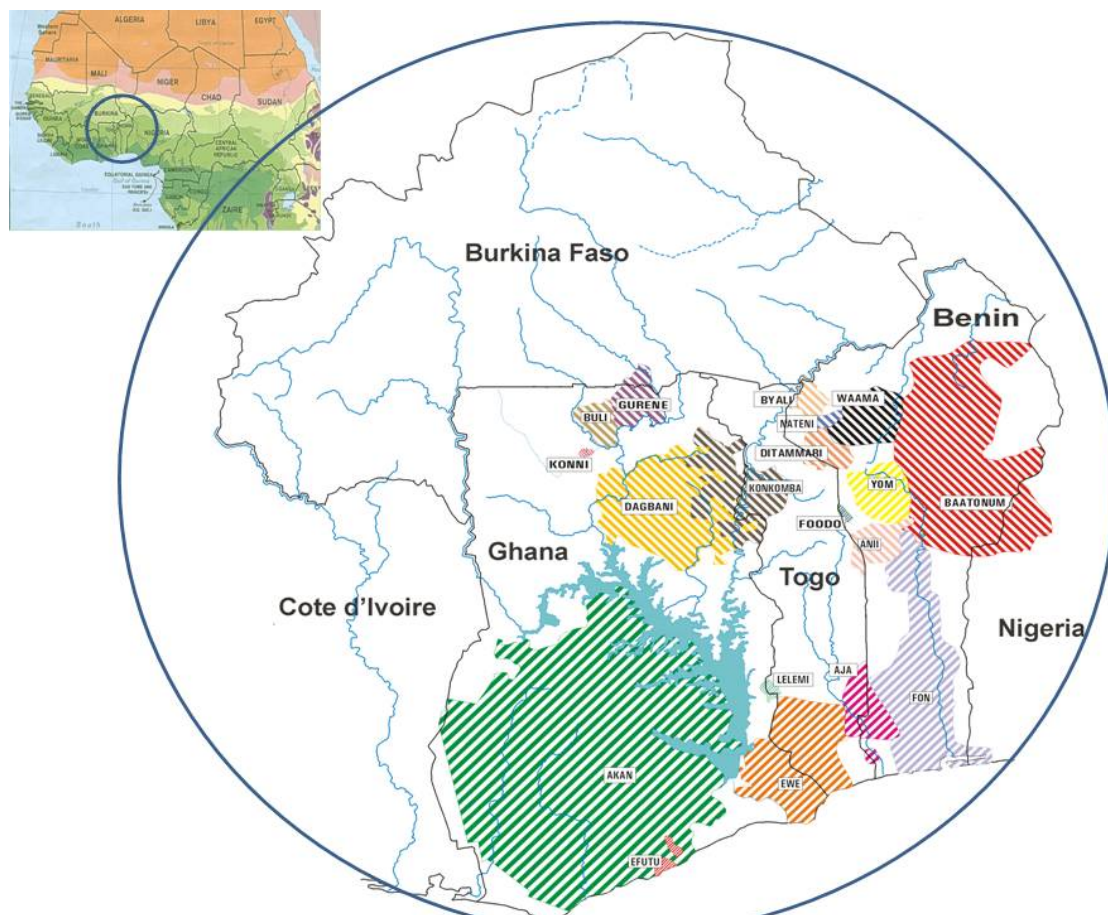


Discourse principles in grammar: Thethetic/categorical dichotomy

This paper deals with communication needs and grammar. For linguists discovering the grammatical systems of languages it is most exciting to find out in which respect the grammars of languages all over the world resemble each other and in which respect they differ, because this gives us further insights into the variety and basic principles of human communication and human language. Sometimes we find a common basis behind quite different grammatical disguises and the range of variation helps us to identify the shared core elements across the languages, as in the case I want to discuss here.

A puzzling particle in Buli

During the last couple of years I have been working on linguistic puzzles in various West African languages, particularly in Ghana. For one of the puzzles I found the key to better understanding only recently.



The language Buli has a basic Subject-Verb-Object word order and makes extensive use of particles, such as *ká*, *lá*, *lē*, *mē* and others, which occur very regularly in different spots within the sentence. Of interest here is the particle *lē*.

The particle *lē* occurs at the beginning of the predicate, i.e. left of the verb, and it does not contribute a specific lexical meaning to the sentence, but remains untranslated.

Subject – (à)lē – Verb – Object

The particle is quite common in conversation and in texts of various genres and speakers insist that it has to be used in various contexts. This suggests that it fulfills certain function. Here are some indications for possible functions which, however, did not prove to be convincing in the end.

a) The particle is particularly frequent and often insisted on after indefinite subject nouns. Does that mean that the particle represents a kind of indefinite article, comparable to English **a**?

Gbáj àlē d̄à tébùlkù zúk.
book PTL lie table:the top
'There is a book on the table.'

There are various arguments against such analysis, among them:

- The particle can also occur after definite subjects (see example below).
- The particle can also occur after pronominal subjects.
- Prosodically, the particle belongs to the verb to its right and not to the subject to its left.

Nípòkbilní lē zà tēṅká.
girl:the PTL stand ground:the
'The girl is standing on the ground.'

b) Some speakers explained that the particle is a past tense marker and indeed, it is present in many utterances reporting past events.

Fí lē pà gōgōwá.
you PTL take watch:the
'You took the watch.'

Again, the analysis as past tense marker does not hold, as soon as the data are thoroughly checked:

- The particle is not restricted to past tense occurrences.
- The particle cannot always occur in past tense environments.

Sírí lē bòrō.
bee PTL exist
'There is a bee.'

c) The particle is required in utterances where the subject represents the most salient, i.e. focal information as well as in information questions concerning the subject. Does that mean that particle *lē* is a postposed subject focus marker, which—as in many languages—has also been generalized to subject information questions?

Wàn lē wì George?
who PTL call G.
'Who called George?'

The major counter arguments against the subject focus marker analysis comprise the following aspects:

- The particle occurs in the same position when the whole sentence and not only the subject can be regarded as focal.
- The particle obligatorily occurs in the same position in adverbial temporal clauses that form the initial clause of a complex sentence and in sentence-initial relative clauses. These are environments in which subjects are typically not focal at all.

Bà lē jè dílá nùè lā, ...
they PTL do that finish PTL
'After they finished doing that, ...'

For several other observations concerning *lē* we even completely lack immediate functional equivalents in more familiar languages like English and corresponding explanations:

- How to account, for instance, for the fact that the particle *lē* is most common at the beginning of all kinds of texts and occurs in titles and final summaries in expository texts?

Mí chúm-dīpō-bōká àlē màà yāā kà jē dī lā ònā.
my future-time-existence PTL I:IPF want it do how PTL this
'This is how I want my future going to be.'

We obviously need an explanation for this particle that goes beyond the above suggestions.

The thetic/categorical distinction

The first clue for a better understanding of the particle came from the study of the particle's distribution in discourse and from the recognition of a philosophical distinction transferred into linguistics.

At the end of the 19th century, the philosophers Brentano and Marty challenged the notion of a unique type of human judgment composed of a subject and a predicate in the Aristotelian sense. The effect that their proposal of a fundamental dichotomy between two basic types of logical statement had on linguistics is outlined in detail by Sasse (1987: 535ff.). Brentano and Marty maintained the classical bipartite subject-predicate structure of a judgment and called it *categorical*. This type of judgment was understood to be constituted of two successive acts: naming an entity and making a statement about it. The opposed *thetic* judgment, on the contrary, was assumed to be logically unstructured, serving the expression of an event or a state or situation. This thetic/categorical distinction on the logical level was further developed, taken up in linguistics and transferred onto the pragmatic level by a number of scholars, such as the functionalist Mathesius of the Prague School and several other linguists such as Kuroda, Kuno, Ulrich, Sasse among others. Despite certain deviations in theory and approach of the scholars, it turned out that utterances can be mentally construed as either unstructured (thetic) or structured (categorical) respectively. In fact, the distinction between categorical and thetic utterances has to be considered a universal

of thought and communications, which is however expressed to different degree and by different procedures in different languages (Ulrich 1988: 388). Regarding the pragmatic level, the bi-partite structure of categorical utterances is made up of a topic and a predicate commenting about it, while in athetic statement the propositional information is encoded as a single, unstructured package.

The thetic/categorical distinction (cf. Ulrich 1988)

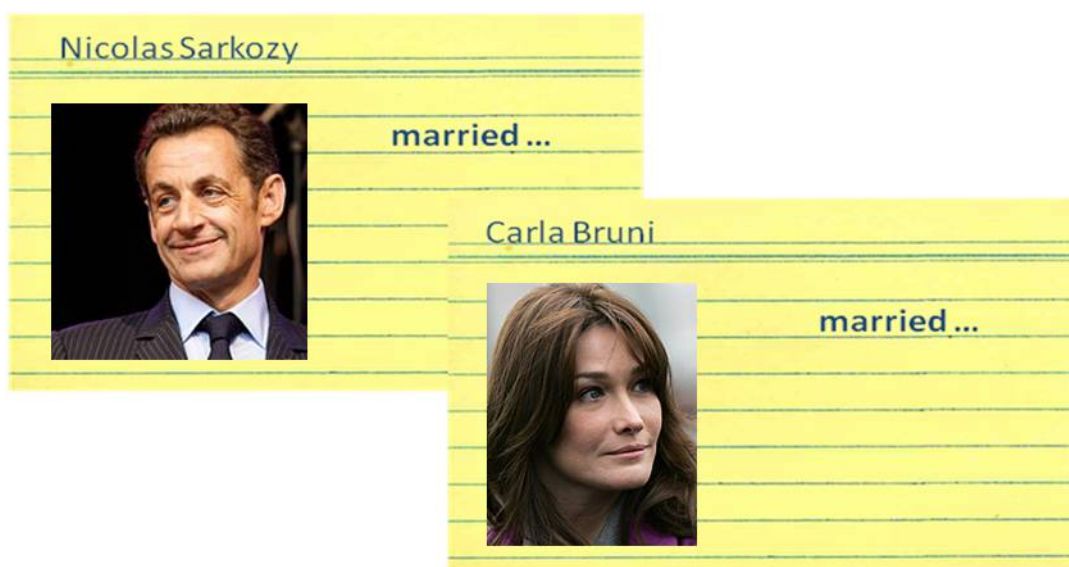
	categorical	thetic
judgement (logical level)	making a statement about an argument	presenting a fact as an inseparable unit
utterance (pragmatic level)	topic-comment structure	unstructured (topic-less)

Most scholars refer to the existence or absence of a topic for distinguishing between structured and unstructured utterances, but this is also a point where controversies start, because the notion of topic varies across linguistic theories and scholars. For the current aim it shall suffice to treat the topic of an utterance as that *what the sentence is about* and to think of the topic as a location or storage address for subsequent information. Tanya Reinhart (1982) introduced the helpful file card metaphor according to which the topic corresponds to the title of a file card: in (a) the comment is stored under the entry Nicolas Sarkozy, and in (b) the comment applies to the topical entry Carla Bruni.

(a) [**Nicolas Sarkozy**]_{Topic} [**married Carla Bruni**]_{Comment}

(b) [**Carla Bruni**]_{Topic} [**married Nicolas Sarkozy**]_{Comment}

(Krifka et al. 2008)



The file card metaphor also illustrates recurrent characteristic features of topics in discourse. The storage address or topic comes typically early in the utterance before the topic is commented on. In languages with basic *Subject-Verb* word order, the

subject in initial position will therefore take over the topic role as often as it can. And if the speaker's communicative goal is to transfer information to other interlocutors, she will not randomly pull out several file cards at the same time what could confuse the hearer and hinder the information processing, but she will rather try to use the already available storage addresses effectively, maintaining one topic for some time span and add as many comments she regards as appropriate, before switching to another one. It follows that topics tend to certain continuity in connected discourse and often represent information already shared between the interlocutors. Many languages provide particular forms, such as definite noun phrases or dedicated pronominal forms, for encoding recurrent topical referents and several languages allow speakers to completely omit topic expressions under certain conditions.

In several intonation languages, including English, the formal distinction betweenthetic and categorical statements relies on accentuation. In the unstructuredthetic utterance, there is only one accent (see capitals) on the subject, while in the bipartite categorical utterance, both subject and predicate are accented.

- a. Thetic utterance (unstructured):

The CHICKen's burnt.

← What the hell is this stench?

- b. Categorical utterance (bipartite topic-comment structure):

The CHICKen's BURNT.

← Where is the roast chicken I expected, why do I have a hamburger on my plate? (Sasse 1987: 529)

The devices best attested for the expression of thethetic/categorical distinction so far comprise accentuation and word order. The so-called subject inversion occurs in Modern Greek, for instance, where the non-topical subject phone follows the verb in thethetic utterance rather than preceding it.

Modern Greek:

- a. ***Xtipise to tilefono.***

rang ART phone

'The PHONE rang.' (thetic)

- b. ***To tilefono xtipise.***

ART phone rang

'The PHONE RANG.' (categorical) (Sasse 1987: 536)

The distinction between categorical andthetic utterances is not always iconically reflected in the grammar of a language. It is possible to have an unstructuredthetic utterance with a binary linguistic structure comprising a subject pronoun and a verb form, such as in the German example below. And it is possible to encode a categorical utterance by just one word, such as in the Romanian example, where the utterance contains just the comment:

- a. Thetic (unstructured) / German:
Es regnet.
'It is raining.'
- b. Categorical (bipartite structure) / Romanian:
Vine.
'She is coming.' (Ulrich 1988: 390)

These two examples illustrate that it is not the surface grammatical structure but the conceptual structure on the pragmatic level that is decisive for the thetic/categorical distinction. The fact that—despite of being conceptualized—the topic of an utterance is not obligatorily represented by an expression of its own in each categorical utterance renders the predicate the most appropriate candidate for encoding the difference between categorical and thetic utterances. We have seen such a case in the English chicken example above: the predicate in the thetic utterance receives no accent, but the one in the categorical utterance does.

We do now know that a number of languages have specific means to distinguish between thetic and categorical utterances but what is that good for? What is the benefit that speakers of these languages gain by a formal thetic/categorical distinction?

We have already heard that the categorical organization into topic and comment particularly serves incremental information transfer in discourse by providing a storage address for new information. In various languages the categorical organization of the utterance represents the dominant structure which is pragmatically less marked than the unstructured thetic organization and it applies in several communicative situations. So what justifies the alternative thetic organization? There seem to be three major forces in favor of the thetic organization.

The first argument pro theticity is immediately linked to the topic requirements that are active in a given language. Some languages have less restrictions on form, reference and grammatical function of topics on the utterance level than others: for instance, some languages always want topics to be encoded as subjects, others don't, some want them to be introduced into discourse via specific constructions, others don't, some languages allow discourse participants and local or temporal information to serve as default topic, others don't. In short, in some languages speakers are forced to employ thetic utterances in those cases where no appropriate topics to establish the dominant categorical utterance type are available.

Secondly, a minor group of predicates does typically NOT comment on a topic across languages. Weather verbs are particularly well-known in this respect. If I communicate weather information, such as ***It is raining. It is cold. or There is a strong wind blowing.*** I do typically not comment about a certain topic, but I just want to report a specific state or event. In a number of languages, this fact is reflected by using the language-specific thetic encoding.

And third, the thetic organization can be deliberately employed for rhetoric and stylistic reasons. The predominant categorical organization can thus be cancelled to

reflect a deliberate break of discourse coherence (for instance, by presenting an unexpected event in episode-initial position).

To sum up,thetic utterances make sense at certain points in discourse formation, too, and they may be grammatically conventionalized to different degree under specific conditions, for specific predicates and in specific languages. In general,thetic utterances can be regarded as presentations of either events or entities.

Let's now come back to Buli and see how the puzzling particle *lɛ̃* fits in.

Thetic utterance in Buli: Linking strategy (particle *lɛ̃*)

With the general outline of the thetic/categorical distinction as background we can now easily recognize that particle *lɛ̃* marks thetic utterances in Buli.

- The particle is more frequent with indefinite subject nouns, because these nouns are less likely to function as storage address, i.e. topic, than definite subject nouns. That notwithstanding, the thetic encoding can also apply to utterances with a definite subject, what clearly distinguishes the particle from an indefinite article.
- The resemblance of the particle with a past tense marker in several environments is based on the fact that it is often employed to present events that suddenly or unexpectedly brought up a change in discourse. Not representing a tense marker, of course, the particle is absent in those past tense cases not intended to bring up an abrupt change in discourse and it is found in non-past tense environments, too.
- The reminiscence of the particle with a subject focus marker derives from the fact that a non-topical subject is more likely to be in the center of attention than a topical subject about which a comment with its own informational peak is made.
- Particle *lɛ̃* is very common at the beginning of texts, because it here serves the introduction of major discourse participants that are then taken up as topics in the subsequent text.
- Likewise, it presents themes and summaries of texts in form of an unstructured title or conclusion to the audience.

Buli ignores the most widespread thetic devices – strong subject accent and subject inversion – and rather employs a dedicated particle at the beginning of the predicate. What is the specific achievement of this predicate-initial device for the thetic encoding?

In fact, there is considerable evidence that *lɛ̃* is not just an arbitrarily chosen particle marking pragmatically unstructured utterances. The concept behind becomes most obvious in environments without an immediate verb (cf. also Schwarz, forthcoming a). Here we realize that the particle marking thetic utterances is only tonally distinguished from the comitative preposition 'with, and'. The comitative preposition in (a) has a low tone and establishes a close link between the expressions for two entities which act as one grammatical argument regarding the following verbal predicate (hier unspecified). The particle encoding thetic utterances has a mid tone. If it is used in an utterance without a completing verbal predicate, as in (b), the link established between the two expressions is so close that they even have to be regarded as coreferential. In the absence of a verbal predicate the particle marking thetic utterances is thus reminiscent of a copular verb like English 'to be', though the copula-like use of the particle is quite restricted.

- | | |
|--------------------------------------|--|
| a. <i>Wá lè</i> [L] <i>nààwā ...</i> | ‘He and the chief ...’ [-coreferential] |
| b. <i>Wá lē</i> [M] <i>nààwā.</i> | ‘He is the chief.’ [+coreferential] |
| compare: <i>Wà kā</i> <i>nààb.</i> | ‘He is a chief.’ |

From these observations, it can be assumed that the particle *lē* marking thetic utterances is etymologically related to the comitative preposition with the meaning ‘with, and’ (*lè*). The particle’s function in thetic utterances is quite similar to that of the comitative preposition, namely to establish a close link between subject and predicate that “repairs” the common bipartite structure between topic and comment. Pragmatically, the linkage has a demoting effect on the second conjunct, the predicate, as in thetic utterances the predicate does not constitute a distinctive focal domain of its own, but has to share it with the subject.

The linking recognition of the thetic-marking particle *lē* can also better account for the regular occurrence of this particle in sentence-initial clauses of various types. Apparently, the linking particle is regularly approached to suppress the prototypical topic-comment structure of subject and predicate and to transfer the subject-predicate string into a single unstructured unit that can occur in initial position in complex sentences.

Thetic utterance in Lelemi: Mimesis strategy (default topics)

Research into other languages has shown that Buli is by far not the only West African language that has a dedicated device for the thetic encoding of utterances. Let me here only briefly add the example of Lelemi, a Ghanaian language which reveals another unique linguistic strategy so far unknown.

Lelemi shares several linguistic features with Buli to which it is also genetically remotely related. In comparison to Buli, however, Lelemi puts more functional load onto the verb. Lelemi is a language with regular subject cross-reference prefixes, i.e. “subject markers” (SM), at the verb:

Categorical:

- | | | | |
|----|-----------------------|--------------|--------------------|
| a. | <i>Ǿ-nàà</i> | <i>í-mò</i> | <i>ǿǿ-dì.</i> |
| | CL(SG)-man | CL(SG)-DET | SM[CL(SG).IPF]-eat |
| | ‘The MAN is EATing.’ | | |
| b. | <i>Bà-nàà</i> | <i>bá-mò</i> | <i>bàà-dì.</i> |
| | CL(PL)-man | CL(PL)-DET | SM[CL(PL).IPF]-eat |
| | ‘The MEN are EATing.’ | | |

In canonical sentences with a topical subject, the prefixed subject marker conflates Tense-Aspect-Mood information and person-number-gender information based on agreement with the topical subject. In the example above, the subject marker thus provides details about the particular singular or plural class and about the imperfective aspect.

Under certain conditions which are indicative for thetic utterances, however, the subject markers differ and are provided by a rather restricted set. While subject

marker set 1 occurring in categorical utterances displays full person-number-gender agreement with the subject, set 2 ofthetic utterances does not allow any person-number-gender agreement. In the example below, the subject marker thus only provides details about the imperfective (IPF) aspect.

Thetic:

a. *Ǿ-nàná* *ǿ-mò* *mò-dì.* (not: *ǿǿ-dì.*)
 CL(SG)-man CL(SG)-DET SM[IPF]-eat
 ‘The MAN is eating.’

b. *Bà-nàná* *bá-mò* *mò-dì* (not: *bàà-dì.*)
 CL(PL)-man CL(PL)-DET SM[IPF]-eat
 ‘The MEN are eating.’

We can thus conclude that Lelemi regularly marks thethetic/categorical distinction at the verb on the basis of its obligatory subject cross-reference (cf. also Schwarz, forthcoming b): Thetic and categorical utterances both require a verb with a subject marker, but the verb does only agree with the topical subject in a categorical utterance:

Set 1	SM _{AGR} -verb	categorical utterance
Set 2	SM _{FIX} -verb	thetic utterance

The subject markers of set 2, i.e. those without person-number-gender agreement, vary only according to Tense-Aspect-Mood and polarity variables. Most interestingly, two of the three available prefixes (**nA-**, **mU-**, and **N-**) are identical with the 1st person singular forms of set 1!

I assume that this is not incidental but that in the absence of a topical subject the speaker installs the most likely storage address or topic she can think of – herself. Hence the use of the 1st person singular subject marker represents a substitutional strategy that mimics the more common categorical pattern as far as possible.

To sum up, we have seen that in Lelemi, only topical subjects trigger person-number-gender agreement. If the subject does not have topic status, the speaker installs herself as the most easily accessible topic, hence we encounter the “irregular” generalization of a 1st person singular form inthetic utterances.

Conclusion

This paper demonstrates that recent research provides evidence for the relevance of thethetic/categorical distinction in the grammar of selected West African languages. Thethetic/categorical distinction is approached for discourse-structuring aims, though the degree to which thethetic organization of an utterance can be deliberately chosen in specific discourse contexts differs across languages. The two Ghanaian languages Buli and Lelemi reveal two uniquethetic strategies which are not recognized in the literature so far (for broad overviews cf. in particular Sasse 1987, 1995 and 2006):

- (i) A linking preposition which repairs the bipartite structure between topical subject and predicate and thus pragmatically demotes the latter (Buli)
 - *Linking strategy* recognizing the distinctive thetic organization: the construction of categorical and thetic utterances differs on first sight
- (ii) The installation of a 1st person default topic via subject cross-reference markers which substitute for the lack of other topics (Lelemi)
 - *Mimesis strategy* following the categorical organization which serves as model for all verbal predicates: yielding a subtle thetic/categorical distinction

We can conclude that less-documented languages of the tropics offer fascinating insights into discourse principles and their grammatical implementation and, having developed their own specific and imaginative thetic expressions, they enrich our understanding of the thetic/categorical distinction.



Kɔnni speakers in Northern Ghana (Photos © Anne Schwarz)

Acknowledgements

The research findings presented in this paper result from work conducted within the frame of the Collaborative Research Centre on Information Structure (SFB 632, <http://www.sfb632.uni-potsdam.de>, Projects B1 and D2) at the University of Potsdam and the Humboldt University Berlin. I am grateful to the German Research Foundation (DFG) for funding the research including generous fieldwork allowances.

References

- Krifka, Manfred, Fiedler, Ines, and Schwarz, Anne. 2008. "[Informationsstruktur. Die sprachlichen Mittel der Gliederung von Äußerung, Satz und Text – Sonderforschungsbereich 632.](#)" *Humboldt-Spektrum* 2008/1:42-48.
- Reinhart, Tanya. 1982. "Pragmatics and Linguistics. An Analysis of Sentence Topics." *Philosophica* 27:53-94.
- Sasse, Hans-Jürgen. 1987. "The Thetic/Categorical Distinction Revisited." *Linguistics* 25:511-580.

- Sasse, Hans-Jürgen. 1995. "Theticity' and VS Order: A Case Study." In *Verb-Subject Order and Theticity in European Languages (Sprachtypologie und Universalienforschung, 48: 1/2)*, eds. Yaron Matras and Hans-Jürgen Sasse. Berlin. 3-31.
- Sasse, Hans-Jürgen. 2006. "Theticity." In *Pragmatic Organization of Discourse in the Languages of Europe*, eds. Giuliano Bernini and Marcia L. Schwarz. Berlin, New York: Mouton de Gruyter. 255-308.
- Schwarz, Anne. Forthcoming a. "To be or not to be? About the Copula System in Buli (Gur)." *Proceedings of the Special WOCAL 6*, eds. Margarida Tadonni Petter and Ronald Beline Mendes. São Paulo.
- Schwarz, Anne. Forthcoming b. "The Semantics/Pragmatics of Lelemi Predicates." *Afrika und Übersee*.
- Ulrich, Miorita. 1988. "Thetisch vs. kategorisch und Informationsstruktur." In *Energieia und Ergon: Sprachliche Variation – Sprachgeschichte – Sprachtypologie*. Eds. John Albrecht, Harald Thun and Jens Lüdtke. Tübingen: Gunter Narr. 387-399.