL   clean
      "…, then combs and puts on clean clothes"

(28) *Désa-desa   ané   paling   sangeta   kena   blabar,   luire*:
      villages   REL   most   strongly   affected flood   that is

Banjar Bali,   Kampung Anyar,   Kampung Kajanan…
Village Bali   village   Anyar   village   Kajanan…

"Villages that are most affected by the flood is; Banjar Bali, Kampung Anyar, Kampung Kajanan…".

marked NP can be interpreted as in one of the four usages listed in section 2. Definition of some usage is vague to some extent, and it is not necessarily easy to classifying each occurrence into one usage category.

anaphoric use	54
associative use	18
situational use	50

We will first focus on only 'anaphoric' use of the suffix -*é*, as to which relatively objective classification is possible by tracing referents expressed in the text, and then look for other usages.

4.2 Anaphoric use of the suffix –*é*

Consider example (28).

(28)(a)   *Ni*   *Sari₁*   *jumah*   *nulungin*   *magarapan*,
          Ms.   Sari   at.home   AO.help   house.works

          nimba   *yéh,*   *ngumbah*   *piring*   *tekén*   *nyakan.*
          ao.take   water   ao.wash   dish   and   cook.rice

(b)   *Di*   *paon*   *ia*   *nungguhang*   *pangedangan₁*
        at   kitchen   3   AO.take   cooking.pot

        isinin-a   *yéh₂*   *atenga,*   *tekepin-a*   *aji kekeb₃,*
        put.in-3ACT   water   half   cover-3ACT   with lid,

        tumuli   *ngendihang*   *api*.
        then   AO.make   fire

(c)   *Suba*   *kéto*   *lantas*   *ia*   *nyemak*   *baas₄*   *di* ***pulu-n-é₅***
      after   like.that   then   3   ao.take   rice   at rice.chest-É

	petang	*kobokan,*	*ingsah-a*		*wadahin-a*		*pané.*
	four	cup	wash-3ACT		put-3ACT		bowl

(d)   *Di subané kedas*   *lantas*   *emeh-a.*
      after   clean   then   leave-3ACT

(e)   *Suba*   *panes*   *yéh-é₂*   *di*   *pangedangan-é₁*
      after   hot   water-e   at   cooking.pot-É

      baas-é ₄   *wadahin-a*   *kuskusan₆*
      rice- É   put-3ACT   steamer

      lantas   *jang-a*   *duur*   *pangedangan-é₁*
      then   put-3ACT   on   cooking.pot-É

      tur   *tepukin-a*   *aji*   *kekeb₃.*
      and   put-3ACT   with   lid

(a) Sari is at home and to help (parents) with the house works, drawing water, washing the dishes, and cooking rice."

(b) In the kitchen, she took a cooking pot, puts water a half, covers it with a lid, and then makes fire."

(c) After that, she takes four cup of rice in the rice chest, wash it, and put it in 'pane'."

(d) After that, she leaves the rice as it is.

(e) After the water in the pot becomes hot, she put the rice in the steamer, and put it on the cooking pot, and covers it with a lid.

Here we can see rough correspondence between previously mentioned referents and the NP marked with the suffix *-é*; among five NPs occurring with the suffix *-é*, four refer previously mentioned referent. We can, however, see the following disagreement between a 'previously mentioned referent' and *-é* marked NP.

(i) *Ni Sari* 'Sari', the referent playing an actor throughout the example is never expressed by *-é* marked NP; it is expressed by either an independent pronoun or a pronominal clitic, or left

unexpressed. This point can be explained by the Balinese referent marking system that exhibits a strong tendency in which so-called "active" referent can not be expressed by a lexical NP. This point will be dealt with in 4.2.1.

(ii) Not all the referents previously mentioned are expressed by –é marked NP. For example, *kekeb* 'a lid' is expressed by a bare noun when it is mentioned for the second time.

4.3　　'Active' NP

Balinese exhibits strong tendency that a referent that is mentioned in the immediately preceding clause, and therefore "active" (Lambrecht 1999: 94) is not expressed by a lexical NP, but expressed by a pronoun or left unexpressed (zero anaphora, hereafter), which are not compatible with the suffix –*é*. We examined only one short articles (article number 2), and counted the occurrence of zero anaphora, pronominal suffix –*a*, independent pronoun *ia*, and NP with or without the suffix –*é* there. The result is shown in table 1.

	The referent is mentioned in			
NP	an adjacent clause in the same sentence	an adjacent clause but not in the same sentence	a non-adjacent clause	Not mentioned previously
zero anaphora	19	1	0	1
pronominal suffix –*a*	8	4	1	0
the third person pronoun *ia*	0	2	0	0
lexical noun with é	0	1	8	4
without é	0	0	9	21

Table 1 occurrence count of zero anaphora, pronominal suffix –*a*, independent pronoun *ia*, and NP with or without the suffix –*é*.

This table shows the points (i) and (ii) below on 'active NP'.

(i) The referent that is already mentioned in the same sentence normally occur as zero-anaphora, or expressed by the pronominal suffix –*a* (-*a* 3ACT) in all the examples.

(ii) The referent mentioned in the immediately preceding sentence is mostly expressed by the

pronominal suffix (-*a* 3ACT) or the independent pronoun (*ia* '3').

et's return to the example (28) above to examine the point (i) and (ii). The example consists of five sentences, which tells how a girl called Sari cooks rice. In (a), the actor Sari is introduced in the form of personal name with the marker *Ni* for the first time, and then expressed by zero anaphora. In (b)-(e), the same referent is expressed by (ii) the pronominal suffix –*a* or (iii) the independent pronoun *ia* when it occurs for the first time in a sentence, and then expressed by the (i) zero anaphora or (ii) the pronominal suffix –*a* in the same sentence.

(29))=(28)(a)   *Ni   Sari₁*   *jumah*   *nulungin* *magarapan*,
               Ms.   Sari   at.home   AO.help   house.works

        φ₁   *nimba yéh,*   *φ₁* *ngumbah piring*   *tekén*   *φ₁*   *nyakan.*
              AO.graw water   AO.wash.dish   and   AO.cook.rice

   (b) *Di*   *paon*   *ia₁*   *nungguhang*   *pangedangan₂*
        at   kitchen   3   AO.take   cooking.pot

        φ₂ *isinin-a₁*   *yéh*   *atenga,*   *φ₂ tekepin-a₁*   *aji kekeb,*
              put.in-3ACT   water   half   *φ₂* cover-3ACT ₁   with lid,

        tumuli   *φ₁* *ngendihang*   *api*.
        then   AO.make   fire

   (c) *Suba*   *kéto*   *lantas*   *ia₁*   *nyemak*   *baas*   *di pulu-n-é*
        after   like.that   then   3   ao.take   rice   at rice.chest-É

        petang   *kobokan,* *ingsah-a₁*   *wadahin-a₁*   *pané.*
        four   cup   wash-3ACT   put-3ACT   ?

   (d) *Di subané* *φ*   *kedas*   *lantas*   *emeh-a₁.*
        after   clean   then   leave-3ACT

(e) | *Suba* | *panes* | *yéh-é* | *di* | *pangedangan-é,* |
|---|---|---|---|---|
| after | hot | water-e | at | cooking.pot-É |

baas-é	*wadahin-a₁*	*kuskusan*
rice- É	put-3ACT	steamer

lantas	*jang-a₁*	*duur*	*pangedangan-é,*
then	put-3ACT	on	cooking.pot-É

tur	*tepukin-a₁ aji*	*kekeb.*
and	put-3ACT	with	lid

(f) Sari is at home and to help (parents) with the house works, drawing water, washing the dishes, and cooking rice."

(g) In the kitchen, she took a cooking pot, puts water a half, covers it with a lid, and then makes fire."

(h) After that, she takes four cup of rice in the rice chest, wash it, and put it in 'pane'."

(i) After that, she leaves the rice as it is.

(j) After the water in the pot becomes hot, she put the rice in the steamer, and put it on the cooking pot, and covers it with a lid.

4.4 Associative use

(30) | *Sekolah-sekolah* | *dasar-é* | *di Bali*, |
|---|---|---|
| school-school | elementary-E | in Bali |

liunan	*jani*	*suba*	*ngelah*	*koperasi*	*Sekolah.*
many	now	already	AV.have	co-op	school

Koperasi Sekolah-é	*ento*	*ka=laksanain*
co-op	shool- E	that	pass=operate

baan	***murid-murid-é***	*muah*	***guru-guru-n-é***.
by	students-E	and	teacher-INS-E

"As for the elementary schools in Bali, many of them already have the school co-op. The School co-op is operated by the students and teachers."

(31) *Dugas* *dina* *Buda* *Pon* *uku Sungsang,*
    time day Buda Pon calendar Sungsang

    tanggal *7* *Januari* *1981,*
    day 7 January 1981

    di Buléléng *ada* *blabar* *gedé.*
    in Buleleng exist flood big

    Kota Singaraja-n-é *ancab-a* *tekén* **yéh-é**.
    city Singaraja-INS-E flood-3act by water-E
    "On the day of Buda Pon, Sungsang, 7 January 1981, there was a big flood in Buleleng, and the Singaraja city was flooded by the water"

4.5 Situational use and idiosyncratic of individual lexical word

The speaker and the addressee may be expressed by a lexical noun in Balinese, and they are marked with the suffix –*é*.

(32) "*Nah* **cerik-cerik-é** *ajak makejang,*
    well children-E all
    nyen *bisa* *nyautin* *patakon* **bapak-É** *ene* ?
    when can av.answer question father-e this
    "Well, all of you (lit. the children), who can answer this question of mine (lit.father)?

Balinese cultural or social things which are familiar to them are often expressed by -*é* marked NP in its first mention, because the text dealt with in this research are targeted to Balinese children. This type of –*é* marked NP often occurs in the begeninng of the text or the paragraph, and announces the setting.

(33)   ***Anak-é   di   Bali***   *di   nuju   Galungan-é*
       people-e in   Bali   on   time   Galungan-e

       makajang pada   ngaturan banten   di   sanngah,
       all   all   offer   offering at   family.temple

       "At Galungan the people in Bali all offer the offerings to the family temples…"

(34)   *Jani   **margan-é di   Bali***   *ngancan ramé   pesan.*
       now   road-e   in   Bali   transport lively   very.much
       "Now, the traffic is very busy in roads at Bali"

Some lexical nouns marked with *-é* in its first mention in the text. Many of them express the natural things and some of them express a specific place in the house.

       natah 'ground'
       tegal 'field'
       surya 'sun'
       punuk-pundukan 'dike (between rice fields)'
       bét-bét 'bush'
       langit 'sky'
       tukad 'river'
       bulan 'moon'
       angin aris 'a gentle breeze'
       pulu 'rice keeper'
       lebuh 'gate of the house'

Most of them are familiar referents among the authors and the expected readers, and therefore we could consider that the suffix *–é* occurs here because the referents are situationally identifiable. But we should note that the occurrence seems to be determined by the lexical property that each individual word has.

Let's return to example (28), which is dealt with in 4.2 above. Here, the lexical noun *pulu* 'rice keeper' is marked with *-é*, while the noun *paon* 'kitchen' is not marked, although both of the

word indicate a specific place in the house.

(35)(=(29)(a)

Ni	*Sari*₁	*jumah*	*nulungin*	*magarapan,*
Ms.	Sari	at.home	AO.help	house.works

nimba	*yéh,*	*ngumbah*	*piring*	*tekén*	*nyakan.*
ao.take	water	ao.wash	dish	and	cook.rice

(b)

Di	***paon-∅***	*ia*	*nungguhang*	*pangedangan*
at	kitchen	3	AO.take	cooking.pot

isinin-a	*yéh*	*atenga,*	*tekepin-a*	*aji kekeb,*
put.in-3ACT	water	half	cover-3ACT	with lid,

tumuli	*ngendihang*	*api.*
then	AO.make	fire

(c)

Suba	*kéto*	*lantas*	*ia*	*nyemak*	*baas*	*di **pulu-n-é***
after	like.that	then	3	ao.take	rice	at rice.chest-É

petang	*kobokan,*	*ingsah-a*	*wadahin-a*	*pané.*
four	cup	wash-3ACT	put-3ACT	bowl

Also consider the following example. This is a sentence occurs in the same text nyakan 'cooking rice' after seven sentence's interval after example (29) above. In this sentence, the action of Sari putting a offering to several places in the house is expressed. Some places are expressed by a –é marked noun, while other are not. We cannot explain the difference from the semantic feature that each word has.

(36)

Di	*suba-n-é*	*lebeng*	*lantas*
at	after	ready	then

Ni Sari	*nanding*	*banten*	*nasi*	*jotan*

Ni Sari   AV.make   offering   rice   ?

lantas   mabanten di   tugu,
then   offer   in   temple

*di sanggah,   di **natah-é**, di sémér, di paon,   di **pulu-n-é***
in family.temple in yard-E   in well   in kitchen in place for keeping rice

*tekén   di **lebuh-é**.*
and   in gate.of.the.house

"When everything is cooked, Sari prepare rice offering then place the offering at *tugu²*, the family temple, in the yard, in the well, in the kichen, the pulu ('place' for keeping rice) till at the from gate of the house".

5. Conclusion

In this paper we have seen that the semantic onditions in which the suffix *-é* occurs. From the elicited examples, we could see that the suffix *-é* Balinese exhibits almost the same semantic range as English definite article does, in that it indicates that the referent of the NP is linguistically or non-linguistically identifiable. From the Balinese translation from the English examples in which the definite article occurs, we can see that the Balinese suffix –é exhibits the three main use of English definite article, that is, the situational use, the anaphoric use, and the associative use. Only the difference is that the –é suffixed NP in Balinese does not refer to a referent that is clearly unfamiliar to the addressee, even though the referent can be identifiable from the larger situation or general knowledge.

From the written data obtained from the elementary school textbook, we could see that the occurrence or absence of the suffix *-é* is determined by the feature that the individual lexical noun has.

---

² special temple buiding in the house for the spirit guarding the house

References

Arka, I Wayan 2005. Speech Levels, Social Predicates and Pragmatic Structure in Balinese: A Lexical Approach. *Pragmatics* 15:2/3. pp. 169-203.

Lyons, Christoper: 2000. *Definiteness*. Cambridge: Cambridge

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik, 1985. *Comprehensive Grammar of the English Language*. London: Longman.

From Borneo to Bantu:
How the Malagasy third person genitive pronoun *-ni may have become a locative suffix in Swahili

Alexander Adelaar
Research Institute for Languages and Cultures of Asia and Africa,
Tokyo University of Foreign Studies
Asia Institute, The University of Melbourne

Abstract

The Proto Malayo-Polynesian 3rd person singular genitive pronoun *-ni=a evolved into *-n(i)ɛ in Southeast Barito languages (South Borneo and Madagascar), where it has maintained the functions of possessive and agent pronoun and is also used after prepositions. In Maanyan and Malagasy reflexes of *-n(i)ɛ also frequently occur in locative adverbial constructions. These constructions as a rule have a locative prefix, but in some cases the latter was deleted. This situation might have paved the way for *-n(i)ɛ to be borrowed as a locative suffix in Swahili and other Bantu languages.

1. Introduction

In this paper[1] I show how the locative suffix *–ni* in Swahili and other Bantu languages in eastern and southern Africa may have evolved from the Southeast Barito 3rd person singular genitive pronoun *-n(i)ɛ (Adelaar and Kikusawa 2014:513). This pronoun in turn derives from the Proto Malayo-Polynesian 3rd person singular genitive pronoun *-ni=a.[2] It became *-ni* or *-nɛ* in various languages belonging to the Southeast Barito language group[3], the members of which are spoken in South Borneo and Madagascar (see maps 1 and 2). In Maanyan and in the Merina dialect of Malagasy[4], *-ni* has maintained the functions that are generally ascribed to *-ni=a and its reflexes in contemporaneous Austronesian languages, namely those of possessive and agent pronoun as well as pronoun after prepositions. Maanyan and Malagasy *-ni* also frequently occurs in locative adverbial constructions. These constructions as a rule have a locative prefix, but in some cases the latter does not surface. When that happens, a situation arises in which the role of *-ni* becomes ambiguous and

---

[1] I would like to express my gratitude to Yuko Abe (Tokyo), Tom Hoogervorst (Leiden), Yuko Kitada (Cologne), Waruno Mahdi (Berlin), Laurent Sagart (Paris), and in particular Martin Walsh (Cambridge), for useful feedback on a previous draft of this paper. The usual disclaimers apply.

[2] See Blust (2013:443), who derives this pronoun from a genitive linker *ni and a 3rd person singular pronominal clitic *a.

[3] Maanyan, Samihim (in South Borneo) and many Malagasy dialects have *-ni*; other Malagasy dialects (especially in the Southwest and South of Madagascar) have *-e*; Dusun Witu (in South Borneo) has *-ɛ* after final consonants and *-nɛ* after final vowels.

[4] Merina Malagasy is spoken in the Merina highlands of central Madagascar. It has been the dialect of central government for the last two centuries and became the basis for the standardization of Malagasy. It is the best described Malagasy dialect and the source for the Malagasy data presented in this paper unless indicated otherwise.

could potentially be re-interpreted as a locative marker. That it has not reached that point in any of the Southeast Barito languages (as far as I know) is most

likely due to the well-established position of *-ni* and its cognates as a personal pronoun. However, the occurrence of *-n(i)ɛ in locative adverbs might have confused outsiders who came in contact with Southeast Barito speakers, a situation which arose when the early Malagasy encountered Bantu speakers in coastal East Africa. This contact might have paved the way for *-n(i)ɛ to change into a locative suffix in Swahili and other Bantu languages (see Map 2). (In this context, the reader is reminded of the likelihood that the first Malagasy migrants stayed on the East African mainland and mixed with Bantu speakers before they finally settled in Madagascar (Deschamps 1960, Adelaar 2007, 2010).

In the pages to follow I draw a scenario in which a change from the Southeast Barito genitive pronoun *-n(i)ɛ into the Bantu locative marker *-ni* or a related form might have taken place. In Section 2 I give an overview of the use of *-ni* in some Southeast Barito languages, and in Section 3, of the use of *-ni* and its related form in Bantu languages. In Section 4 I give arguments for the assumed relationship between Southeast Barito *-n(i)ɛ and Bantu *-ını (see below). I make some concluding remarks in Section 5.
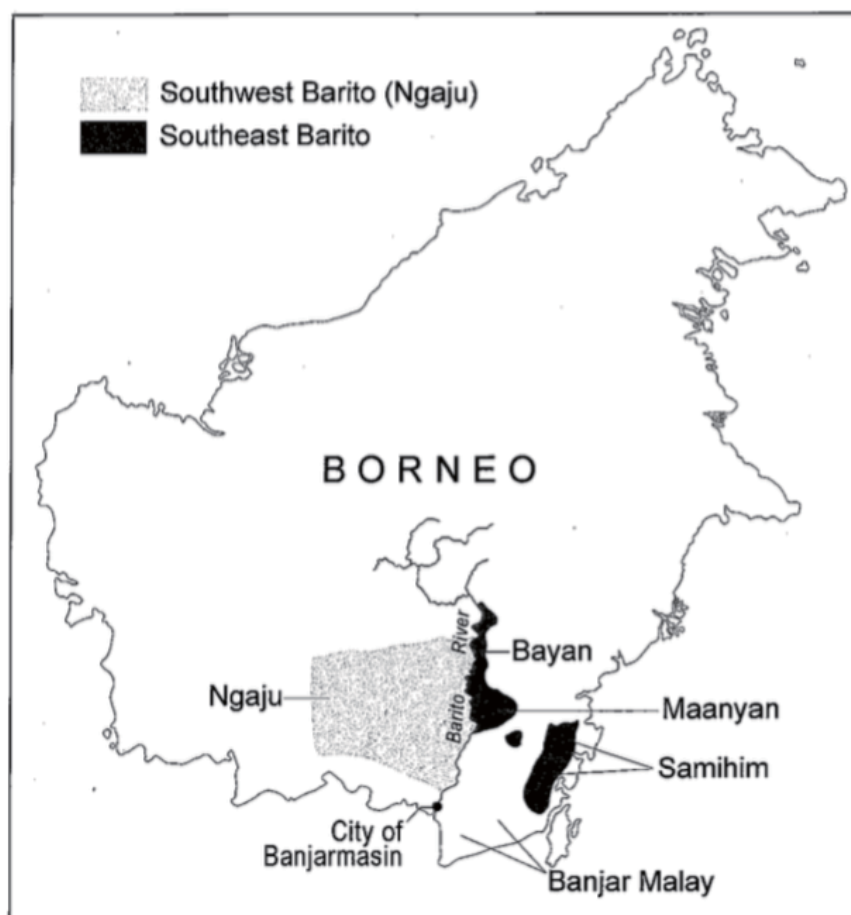
MAP1 Borneo Island showing areas where Maanyan and Samihim are
       spoken

Sabaki
languages
(Zones E, G,
incl. Swahili,
Comorian

COMOROS

M
A
D
A
G
A
S
C
A
R

Zone P languages
incl. Makhuwa
(N. Mozambique)

South Bantu languages
(Zone S), incl. Zulu,
Xhosa, Tswana, Ndebele
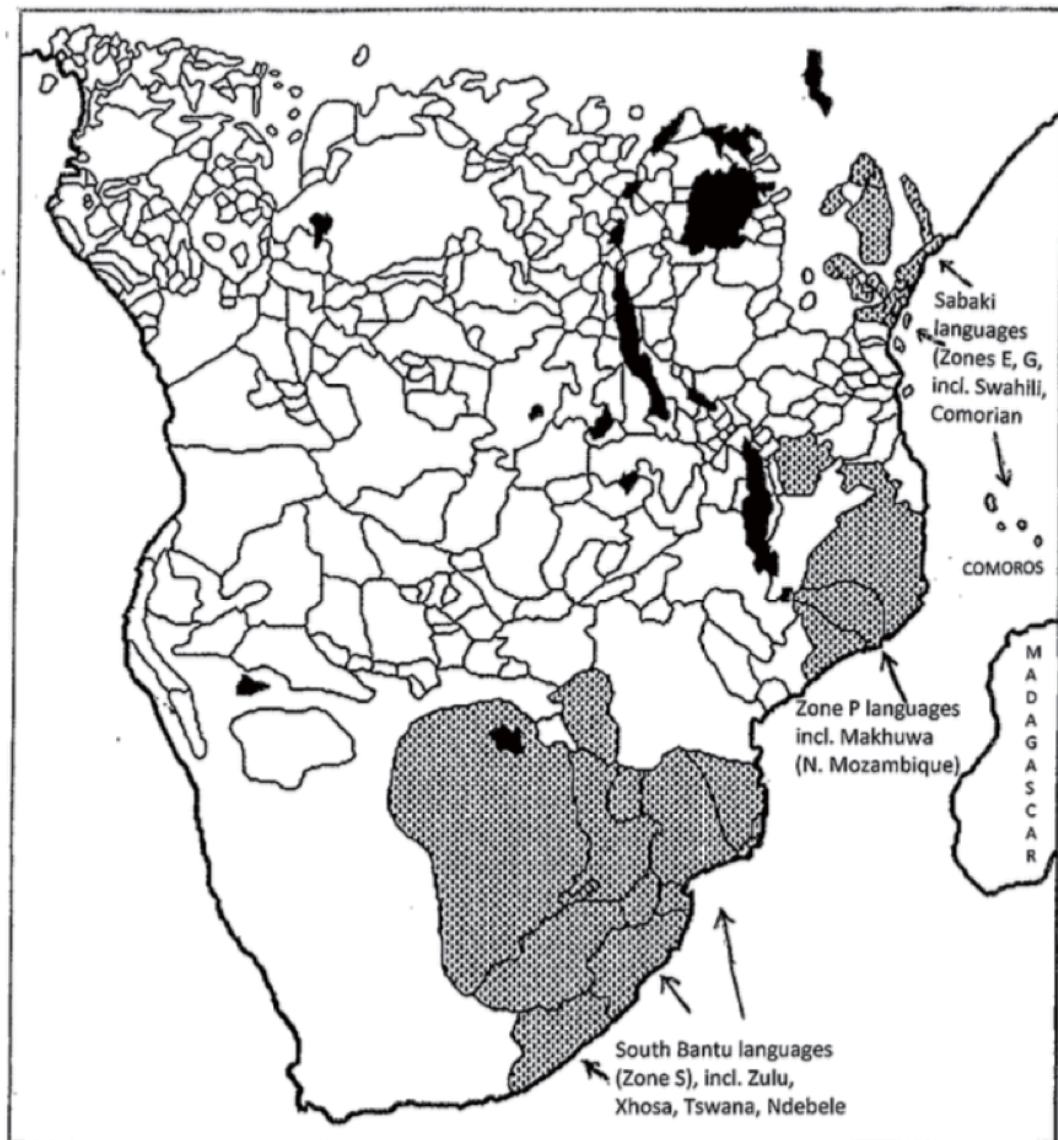
MAP 2 The distributeon of the locative suffix *-ini in sub-saharan Africa (Adapted from Schadeberg 2003)

2. The use of *-ni* in Southeast Barito languages

As indicated in the footnote 1, Proto South East Barito: *-n(i)ɛ became Maanyan and Samihim *-ni*. It also became *-ni* in Malagasy; however, since the southern and southwestern dialects of Malagasy[5] have a corresponding *-e*, it is very likely that Proto Malagasy still had a mid-vowel in this suffix. This is also in accordance with the phonological rule that southern and southwestern Malagasy dialects still distinguish between Proto Malayo-Polynesian *-a and *-ay (reflected in them as -e) and *-i (reflected as *-i), whereas standard Malagasy merged all these final vowels to -i.

Proto Malayo-Polynesian *-ni=a →
     Proto South East Barito: *-n(i)ɛ →
         Proto-Malagasy *-ne   →
           (Merina) Malagasy *-ni*, Sakalava Malagasy *-ne*

The roles performed by Maanyan, Samihim[6] and (Merina) Malagasy *-ni* are comparable to those of the Malay/Indonesian cognate *-ña*. Below I give examples of the use of *-ni* in Maanyan and Malagasy.

2.1 *-ni* in Maanyan (and Samihim)

In Maanyan, *-ni* has the following roles:

It is a possessive pronoun, e.g.

 (1) *tenga-ni* 'her/his/their body'
   uruwawa-ni 'his/her/their face'

It is an agent pronoun, as in *kuta-ni* 'eaten by him' in the following sentence:

(2) Dami	*mandru*	*palus*	*kuta-ni*	*re-era-e*	*dahulu*
after	cook	right-away	eat-3GEN	alone	all

puang	*ka-andrei*	*ineh-ni*	*teka*	*ume*
not	N.VOL-wait	mother-3GEN	come.from	field

'After cooking it he ate it all straight away without waiting for his Mum to come back from the field'

---

[5] These seem to form a first-order branch within the classification of Malagasy dialects (Adelaar 2013).

[6] Samihim is very similar to Maanyan. It is also very poorly documented, and the only Samihim prose available are actually liturgical texts prepared by Protestant missionaries around the turn of the 20th century. Therefore, apart from presenting the occasional example, I will give not discuss this speech form extensively.

It can be an adverbial suffix, e.g.

(3) *jujur* 'honest, straight'   *sabujur-ni* 'in fact, really'
    akhir 'end'   *akhir-ni* 'in the end, finally'
    ma 'LOC'; *andraw ka-telu* '3rd day' *ma andraw katelu-ni* 'on the 3rd day'
    hampe 'until'; *ka-pitu* '7th'   *hampe andrau kapitu-ni* 'up to 7 days;
                                            until day 7'

Compare also the following example, in which the possessive role of *-ni* is not particularly obvious:

(4) *Hu-an*   *hampe*   *lawit* *kia-ni*
    not.yet   until   far   road;walk-3GEN
    'he hadn't yet gone far [along the road], …', 'not far on the way, …'

Finally, in the following two instances, -ni is an object pronoun. (The number of instances with -ni assuming this role is small):

(5) *Bagamat*   *hi*   *Gayuhan*   *nyan-ri'et-ni*   *dan*
                                     N-saN-near-ni
    quickly   ART   G.   AV-APPL-near-3GEN   and (=Indonesian)

    nawut   *gayung*   *hang* *tandruk*   *kawawe*
    n-tawut
    AV-throw   bucket   LOC   horn   deer
    'Gayuhan quickly ran to the deer and threw a bucket over its horns'

In the following Samihim instances (taken from *Soerat Pangajaran* n.d.), the adverb ha-huang-ni has a locative preposition, whereas huang-ni, which is otherwise identical, lacks it:

(6) *Hi*   *Noah* *kaluar*   *hengka*   *banawa*   *andi*
    ART   Noah come-out   from   boat   and

    sagala   *eha*   *he*   *ha-huang-ni*
    all   animals REL   LOC-in(side)-3GEN
    'Noah came out of the boat with all the animals that were inside'

(7) (About the Garden of Eden):
    Hi   *Adam* *hire*   *rueh*   *Hawa* *na-huyu*   *muneng huang-ni*
    ART   Adam   they   two   Eve   UV-order, tell stay   in(side)-3GEN
    'Adam and Eve were told to live in it'

2.1 *-ni* in Malagasy

The Malagasy 3rd person genitive pronominal suffix is used in the following ways:

As a possessive pronoun :

(8) *tena* 'body'   *tena-ni* 'her/his/their/its body'
    satruka 'hairdo'   *satru-ni* 'his/her/their/its hairdo'
    lakana 'boat'   *laka-ni* 'his/her/their boat' (Dez 1982:38)

As an agent pronoun:

(9) Gaga   *ami-ni*   *n-a-tau-ni*
surprised   PREP-ART   PST-UV-do-3GEN
'I'm surprised by what he did' (Malzac 1960:108)

(10) N-an-datsah-a-ni   *irai-m-bilanja*   *ni*   *karama-ku*
PST-CIRC-fall-CIRC-3GEN   one-LNK-franc   ART   wages-1s.GEN
'He shortened my pay by one franc' (Malzac 1960:73)

In conjunction with prepositions:

(11) *ami-ni* 'with her/him, at his/her place, among them' etc. (*ami(n)-* is a multipurpose preposition).

In adverbs, especially ones indicating location in space and time. These adverbs usually also have one of the locative prefixes i- or a(N)- :

(12) *i-vela-ni* [LOC-leaving,abandoning+3GEN] 'outside, exterior, foreign'
    an-dani-ni [LOC-end-3GEN] 'on one side, on the one hand'
    an-ivu-ni [LOC-centre-3GEN] 'in the middle'
    ara-kevi-ni 'approximately' [following-thinking-3GEN]
    aman'-etsi-ni 'by the hundreds of thousands' [*aman'* 'with', *hetsi* 'hunderd thousand']
    *ambu-n*i 'above, on top' (there is no *ambu but compare *avu* 'elevated')
    *amba-n*i 'below, beneath' (no *amba, but compare *ava* 'downstream')
    fara-ni [posterity;final,last-3GEN] 'last; finally, in the end, at last'
    eu akaiki-ni 'next to' [*eu* 'here (in sight)', *akaiki* 'near']

Notice that the genitive (possessive) notion of *-ni* in locative adverbs is not very prominent, especially in the last four examples, which do not have a locative prefix (in *fara-ni* and *eu akaiki-ni*) or in which the locative prefix has become lexicalized (as in *ambu-ni* and *amba-ni*).

3. The use of -(i)ni in East and South Bantu languages

3.1 Swahili -ni

In Swahili, *-ni* indicates location in general. For this language to have a single locative marker which is suffixed is remarkable because Bantu languages generally agree in using three prefixes to indicate location, namely *pa-* (for definite location), *ku-* (for indefinite location), and *mu-* (for location inside).[7] (These forms and their associated notions still exist in Swahili locative adverbs,

---

[7] Referred to as Class 16, 17 and 18 prefixes respectively in Bantu linguistic literature.

compare *pale* 'there, in that place' and *kule* 'there, in that direction'). In Schadeberg's (2003:158) words, "its etymological history is a bit of a mystery". In contrast to *pa-*, *ku-* and *mu-*, *-ni* can be affixed to nouns of any class, although it cannot usually be tacked on to proper nouns and nouns referring to persons and animals. Here follow some examples of Swahili *-ni* taken from Wilson (1970:28):

(13)	*mfereji* 'ditch'	*mfereji-ni* 'in the ditch'
	mguu 'leg'	*mguu-ni* 'on the leg'
	mwisho 'end'	*mwisho-ni* 'at the end'
	mlima 'mountain'	*mlima-ni* 'on the mountain'
	soko 'market'	*soko-ni* 'on/at the market'
	nyumba 'house'	*nyumba-ni* 'in the house; at home'
	meza 'table'	*meza-ni* 'on the table'

In the following instance *-ni* derives temporal adverbs from nouns referring to time (Martin Walsh p.c.):

| (14) | *mwaka* 'year' | *mwaka-ni* 'next year' |

There are several adverbs formed with *-ni* for which the corresponding root is no longer extant as a free form in Standard Swahili (Martin Walsh p.c.):

(15)	*jio	*jioni* 'at evening'
	*nda	*ndani* 'inside' (but cf. *unda* 'womb')
	*chi	*chini* 'below' (cf. *nchi* 'country, land')

Toponyms normally do not receive *-ni*. For instance, in the contrastive sentences below, *-ni* is suffixed to the noun *mji* 'town' but not to the toponym *Nairobi* (Wilson 1970:28):

(16) *Mzee*   *yule*   *a-na-kaa*   *mji-ni*
      Old man   that   3s-PROG-reside town-LOC
      'That old man lives in the town'

(17) *Mzee*   *yule*   *a-na-kaa*   *Nairobi* (*Nairobi-ni)
      Old man   that   3s-PROG-reside Nairobi
      'That old man lives in Nairobi'

Nevertheless, as Samsom and Schadeberg (1994:135) point out, *-ni* does occur as a fossilized suffix in some place names. It still occurs in *Quelimani*, the name of a seaport in Mozambique, which stands for *kilima-ni* and literally means 'on/at the hill'. It also shows up in *Moroni*, the name of the capital of the Comorian Island republic, which is derived from Comorian *moro-ni* 'in the fire' (possibly in reference to the location of the city nearby a volcano). Other such examples are current toponyms like *Mbuyuni* (literally 'at the baobab tree') and *Maweni* ('at the rocks; a stony place') (Martin Walsh p.c.). Although in sentence (17) -ni cannot be combined with Nairobi, there is no such embargo on its application in sentence (18), in which Uzunguni 'the European quarter' is derived with -ni:

(18) *Mzee*   *yule*   *a-na-kaa*   *U-zungu*-ni
      Old man   that   3s-PROG-reside   ABSTRACT.NOM-European-LOC
      'That old man lives in the European quarter' (M. Walsh p.c.)

3.2 Reflexes of *-inı elsewhere

As already apparent in the Comorian *moro-ni* instance, Swahili is not the only Bantu language to exhibit this locative suffix or a form related to it. It is also found elsewhere in the Bantu language family, especially in members of the Sabaki group, which are spoken in Guthrie's Zones E and G and include Swahili and Comorian languages[8]. It is furthermore attested in some languages neighbouring the Sabaki group including Gikuyu (in Kenya). Finally, it features in Bantu languages further south: in northern Mozambique it occurs in languages in the P Zone, such as Makhuwa, once the major lingua franca among slaves in Madagascar. In southern Africa, it is present in languages in the S Zone, a rather large area including Botswana, Swaziland, Zimbabwe, Lesotho, southern Mozambique and the eastern half of South Africa. Major languages in this zone are (among others) Xhosa, Zulu, Ndebele, Tswana and Sotho (Gowlett 2003;609-611).

In many of the languages in which the suffix occurs it has an additional initial *i* and is disyllabic (e.g. Zulu *-ini*), one of the reasons why Samsom and Schadeberg (1994) derive it from an original *-inı. The Zulu reflex is applied in conjunction with the prefix *e-* ~ *o-*, which also marks location[9], and when a host noun is suffixed with *-ini*, the following morphophonemic changes take place:

(19) final -/*i* + -*ini* → -/*ini*
      final -/*e,a* + -*ini* → -/*eni*
      final -/*o* + -*ini* → -/*weni*

This is shown in the following examples:

(20) *sikebhe* 'ship'   *e-sikekbheni* 'on the ship'
      nyama 'meat'   *e-nyameni* 'in the meat'
      mfula 'river'   *e-mfuleni* 'from the river'
      nyoni 'bird'   *e-nyonini* 'to the bird'
      cingo 'telephone'   *o-cingweni* 'on the telephone'

Note that *-inı still occurs with one of the original prefixes (c.q. *ku-*) in the Sabaki languages Elwena and Lower Pokomo (Nurse and Hinnebusch 1993:364).

3.3 Previous explanations for the origin and spread of *-inı

---

[8] Sabaki languages form a well-established genetic subgroup within the coastal East Bantu subgroup (Nuse and Hinnebusch 1993). They have their traditional homeland mainly in Kenya and Tanzania in the so-called Bantu E and G Zones. However, their current spread is much wider, and Swahili is gaining considerable terrain in other eastern and central Africa states. The division into geographical zones referred to by a letter is a widespread Bantuist convention to locate Bantu languages introduced by Malcolm Guthrie (1948).

[9] This is the Class 25 prefix, which has a much more limited occurrence than *pa-, *mu- and *mu-. The distribution of the *-e and *-o* allophones is not made explicit in Spuy (1995): one could speculate that *o-* is triggered when the host has a final *o* (as in *cingo*).

How did locative *-inı spread among such a disparate selection of Bantu languages? It is noteworthy that neither all Sabaki languages nor all languages in Zone P or Zone S have a reflex of it. For instance, it is missing in Mwani (a Sabaki language) and Shona (a language in Zone S) (Schadeberg 2003:158). In other words, in so far as Sabaki languages and Zone S languages constitute genetic subgroups, neither of these seem to be defined as such by *-inı. According to Samsom and Schadeberg, this suggests that the current spread of *-inı reflexes must be due to diffusion through contact rather than to inheritance from a remote proto-stage. They believe that *-inı was originally borrowed into Swahili from a language outside the Bantu group and from there on spread to neighbouring Bantu languages (Samsom and Schadeberg 1994:136). In their estimate, this is more convincing than tracing it back to Proto Sabaki, or, as Meinhof (1941-42) did before, deriving it from the locative prefix *mu-. Nurse and Hinnebusch (1993:349, 364, 474) also generally believe in an areal spread of *-inı; however, in contrast to Samsom and Schadeberg (1994), they also consider the suffix critical for the definition of Sabaki as a genetic subgroup.[10]

As to the original source of *-inı, Samsom and Schadeberg (1994) support Sacleux, who argued that it is a grammaticalised form of the noun *inı which means 'liver', an organ traditionally considered to be the centre of senses and feelings in Bantu languages. This etymology is also endorsed by Güldemann (1990). However, Schadeberg (2003) no longer refers to this hypothesis.

Güldemann (1999) made the most comprehensive study on the development of nominal suffixes in Bantu languages so far. He tries to understand how some Bantu languages ended up with nominal suffixes while belonging to a language family that is so obviously head-initial and prefix-prone. In so doing he approaches the issue from a typological perspective, treating the development of *-inı together with that of a diminutive suffix (derived from *-yana 'child') and a gender/ associative suffix (derived from *-kadi 'wife, female'). The latter two occur exclusively in southern Bantu languages. Significantly, while he is able to explain convincingly that they are the result of contact with Khoisan languages, he is left with no real answer as to how *-inı could have become a suffix. There is no precedent for it in the Khoisan languages (which moreover would not have provided an explanation for its occurrence in Bantu languages outside Zone S). He also observes that as derivational suffixes, *-yana and *-kadi syntactically different from the locative suffix*-inı. Although "[t]here is no evidence that there has ever existed a productive head-final nominal syntagm in the Bantu family" (pp62-62), *-yana and *-kadi can still be interpreted as "having developed out of modifiers in associative or compound-like structures". This makes them more adaptable to the structure of Bantu, in which modifiers follow the nominal head. However, such an interpretation does not work for *-inı because its lexical source was a head. Apart from Khoisan languages, Güldemann also considered other non-Bantu languages adjacent to the Bantu region as a possible source for *-inı. He discounted Malagasy on account of the fact that in this language the locative prepositions (*i-* and *aN-*) are prefixed. An obvious conclusion, but it is not the only one possible, as will be demonstrated below.

---

[10] Nurse and Hinnebusch believe that Mwani has maintained some traces of the suffix and therefore must have had it at some earlier stage. It must have become obsolete under the influence of Makonde and/or Maviha, which are nearby Bantu languages in northern Mozambique (Nurse and Hinnebusch 1993:526).

4. Scenario for a borrowing from Malagasy

4.1 Paving the way: the presence of -ni in South East Barito locatives

The Maanyan, Samihim and Malagasy examples above show that *-ni* often occurs in locative adverbs. In some of these adverbs the original locative prefixes have become opaque through lexicalization, as in Malagasy *ambu-ni* 'above, on top' (< *aN-wau + *-n(i)e) or have been lost altogether, as in Samihim *huang-ni* 'inside', and Malagasy *fara-ni* 'last; finally'. It seems that in these languages the way has been prepared for a semantic shift from a genitive (possessive) pronoun to a locative postposition.

The shift itself admittedly never took place in South East Barito languages, where *-ni* has not ceased to be a genitive pronoun. This is not surprising given the high functional load of 3ʳᵈ person genitive pronouns in general and the fact that *-ni* in its capacity as a pronoun has never been threatened by lexical replacement in these languages.

On the other hand, such a shift may have become possible in the process of borrowing in a contact situation involving speakers of non-Austronesian languages, who would have perceived *-ni* as a locative suffix. Several of the Maanyan examples above allow a locative interpretation of *-ni*. Compare:

(21) *Panan tane ngaworiong-ni*
      lit. 'all land surrounding it' →   'all land in the surroundings'

      ruampulu taun lawah-ni
      lit. 'of twenty years length' →   'twenty years long', 'in twenty years'

      sabujur-ni
      lit. 'the truth of it'   →   'in truth'

A possible association of *-ni* with location is even more obvious in the Samihim examples involving *ha-huang-ni* and *huang-ni*. Both literally mean 'at-inside-of.it'; however, while the former still has a locative prefix, the latter has lost it.

The implied shift is also clear in the Malagasy locative adverbs shown above. Although in these adverbs location is primarily indicated with *aN-*, *i-* or Ø, they all have the suffix *-ni* in common, creating a situation in which the latter can easily be mistaken for a locative suffix. Compare:

(22) *i-vela-ni* (*lit. 'at-outside-of.it'*   →   *'outside, exterior, foreign'*
      an-dani-ni (lit. at-one.side-of.it'   →   'on one side, on the one hand'
      Ø-fara-ni (lit. 'last-of.it')   →   'at last, in the end, finally'

If *-ını was indeed borrowed from Malagasy, Samsom and Schadeberg's assumption that Swahili became the point of its dispersal into other Bantu languages makes good sense, as there is more evidence that this language was in contact with Malagasy in the past (see Section 4.4).

Furthermore, in the history of Swahili in particular, the borrowing of *-ni* as a locative suffix into an early form of Swahili (or a late form of Proto Sabaki?) might have facilitated its evolution into a lingua franca.[11] Being used by L2 speakers, linguae francae tend towards morphological simplification, and it is easy to see how in the case of Swahili L2 speakers might have preferred the use of a single suffix over maintaining the original Bantu prefixes *pa-*, *ku-* and *mu-*, the distribution of which is complementary, and therefore bound to rules.

4.2 A possible explanation for the evolution of *-inı

The explanations proposed in this section remain speculative at best.[12]

While Malagasy (Maanyan and Samihim) have -ni, it should be kept in mind that this suffix evolved from a Proto Malayo-Polynesian pronominal suffix *-ni=a. As can be expected, this suffix became reduced to *-ña in various Austronesian daughter languages (including Malay/Indonesian). In modern Malagasy dialects, *n and *ñ have merged to n. The same might have happened in Maanyan and other Proto South East Barito languages, although this is not obvious because ñ was often re-introduced in these languages through influence from Malay. (As noted above, the corresponding Proto South East Barito form must have been *-n(i)ε). Whether or not South East Barito languages developed a corresponding palatal ñ remains unclear. But whatever the case may have been, it does not exclude the possibility that it was still present in Proto South East Barito and in early Malagasy. An early Malagasy *niε, (*nyε) or *-ñε is totally conceivable. Nowadays most Malagasy dialects have -*ni*, but some phonologically conservative dialects in the south and west of Madagascar have retained a mid-vowel in this suffix.[13]

From that vantage point, early Swahili *-inı can be derived from early Malagasy *-niε, with subsequent vowel metathesis and vowel height leveling, as follows:

early Malagasy *-niε > early Swahili *-inε > Bantu *-inı

It could possibly even be derived from an early eastern Malagasy *-nii, provided that an eastern Malagasy variant had already began to diverge in the earliest stages of the language after its speakers had left Borneo island:

early eastern Malagasy *-nii > > early Swahili *-inı

In case early Malagasy had *-ñε instead, Sabaki*-inı can be explained as the result of vowel epenthesis with concomitant feature metathesis (in this case, metathesis of the feature +palatal) and vowel height leveling, causing the emergence of a high-front vowel *i:

early Malagasy *-ñε (> early Merina Malagasy *-ñi?)

---

[11] Swahili has been a trade language on the East African coast (and in islands off that coast) since the 9th century AD (Mufwene 2003:205).

[12] For tracing phonological changes in Bantu languages I am totally dependent on the expertise of Bantu comparative-historical linguists. Their input will be highly appreciated.

[13] See fn1. Note that the final high vowel in Merina Malagasy -*ni* is no evidence for the presence of a preceding palatal nasal, as word-final *ε and *i as a rule merged to -*i* in this dialect.

> early Swahili *-Vñı > *-inı

If vowel metathesis took place, this may have been motivated by stress which in Sabaki languages as a rule falls on the penultimate syllable: this pattern might have disallowed the development of a (*-niɛ >) *-nii suffix. Being disyllabic, the latter would have caused stress to shift across two syllables, and hence from the root to a suffix.

4.3 A possible explanation for the spread of *-inı in Zone S languages

Contact between Sabaki languages (originally in Kenya and Tanzania) and Bantu languages in southern Africa may seem far-fetched, given the distance involved. However, only about 1200 years ago these languages were spread very differently, and they may have been in contact.

On the one hand, Swahili settlements originally extended much further south along the Indian Ocean than they do today, as noted by Nurse and Hinnebusch (1993:527). They refer to Sinclair (1982), who found an archaeological site nearby Chibwene in southern Mozambique[14] exhibiting cultural traits very similar to those found in coastal Swahili sites. Sinclair estimates this site to be from the late 8th or early 9th century AD. Today different Bantu languages are spoken in this area.[15]

On the other hand, the migration of speakers of Bantu languages to southern Africa is a relatively late phenomenon, and their languages generally show great mutual similarity. For instance, the settlement of Zulus and Xhosas probably does not date further back in time than the 12th century AD. Equally important is that the migration route that they followed over the centuries first must have taken them from the Bantu homeland in Cameroon to eastern Africa before they finally moved downwards into southern Africa. Grollemund et al. (2015) argue for a route initially going in southeastern direction from the homeland in Cameroon to East Africa via a corridor of Savannah land (the so-called Sangha River Interval) through the forest occupying the Congo Basin; according to them this corridor opened up a number of times in the past. Blench (2012) also assumes a migration route to via East Africa, but he speculates that

---

[14] This is a coastal city halfway between the estuaries of the Zambezi and Limpopo Rivers.

[15] Swahili culture and society have a strong Islamic signature and is heavily overlaid with Arab and (later) Portuguese influence. The assumptions used to be (and have not entirely been put to rest) that Swahili culture is basically the result of contacts between Bantus and Arabs in coastal East Africa, and that the Swahili maritime disposition, which is a-typical of Bantu societies in general, was due to Arab influence. However, in the last three decades or so it has become obvious that these assumptions are not justified (compare Nurse and Spear [1985]). Arabs came relatively late to the Indian Ocean scene. Greek sources predating Islam mention the existence of a thriving culture along Africa's east coast all the way to Mozambique in the first half of the first millennium. Mahdi (1999:161-163) argues that Arabs were preceded by insular South East Asians. Arab authors from the 10th and 12th century A.D. refer to the presence of traders and pirates from (presumably) insular South East Asia and the lack of maritime skills among coastal East Africans. Mahdi 1999 calls attention to the use of double outriggers and the occurrence of oculi fore and aft on some East African boat constructions, particularly the Swahili *mtepe*. Blench (2012) believes that the Swahili acquired the skills for long-distance sailing from South East Asians, enabling them to expand in various directions, including southwards to Sofala and Chibwene in Mozambique, and eastwards to the Comoros in the Indian Ocean. While the initially much wider spread of Swahili societies along Africa's east Coast is now largely obliterated by later influences, it can still be traced through archaeological research (Blench 2012).

the route initially went due south along Africa's west coast; it then turned eastwards towards coastal east Africa (northern Mozambique), before reaching southern Africa. If southern Bantu speakers were still in coastal East Africa before the 12th century AD, they must have been in contact with speakers of Swahili[16], especially if at that time Swahili settlements stretched all the way to Chibwene, as suggested by Sinclair (see Map 3). Swahili had presumably already become a lingua franca in the 8th century AD Nurse and Hinnebusch (1993:2).

MAP 3 Spread of Swahili and South Bantu languages in the past

---

[16] Samsom and Schadeberg (1993:136) note the occurrence of Swahili loanwords in Zulu.

4.4 Other evidence for Austronesian influence

As far as Austronesian influence on Bantu languages is concerned, the borrowing of Malagasy *-ni* into Swahili does not stand in isolation, as there are many Malay and Malagasy loanwords that have been identified in Swahili. An extensive collection of Malay loanwords is Walsh[17]; other sources (for Malay and Malagasy loanwords) are Adelaar (2007) and Hoogervorst (2013). These loanwords clearly illustrate historical contacts between Sabaki speakers and speakers of Malay and Malagasy, which must have persisted over a long period of time. While part of the proposed loanwords cannot be dated or are demonstrably recent, other loanwords are definitely old. Among the latter we find the following ones:

From Malay:
Malay *tuba* 'fish poison', Swahili *u-tupa* 'id.' (Walsh)
Malay *kəladi* 'taro (colocasia)', Swahili *kiazi* 'tuber' (Walsh)
Malay *kota* 'fortified place, town', Swahili *u-kuta* 'stone walls of a house'
    (ultimately < Dravidian; Adelaar, Hoogervorst)

From Malagasy or Malay:
Malagasy *sàmbu* 'ship', Maanyan *sam^baw* 'celestial ship', Malay (7th century
    inscription) *sambaw*, 'ship', dialectal Swahili *sambo* 'id.' (Walsh);
Malagasy *farìhi* 'lake', Malay *pərigi* 'well, spring', and Swahili *m-fereji* 'ditch,
    furrow' (ultimately < Dravidian)
Malagasy *sùratră*, Malay *surat* 'writing; letter, document', Swahili *chora* 'to
    carve, draw'

From Malagasy:
Malagasy *èvukă* 'steam, gas, vapour', Swahili *m-vuke* 'vapour, steam'[18]
Malagasy *tùrakă* 'to launch (spear, etc.), throw', Swahili *tora* 'fishing speer'
Malagasy *vàri* 'rice (in general)', Swahili *wali* 'cooked rice' (ultimately <
    Dravidian)
Malagasy *vuài* (Malay *buaya*) 'crocodile', Swahili *m-buai* 'savage, rapacious'

That there existed a contact situation involving Bantu languages and Austronesian languages has become increasingly obvious with the awareness that the Southeast Barito speakers who settled Madagascar must first have sailed to mainland East Africa (and stayed there long enough to mix with local Bantu speakers) before they ended up on the island (Deschamps 1960); Adelaar 2007; Blench 2007, 2008). It has also become more obvious given the nature and extent of Bantu influence in Malagasy (Adelaar 2010; 2012). The awareness is based on the fact that both Bantu linguistic influence on Malagasy dialects and Bantu and Asian DNA in the Malagasy population turn out to be spread much more evenly over the entire island than hitherto assumed (Hurles et al. 2005; Adelaar 2013). Such a linguistic unity and even spread of Asian and Bantu DNA strands suggests the formation of a mixed Austronesian-Bantu population prior

---

[17] Walsh (unpublished); this is the source of items listed in Blench 2012, Table 3).
[18] The retention of *-k(e)* in *m-vuke* is unexpected and needs further explanation.

to the settlement of Madagascar (and hence most likely in coastal East Africa), rather than in Madagascar itself.

5. Concluding remarks

Deriving the eastern and southern Bantu reflexes of the locative suffix *-ɪnɪ via Swahili from early Malagasy *-n(i)ɛ or *-ñɛ provides an attractive explanation.

The occurrence of *-ɪnɪ remains an unresolved problem in Bantu linguistic literature. As a suffix, it is unusual in the Bantu family as a whole, in which the locative is typically expressed by a set of prefixes. It is also odd from a morphotactic viewpoint. If it was initially a noun, as is often assumed in Bantu linguistics, it would have been the nominal head in a genitive construction. However, for a head to come at the end of a noun phrase is rather counter-intuitive in Bantu languages, which are overwhelmingly head-initial.

Although Malagasy lacks a locative suffix as such, it does have a 3ʳᵈ person genitive suffix -ni (< early Malagasy *-n(i)ɛ, *-ñɛ) which matches *-ɪnɪ to a significant extent (synchronically as well as diachronically). The extended use of this suffix in Malagasy provides a clear context in which it could have been re-interpreted as a locative suffix by L2 speakers in a language contact situation.

As to the historical conditions for an encounter between Austronesian and Bantu speakers on the East African mainland, the evidence of Malay and Malagasy loanwords in Swahili leaves no doubt that contact took place. Furthermore, recent multidisciplinary research in East Africa and Madagascar has provided an increasingly clear context for such contact. It is essential to acknowledge this in both African and South East Asian studies.

Explaining *-ɪnɪ as an ultimately non-Bantu borrowing might help to solve some crucial problems in the classification of eastern and southern Bantu languages. Provided that this suffix was originally a loan morpheme and that Nurse and Hinnebusch (1993) are correct in their assessment that Mwani once did share it, it might turn out to be critical for the establishment of Sabaki as a lower-order genetic subgroup. Conversely, its elimination from the inventory of inherited morphemes in the case of P languages and S languages might help to establish affiliations among these languages that are genuinely genetic.

Furthermore, tracing Bantu *-ɪnɪ to early Malagasy *-n(i)ɛ or *-ñɛ does in fact not disqualify Sacleux's proposition that it derived from *ɪnɪ ('liver') in its notional meaning as the centre of one's senses and feelings. The acceptance of early Malagasy *-n(i)ɛ or *-ñɛ into Swahili was conceivably facilitated by the similarity in form and meaning that exists between this suffix and Sabaki *ɪnɪ.[19]

---

[19] Incidentally, in Malagasy the liver is also the centre of the senses and feelings, and its signifier is also used in a locative adverbial construction: *àti* 'liver' in combination with the locative preposition *aN-* yields *an-àti* 'inside'.

REFERENCE LIST

Adelaar, Alexander. 2013. 'Malagasy dialect divisions: genetic versus emblematic criteria', *Oceanic Linguistics* 52/2:457-480.

Adelaar, Alexander. 2012. 'Malagasy phonology history and Bantu influence', *Oceanic Linguistics*   51/1:124-160.

Adelaar, Alexander. 2010. 'The amalgamation of Malagasy', in: *A Journey through Austronesian and Papuan Linguistic and Cultural Space: Papers in Honour of Andrew K. Pawley* (John Bowden, Nikolaus P. Himmelmann, Malcolm D. Ross eds). 161-178. Canberra: Pacific Linguistics.

Adelaar, Alexander. 2007. 'Language contact in the Austronesian Far West', Paper: 3rd Conference on Austronesian Languages and Linguistics, London: School of Asian and African Studies.

Adelaar, Alexander; Kikusawa, Ritsuko. 2014. 'Proto-Malagasy personal pronouns: a reconstruction', *Oceanic Linguistics* 53/2:480-516.

Blench, Roger. 2012. 'Two Vanished African Maritime Traditions and a Parallel from South America', *African Archaeology Review* 29(2-3), 273-292.

Blench, Roger. 2008. 'The Austronesians in Madagascar and their interaction with the Bantu of the East African coast: surveying the linguistic evidence for domestic and translocated animals', *Studies in Philippine Languages and Cultures* 18:18-43.

Blench, Roger. 2007. 'New palaeozoogeographical evidence for the settlement of Madagascar', *Azania* 42:69-82.

Blust, Robert. 2013. *The Austronesian languages*. (2nd and revised edition). Asia Pacific Linguistics 008. Canberra: Australian National University, College of Asia and the Pacific.

Deschamps, Hubert. 1960. *Histoire de Madagascar.* Paris: Éditions Berger-Levrault.

Dez, Jacques. 1980. *Structures de la langue malgache.* Paris: Publications Orientalistes de France (POF).

Gowlett, Derek. 2003. 'Zone S', in (Derek Nurse and Gérard Philippson eds), 609-638.

Grollemund, Rebecca; Simon Branford; Koen Bostoen; Andrew Meade; Chris Venditti; Mark Pagel. 2015. 'Bantu expansion shows that habitat alters the route and pace of human dispersals', Proceedings of the National Academy of Sciences, www.pnas.org/cgi/doi/10.1073/pnas.1503793112.

Güldemann, Tom. 1999. 'Head-initial meets head-final: nominal suffixes in eastern and southern Bantu from a historical perspective', *Studies in African Linguistics* 28/1:49-91.

Guthrie, Malcolm. 1948. *The classification of the Bantu languages.* London: International African Institute.

Hoogervorst, Tom. 2013. *Southeast Asia in the Ancient Indian Ocean World.* Oxford: Archeaopress Publishers of British Archaeological Reports (BAR).

Hurles, Matthew E., Bryan C. Sykes, Mark A. Jobling, and Peter Forster. 2005. The dual origin of the Malagasy in island Southeast Asia and East Africa: evidence from maternal and paternal lineages, *American Journal of Human Genetics*.

Mahdi, Waruno. 1999. 'The dispersal of Austronesian boat forms in the Indian Ocean', in (Roger Blench and Matthew Spriggs eds): *Archaeology and language III: Artefacts, languages and texts.* 144-179. London and New York: Routledge.

Malzac, (R.P.) Victorin. 1960. *Grammaire malgache* (4th edition). Paris: Société d'Éditions Géographiques, Maritimes et Coloniales.

Meinhof, Carl. 1941-1942. 'Entstehung und Gebrauch der Lokativendung in Bantusprachen', *Zeitschrift für Eingeborenen-Sprachen* 32:161-164.

Mufwene, Salikoko S. 2003. 'Contact languages in the Bantu area', in (Derek Nurse and Gérard Philippson eds): 195-208.

Nurse, Derek; Thomas J. Hinnebusch (eds). 1993. *Swahili and Sabaki: a linguistic history.* Berkeley, Los Angeles and London: University of California Press.

Nurse, Derek; Gérard Philippson (eds): *The Bantu languages.* Routledge Family Series. London: Routledge.

Nurse, Derek; Thomas Spear. 1985. *The Swahili: reconstructing the history and language of anAfrican society, 800-1500.* Philadelphia: University of Pennsylvania Press.

Sacleux, Ch. 1939. Dictionnaire Swahili – français. Paris: Institut d'Ethnologie.

Samsom, Ridder; Thilo C. Schadeberg. 1994. 'Kiinamacho cha mahali: kaimbishi tamati cha mahali ni' (on the locative suffix -*ni*), *Afrikanistische Arbeitspapiere* (Cologne) 37:127-138
http://www.qucosa.de/fileadmin/data/qucosa/documents/9802/1_07_schadeberg.pdf

Schadeberg, Thilo C. 2003. 'Historical linguistics', in (Derek Nurse and Gérard Philippson eds), 143 -163.

Sinclair, Paul. 1982. 'Chibuene: an early trading site in southern Mozambique', in: (J. de Vere Allen and T. H. Wilson eds): *From Zinj to Zanzibar.* (Paideuma: Mitteilungen zur Kulturkunde 28). Wiesbaden: Franz Steiner Verlag.

Soerat pangadjaran agama olon kristen huang bahasa Dajak Samihim. n.d. Banjarmasin (Indonesia): Bandjermasinsche drukkerij. (Copy of this text at Leiden University Library).

Spuy, Andrew van der. 1995. 'Morphological case-marking in Zulu', in (Anthony Traill; Rainer Vossen; Megan Biesele eds): *Papers in memory of Patrick J. Dickens.* 297-317. Cologne: Rüdiger Köppe Verlag.

Walsh, Martin. (Unpublished) List of Malay loanwords in Swahili.

Wilson, Peter M. 1970. *Simplified Swahili.* London: Longman.

Definiteness and specificity in Abui

František Kratochvíl and Benidiktus Delpada[1]

1. Introduction

This paper discusses referential properties of articles in Abui. Abui deictic system (from which the articles were grammaticalised) makes a three-way contrast and alternates the viewpoint between the speaker and the addressee/hearer.[2] Discussions of definiteness revolve around the role of familiarity, and its status as a defining feature for definiteness and its relationship to uniqueness or identifiability (Gundel et al. 1993; Lyons 1999; Roehrs 2009, and many others).

Abui is a language with a deictic contrast in the definite articles (cf. Lyons 1999:55-56). The deictic contrast indicates whether the referent is discourse-immediate or not. Additional pragmatic functions, not unlike those in Bella Coola (Newman 1968; Davis and Saunders 1975) are also available, mapping the location of the referent in other dimensions, primarily in time and stance.

Systems with hearer-oriented articles are cross-linguistically rare but indicate that definiteness can be combined with other categories (Dryer 2014:241). Abui offers some insights about how the category of 'familiarity' may be structured. While

---

Gundel et al. (1993) rank familiarity above the plain definiteness in their Givenness Hierarchy, in Abui, familiarity seems to run in parallel to definiteness and possibly extends to specificity. Another noteworthy property of the Abui system is the three-way distinction in indefinites, where specificity is marked and allows the viewpoint alternation as well.

Before presenting the Abui data, we first address the terminology here, in light of the recent discussion about the distinction between articles and demonstratives (Davis et al. 2014; Dryer 2014). We then present the most important points about definiteness and specificity debated in the literature. Section 1.2 places the Abui system in the context of the Timor-Alor-Pantar family, showing that complex systems are the norm for this family. Section 2 describes the referential properties of Abui NPs in various constructions discussed in connection to definiteness in the literature. We base our discussion on naturally occurring sentences, but systematically manipulate the articles to obtain full paradigms and to show the basic contrasts. The behavior of Abui articles in natural discourse is not discussed here, but has been discussed briefly in Kratochvíl (2015) and will be elaborated on in a separate paper.

1.1. Terminology and tests

The line between demonstratives and articles is blurred and subject of vivid discussion lately (Dryer 2013a-b, 2014; Davis et al. 2014). Taking these two positions as representative of the range of views, Table 1 highlights the overlaps and differences. It should be noted that the defining criteria for articles in Dryer's (2013a-b, 2014) are looser than those listed in Davis et al. (2014).

FEATURE	Dryer 2013a-b, 2014	Davis et al. 2014
MORPHOLOGY	free or bound form	bound form
SYNTAX	obligatory or optional	obligatory
FAMILIARITY	familiar	familiar
UNIQUENESS	unique	unique
DEMONSTRATIVES	included	excluded
CONTRASTS	anaphoric vs. shared knowledge	n.a.
SEMANTIC TESTS	n.a.	maximality (plurals and mass)

Table 1. Defining criteria for definite articles

If the definition by Davis et al. (2014) were followed here, the Abui forms could not be considered articles, because they are not obligatory. It will be shown below, that the use of Abui articles interacts with case marking and agreement, both of which are sensitive to referential properties of the marked arguments.

The broader definition used by Dryer (2013a, b; 2014) accommodates for the situation in many languages where the spatial-deictic forms (accompanied by a pointing gesture) are also used with nominals in contexts where English would use a definite article, rather than a demonstrative (Dryer 2013a). The definition of indefinite articles in Dryer (2013b) is also more flexible; it includes also the numeral 'one' marking NPs to signal indefiniteness (referent not known to the hearer).

Although Dryer's framework does not explicitly list semantic tests, the typology of articles is constructed with reference to both individual languages as well as literature dealing with definiteness and specificity. We therefore examine the behavior of the Abui articles using the known tests (Partee 1972; Enç 1991; Gundel et al. 1993; Matthewson 1998; Lyons 1999; Von Heusinger 2002; Abbot 2004; Levinson 2004; Roehrs 2009, and others).

Definiteness is traditionally defined with reference to *uniqueness* and *familiarity* (Lyons 1999; Abbot 2004, and others). Noun phrases marked as definite (in English with *the*) denote entities known to both speaker and hearer (Lyons 1999:3). In earlier accounts of definiteness, such entities would be described as familiar, but the current consensus seems to be to anchor definiteness in a more general notion of *identifiability* (Lyons 1999:5). Definite noun phrases denote entities, which the hearer

can unambiguously identify. Since some definite noun phrases can refer to entities that are not identifiable, the notion of uniqueness has been introduced.

The notion of uniqueness has seen a similar revision. Initially, uniqueness was understood as the existence of one and only one entity matching a definite description (Abbot 2004:125). However, since Hawkins (1978), *uniqueness* has been replaced with *inclusiveness* – reference to the totality of entities or matter to which the descriptive content of the NP applies (Lyons 1999:11; Abbot 2004:126).

Certain syntactic environments have been shown to be sensitive to definiteness or specificity. We will discuss some of those environments here and some when presenting the relevant Abui data. Lyons (1999:16-17) discusses possessives, partitives, superlatives and existentials; in these constructions definite NPs pattern in the same way as proper names, possessed NPs, pronouns and NPs with universal quantifiers.

Definiteness is contrasted with indefiniteness and in some accounts also with specificity (Enç 1991, von Heusinger 2002). There are different views of the categorical status of *specificity*. Abbot (2004:144) considers specificity to be merely a pragmatic effect, while the underlying distinction is one of definiteness and indefiniteness. Specific description is defined as one where the speaker has a particular individual in mind (Abbott 2004:145). The classical example has two readings (from Partee 1972 via Abbott 2004:146).

(1)   John would like to marry a girl his parents don't approve of.

In the wide-scope reading, John's parents happen to dislike his girl. In the narrow-scope reading, John has picked a girl to offend his parents. The first reading has a more specific reference than the second. The ambiguity remains also when the tense and modality is manipulated:

(2)   John succeeded in marrying a girl his parents don't approve of.

(3)   John married a girl his parents don't approve of.

We will discuss the Abui equivalents of the above constructions in section 2.7. Von Heusinger (2002:246) summarizes informal characterizations of specificity in the

literature. The paper highlights speaker's certainty about the identity of the referent and the differences in identifiability between definite and specific descriptions (p. 249). Von Heusinger (2002:268) proposes that specific descriptions are referentially anchored to another object in the discourse. Further, specific NPs are not dependent on the matrix predicate and operators such as modal verbs and combine with *a certain* (Von Heusinger 2002:272). Lyons (1999:59) converges with the above descriptions and observes that marking of specificity is fairly widespread.

1.2.   Demonstratives and articles in the TAP languages

In the Timor-Alor-Pantar family (TAP), the differences in ordering of NP constituents are minimal (Schapper 2014a:14). Uniformly, the NP final slot is reserved for referent tracking purposes by words labeled as demonstratives or articles. A prenominal deictic slot seems to be unique to Abui, and is therefore possibly a recent innovation.

(1) proto-AP NP Template: Gen **N** Attr Num/Quant **Dem**

A number of features can be encoded by the Alor-Pantar demonstratives. In terms of Diessel's typology of demonstratives, features associated with deixis are the norm, but features of quality (ontology, animacy, humanness, sex, number, etc.) are absent. Visibility, knowledge, and viewpoint features embed perspectives of the speaker and hearer (Holton 2014:58; Schapper 2014b:310; and Kratochvíl 2014a:379).

language	DISTANCE	ELEVATION	VISIBILITY	KNOWLEDGE	VIEWPOINT	source
Western Pantar	+ (3-way)	+	+	-	-	Holton 2014:57-59
Teiwa	+ (2-way)	-	-	-	-	Klamer 2010:130-138
Kaera	+ (2-way)	?	?	?	?	Klamer 2014:117
Blagar	+ (2-way)	+	+	-	+	Steinhauer 2014:181
Adang	+ (2-way)	+	-	-	-	Robinson and Haan 2014:256-257
Abui	+ (3-way)	+	-	-	+	Kratochvíl 2011
Klon	+ (2-way)	+	?	?	?	Baird 2008:58-61
Kamang	+ (2-way)	-	-	+	-	Schapper 2014b:310
Sawila	+ (2-way)	-	+	-	-	Kratochvíl 2014a:376-377
Wersing	+ (2-way)	-	-	-	-	Schapper and Hendery 2014:469-470

Table 2. Features encoded by deictic words in Alor-Pantar languages

Western Pantar, Kamang, and Wersing have developed articles, transparently derived from demonstratives, typically tracking the definite-indefinite distinction. Definite and specific indefinite articles are known to originate in demonstratives (Diessel 1999:128; Heine and Kuteva 2002:109-111).[3] During the grammaticalisation, the generalization of meaning from spatial to discourse deixis often goes together with phonological reduction of the source form. The linear order however is universally preserved (Moravscik 2011:81-82). Because of this diachronic change pattern I conclude that the Abui prenominal spatial demonstratives are an innovation.

2. Referential properties of Abui Noun Phrases

In the context of complex systems of the Timor-Alor-Pantar family outlined above, the Abui system appears of to be of moderate complexity. This section outlines the referential properties of Abui Noun Phrases and marking of classic referential distinctions discussed in the literature, including Givenness Hierarchy (Gundel et al. 1993:275), possessed NPs, partitives, and combinations with Abui case markers (agentive *di*) and verbal agreement (head-marking).

Abui NP contains two slots that can be filled with deictic words, as shown in (4). Spatial-deictic forms occur in the prenominal slot (DEICT). Forms in the post-nominal slot point in discourse and interact with hearer's knowledge. A single NP may contain both DEICT and ART.[4]

(4)   Abui NP Template: **DEICT** GEN N   ATTR   NUM/QUANT **ART**

---

[3] The rise of articles in some languages coincided with the disappearance of the case marking, but no direct causal relation between the two seems to exist (Lyons 1999:324-325). In AP languages, case generally not marked on demonstratives, so this is not relevant.

[4] The double-marked form is somewhat similar with the English 'this X here', except that the Abui forms in both slots must come from the paradigms given in Tables 3 and 4 respectively, while English uses the adverbial form here for the spatial deictic function. Abui possesses a separate paradigm of adverbial demonstrative forms (*ma* 'be.PROX', *ta* 'be. PROX.AD', *la* 'be.MD', *fa* 'be.MD.AD', and *ya* 'be.DST') which cannot occur in the DEICT slot of the NP.

A three-way distinction is made (proximal, medial, distal) with an additional dimension of viewpoint alternating between speaker and addressee (Kratochvíl 2007:162-163; Kratochvíl 2011:763-766; Dryer 2014:241). Elevation applies only to medial and distal forms.

2.1. Spatial deixis

The prenominal slot encodes spatial location of referents (Kratochvíl 2007:163). The inventory of deictic words that can occur in the prenominal position can be seen in Table 3. These forms combine with pointing gestures accomplished with fingers, hands, lips, eyebrows or chin (Levinson 1983:65-68; 2004:108). The shaded cells contain forms, which also occur in the phrase-final slot and will be discussed in 2.2.[5]

	VIEWPOINT (V)		ELEVATION	
DISTANCE	SPEAKER	ADDRESSEE	LOW	HIGH
PROXIMAL	*do* (PRX)	*to* (PRX.AD)	*	*
MEDIAL	*o, lo* (MD)	*yo* (MD.AD)	*ò* (MD.L)	*ó* (MD.H)
DISTAL	*oro* (DST)		*wò* (DST.L)	*wó* (DST.H)

Table 3: Abui prenominal deictic words

2.2. Articles

Abui articles are NP-final enclitics (in the current orthography, only the medial *o* is written together with the noun), which encode the referential status of the NP as either definite, specific, or indefinite and non-specific.[6] The articles are listed in Table 4.

---

[5] Although split-viewpoint systems are cross-linguistically rare, they are an alternative to systems conceptualizing space from egocentric perspective or from fixed coordinates of the environment (cf. Diessel 2014). The egocentric frame is multiplied to include the frame of the hearer, and presupposes therefore social cognition. The hearer frame is structured in the same way as the egocentric frame and could be therefore thought of as secondary (Diessel 2014:129-130). It will be interesting to study how the hearer frame is acquired by Abui children and affected in the ongoing language shift.

[6] Articles may undergo lengthening when the NP expresses a topical or agentive argument. The definite articles are grammaticalised from demonstratives; the specific articles from equative/similative demonstratives similar to the English *such* or the Slavic *kak ~ jak* (Arsenijević et al. 2014; Anderson and Morzycki 2015).

The definite articles preserve the deictic contrast between proximal and medial forms inherited from the demonstrative forms from which they were grammaticalized. Such systems, although not very common, are known in the definiteness literature. Lyons (1999:55-56) notes that distinctions of distance from the speaker and association with different person are occasionally found in simple definites, indicating the existence of [±PROX] feature independently of [±DEF]. Abui is a language where the same deictic features appear both on demonstratives and on the definite articles (grammaticalised from them).[7]

The addressee-based forms are part of the paradigm, but they no longer indicate that the deictic center is based in the addressee but rather familiarity, empathy, or noteworthiness, which will be discussed in more detail later. The indefinite cardinal article *nuku* originating in the numeral 'one' is also listed in the paradigm.[8] In addition, bare nouns can also have indefinite, non-specific reference. The definite and specific articles have undergone grammaticalisation into markers of relative and absolute tense, evidentiality and stance and became associated with the clause-final and sentence-final position when marking those categories (Kratochvíl 2011).[9]

IDENTIFIABLE FOR:	SPEAKER	ADDRESSEE	NEITHER
DEFINITE	*do* (PROX)	*to* (PROX.AD)	
	o (MD)	*yo* (MD.AD)	
SPECIFIC	*nu* (SPC)	*hu* (SPC.AD)	
INDEFINITE			*nuku*

Table 4: Abui articles

[7] Newman (1969:304-305) shows that the feature [±VISIBLE] is maintained in both the article and demonstrative paradigms (weak and strong) of Bella Coola (also known as Nuxalk, Salishan, British Columbia, Canada). Davis and Saunders (1975:850-851) show that in addition to visibility, the Bella Coola contrast may encode deictic time.

[8] The numeral *nuku* 'one' is a regular reflex of the Proto-Alor-Pantar form *nuk* 'one' (Holton and Robinson 2014:75). The specific articles are related to the equative and simulative demonstrative paradigm (ESD) *n-* 'like.PROX' ~ *w-* 'like.MD' ~ *h-* 'like.DST'. The similarity between nuku and nu seems coincidental.

[9] Lyons (1999:60-62) discusses other languages that use definite articles to nominalize and to mark subordinated clauses. The grammaticalisation path from *determiners* to *tense markers* and higher categories is common and well attested (see for example Yap et al. 2011 and papers therein).

Demonstratives are considered definite, because their referent is identifiable (Lyons 1999:21). Their reference is not inclusive; it involves a contrast, clear or implied, between the intended referent and the potential set (Lyons 1999:17). The Abui definite articles mark not only that their referent is accessible to the hearer in the immediate or non-immediate context, but also whether or not the referent is familiar. Lyons (1999:21) points out that interpreting demonstratives is easier than inferencing simple definites because the speaker does the referent-identification for the hearer. Familiar articles in Abui highlight that the speaker may choose to do the referent-identification work from hearer's perspective. Familiarity, given a prominent role in the early accounts of definiteness, is given a systematic treatment in Abui.

In principle, we understand the notion of familiarity in terms of Gundel et al. (1993:278), as a special cognitive status, such that the hearer already has a representation in memory: in long-term memory if it has not been recently mentioned or perceived, or in short memory if it has. Referents marked as familiar may be previously mentioned by the hearer, or somehow associated with hearer's perspective. The interpretation is context-dependent and also interacts with the referential type (proper names vs. common nouns). The speaker may choose a familiar article to draw hearer's attention to a referent so that the common ground is updated with the presented information, or simply as an invitation to fill out the speaker's meaning (as the English *you know*).[10] It should be noted that familiarity is not treated as a scalar category in Abui, but the proximal and medial forms map the difference between immediate and non-immediate (usually past) context. The range of discourse uses of Abui familiar articles will be discussed in a separate publication.

---

[10] Stubbe and Holmes (1995: 69) define the colloquial New Zealand English as a pragmatic device with two broad interactive meanings, quite similar to the Abui uses, namely: (i) a marker expressing speaker's confidence that the hearer shares the relevant knowledge, or reassuring the hearer of the validity of the proposition, and (ii) a marker of uncertainty about hearer's attitude or the linguistic precision of the description. Macaulay (2002) in a follow-up study of Scottish English highlights the fact that the use of you know may be quite idiosyncratic, and perhaps attributed to a personal speech style and the rhythmic organization of utterances (p. 765). The same study also shows that the use of you know does not appear to be primarily based on the assumption of shared knowledge (ibid.).

2.3. Referential properties of bare nouns, case marking and agreement

Referential properties of arguments and information flow are known to interact with their grammatical expression and trigger alternations in voice, case, and agreement marking (Hopper and Thompson 1980:253).[11] In Abui, only the specific and definite undergoer arguments (i.e. identifiable at least for the speaker) are indexed on the verb. Because the person indexing is obligatorily, the specific or definite reference of a noun can be inferred from the marking on the verb, as shown in (5). On the other hand, bare nouns without indexing have generic reference.

(5)   Verbal agreement and definiteness

a.   *maama* ***bataa***   *faaqda*

        father   [wood   chop.IPFV]

        'father chops wood OR father is chopping wood'

b.   *maama*   ***bataa***   ***he-****faaqda*

        father   [wood   3.LOC-chop.IPFV]

        'father is chopping **the** wood, (a certain/known quantity of wood)'

c.   ***kaai*** *diking*   *pee=ng*   *mareei=ba*   ***arui***   *kafia*

        dog   fire.place   near=SEE   go.up.PFV=SIM   [ashes   scrape.IPFV]

        'the dog went up to the fireplace to scrape in ashes'

d.   ***kaai*** *diking*   *pee=ng*   *mareei=ba*   ***arui***   *he-kafia*

        dog   fire.place   near=SEE   go.up.PFV=SIM   [ashes   3.LOC-scrape.IPFV]

        'the dog went up to the fireplace to scrape the ashes'

Undergoer arguments, which are possessed or contain a relative clause are interpreted as definite and have to be indexed on the verb, as in (6). Note than in (b), the structure is interpreted as a topic + comment sequence describing the settings of the prohibition expressed in the second clause. The leg is no longer an argument of the verb *kafia* 'scratch', although its reference remains definite.

---

[11] Hopper and Thompson subsume definiteness and referentiality under their super-type of *Individuation* which also includes other contrasts such as proper vs. common nouns, animate vs. inanimate, concrete vs. abstract, singular vs. plural, and count vs. mass (1980:253).

(6)   Verbal agreement and N + Rel

a.   ***e-toku   ba yokung   nu   he-****kafia   naha!*

[2SG.AL-leg   REL be.inflamed SPC] 3.LOC-scratch.IPFV not

'don't scratch your inflamed leg!'

b.   *e-toku   ba yokung   nu   kafia   naha!*

2SG.AL-leg TOP be.inflamed SPC_{RT}   scratch.IPFV not

'your leg is inflamed (so) don't scratch (anywhere)!'

The same holds for the agentive arguments. Only specific arguments are marked with the agentive *di* (3AGT).[12] The adnominal pronoun functions as a personal determiner (Lyons 1999:141-145) and is interpreted as definite, being readily identifiable to the hearer.[13] The unmarked NP *moku loku* 'children' in (a) has a generic reading.

(7)   Distribution of the agentive pronoun *di* (N *di*)

a.   ***moku   loku*** *kuul   sakola   he-sei*

[kid   PL]   must   school   3.LOC-come.down.IPFV

'children must attend school'   B3.3.2

b.   ***moku   loku   di***   *kuul   sakola   he-sei*

[kid   PL   3AGT] must   school   3.LOC-come.down.IPFV

'the children must attend school'   B3.3.2

The primary function of the pronoun *di* (3AGT) is to mark arguments characterized by volition and control, almost always animate. When these two conditions are not met,

---

[12] The third person pronoun *di* is innovated in Abui. The reconstructed pAP form is **ga* (Holton et al. 2012:115), whose regular reflex in Abui should be **ha. It is possible that *di* (and *da* in some dialects) originates in the proximal demonstrative *do*.

[13] Third person adnominal pronouns *dia* (SG) and *dorang* (PL) are also found in local Alor Malay, calquing the structures of the local Papuan languages. Similar uses, although including also pronouns of first and second person are found also in Papuan Malay (Kluge 2014:333, 383).

the third person pronoun *hedo*, belonging to a 'plain-case' paradigm, can be used instead.[14] Further, the pronoun *di* is not compatible with generic reference, as in (8).

(8)   Generic reference and the agentive pronoun *di* (N *di*)

a.   *war-tama maiye,* **kumal   tafuda**   *mong-e*

dry.season when   [mosquito be.all]   die.IPFV-PROG

'when it is dry season, mosquitoes will be dying'   B7.61.2

b.   *tuntama do,*   **kumal   di**   *ne-l=takei*

night   PROX   [mosquito 3AGT] 1SG.LOC-GIVE=bite.IPFV

'last night, the mosquitoes were biting me'   B7.34.3

Topical, fronted NPs may combine with demonstratives, in addition to the topic marker *hel* (3.TOP), but remain to be indexed on the verb.

(9)   Topical NPs marked with demonstratives

kawen do,   **a**   **ha-***komangdi-a*   *naha!*

machete PROX   2SG.AGT 3.PAT-blunt.PFV-CONT   not

'don't you blunt <u>the machete</u>'   EVY.1114

These and other types of differential argument marking are described in detail in Kratochvíl (2014c).

2.4.   Possessives and partitives

Abui marks possessed nouns with a prefix indexing the person and number of the possessor. Possessors marked in this way are animate or individuated and the construction has a definite reference. A juxtaposition of two nouns, shown in (a) below, has a similar meaning but is analyzed as a compound whose reference is not restricted.

---

[14] As the term 'plain' suggests, the *CVdo* pronominal paradigm is not marked for agentive case, and is compatible with both A and U arguments. As A argument, it is followed by *di* (3AGT); as U argument, it may be indexed on the verb with a person prefix.

(10)   Possessor marking

a.   *fala   ameeng*

house thatch

'house thatches, roofing grass'

b.   *fala   **he**-ameeng*

house   3.AL-thatch

'the house roof, thatches (for the house)'

c.   *mok-fala   **he**-cet   **he-t**-adafi*

church   3.AL-paint 3.LOC-DISTR.PAT-peel.off.PFV

'the paint on the church (in this village) is peeled off'

Other types of possessive constructions exist, which do not require the possessed entity to be definite. The reference is inferred using the possessive marking and articles. This is shown in (11), where the NP *hemayol* 'their wifes' in (a) is definite, the NP *seeng nu* 'money' in (b) is indefinite and specific, and the NP *seeng* 'money' in (c) is indefinite and non-specific.

(11)   Other possessive constructions

a.   *Ne-naana   loku tafuda **he**-mayol   ho-pa.*

1SG.AL-older.sibling PL   be.all   3.AL-wife   3.REC-have.IPFV

'All my older brothers have their wives.'   D.HOPA1

b.   *Seeng **nu**   ne-i   naha.*

money SPC   1SG.LOC-own.IPFV not

'(That certain) money is not mine.'   D.NEI.1

c.   *Seeng no-pa   naha.*

money 1SG.REC-have.IPFV not

'I do not have (any) money.'   D.NOPA1

Abui partitives consist of two juxtaposed phrases: the first is an NP defining the set, the second one a quantifier phrase specifying the part to be taken. The set is always definite, but the marking is variable. In principle, the article is optional in partitives with possessive marking, but obligatory in the remaining cases. The Abui corpus contains examples of partitives with both proximal and medial forms: the proximal form is common with *tafuda* 'all'. However, even the addressee-based medial form *yo*

can be used, as shown in (12), from which I conclude that this form marks definiteness or at least is compatible with it, as we will discuss later.

(12)   Abui partitives

a.   *na   e-feela   nuku hoo-k=siyeei*
       1SG.AGT [2SG.AL-friend]~set~ one   3.GOAL-BRING=meet.PFV
       'I met one of your friends.'   E15BDD5

b.   *na   ne-feela   loku **yo**   mingwaaha hoo-k=siyeei*
       1SG.AGT [1SG.AL-friend PL   MD.AD]~set~ some   3.GOAL-BRING=meet.PFV
       'I met several of my friends.'   E15BDD7

c.   *na   feela   loku **yo**   nuku hoo-k=siyeei*
       1SG.AGT [friend PL   MD.AD]~set~ one   3.GOAL-BRING=meet.PFV
       'I met one of the (those) friends.'   E15BDD9

2.5.   Abui articles used with unique reference nouns

Nouns with unique reference in the given context (both situational and general) such as *tuong* 'priest' or *raha* 'chief, regent' occur with and without articles. The reference is established in the given context. While the addressee-based forms highlight the familiarity of the referent (immediate or established in the past), the forms *nu* and *hu* force indefinite readings and presuppose a set containing other types. The form *hu* indicates, that the type is familiar, implying, that the hearer is familiar with the composition of the larger presupposed set.

(13)   *Yaal do na mook heesiyeei naha.* 'I didn't go to the church today.'

a.   *Tuong   ha-riik-e.*
       priest   3.PAT-ill-PROG
       'The priest is ill.'   E15BDD55

b.   *Tuong **do**   ha-riik-e.*
       priest   PROX   3.PAT-ill-PROG
       'The priest is ill.'   E15BDD56

c. *Tuong* ***to***   *ha-riik-e.*

    priest   PROX.AD   3.PAT-ill-PROG

    'The priest (familiar) is ill.'   E15BDD57

d. *Tuong=**o***   *ha-riik-e.*

    priest=MD   3.PAT-ill-PROG

    'That priest (discourse-old) is ill.'   E15BDD58

e. *Tuong* ***yo***   *ha-riik-e.*

    priest   MD.AD   3.PAT-ill-PROG

    'That priest (familiar, discourse-old) is ill.'   E15BDD59

f. *Tuong* ***nu***   *ha-riik-e.*

    priest   SPC   3.PAT-ill-PROG

    'A priest (among other people responsible for the service) is ill.'   E15BDD60

g. *Tuong* ***hu***   *ha-riik-e.*

    priest   SPC.AD   3.PAT-ill-PROG

    'The only priest (familiar type) is ill.'   E15BDD61

The same contrasts are found with other unique reference nouns such as *raha* 'king, chief, regent'.

2.6.   Abui articles used with proper names

Abui definite and specific articles can be used with proper names. For specific articles we have to assume that they are underspecified for definiteness given their compatibility with proper names, and should not be analyzed as true indefinite articles (cf. Lyons 1999:51). In (14) we give an example of a question where the article following the proper name is systematically manipulated and we indicate the consequences for the context in the translation line. The hearer-oriented articles indicate familiarity with the person and could also be translated as *our Fani*. Another possible context is one in which the person has just been mentioned. The specific articles force a set interpretation, where *Fani* is a member of a group and is coming.

(14)   Abui articles with proper names - questions

a.   *Ma, Fani **do** yaal ko di sei?*

      PART PN   PROX today FUT 3.AGT come.down.IPFV

      'By the way, will Fani come down today?'   EVY.1310

b.   *Ma, Fani **do** yaal ko di sei?*

      PART PN   PROX.AD today FUT 3.AGT come.down.IPFV

      'By the way, will (our) Fani come down today?'   EVY.1310A

c.   *Ma, Fani=**o** yaal ko di sei?*

      PART PN=MD   today FUT 3.AGT come.down.IPFV

      'By the way, will Fani (we talked about) come down today?' EVY.1310B

d.   *Ma, Fani **yo** yaal ko di sei?*

      PART PN   MD.AD today FUT 3.AGT come.down.IPFV

      'By the way, will Fani (we know) come down today?'   EVY.1310C

e.   *Ma, Fani **nu** yaal ko di sei?*

      PART PN   SPC today FUT 3.AGT come.down.IPFV

      'By the way, will Fani (among others) come down today?'   EVY.1310D

f.   *Ma, Fani **hu** yaal ko di sei?*

      PART PN   SPC.AD today FUT 3.AGT come.down.IPFV

      'By the way, will Fani (and no one else) come down today?' EVY.1310E

Another example of a proper names combined with articles can be seen in the comparative construction in (15). We find similar effects as above. In this context, the proximal articles may be also used in the situation where *Maifan* is physically present when the comparison is made, or that the speaker has some special connection with *Maifan*. Both specific articles presuppose a set of other children to which *Maifan* belongs.

(15)   Abui articles with proper names - comparatives

a.   *Lema moku fila, Maifan **do** fing.*

      PN   kid   be.young PN   PROX be.eldest

      'Lema is older than (our - exclusive) Maifan.'   EVY.700A

b.   *Lema   moku   fila,   Maifan   **to**   fing.*

    PN   kid   be.young PN   PROX.AD be.eldest

    'Lema is older than (our - inclusive) Maifan.'   EVY.700B

c.   *Lema   moku   fila,   Maifan=**o**   fing.*

    PN   kid   be.young PN=MD   be.eldest

    'Lema is older than Maifan (earlier mentioned or less well-acquainted).'

                                                 EVY.700C

d.   *Lema   moku   fila,   Maifan   **yo**   fing.*

    PN   kid   be.young PN   MD.AD be.eldest

    'Lema is older than Maifan (less well-acquainted to us).'   EVY.700D

e.   *Lema   moku   fila,   Maifan   **nu**   fing.*

    PN   kid   be.young PN   SPC be.eldest

    'Lema is older than Maifan (among other kids).'   EVY.700E

f.   *Lema   moku   fila,   Maifan   **hu**   fing.*

    PN   kid   be.young PN   SPC be.eldest

    'Lema is older than only Maifan (among other kids).'   EVY.700F

We conclude that the proper names show similar effects with specific articles as nouns with unique reference discussed in section 2.5. The combinatory properties of common nouns with articles will be discussed in detail in section 2.8.

2.7.   Indefiniteness and specificity

As discussed in section 1.1, indefinite and specific referents are both not identifiable to the hearer. Lyons (1999:49-51) claims that languages that only mark definiteness are the most common. Marking of indefiniteness only, or of both indefiniteness and definiteness is also a frequent pattern. Markers of indefiniteness are commonly derived from the numeral 'one' and encode cardinality. In case of the English *sm* (reduction of *some*), vague number is encoded, in addition to indefiniteness (Lyons 1999:50).

In Abui, indefinite non-specific referents are expressed either as bare nouns, or with the indefinite article *nuku*. Indefinite specific referents take *nu* or *hu*.[15] An example of the bare noun expression contrasted with one with specific reference can be seen in (16). In both cases the speaker is looking for a *taangwaala* 'mediator' whose role in the Abui traditional legal system is similar to that of an attorney. In the first example, the speaker has no particular mediator in mind, any will do (narrow scope). In the second case, the speaker knows precisely which mediator, but does not expect the addressee to know (wide scope).

(16)   Indefinite reference of bare nouns

a.   *Na   taangwaala tahai.   Na   dara nuku h-ieng   naha.*
      1SG.AGT mediator   search.IPFV 1SG.AGT still   one   3.PAT-find.IPFV not
      'I am looking for a mediator. I have not found any yet.'   E15BDD77

b.   *Na   taangwaala **nu** hee-l=tahai.   Na   dara*
      1SG.AGT mediator   SPC   3.BEN-GIVE=search.IPFV 1SG.AGT still
      h-ieng   naha.
      3.PAT-find.IPFV   not
      'I am looking for a mediator. I have not found him yet.'   E15BDD76

Another set of examples (paraphrases of Partee's examples 1972), illustrating the difference in marking of specific and indefinite referents in Abui is given in (17). The indefinite referent of *Flores mayool* 'Flores girl' is marked with the indefinite article *nuku* which originates in the numeral *nuku* 'one', matching a common cross-linguistic pattern, pointed out by Lyons (1999:50). Where the speaker has a specific referent in mind, the specific *nu* is used. Despite its superficial similarity with *nuku*, this form has a different source, and is related to the demonstrative root *n-* with a meaning similar to the English 'certain, such'.

---

[15] In the literature, the two types of indefinites are sometimes referred to as narrow-scope vs. wide scope indefinites. We present equivalents of the scope contrasts discussed in Matthewson (1999:88-92), showing that the Abui articles encode a similar contrast.

(17)   Indefinite reference with *nuku*

a.   *Ne-noo-mi=ng*   *maraang   na   Flores   mayool **nuku***
      1SG.AL-1SG.GOAL-be.inside=SEE come.up.IPFV 1SG.AGT place   woman   INDEF

      hee-l=mia.   *Na   dara nuku kaang baai*
      3.BEN-GIVE=marry.IPFV 1SG.AGT still   one   be.good also

      hoo-k=sei   naha.
      3.GOAL-BRING-meet.IPFV   not

      'I want to marry a girl from Flores. I haven't met one yet.'   E15BDD83

b.   *Ne-noo-mi=ng*   *maraang   na   Flores   mayool **nu***
      1SG.AL-1SG.GOAL-be.inside=SEE come.up.IPFV 1SG.AGT place   woman   SPC

      hee-l=mia.   *Nido   te-feela   wan   tuung yeekna.*
      3.BEN-GIVE=marry.IPFV 1PL.EXCL.FOC DISTR.AL-friend already year   be.five.IPFV

      'I want to marry a girl from Flores. We have been friends for five years.'

                                                        E15BDD84

Definite and specific referents differ in whether or not the speaker is certain about the identity of the referent. Similarly to the definites, specifics are linked to previously established discourse referents, but indefinites and non-specifics are not (Enç 1991:9). Further, specificity entails existence, which is why in some languages, such as Turkish, specific referents are incompatible with negative existentials (Enç 1991:14-16). Abui specific articles may occur in negative existentials but they always presuppose a set and are therefore similar to the Turkish determiners such as *birkaç* 'some' and *hiçbir* 'any' (Enç 1991:15). We interpret these constructions as partitives, where the article follows the set-defining phrase and is followed by an empty quantifier, which can be made overt.

(18)   Negative existentials and specific articles

a.   *Faring   wiil   ha-du   naha.*
      many   child   3.PAT-have.PRF   not
      'Many (people) did not have (any) child(ren).'   SULTAN59

b.   *Faring   wiil loku   ha-du   naha.*
      many   child PL   3.PAT-have.PRF   not
      'Many (people) did not have (any) children.'   SULTAN59A

c. *Faring wiil **nu** Ø hadu naha.*
    many [child SPC]SET 3.pat-have.prf not
    'Many (people) did not have (such) children.' SULTAN59B

d. *Faring wiil **hu** Ø ha-du naha.*
    many [child SPC.AD]SET 3.PAT-have.PRF not
    'Many (people) just did not have children (had other things).' SULTAN59C

e. *Faring wiil **nu** nuku wala ha-du naha.*
    many [child SPC.AD]SET one only 3.PAT-have.PRF not
    'Many (people) just did not have a single one of such children.' SULTAN59D

Contextually identifiable referents (cf. Lyons 1999:3-6) are marked in Abui also with the specific article *nu*, as in (19). English uses in these contexts the definite article although the referents are not familiar to the hearer, but the physical situation makes them identifiable.

(19) Situational uses of the specific article

a. *Tila **nu** latukoi ming-fikda naha, di sik-i=he!*
    rope SPC very.much APPL-tighten.IPFV not 3.AGT snap-PFV=PROH
    Don't tie the rope too tightly, don't let it snap! EVY.689

b. *Lukai-isi **nu** he-bakon-te!*
    pepper-fruit SPC 3.LOC-pluck.off.PFV-PRIOR
    'Pluck those peppers!' EVY.707

We have shown in section 2.3, that generic reference is encoded by bare noun phrases in Abui. Non-specific and specific indefinites are systematically distinguished by *nu*/*hu* vs. *nuku*. Situationally identifiable referents (definite in English) are marked as specific in Abui.

2.8. Givenness hierarchy

Gundel et al. (1993) consider familiarity and uniqueness to be distinct cognitive statuses, which are part of Givenness Hierarchy. The cognitive status of typical definite NPs is lower than that of *activated information* (marked with demonstratives)

and *information in focus* (marked with pronouns). The hierarchy is reproduced in Table 5.

COGNITIVE STATUS	English	Definition
IN FOCUS	*it*	the referent is in short-term memory and at the current center of attention
ACTIVATED	*that*, *this*, *this* N	the referent is represented in current short-term memory
FAMILIAR	*that* N	the hearer already has a representation in memory (in long-term memory if it has not been recently mentioned or perceived, or in short memory if it has)
UNIQUELY IDENTIFIABLE	*the* N	the hearer can identify the intended object on the basis of the nominal alone, but the identifiability does not have to be based on previous familiarity
REFERENTIAL	indefinite *this* N	the speaker intends to refer to a particular object; the hearer has or is able to construct a representation
TYPE IDENTIFIABLE	*a* N	the hearer is able to access a representation of object described

Table 5: Givenness Hierarchy (Gundel et al. (1993:275-280)

The contrasting sentences in (20) show the Abui equivalents of the dog-sentences from Gundel et al. (1993) exemplifying the Givenness Hierarchy. The sentences are modified to explore all possible contrasts in Abui. We start at the bottom of the hierarchy, with the examples of non-uniquely identifiable objects. Indefinite reference is marked with the article *nuku*. Specific referents (known to the speaker only) are marked with the articles *nu* and *hu*. The addressee-based form *hu* indicates that the hearer is able to identify the type and that the type is in some way noteworthy. This contrasts with the function of the specific articles with proper names and with nouns with unique reference, where the specific articles presupposed a set. With 'noteworthiness' I mean a category similar to the Slavic 'dative of empathy' (DE), related to the 'ethical dative' and 'dative of interest' in some Indo-European languages (Fried 2011).[16] The category has an interpersonal function and highlights to

---

[16] Dative of empathy (DE) is a special type of dative, resembling ungoverned datives (Dative of Interest), which 'always mark human referents with some interest in the reported event' (Fried 2011: 4). DEs 'serve a discourse-deictic function in speaker-hearer relations' (Fried 2011: 5), but unlike the Abui addressee-based forms, DEs are speaker-oriented.

the hearer what the speaker deems to be worth of hearer's interest, eliciting hearer's empathy and attention. The information in (e) below has a flavor of surprise on the side of the hearer and puzzlement about what happened and invites the hearer to reason about the event.

(20)   *El tuntama na taa beeka.* 'Last night I couldn't sleep.' (indefinite)

a.   *Kaai fala baleekna   mia   panen=ba   n-ieng   moopi   naha.*
       [dog house surround.IPFV be.in]_{RC} make.PFV=SIM   1SG.INAL-eye sleepy.PFV not
       'A dog (next door) kept me awake.'   E15BDD15

b.   *Kaai (**nuku**)   panen=ba   n-ieng   ariidi.*
       [dog   INDEF]_{NP}   make.PFV=SIM   1SG.INAL-eye   open.up.PFV
       'A dog kept me awake.'   E15BDD16

c.   *Heel   kaai (**nuku**) (**do**)   panen=ba   n-ieng   ariidi.*
       [3TOP   dog   one   PROX]_{NP} make.PFV=SIM   1SG.INAL-eye   open.up.PFV
       'This dog (next door) kept me awake.'   E15BDD18-9

d.   *Kaai **nu**   (di)   panen=ba   n-ieng   ariidi.*
       [dog   SPC]_{NP} 3.AGT   make.PFV=SIM   1SG.INAL-eye   open.up.PFV
       'A certain dog kept me awake.'   E15BDD24

e.   *Kaai **hu**   (di)   panen=ba   n-ieng   ariidi.*
       [dog   SPC]_{NP} 3.AGT   make.PFV=SIM   1SG.INAL-eye   open.up.PFV
       '(Imagine) a dog kept me awake.'   E15BDD25

As shown above, the definite NPs encode two different cognitive statuses: uniquely identifiable and familiar. Abui data shows that a further division is possible. The proximal forms seem to mark representations available in the 'immediate context', be it short-term memory or general knowledge. Medial forms portray the uniquely identifiable object as not being in the immediate context. The familiarity distinction is available to both.

(21) *El tuntama na taa beeka.* 'Last night I couldn't sleep.' (definite)

a. *Kaai **do**   (di)   panen=ba   n-ieng   ariidi.*
    [dog   PROX]_{NP} 3.AGT   make.PFV=SIM 1SG.INAL-eye open.up.PFV
    'The dog kept me awake.'   E15BDD20

b. *Kaai=**o**   (di)   panen=ba   n-ieng   ariidi.*
    [dog=MD]_{NP}   3.AGT   make.PFV=SIM   1SG.INAL-eye open.up.PFV
    'That dog (previously mentioned) kept me awake.'   E15BDD22

c. *Kaai **to**   (di)   panen=ba   n-ieng   ariidi.*
    [dog   PROX.AD]_{NP} 3.AGT   make.PFV=SIM   1SG.INAL-eye open.up.PFV
    'The (familiar) dog kept me awake.'   E15BDD21

d. *Kaai **yo**   (di)   panen=ba   n-ieng   ariidi.*
    [dog   MD.AD]_{NP} 3.AGT   make.PFV=SIM   1SG.INAL-eye open.up.PFV
    'That (familiar) dog kept me awake.'   E15BDD23

The notions of 'immediate' vs. 'non-immediate context' appealed to above will be discussed in a separate publication focusing on natural discourse, and were sketched in Kratochvíl (2015). At this stage it will suffice to say that representations in 'non-immediate context' may also be created by accommodation of presupposition associated with the 'familiar' forms.[17]

Forms indicating activated cognitive status are shown in (22). One possibility is to use any of the deictic pronouns, discussed in section 2.1, providing the dog is visible, as in (a). The other option involves using a headless relative clause consisting of the classificatory posture verb *it* 'be (for non-humans)' followed by an article.[18] With proximal and medial articles, the dog is visible. With the specific article, the dog is absent, but the speaker is nodding towards dog's location in the night, when it was barking. Finally, the agentive pronoun di can fully replace the NP, as in the English equivalent.

---

[17] 'If at time *t* something is said that requires presupposition *P* to be acceptable, and if *P* is not presupposed just before *t*, then – *ceteris paribus* and within certain limits – presupposition *P* comes into existence at *t*.' (Lewis 1979:340, via Abbott 2004:134)

[18] Many Papuan languages use posture verbs to classify nouns in context where English would use simply the verb 'to be' as in *there is a dog*… A recent summary of the known systems can be found in Rumsey (2002). The Abui facts are briefly outlined in Kratochvíl (2007:10).

(22) *El tuntama na taa beeka*. 'Last night I couldn't sleep.' (activated)

a. ***Do***   *panen=ba*   *n-ieng*   *ariidi.*

    PROX   make.PFV=SIM   1SG.INAL-eye   open.up.PFV

    'This (pointing at the proximate dog) kept me awake.'   E15BDD26

b. ***It***   ***to***   *di panen=ba*   *n-ieng*   *ariidi.*

    NON.HUM   PROX.AD   3.AGT make.PFV=SIM   1SG.INAL-eye   open.up.PFV

    'This (animal) (you know) kept me awake.'   E15BDD28

c. ***It***   ***nu***   *di panen=ba*   *n-ieng*   *ariidi.*

    NON.HUM   SPC   3.AGT make.PFV=SIM   1SG.INAL-eye   open.up.PFV

    'That (animal) (wherever it is now) kept me awake.'   E15BDD30

d. ***Di***   *panen=ba*   *n-ieng*   *ariidi.*

    3.AGT   make.PFV=SIM   1SG.INAL-eye   open.up.PFV

    'It (the dog) kept me awake.'   E15BDD26

The Abui forms exemplified above are arranged in a customized Givenness Hierarchy in Table 6.[19] For Abui, uniquely identifiable objects can be marked as familiar with the addressee-based forms. For specific reference (lower end of the same hierarchy), addressee-based form mark a form of familiarity which we term here as 'noteworthiness'.

---

[19] The shorthand CLV stands for classificatory verbs (*it* 'be put flat', *mihi* 'be put upright', *taa* 'lie', *mit* 'sit', *natet* 'stand' and *tili* 'hang').

COGNITIVE STATUS	Abui [speaker-viewpoint]	Abui [addressee-viewpoint]	Modified definition
IN FOCUS	*di*		the referent is in short-term memory and at the current center of attention
ACTIVATED [DEMONSTRATIVE]	*do, lo/o, ò, ó, wò, wó, oro*	*to, yo*	the referent is represented in current short-term memory and visible
ACTIVATED [RELATIVE CLAUSE]	CLV *do*, CLV=*o*, CLV *nu*	CLV *to*, CLV yo, CLV *hu*	the referent is represented in current short-term memory and visible or invisible
UNIQUELY IDENTIFIABLE (IMMEDIATE)	N *do*	N *to*	the hearer can identify the intended object on the basis of the nominal alone, but the identifiability does not have to be based on previous familiarity; the intended object is located within the immediate context (including general knowledge)
UNIQUELY IDENTIFIABLE (NON-IMMEDIATE)	N *o*	N *yo*	the hearer can identify the intended object on the basis of the nominal alone, but the identifiability does not have to be based on previous familiarity; the intended object is located outside the immediate context (but within the general knowledge)
REFERENTIAL	N *nu* *heel* N *do* *heel* N *o* *heel* N *nu*	N *hu* *heel* N *to* *heel* N *yo* *heel* N *hu*	the speaker intends to refer to a particular object and can indicate where it is located in the context; the hearer has or is able to construct a representation which can be marked as noteworthy
TYPE IDENTIFIABLE	N (*nuku*)		the hearer is able to access a representation of object described

Table 6: Givenness Hierarchy (Gundel et al. 1993) for Abui

Arkoh and Matthewson (2013) discuss the uses of the familiar article *nʊ* in Akan. In Akan, the familiarity is only compatible with the definite reference, and incompatible with indefinites, where a dedicated article *bí* 'a certain' is used (p. 7). There is also a dedicated definite article *nó* covering the 'uniquely identifiable' category in the Givenness Hierarchy (Akhoh and Matthewson 2013:7). Unaware of another language with a similar split in the specific category, we use the label 'noteworthiness' rather than 'familiarity', which is traditionally understood as being restricted to uniquely identifiable referents.

3.   Discussion

This paper mapped the uses of the Abui demonstratives and articles. The article paradigm encodes in a systematic manner definiteness, indefiniteness, and specificity. For definite reference, proximal and medial forms distinguish two degrees of distance in the 'immediate context'. In addition, familiarity and noteworthiness are marked by hearer-oriented articles. Abui articles may co-occur with proper nouns and with some pronouns. The hearer-oriented forms have a variety of interesting discourse uses, often related to stance, which will be discussed in a separate paper.

The Abui system offers an insight into the category of familiarity, which is often taken to be another dimension of definiteness (Lyons 1999:6). Familiarity seems to be a separate category possibly also available for specific reference, allowing the speaker to express confidence that the hearer shares the relevant knowledge, or to reassure the hearer that he can do so. While in Akan, the familiar article *nʊ* introduces a presupposition that the relevant discourse referent is present in the common ground between speaker and hearer (Arkoh and Matthewson 2013), the Abui familiar articles have a greater range of functions. As pointed out by Stubbe and Holmes (1995) and Macaulay (2002) the English hearer-oriented 'you know' sometimes marks speaker's uncertainty about hearer's attitude, or about the precision of the description and their use could be quite idiosyncratic and dependent on personal speech styles. The same seems to be true for the Abui familiar forms.

Finally, Abui makes a three-way contrast in indefinites distinguishing formally indefinite non-specific referents (a.k.a. narrow scope indefinites, marked with *nuku*) from specific indefinites (wide-scope indefinites, marked with *nu* or *hu*).

References

Abbot, Barbara. 2004. Definiteness and Indefiniteness. In Laurence R. Horn and Gregory Ward (eds.) *The Handbook of Pragmatics*, 122-149. Malden (Ma.): Blackwell.

Anderson, Curt and Marcin Morzycki. 2015. Degrees and kinds. *Natural Language & Linguistic Theory* 33(3): 791-828.

Arkoh, Ruby, and Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123: 1-30.

Arsenijević, Boban, František Kratochvíl, and Joanna Ut-seong Sio. 2014. Equative/similative demonstratives and the semantics of indirect modification. Manuscript. Nanyang Technological University, Singapore.

Baird, Louise. 2008. *A grammar of Klon: a non-Austronesian language of Alor, Indonesia*. Canberra: Pacific Linguistics.

Davis, Phillip W. and Ross Saunders. 1975. Bella Coola Nominal Deixis. *Language* 51(4):845-858.

Davis, Henry, Carrie Gillon, and Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90 (4):e180-e226.

Diessel, Holger. 1999. *Demonstratives: form, function and grammaticalization* [Typological Studies in Language, vol. 42]. Amsterdam: John Benjamins.

Diessel, Holger. 2014. Demonstratives, Frames of Reference, and Semantic Universals of Space. *Language and Linguistics Compass* 8/3 (2014): 116–132.

Dryer, Matthew S. 2013a. Definite Articles. In Matthew S. Dryer & Martin Haspelmath (eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/37, Accessed on 2014-12-24.

Dryer, Matthew S. 2013b. Indefinite Articles. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/38, Accessed on 2014-12-24.)

Dryer, Matthew S. 2014. Combining methods for uncovering linguistic diversity: The case of definite and indefinite articles (Commentary on Davis, Gillon, and Matthewson). *Language* 90(4): e232-e249.

Enç, Mürvet. 1991. The Semantics of Specificity. *Linguistic Inquiry* 22(1):1-25.

Fried, Mirjam. 2011. The notion of affectedness in expressing interpersonal functions. In Marcin Grygiel and Laura Janda (eds.) *Slavic Linguistics in a Cognitive Framework*, 121-144. Frankfurt & New York: Peter Lang Publishing Company.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2): 274–307.

Hanks, William F. 2005. Explorations in the deictic field. *Current Anthropology*. 46 (2): 191-220.

Hawkins, John. 1978. *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.

Heine, Bernd, and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

von Heusinger, Klaus. 2002. Specificity and Definiteness in Sentence and Discourse Structure. *Journal of Semantics* 19: 245-274.

Holton, Gary, and Laura C. Robinson. 2014. The internal history of the Alor-Pantar language family. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology*, 55-97. Berlin: Language Science Press.

Holton, Gary, Marian Klamer, František Kratochvíl, Laura C. Robinson, and Antoinette Schapper. 2012. The historical relations of the Papuan languages of Alor and Pantar. *Oceanic Linguistics*. 51 (1): 86-122.

Holton, Gary. 2014. Western Pantar. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 23-95 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Hopper, Paul J., and Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56 (2): 251-299.

Klamer, Marian. 2010. *A grammar of Teiwa*. Berlin: De Gruyter Mouton.

Klamer, Marian. 2014. Kaera. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 97-146 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Kluge, Angela. 2014. *A grammar of Papuan Malay*. Ph.D. dissertation, Leiden University. LOT: Utrecht.

Kratochvíl, František. 2007. *A grammar of Abui*. Ph.D. dissertation, Leiden University. LOT: Utrecht.

Kratochvíl, František. 2011. Discourse-structuring functions of Abui demonstratives. In Foong Ha Yap and Janick Wrona (eds.) Nominalization in Asian Languages:

Diachronic and Typological Perspectives. Volume 2: Korean, Japanese and Austronesian Languages, 761- 792. Amsterdam/Philadelphia: John Benjamins.

Kratochvíl, František. 2014a. Sawila. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 351-438 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Kratochvíl, František. 2014b. Number in Abui and Sawila. In Marian Klamer and František Kratochvíl (eds.) *Number and quantity in East Nusantara*, 123-151 [A-PL 012 / SAL 001]. Asia-Pacific Linguistics: Canberra.

Kratochvíl, František. 2014c. Differential argument realization in Abui. *Linguistics* 52 (2): 543–602.

Kratochvíl, František. 2015. Definiteness and referent tracking in Abui. Paper read at the Second International Workshop on Information Structure in Austronesian languages, Tokyo University of Foreign Studies, Japan, February 11-13.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2004. Deixis. In Laurence R. Horn and Gregory Ward (eds.) *The Handbook of Pragmatics*, 97-121. Malden (Ma.): Blackwell.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8: 339-359.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Macaulay, Ronald. 2002. You know, it depends. *Journal of Pragmatics* 34:749–767.

Matthewson, Lisa. 1998. On the interpretation of wide-scope indefinites. *Natural Language Semantics*, 7(1), 79-134.

Moravcsik, Edith A. 2011. Explaining Language Universals. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 69-89. Oxford: Oxford University Press.

Newman, Stanley. 1969. Bella Coola Paradigms. *Intrnational Journal of American Linguistics* 35(4):299-306.

Partee, Barbara H. 1972. Opacity, coreference, and pronouns. In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 415–441. Dordrecht: Reidel.

Roehrs, Dorian. 2009. *Demonstratives and definite articles as nominal auxiliaries*. Amsterdam: John Benjamins.

Rumsey, Alan. 2002. Men stand, women sit - On the grammaticalization of posture verbs in Papuan languages, its bodily basis and cultural correlates. In John Newman (ed.), *The Linguistics of Sitting, Standing and Lying*, 179-211. Amsterdam: John Benjamins.

Schapper, Antoinette (ed.). 2014a. *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars* [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Schapper, Antoinette. 2014b. Kamang. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 285-349 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Schapper, Antoinette and Rachel Hendery. 2014. Wersing. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 439-503 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Steinhauer, Hein. 2014. Blagar. In Antoinette Schapper (ed.) *The Papuan Languages of Timor, Alor and Pantar: Volume 1 Sketch Grammars*, 147-219 [Pacific Linguistics Vol. 644]. De Gruyter Mouton: Berlin.

Stubbe, Maria and Janet Holmes. 1995. You know, eh and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication* 15(1):63-88.

Yap, Foong Ha, Karen Grunow-Hårsta, and Janick Wrona. 2011. *Nominalization in Asian languages diachronic and typological perspectives*. Amsterdam: John Benjamins.