Analyzing questions under discussion and information structure in a Balinese narrative

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.

Keywords: annotation, focus, linguistic universals, narrative, Question under Discussion

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.
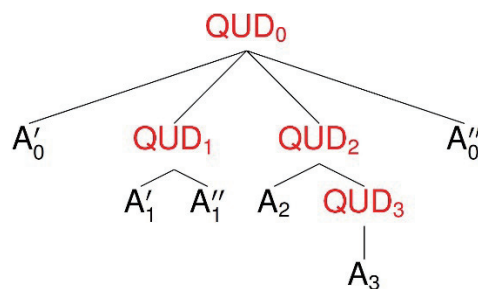
Figure 1 Discourse tree with Questions under Discussion

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 A_0'. (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.

$$q_{15}$$
$$a'_{15} \quad a''_{15}$$
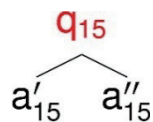
Figure 2 A Question under Discussion with two partial answers

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.