Review Article

∴

Information Structure and Reference Systems: Toward a Non-Aprioristic Typology

Reading Fernandez-Vest and Van Valin 2016

Zygmunt Frajzyngier

University of Colorado at Boulder

zygmunt.frajzyngier@colorado.edu

Abstract

The present article proposes a non-aprioristic approach to analyzing the domains of information structure and reference systems. The article is inspired by the papers in *Information structuring of spoken language from a cross-linguistic perspective* (Fernandez-Vest and Van Valin (eds.), 2016), and from my own research on languages for which only spoken data exist. As an outcome of this study it may turn out that ‘information structure’ and ‘reference system’ each constitute a distinct functional domain in some languages. The study addresses some of the most interesting findings in languages discussed in the volume, supplemented by my own findings on a variety of languages.

Keywords

topicalization – focus – systems of reference – detachment

1 Introduction

1.1 Background

The volume that inspired this paper, *Information structuring of spoken language from a cross-linguistic perspective* (Fernandez-Vest and Van Valin (eds.),

2016), deals with two areas: information structure, as traditionally understood over the last 40 years; and reference systems, which have yet to be recognized as an important domain in the functions coded in languages. In the volume under discussion these two domains are often not distinguished, and in many papers functions belonging to the system of reference are treated as constituting part of the information structure.

The editors have divided the volume into four parts: (1) Theoretical approaches to information structure (four papers); (2) Information structure and spoken language (six papers); (3) Information structure and discourse particles (two papers); and (4) Information structure and language contact, an area of particular interest to the readers of this journal (two papers).

Although the title of the book states that it deals with spoken languages, three papers, namely Pan's paper on Mandarin, Etxepare's paper on Basque, and Bentley, Ciconte, Cruschina and Ramsammy's paper on Italian dialects (not dialects of Italian), are entirely based on sentences constructed for the purpose of investigation. While constructing sentences for syntactic investigation can be a legitimate method in the study of rules of sentence formation, such methodology has little value for the study of information structure, i.e. the domain that, by its very scope, should study the speaker's formulation of utterances to address the listener's state of knowledge. Similarly, the paper by Shimojo on Japanese is based on written sentences elicited with the help of pictures and a film clip. With respect to the issue of information structure it represents an artificial set of data, especially when the subjects are told to produce 'one sentence per picture' (Shimojo, 2016: 63). Natural discourse does not always contain complete sentences, even if the entity 'sentence' can be postulated for the language in question (Mithun, 2004).

Of particular interest to readers of this journal are the effects of language contact on information structure. Two papers deal with information structure in languages that emerged as products of language contact. One is Fida Bizri's study of Pidgin Madam, spoken by domestic workers of Sri Lankan origin in Lebanon, and the other is Peter Sломanson's study of Sri Lankan Malay. Bizri's study is based on a very large corpus of natural spoken language. Language data in Sломanson's study of the role of information structure in the restructuring of language through language contact consist of a few sentences, whose sources are not given, and their permutations, and do not appear to document natural speech that takes into consideration the listener's cognitive state.

Bizri's analysis of Pidgin Madam deals with three formal means: a prosodic marking, whereby one part of the utterance, here called form A, is separated through intonation and a short pause from another part, here called B. This formal means, which is frequently encountered in other languages and includes left detachment in French, gives the listener a wide range of possible

inferences with respect to the relationship between the two parts of the utterance. The second formal device is the reiteration of the first formal means within the utterance, such that the sequence A B is followed by another sequence A B. The third formal device, which Bizri calls *verum focus* after Höhle (1988), is the expression *bade*, lit. 'I want' (borrowed from Arabic), which to this reader appears to be an intensifier or a focus marker with the whole proposition in its scope. The interest of Bizri's study is that all three formal means are innovations in Pidgin Madam, in comparison with Sinhala, the native language spoken by speakers of Pidgin Madam.

In my discussion below I focus on findings from natural rather than elicited data pertaining to the issues of what belongs and what does not belong to information structure, the issues of topicalization and focus, and the systems of reference.

1.2 *A Few Questions Regarding Information Structure*

One of the necessities with respect to the study of information structure is to delimit its boundaries and, more specifically, to provide an explicit set of criteria that distinguish information structure from other functional domains in the language. A different but related question is how to determine what form-function pairing in the language is a part of information structure and what is not a part of information structure. The papers in the volume under discussion are mute with respect to these two questions.

For many years the study of information structure has assumed that all languages code the functions theme and rheme, or topic and comment, and the function of focus. Consequently, the research questions have focused on discovering how languages code these categories. The increase in the number of languages studied, along with more nuanced analyses, has increased the number of functions to include 'narrow' and 'broad' focus, contrastive topic or contrastive focus, anti-topic and a few others. Some linguists, especially those working on natural discourse data from a wide variety of languages, have increasingly become aware that these categories do not necessarily occur in all natural languages, and that the established categories do not account for all of the functions in the domain of information structure that are actually coded across languages (cf. Matic' and Wedgewood, 2013).

In the description of individual languages, the theoretical question with respect to information structure is not how a given aprioristic function is coded but rather what functions in the domain of information structure are actually coded in individual languages; what are the properties and syntactic implications of the categories coded; and what kind of evidence one can provide for the existence of a given function. The answer to these questions requires rigorous methodologies in the discovery of functions. Once the

methodological questions are resolved, they will provide a non-aprioristic tool to discover functions and to construct a typology of functions within the domain of information structure. The volume edited by Fernandez-Vest and Van Valin presents a number of interesting studies, based on a variety of languages, that describe phenomena that do not always fit neatly into the dichotomies of topic-comment, theme-rheme, or focus-comment, and leave many questions open with respect to reference systems.

One of the basic questions with respect to information structure is what constitutes part of the information structure and what does not. If one accepts Lambrecht (1994) and Van Valin (2016), who incorporates Lambrecht's approach to information structure in his Role and Reference Grammar, every utterance has some elements of information structure. As Van Valin presents it, Lambrecht distinguishes between predicate focus, sentence focus and narrow focus. An example of predicate focus is the following answer to the question 'What happened to your car?':

My car broke DOWN.

According to Van Valin, predicate focus corresponds to the traditional topic-comment distinction.

Sentence focus, according to Van Valin, is exemplified by the following answer to the question 'What happened?':

My CAR broke down.

Narrow focus is exemplified in answer to the question 'Did your motorcycle break down?'

No, my CAR broke down or
It's my car that broke down

(VAN VALIN, 2016: 79)

The approach to information structure that Van Valin advocates thus assumes that every utterance contains some function belonging to the domain of information structure. Empirical facts across languages do not support this assumption. For example, an indicative sentence may have a topic, but does not have to. The evidence is provided by languages that have markers of topicalization and where these markers are used only when topicalization is involved. Consider Mina, a Central Chadic language, which has a variety of forms coding topicalization.

In Mina, the discourse topic is typically established at the beginning of the discourse. Nouns that are marked as discourse topics have not been mentioned before in discourse. The main topicalizing means is the use of the demonstrative *wa*, or its phrase-final form *wàcín*, following the noun. The two clauses below are equational. Both are the first line of a story and set the topic of the narrative:

- (1) *ngàlámbar wàcín ngàlámbar tá kwáyàn*
 story DEM story GEN squirrel
 'This story is the story of the squirrel.'
- (2) *hìd-yû wà í tàtà màkád*
 man-PL DEM 3PL 3PL three
 'There were three men.'

A sentence without a topic marker does not code a topic-comment distinction and also does not constitute predicate focus (Mina has a separate means of coding focus on the predicate. Focus on the predicate has different forms, depending on the aspect of the clause. In both past and present, focus on the predicate is coded by the dependent aspects.):

- (3) *í ndà ká bèr-é cìkíd' bùhù ntá*
 3PL go INF sell-GO sesame bag (F.) one
 'They were going to sell one bag of sesame seeds.'

Topicalization of the subject in Mina may also involve use of the full noun and a subject pronoun coding the number of the noun, the pronoun which precedes the verb. The following example contains both means of topicalization: *wèhín* (variant of *wàcín*), in the matrix clause, and the subject noun followed by the subject pronoun in the embedded clause. The first topicalized constituent, *hìdì wèhín* 'this man' refers to one of three men who are protagonists in the story. It is a topic of a paragraph. Other paragraphs dealing with other protagonists have their own topics. The noun *ván* 'rain' followed by the subject pronoun represents the topicalization of the subject:

- (4) *hìdì wèhín à zá ván á n ká*
 man DEM 3SG COMP rain 3SG PREP INF
dā á gār kà nd-á-k kàsám skù
 fall 3SG want INF touch-GO-1SG body NEG
 'This man said, "Rain, when it falls, will not touch me."'

Proper names, even when coded by common nouns, as well as titles and the name for God, cannot be followed by the demonstrative *wa*. In such a case, the only available marker of topicalization of the proper name is the third-person pronoun preceding the verb. This is illustrated in the following example in both the main and the embedded clause:

- (5) *mímèŋ à zá àmmá bìḡán à*
 leopard 3SG COMP truly God 3SG
mbál-á-kù nd-á ḡì gwád á
 like-GO-1SG go-GO meat plenty PRED
bíŋ nàŋ
 room 1SG
 'Leopard [proper name for a protagonist in the story] said, "God truly loves me, as there is a lot of meat in my room."'

Topicalization of a pronominal subjects can be coded through several means. For the third-person plural subject, topicalization is coded by the independent pronoun *tàtà*, realized as *tàtà* in phrase-internal position:

- (6) *tàtà gwád á bíŋ á màcín*
 3PL plenty PRED room PRED DEM
 'As far as they are concerned, they are numerous in that room over there.'

Compare the non-topicalized subject pronoun, whose form is different from the independent pronoun. In the following sentence none of the components is marked as topicalized:

- (7) *ì gwád á bíŋ á màcín*
 3PL plenty PRED room PRED DEM
 'They are numerous in that room over there.'

The pronominal subject can also be topicalized by the deduced reference marker *ta*. The subject pronoun must be used again before the verb:

- (8) *hà tán tùwár á nà fálà tàtàn*
 2SG DEM suffer PRED PREP among 3PL
ḡìŋ wà ḡàzàd tótàn vù
 time DEM work good Q
 'You suffer a lot among them, is it then a good work?'

Topicalization of the direct or indirect object is marked through the placement of noun phrase before the verb (topicalized noun phrases are bolded):

- (9) à zá ngùl-yû **ǵámáy** tá màcín
 3SG COMP husband-PL stick GEN DEM
 lùw-á-ŋ màk
 say-GO-3SG would you
 ‘She said, my husband, this stick, talk to it, would you’
- (10) kwík **kwàlkwàl-yû** bákàhà ká màl tà
 kwik leper-PL today INF seize 3PL
 páť í n kà rèh-é í
 tomorrow 3PL PREP INF escape-GO PREP
 tiki nók tìyú
 where 1PL.INCL see:3SG
 ‘Lepers. Today we will catch them. Tomorrow we will see from
 where they will escape.’

Mina has different means of topicalization for different arguments, and for different classes of nouns.

1.3 *Theoretical Status of Focus*

Lambrecht’s proposal, as presented by Van Valin, implies that answers to questions of the type ‘what happened’ represent sentence focus. The issue here is whether indeed all such answers represent sentence focus. The purpose of what follows is to demonstrate that the focus function is independent of questions and answers to questions, and that questions and answers do not necessarily imply focus. The scope of focus in a given language is determined by the way focus is marked in that language. The general principle of the present approach is as follows: If the language has a means of marking focus, and if a question can, but does not have to, be marked by this means, then question represents one function and focus represents another function. The same is true for the answer to a question: If the answer can, but does not have to, be marked for focus, then answer and focus represent two separate functions.

Consider Wandala, a Central Chadic language (Frajzyngier, 2012). In the pragmatically neutral clause, the subject pronoun must precede the verb in some aspects and must be suffixed to the verb in other aspects, while nominal arguments follow the verb. Focus on any constituent is marked by the particle *bà*, which precedes the focused constituent. Answers to some questions do

not have the focus marker, while answers to other questions do have the focus marker. Here is an example of the question-answer sequence without the focus marker:

- (11) kwá tà-tèyà wà kùrè
2PL search-search WHAT 2PL
'What are you looking for?'

- (12) àa, ɲà bàkà ɲgè
well 1PL.EXCL make hunt
'Well, we are hunting.'

Here is another example where the question is followed by an answer without the focus marker:

- (13) kùr kwà sú hùn kwà sè fì-yà źw mtú
2PL 2PL come here 2PL come bother-1SG bad lack or
kwà s kàtà-w kùrè
2PL come want-what 2PL
'You, you came here to bother me, or what is that you want?'

- (14) hà bà màl-á-tàr nàwá ɲrè à fà-ɲr-à-r nàvìrè úksòɲìrè
ha say chief-GEN-3PL COM 1INCL 3SG put-1INCL-ON-ADD.M rabbit idiocy
'Ha, said their chief, it is the rabbit that cheated us.'

The focus marker may be used in answers, if the speaker wants to code focus. In the following exchange, the focus marker *bà* precedes the subject pronoun and has in its scope the whole proposition:

- (15) má màgà nó stàrà é únù wá
1INCL:FUT do PREP how eh DEM COM
'So, what are we going to do, and this

- (16) è šàgrà bà má wòvà dàlyè mà mà šà-n-š
eh good FOC 1INCL try again HYP 1INCL find-3SG-find
ùml ár gè ɲàné
other ON TO 3SG
'Well, we have to try again, to see if we can find others in addition to that one.' (data from Frajzyngier's 2013 fieldwork)

Similarly, in some studies concerning focus, including studies in the volume under discussion, it is assumed that content questions are automatically focused (as in the traditional analysis of Basque, as reported in Etxepare, 2016: 122). Again, Wandala provides the evidence that some content questions are focused and others are not. Wandala has a contrastive focus marker *kɛ̀nì* or *kánà*, whose function is to contradict the listener's presupposition. This marker may be used with the content question marker, in this case *ázàr* 'what':

- (17) *ázàr kɛ̀nà nùbá n gə̀ k á-m úunà*
 what C.FOC part DEM TO 2SG PRED-IN DEM
 'What is your part in it?' (said Hyena)

- (18) *ɓákà nùb-á-rù stàrà bà nàvirè*
 NEG.EX part-GEN-1SG how come say rabbit
 "How come there is no part for me?" said the rabbit

The answer to such a question does not have to have a focus marker:

- (19) *yò í s màrà-kà né nì ɓákà nùb-á-rù*
 well 1SG come show-2SG DEM INTENS NEG.EX part-GEN-1SG
á-m kwà
 PRED-IN Q
 'It is me who came to show you, and there is no part for me?'

- (20) *á ɓákà*
 ah NEG.EX
 'Ah there is not.'

The question marker *ázàrà* can occur without the contrastive focus marker:

- (21) *ázàrà dàbàrì-á-kùrè*
 what reason-GEN-2PL
 'What is your reason?'
- (22) *à dàbàrì-á-ɲr áɲkwà*
 ah reason-GEN-1INCL exist
 'Ah, we have our reason.'

2 Detachment Constructions: To What Domain They Belong?

The theoretical part of the volume opens with a programmatic paper by Fernandez-Vest, who postulates something called Detachment Linguistics, which is simply a study of detachment constructions. No new theoretical principles or methodological procedures are postulated. Fernandez-Vest postulates that dialog should be the primary source of data in the study of information structure. The Initial Detachment and Final Detachment are essentially language utterances consisting of two parts, where one part is not a syntactic component of another part, e.g. French:

- (23) Les fenêtres, elles ne s'ouvrent pas, señor
 DEF windows 3PL.F NEG REFL-open:PL NEG, señor
 'The windows, they do not open, señor.'

and

- (24) Ça part vite, ces machins-là
 DEM go fast DEM.PL thing.PL-DEM.LOC
 'It goes off quickly, these things' (Fernandez-Vest, 2016: 11, glosses mine)

Fernandez-Vest analyses these constructions within the framework of theme and rheme contrasts, essentially not too different from the topic-comment functions. Such constructions are very frequent in spoken French, and they have been observed in other languages. Fernandez-Vest asserts, correctly, that detachments are 'essential elements of oral grammar' and that they play a role in 'situation anchoring, relevance and planning of the conveyed information' (Fernandez-Vest, 2016: 30). While all of these are reasonable functions, Fernandez-Vest does not offer a methodological framework to discover the functions of detachments and the kind of argumentation that one can provide for a given hypothesis. Consider the following examples, all of which contain the first-person pronoun *moi* in the 'detached' part:

- (25) Moi c'est yoga et muscadet bio.
 1SG DEM:COP yoga CONJ muscadet organic
 'Me, it's yoga and organic muscadet' (Gourio, 2015: 157)

One can analyze the function of *moi* '1sg' as theme or topic, but that would be an arbitrary assignment of a function. The speaker of French has several other forms that she/he could have chosen, including *pour moi* 'for me':

- (26) Pour moi, les municipales, c'est pas de la politique
 FOR 1SG DEF:PL municipal DEM:COP NEG ART DEF politics
 'For me, municipal elections, it's not [real] politics' (Gourio, 2015: 157)

Here, the topic is most probably 'municipal elections', and the comment is 'it is not politics'.

Topicalization, however, is not necessarily a function of the detachment constructions. An utterance can have several detachment constructions, but only one noun in those constructions can serve as a topic:

- (27) Moi ma mère elle a bercé toute ma
 1SG 1SG.F mother3F AUX cradle:PTCPL.PAST entire:F 1SG:F
 jeunesse en me tirant sur les cheveux.
 youth by 1SG.OBJ pull:GER ON DEF:PL hair
 'Me, my mother, she cradled me all throughout my youth by pulling on my hair' (Gourio, 2015: 144)

Without strict theoretical and methodological criteria, one could arbitrarily assign the role of topic to *moi* 'me', *ma mère* 'my mother', or the way of cradling. The point here is that one cannot assign the function of a given form by mere analysis of the inference from a given sentence or a given fragment of discourse. Fernandez-Vest discusses the similar and famous expression by F. Hollande: *Moi Président de la République* 'Me, as the president of the republic' (lit. 'Me, the president of the republic'), but the discussion consists of inferences drawn from the individual sentences rather than a methodology of discovery of the function of left and right detachment. The discovery of the function should take into consideration all formal means available in the language, discover what functions they code, discover which functions belong to which domain, and then describe each function in relationship to other functions. Fernandez-Vest rightly states that detachment constructions in French constitute an area in need of analysis. It may be necessary to include in the study constructions that are syntactically integrated with the clauses where they occur but which leave the inferences about semantic connection to the listener's knowledge of the non-linguistic universe, as in the following example:

- (26) Mon mari est mort, c'est pour ça que
 1SG.POSS husband be:3SG.PRES dead DEM-be.PRES for DEM COMP
 je vais plus en Bourgogne,
 1SG go anymore in Burgundy

c'était lui la Bourgogne.
 DEM-be.PAST 3SG DEF.F Burgundy
 'My husband is dead, and that's why I don't travel anymore to Burgundy.
 Burgundy was his thing' (lit.) 'It was him – Burgundy.' (Gourio, 2015: 214)

One of the questions is whether the detachment constructions even belong to the domain of information structure or whether they perhaps belong to some other domain, such as changing the perspective of the conversation, e.g. from listener or third person to the speaker, or from speaker to second or third person. Whether or not the functions are differentiated should be tested by the co-occurrence, or lack of co-occurrence, of various coding devices for information structure within the same utterance, clause, or sentence.

The study of Finnish detached noun phrases, by Ritva Laury and Marja-Liisa Helasvuo, in fact demonstrates that detached constructions in Finnish, involving relative clauses, are coding means to assure referent continuity. They appear to be coding means within the system of reference rather than within the system of information structure. In Finnish, one cannot establish whether the construction is left detached or right detached.

It is not clear at all that the French left and right detached constructions function within the system of information structure. Function(s) of these constructions have yet to be discovered.

One of the open questions with respect to detachment constructions is whether there any constraints on detachments; in particular, whether there are languages where detachments are more frequent and whether there are languages where detachments do not occur at all. The obverse side of Fernandez-Vest's study implies that literary French, whether spoken or written, has fewer detachment constructions than non-literary spoken French.

3 The Hindi (and Polish) Particle *to*. What Domain Do They Belong To?

Annie Montaut's paper deals with the particle *to* in Hindi, which is variously described as a conjunction and a discourse particle, and which is translated into English by a wide variety of forms such as 'sure, well, at least, finally, will you, but' (Montaut, 2016: 263). This particle in Hindi can occur at the beginning of the clause, or it can occur alone. The widely different translations indicate that this particle has multiple functions. In fact, one could reach such a conclusion if one were to take inferences from individual sentences to be

the evidence for different functions. Montaut aims at finding an overarching function of the particle *to*, which, in different syntactic positions and in correlation with different categories, triggers the different inferences.

At the beginning of the clause, *to* functions as a conjunction, but not as a coordinating conjunction. It introduces ‘a new turn or phrase of the story in narrative contexts’ (Montaut, 2016: 264). When the particle *to* follows a noun in clause-initial position, it functions as complementizer. The other functions of *to* that Montaut describes are: constructing the theme as contrastive, essentially a combination of topic and focus; contrasting one proposition with another (‘re-qualification of an already mentioned term’ (Montaut, 2016: 272); and an argumentative particle with a variety of uses. Montaut summarizes her findings by claiming that the diversity of functions of *to* result from its abstract function: ‘*to* seems to convey the speaker’s judgment on a term or sequence on which it has scope, in such a way that *to*-P triggers the implicit or explicit representation of P’ (P: non P or other than P)’ (Montaut, 2016: 279).

The reason I have reported at length on the Hindi particle *to* is that it does not appear to fit well into the traditional functions within the information structure. The description of *to* raises thus the question to what functional domain the function of *to* belongs. The other reason for which I have reported on the Montaut study is that, for the majority of examples that she cites from Hindi, the Polish equivalent also has the particle *to*, and in the same position in which the Hindi particle occurs. The only difference is that the Polish *to* cannot occur as the only constituent of a clause. Like Hindi *to*, Polish *to* also has a demonstrative origin, and the functions of Polish *to* are very similar to the functions of Hindi *to*. There are too many points of similarity for it to be accidental.

An alternative possibility for the function of the particle *to* in both languages is that it marks a comment clause (Frajzyngier, 2010). The comment can be on an individual lexical item, resulting in topicalization inference; on what the speaker said before; or on what another speaker said. In all cases, the clauses marked by *to* are pragmatically dependent, i.e. have to be interpreted in connection with some other lexical item or a proposition.

4 Topicalization

In the next three sections I raise some questions and potulates for the non-aprioritic approach to the issues of topicalization, focus, and reference.

The following are some of the questions with respect to topicalization:

- (1) Does the language under investigation have a topicalization function distinct from other functions within or outside the information structure? Japanese is one of the languages that have traditionally been analyzed as having the category topic, marked by the particle *wa* following a noun phrase. So here at least we have a formal means of topicalization. Shimojo's study does not challenge the topicalization function of *wa*. Similarly, Mina, as illustrated in the present study, has a distinct formal means of coding topicalization. So far, we do not have an explicit statement to the effect that a given language does not have a topicalization construction.
 - (2) What type of syntactic unit is in the scope of topicalization, e.g. narrative, paragraph or a similar unit, sentence? Here the available information is not sufficient for a large number of languages. The English phrases 'speaking of ...', 'in what concerns X ...', appear to introduce topics of units potentially larger than a sentence. In Mina, the demonstrative *wà* after the noun phrase in clause-initial position can introduce topic of the narrative.
 - (3) Does any clause have to have topic? This is an open question in linguistic theory and many descriptive grammars. If a clause does not have a topic marker or a comment clause marker (Frajzyngier, 2010), then indeed there is no basis for postulating that an indicative clause has topic and comment constituents. In Lele (East Chadic) there appears to be no means of topicalization of the nominal subject. Postulating the category topic with the nominal subject in clause-initial position would then be motivated only by the tendency to look for a topic in any indicative sentence, as is done in Frajzyngier (2001). There are morphological means of topicalizing the pronominal subject. There are also means of topicalizing the object, by its position at the beginning of the clause and by the definite marker following it:
- (28) è-gé cáaní **bayndí-ŋ** kalo se yír-íy
 go-3PL bush man-DEF snake INCEPT bite-3M
 'When they went into the bush, a snake bit the man.'
- (4) Can a clause have more than one topic? If indeed 'aboutness' is to be retained as one of the functional features of the topic, then it is difficult to conceive in what way a sentence may have several topics. Etxepare (2016: 123) states that 'All elements to the left of the preverbal focus constituents are interpreted as topics'. It would be interesting to know what is the

understanding of the term topic, and who is interpreting the constituents preceding the focus as topics.

5 Focus

Some of the open questions with respect to focus are:

- (1) How the coding of focus differs from the coding of other functions within the domain of information structure or indeed from other functions coded in the language. Several studies in the volume explicitly state what are the formal means that carry the function of coding focus. This is the case with Van Valin (2016) for three Amazonian languages: Forker and Belaev demonstrate the use of interrogative particles as focus markers in Nakh-Dagestanian languages; Etxepare 2016 describes that the preverbal position of an argument in Basque is considered a focus marker; and Sahkai (2016) demonstrates the use of demonstratives as predicate and narrow focus markers in Estonian. Matic (2016) demonstrates the grammaticalization of tag questions into focus markers in Even.
- (2) What types of functions or constituents can be in the scope of focus, e.g. noun phrase, verb or verb phrase, adverb? It appears that in languages that code focus, any constituent can be marked for focus. If one wants to postulate the existence of sentence focus, the formal means of coding sentence focus must be stated and contrasted with sentences which have no sentence focus.
- (3) What are the systemic consequences of coding focus? Some languages make a distinction between pragmatically dependent and pragmatically independent clauses. A pragmatically dependent clause must be interpreted in connection with some other clause or in connection with some event, while a pragmatically independent clause may be interpreted on its own. In English, clauses marked by the complementizer 'that' are pragmatically dependent.

Scholars of African languages have noticed the presence of two aspectual and tense systems (Gouffé, 1966, 1967, 1968; Hyman and Watters, 1984; Caron, 1986 and 2000; Robert, 1991; Pawlak, 1993; Jungraithmayr, 1994 and references there; Newman, 2000 and numerous other grammars of Chadic languages e.g. Frajzyngier with Shay, 2002; Frajzyngier and Johnston, with Edwards, 2005; Frajzyngier, 2008). In some Chadic languages, tense and aspectual systems mark the distinction between pragmatically dependent and pragmatically

independent clauses. The dependent aspectual system is used in comments on the constituents in focus, in clauses with content questions, and in relative clauses, but not with comments on a topicalized constituent or with polar questions. Here is an example of the use of the dependent completive aspect labeled ‘preterit’ in Newman (2000):

- (29a) *jìyà sun sana-ř dà mu*
 yesterday 3PL:COMPL know-CAUS ASSC 1PL
 ‘Yesterday they informed us.’

- (29b) *jìyà suka sana-ř dà mu*
 yesterday 3PL:PRET know-CAUS ASSC 1PL
 ‘Yesterday they informed us.’ (Newman, 2000: 572, glosses Z.F.)

The pragmatic dependency of the comment-on-focus clause lies in the fact that the focused element may be conceived as a predication of its own, as evidenced by the frequent grammaticalization of the focus marker as a clausal expression corresponding to ‘it is X that ...’

The importance of the pragmatic dependency linked with focus marking is that in languages in which this distinction is marked by tense and aspectual systems, focus is a category distinct from topic, and the comment on focus is distinct from the comment on topic.

- (4) Can there be more than one focus within a given clause or sentence? Recall that the question of whether there can be more than one topic in the clause had to be answered in the negative. With respect to focus, our answer must be much more nuanced. It appears that when focus is on arguments or adjuncts, several foci are possible within one clause, as in this example from Polish:

- (30) *ty mnie, na polskiej ziemi, polskiego*
 2SG:NOM 1SG:ACC ON Polish:ADJ:F soil Polish:ADJ:M:GEN
słowa zabronisz
 word:GEN forbid:FUT:2SG
 ‘Are you going to forbid me the use of a Polish word on Polish soil?’
 Sources for the Polish Frequentative Dictionary.

In Kapsiki (Central Chadic) there exists a particle *náx* (*náxà*) whose function appears to be that of coding focus on affirmative modality (all data on Kapsiki from Frajzyngier field notes):

- (30) *náxə zrá kə dzí šəkw*
 AFF Zra R.PAST go market
 'Zra went to the market'

The affirmative marker may be omitted, in which case the subject occupies clause-initial position:

- (31) *zrá kə dzí šəkw*
 Zra R.PAST go market
 'Zra went to the market'

The marker *náx* may occur in polar questions:

- (32) *náx zrá kə sə li šəkw ɓəší mášikə ná*
 AFF Zra R.PAST be:PAST PREP market today morning PQ
 'Did Zra go to the market this morning?'

The evidence that *náx* is the focus on affirmative modality marker is provided by the fact that it cannot occur in negative clauses:

- (33a) *kə dzí šəkw zrá w*
 R.PAST go market Zra NEG
 'Zra didn't go to the market'
- (33b) **náx kə dzí šəkw zrá w*
 AFF R.PAST go market Zra NEG
 'Zra didn't go to the market'

The marker of focus on affirmative modality cannot occur in content questions:

- (34) *tə má sə ná wà*
 PREP there be:PAST 2SG CQ
 'where were you?'

The focus on affirmative modality cannot occur in imperative modality:

- (35) *təkə ɣná psikí*
 learn language psiki
 'learn the language of Psiki!' 'learn Kapsiki'

Focus on arguments is marked by a construction corresponding to ‘it is X that ...’. The relevant construction in Kapsiki has the form ‘NP rel’, where the relative marker is *ndə* for the singular head and *ká* for the plural head:

- (36) *zrá ndə pá-v hlá*
 Zra REL.SG buy-VENT cow
 ‘It is Zra that bought a cow’ [and not somebody else]

The data indicate that, in Kapsiki, one cannot have two foci in different domains in one clause. This is evidenced by the fact that if an argument is in focus, the focus on affirmative modality cannot be added:

- (37) **náx zrá ndə pá-v hlá*
 AFF Zra REL.SG buy-VENT cow
 ‘It is Zra that bought a cow’

Questions about the subject are marked in the same way as the argument in focus, viz. the question word is followed by the relative marker:

- (38) *wá ndə dzí šikw wà*
 who REL.SG go market CQ
 ‘who went to the market’

The focus on affirmative marker cannot be added to such a clause.

An interesting question for cross-linguistic study is whether the constraint on two foci in different domains obtains in other languages.

6 The Reference System

Systems of reference remain one of the areas in need of thorough descriptive and cross-linguistic studies. It is quite telling that only two categories of the reference system, definite and indefinite articles, are investigated in WALS (Dryer, 2005). Given the fact that only a fraction of languages code the distinction between definite and indefinite marking, there is a need to come up with theoretical and methodological approaches that will not be based on the categories of Western Indo-European languages and that could serve in the analysis of a large number of languages without imposing categories that do not exist in individual languages.

Several papers in the present volume explicitly deal with the system of reference. Gundel and Hedberg's programmatic paper focuses on the cognitive status of referents and on defending the Givenness Hierarchy against some past criticism. Gundel and Hedberg's assumption is that the Givenness Hierarchy is universal and that elements of the reference system follow the Givenness Hierarchy, which has six cognitive statuses where each status entails lower statuses' 'in focus > activated > familiar > uniquely identifiable > referential > type identifiable here illustrated in English:

In			uniquely		type
Focus >	activated >	familiar >	identifiable >	referential >	identifiable.
It	this/that/this NP	that NP	the NP	indefinite this NP	a NP

The term 'it' represents an unstressed pronoun. (Hedberg and Gundel, 2016: 35)

The six cognitive statuses are also supposed to 'specify procedural information to the addressee about how to mentally access representations associated with the intended interpretations of nominal expressions' (Gundel and Hedberg, 2016: 50). There are two components to this description. One component is the cognitive state of listeners, which may or may not exist. Whether such cognitive statuses exist should be subject to testing through the methods that study cognitive states of speakers. The other component, viz. the 'procedural information to the addressee about how to mentally access representations associated with the intended interpretations of nominal expressions', is a component that can and needs to be tested through linguistic methods, and more specifically through the examination of the formal means existing in individual languages and functions coded by these means. What remains to be studied, with respect to the reference systems, is what kind of functions are encoded in individual languages, and what functions are encoded more often and what functions are encoded less often across languages. The study of the functions encoded in the reference system obligatorily involves the study of the formal means deployed in the coding of reference. Here is an illustration of the formal means and the structure of the reference system in Polish:

Formal means of coding within the system of reference:

Gender system: Three genders in the singular: masculine, feminine, and neuter; two genders in the plural: human masculine and all other nouns.

Number system: Singular and plural (number plays a role in the reference system)

Bare nouns

Pronouns

'Proximate' demonstratives and determiners of the series *t*: *ten* (m.), *ta* (f.) *to* (n.) *ci* (human m.pl.), and *te* (all others).

'Agreement' on verbs, numerals, adjectives

Remote demonstratives and determiners: *tam-ten tam-ta, tam-to* ('that one there')

Place: *tu* 'here', *tam* 'there'

Demonstratives and determiners of the series *ow*: *ów* (m.), *owa* (f.), *owo* (n.). *owi* (pl.m. human) and *owe* (all others).

Unknown reference markers of the series *jakiś* 'some' (m.), *jakaś* (f.), *jakiś* (n.pl., f.pl.), *jacyś* (human m.pl.)

These formal coding means interact in the coding of the following functions:

Here is an explanation of the various functions: *coref.* stands for co-reference and instructs the listener to identify the noun through coreference with the previously mentioned noun. Obviative reference tells the listener that the noun is necessarily different from the previous mentioned noun. Deduced reference instructs the listener to identify the noun through analysis of the proximate context in space and time or the remote context in space and time. The term 'time' refers here to proximate and remote mention in discourse.

Other languages have different coding means and code different functions. For example, a number of Chadic languages, such as Mupun (Frajzyngier, 1993) and Hdi (Frajzyngier with Shay, 2002), code the function ‘previous mention’, which is distinct from deixis and different from the category ‘definite’ in English and French. The structure of the reference system in Hdi is as follows:

Languages that have logophoric pronouns code additional functions in the reference system.

Examination of the formal means existing in individual languages demonstrates existence of significant differences across languages with respect to functions coded in the reference system.

7 Conclusions

Stemming from the Fernandez-Vest and Van Valin (eds.), 2016 volume, the present study raises the following questions:

- (1) How does one determine which function belongs to the domain of information structure and which one does not?
- (2) If an utterance does not code topic or focus, can one speak of the existence of those functions in the utterance?
- (3) Can a clause have more than one topic?
- (4) What is the scope of topicalization?
- (5) Can a clause have more than one focus?

The study also proposes that, in addition to the information structure, languages code other functions pertaining to the interaction of speaker and listener, functions that have yet to be clearly delineated. These were the functions documented in the rich and stimulating papers by Fernandez-Vest and Montaut in Fernandez-Vest and Van Valin (eds.) 2016.

The present paper also proposes separation of the domain of reference from the domain of information structure. Moreover, the study proposes that, just like in any other functional domain in language, the functions within the domain of reference must be discovered rather than asserted (Frajzyngier with Shay, 2016).

Acknowledgments

The data on which this study is based were gathered over the course of many years with the support of the National Science Foundation, the National Endowment for the Humanities, and the University of Colorado. Data for Kapsiki come from my work with Deli Basil, born in Sir, Cameroon. I am grateful to Bernard Comrie for an important comment regarding terminology, and especially to Erin Shay for her comments on the substance and form of this paper.

References

- Bentley, Delia, Francesco Maria Ciconte, Silvio Cruschina, and Michale Ramsammy. 2016. Micro-variation in information structure: Existential constructions in Italo-Romance. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.),

- Information structuring of spoken language from a cross-linguistic perspective*, 95–120. Berlin: De Gruyter Mouton.
- Bietak, Manfred, Johanna Holaubek, Hans Mukarovsky, and Helmut Satzinger (eds.). 1994. *Zwischen den Beiden Ewigkeiten. Festschrift Gertrud Thausing*. Wien: Institute für Ägyptologie der Universität Wien.
- Bizri, Frida. 2016. Information structuring strategies in Pidgin Mandam. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 285–303. Berlin: De Gruyter Mouton.
- Bril, Isabelle (ed.). 2010. *Clause Linking and Clause Hierarchy. Syntax and pragmatics*. Amsterdam: Benjamins.
- Caron, Bernard (ed.). 2000. *Topicalisation et focalisation dans les langues africaines*. Louvain-Paris: Peeters.
- Caron, Bernard. 1986. Les accomplis I et II du haoussa et la subordination. Aspects, modalité: Problèmes de catégorisation grammaticale. *Collection ERA* 642. Université Paris 7, 103–120.
- Caron, Bernard. 2000. Assertion et préconstruit: topicalisation et focalisation dans les langues africaines. In Bernard Caron, *Topicalisation et focalisation dans les langues africaines*, 7–44. Louvain-Paris: Peeters.
- Dryer, Matthew. 2005. Definite articles. In Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, with the collaboration of Hans-Jörg Bibiko, Hagen Jung, and Claudia Schmidt, *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Ettxepare, Ricardo. 2016. How does adjacency arise? Grammatical condition on focus-verb adjacency in Basque. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 121–147. Berlin: De Gruyter Mouton.
- Fernandez-Vest, M.M. Jocelyne. 2016. Detachment linguistics and information grammar of oral languages. Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 7–32. Berlin: De Gruyter Mouton.
- Fernandez-Vest, M.M. Jocelyne and Robert D. Van Valin, Jr. (eds.). 2016. *Information structuring of spoken language from a cross-linguistic perspective*. Berlin: De Gruyter Mouton.
- Forker, Diana, and Oleg Belaev. Word order and focus particles in Nakh-Dagestanian languages. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 239–261. Berlin: De Gruyter Mouton.
- Frajzyngier, Zygmunt. 2001. *A Grammar of Lele*. Stanford Monographs in African Linguistics. Stanford: CSLI.

- Frajzyngier, Zygmunt. 2004. Tense and aspect as coding means. In *Journal of West African Languages* 30(2): 53–67. Special issue, ed. by Bernard Comrie and Ekkehard Wolff.
- Frajzyngier, Zygmunt. 2008. *A Grammar of Gidar*. Frankfurt: Peter Lang.
- Frajzyngier, Zygmunt. 2010. A comment predication: crossing the boundaries between the simple and the complex sentence. In Isabelle Bril (ed.), *Clause-hierarchy and Clause-linking: the Syntax and pragmatics interface*, 313–332. Amsterdam: Benjamins.
- Frajzyngier, Zygmunt. 2012. *A grammar of Wandala*. Berlin / New York: Mouton de Gruyter.
- Frajzyngier, Zygmunt, and Eric Johnston, with Adrian Edwards. 2005. *A Grammar of Mina*. Berlin / New York: Mouton de Gruyter.
- Frajzyngier, Zygmunt, with Erin Shay. 2002. *A Grammar of Hdi*. Berlin: Mouton de Gruyter.
- Frajzyngier, Zygmunt, with Erin Shay. 2016. *The role of functions in syntax: a unified approach to language theory, description, and typology*. Amsterdam: Benjamins.
- Gouffé, Claude. 1966. Les problèmes de l'aspect en haoussa. *Compte Rendu du Groupe Linguistique d'Études Chamito-Sémitiques*, GLECS 10: 151–165.
- Gouffé, Claude. 1967. Les problèmes de l'aspect en haoussa. *Compte Rendu du Groupe Linguistique d'Études Chamito-Sémitiques*, GLECS 11: 29–67.
- Gouffé, Claude. 1968. Les problèmes de l'aspect en haoussa. *Compte Rendu du Groupe Linguistique d'Études Chamito-Sémitiques*, GLECS 12: 27–51.
- Gourio, Jean-Marie. 2015. *Le petit troquet de brèves de comptoir*. Paris: Robert Laffont.
- Gundel, Jeanette K. and Nancy Hedberg. 2016. Reference and cognitive status: scalar inferences and typology. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 33–53. Berlin: De Gruyter Mouton.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, with the collaboration of Hans-Jörg Bibiko, Hagen Jung, and Claudia Schmidt. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Höhle, Tilman N. 1988. Vorwort und Nachwort zu Verum-Focus. *Sprache und Pragmatik* 5: 1–7.
- Hyman, Larry, and John Watters. 1984. Auxiliary focus. *Studies in African Linguistics* 15(3): 233–273.
- Jungrathmayr, Herrmann. 1994. 'Zweite Tempora' in Afrikanischen Sprachen-Ägyptisch-Tschadische Gemeinsamkeiten. In Manfred Bietak, Johanna Holaubek, Hans Mukarowsky, and Helmut Satzinger (eds.), *Zwischen den Beiden Ewigkeiten*. Festschrift Gertrud Thausung. Wien: Institute für Ägyptologie der Universität Wien.
- Lauri, Ritva and Marja-Liisa Helasvuo. 2016. Detached NPs with relative clauses in Finnish conversations. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin

- (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 149–165. Berlin: De Gruyter Mouton.
- Matić, Dejan. 2016. Tag questions and focus markers: evidence from Even. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 167–189. Berlin: De Gruyter Mouton.
- Matić, Dejan, and Daniel Wedgwood. 2013. The meaning of focus. The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49: 127–163.
- Montaut, Annie. 2016. The discourse particle to and word ordering in Hindi: from grammar to discourse. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 263–282. Berlin: De Gruyter Mouton.
- Newman, Paul. 2000. *The Hausa language*. New Haven / London: Yale University Press.
- Pan, Victor. 2016. Syntactic and prosodic marking of contrastiveness in oral Mandarin. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 191–210. Berlin: De Gruyter Mouton.
- Sahkai, Heete. 2016. Demonstratives and information structure in spoken Estonian. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 211–235. Berlin: De Gruyter Mouton.
- Schimojo, Mitsauki. 2016. Saliency in discourse and sentence form: zero anaphora and topicalization in Japanese. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 55–75. Berlin: De Gruyter Mouton.
- Slomanson, Peter. 2016. New information structuring processes and morphosyntactic change. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 305–325. Berlin: De Gruyter Mouton.
- Van Valin, Robert D., Jr. 2016. An overview of information structure in three Amazonian languages. In Jocelyne M.M. Fernandez-Vest and Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 77–92. Berlin: De Gruyter Mouton.

Abbreviations

1	first person
2	second person
3	third person

ACC	accusative
AFF	affirmative (focus)
ASSC	associative
AUX	auxiliary
C.FOC	contrastive focus
CAUS	causative
COMP	complementizer
COMPL	completive
COP	copula
CQ	content question
DAT	dative
DEF	definite
DEM	demonstrative
EE	end-of-event
EP	epenthetic
EXCL	exclusive
F	feminine
FOC	focus
FUT	future
GEN	genitive
GER	gerundive
GO	goal
HABIT	habitual
HYP	hypothetical
IMPF	imperfective
INCEPT	inceptive
INCL	inclusive
INF	infinitive
INTENS	intensive
LOC	locative
M	masculine
NEG	negative
NOM	nominative
OBJ	object
PAST	past
PL	plural
PQ	polar question
PRED	predicator
PREP	preposition
PRF	perfective

PROG	progressive
PTCPL	participle
Q	question
REFL	reflexive
RP	referential past
SG	singular
VENT	ventive