

The interaction between topicalization and structural constraints:

Evidence from Yucatec Maya¹

Stavros Skopeteas and Elisabeth Verhoeven

University of Potsdam, University of Bremen

1. Preliminaries

It has been cross-linguistically observed that topics occur sentence initially (see Vallduví 1992, Lambrecht 1994, Neeleman and Van De Koot 2008). It has also been shown that this generalization is functionally motivated: topics indicate the address in the common ground to which the asserted information is related, and it is an advantage for the hearer if this information is introduced early in the utterance (Clark and Clark 1977; Clark and Haviland 1977). This rationale motivates the linear precedence of topic constituents without referring to particular positions in constituent structure. In line with this generalization, recent research on the discourse configurationality of constituent structure has shown that many languages display a syntactic position that is reserved for topics and this position is typically at the beginning of the sentence (see, e.g., the syntactic analysis

¹ We are very grateful to Geertje Van Bergen, Gisbert Fanselow, Helen De Hoop, Christian Lehmann, and two anonymous reviewers for discussions about the grammar of Yucatec Maya as well as comments on previous versions of this article. Grateful thanks are due to our consultants, in particular Amedee Colli Colli, Ramón May Cupul, and Ernesto May Balam. The study on Yucatec Maya as well as the field work were supported by the research center 632 *Information Structure* (project D2, Typology of Information Structure) at the University of Potsdam for Stavros Skopeteas and by the Universities of Erfurt (2004) and Bremen (2008) for Elisabeth Verhoeven.

of Hungarian in É. Kiss 1998: 256). There is certainly an overlap between the generalization that topics precede comments in strict linear terms and the assumption of a sentence-initial position for topics in the constituent structure of particular languages and hence, it is reasonable to ask whether these statements are independent of each another. In particular, the question is whether (a) a particular syntactic position is associated with the information structural concept of topic or (b) is frequently (but not exclusively) used for the encoding of discourse topics, just because it has a linear position that is appropriate to satisfy cross-linguistic discourse preferences (such as the preference for topics to precede comments).

Mayan languages present an interesting case in this respect, since most of them display sentence-initial positions that bear particular morphological marking (right edge enclitics), and are traditionally called “topic positions”. The most elaborated account on the interaction between syntax and pragmatics in this language family has been proposed by Aissen (1992), who argues for a distinction between a clause external position for topics (i.e., external topics) and a position of topics that are moved out of the core clause (i.e., internal topics). Moreover, Aissen (1992) claims that this difference in constituent structure has implications for the information structural properties of topics. External topics are motivated in discourse by a marked contextual condition: a deviation from the expected pragmatic configuration, which is the topic continuity. By consequence, external topics are expected to occur in cases of topic shift, while internal topics are less restricted and may also constitute continuing topics.

Our article deals with Yucatec Maya, a Mayan language spoken in the Mexican states of Yucatán, Quintana Roo, and Campeche, as well as in neighboring parts of Belize and

Guatemala (700,000 speakers according to the 1990 census).² The canonical order in Yucatec Maya is V-initial (see Norman and Campbell 1978: 144; Lehmann 1990: 44, 2003: 28), as illustrated in (1). It is important to note that this order involves ambiguity. First, Yucatec Maya is a head marking language: a so-called set A clitic, which is attached to the aspect auxiliary in (1), cross-references the agent constituent of transitive verbs or the single argument of intransitive verbs in the incompletive status; a set B suffix is attached to the lexical verb and cross-references the patient constituent of transitive verbs and the single argument of intransitive verbs in the completive and subjunctive status (see details in Lehmann 2003: 21; Bohnemeyer 2004: sect. 3; Verhoeven 2007: 110). Crucially, dependencies are not coded through case affixes and the relative order of postverbal arguments is not rigid, which implies that two third person arguments are ambiguous (see also De Swart 2007: 86). There is a preference for the OS order, but this preference may be overridden by asymmetries in animacy, definiteness or weight (see Skopeteas and Verhoeven 2005; Gutiérrez Bravo and Monforte y Madera 2008b; Bohnemeyer 2008). Hence, the linear order of the arguments only gives a probabilistic cue for their function in the clause.

² The evidence presented in this article has been collected in interviews and experimental studies with native speakers living in Yaxley and Felipe Carrillo Puerto (Quintana Roo; Mexico), December 2004, August 2006, and March 2008.

- (1) *T-u* *hàant-ah* *òon* *Pèedróoh.*
 PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado Pedro
 ‘Pedro ate avocado.’³

The construction that is dealt with in this article is exemplified in (2). The sentence-initial constituent is accompanied by a right-edge clitic *-e* ‘D3’, which belongs to a class of enclitics that encode deictic concepts. Previous accounts on Yucatec Maya assume that the sentence initial constituent in this construction occupies a “topic position” (Lehmann 2003: 28, Bohnemeyer 1998a, 1998b).

- (2) *Pèedróoh-e* *t-u* *hàant-ah* *òon.*
 Pedro-D3 PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado
 ‘Pedro ate (an) avocado.’

The construction in (2) contrasts with another construction involving a preverbal constituent, which is exemplified in (3). The preverbal constituent in this construction is not enclosed by a right-edge clitic and the verb – under particular morpho-syntactic conditions, i.e., when the preverbal constituent is the agent of a transitive verb – appears in a special morphological form, which is characterized by the drop of the set-A clitic and the aspect/mood auxiliary and by special status marking. In the following, we assume that

³ A= person clitic, class A; AN= animate; B= person affix, class B; CL= class; CMPL= completive; CNJ= conjunction; D1= 1st person deixis; D2= 2nd person deixis; D3= 3rd person deixis; D4= locative/negative clitic; DEB= debitive; DEF= definite; F= feminine; HESIT= hesitant; IMM.FUT= immediate future; INAN= inanimate; INCMPL= incompletive; INTRV= introversive; IPFV= imperfective; M= masculine; PART= participle; PASS= passive; PF= perfect; PFV= perfective; PL= plural; PROC= processive; PROG= progressive; PRSV= presentative; REL= relationalizer; SG= singular; SUBJ= subjunctive; TERM= terminative; TRR= transitivizer; 0= meaningless element; 1= 1st person; 3= 3rd person.

the preverbal constituent occupies a focus position in this construction (see Lehmann 1990, 2003: 29; see also Bricker 1979; Bohnemeyer 1998b, 2008; Tonhauser 2003, 2007 for an analysis of the same structure as a cleft construction).

- (3) *Pèdróoh hàant òon.*
Pedro eat:TRR(SUBJ)(B.3.SG) avocado
'It was Pedro who ate (an) avocado.'

Aim of this article is to account for the discourse properties of the so-called “topic-position” introduced in example (2). In order to avoid circularity when using a pragmatic term for a syntactic entity, we refer to this configuration as left-dislocation. Section 2 outlines the syntactic properties of left dislocation and provides evidence that left dislocated material is clause external. Section 3 deals with the information structure of these constituents and shows that there are two independent motivations for left dislocation. Based on our data, it holds true that “topic → left dislocation”, but not that “left dislocation → topic”. The implications of these findings are discussed in Section 4.

2. Structural properties of left dislocation

All kinds of arguments and adjuncts may be left dislocated in Yucatec Maya (see examples of different arguments in (4), see also (11) and (14) below for left dislocated local adjuncts), hence it is clear that this structural possibility is not reserved for a particular type of dependents. As mentioned in Section 1, left dislocation differs from focus fronting (see (3)) in that it does not have any interaction with the verb morphology

and it is characterized by the fact that a deictic enclitic occurs at the right edge of the left dislocated material.

- (4) a. *Pèdróoh-e'* *t-u* *hàant-ah* *òon.*
 Pedro-D3 PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado
 'Pedro ate (an) avocado.'
- b. *Òon-e'* *t-u* *hàant-ah* *Pèdróoh.*
 avocado-D3 PFV-A.3 eat:TRR-CMPL(B.3.SG) Pedro
 'Pedro ate (an) avocado.'
- c. *Pèdróoh-e'* *h* *hàan-ih.*
 Pedro-D3 PFV eat(CMPL)-B.3.SG
 'Pedro ate.'

When both a left dislocated constituent and a focus constituent are present in a sentence, then the only possible order is “ $XP_{LD} \prec YP_{FOC}$ ”. The constituent enclosed by the enclitic *-e'* in (5a) is left dislocated, while the subsequent preverbal constituent occupies the focus position and is not accompanied by an enclitic. The opposite option is not possible as shown in (5b). Furthermore, the negation particle precedes the focus constituent and follows the left dislocated constituent (see Skopeteas and Verhoeven 2007 and references therein).

- (5) a. *Pèdróoh-e'* *òon* *t-u* *hàant-ah.*
 Pedro-D3 avocado PFV-A.3 eat:TRR-CMPL(B.3.SG)
 'It was avocado that Pedro ate.'
- b. **Òon Pèdróoh-e'* *t-u* *hàant-ah.*

Finally, while there is exactly one position for focused material, it is possible to have more than one left dislocated constituent. Example (6a) illustrates the possibility to form

sentences with more than one left dislocated constituents, and (6b) simulates a (non-possible) double foci construction with two preverbal NPs without right edge enclitic.

- (6) a. *Pèedróoh-e' òon-e' t-u hàant-ah.*
 Pedro-D3 avocado-D3 PFV-A.3 eat:TRR-CMPL(B.3.SG)
 'Pedro ate avocados.'
 b. **Pèedróoh òon t-u hàant-ah.*

A focused constituent is a clausal constituent that could alternatively occur in situ. The assumption that this constituent is fronted in the focus position is supported by the fact that it cannot co-occur with a co-referent element in situ, as shown in (7a). This restriction does not hold for left dislocated constituents that may co-occur with an element in situ, as shown in (7b). The version of (7b) with a pronoun in situ invokes variable reactions across speakers, since it is difficult to reconstruct the contextual conditions under which an emphatic pronoun would be used in situ (and not in the focus position). However, there is a clear difference in the acceptability of (7a) and (7b); in spontaneous discourse, we find plenty of utterances in which an emphatic pronoun that is co-referent with the left dislocated constituent is realized in the focus position, as exemplified in (8).

- (7) a. **Pèedróoh_j táan u bin màan leti'_j.*
 Pedro PROG A.3 go buy:INTRV that.one
 'It is Pedro_j, that he_j goes shopping.'
 b. *Pèedróoh_j-e' táan u bin màan leti'_j.*
 Pedro-D3 PROG A.3 go buy:INTRV that.one
 'Pedro_j, he_j goes shopping.'

- (8) *Le ah kòonol-o' leti' túun y-áalkab.*
 DEF M seller-D2 that.one PROG:A.3 0-run
 'The seller, it is him that is running.'

Furthermore, some left dislocated constituents do not correspond to constituents that could appear in situ, as illustrated in (9). Locative constituents are PPs in Yucatec Maya (apart from some instances of NP locatives that occur only in restricted constructions, e.g., with verbs of motion governing a locative complement). Hence, if the left dislocated NP in (9) was realized as a clause-internal locative adjunct, it would be obligatorily introduced by a locative preposition.

- (9) *In nah-il-e' yan u yàantal hun-p'éel cha'n sáamal-i'.*
 A.1.SG house-REL- D3 DEBA.3 exist:PROC one-CL.INAN spectacle tomorrow- D4
 'As concerns my house, there will be a party tomorrow.'

The facts presented in this section suggest that left dislocated material is projected in a layer that is outside the domain that contains the basic clause and the focus position. Left dislocated material does not occupy a unique position, as shown by the fact that this configuration may be recursively used. Furthermore, it is not the result of extraction out of the clause since it may co-occur with co-referent postverbal material and it does not require a syntactic relation to a clausal constituent.

3. Information structural properties of left dislocated constituents

In section 2, we have shown that left dislocated material is realized in the outer layer of the constituent structure of a sentence in Yucatec Maya and that it is distributionally distinct from the focus position, which is immediately preverbal. In this section, we will address the question of what licenses left dislocation. Our data shows that there are at

least two independent licensing conditions for left dislocation, one of them being pragmatic, the other one being syntactic in nature.

3.1. *Subjects of intransitive vs. subjects of transitive verbs*

3.1.1. Method. In order to identify the information structural properties of left dislocation, we first discuss a data set of semi-spontaneous narratives elicited with four native speakers (coded as F, L, N, and R in the following).⁴ The narratives are elicited by means of videos and picture sets presenting sequences of actions that are designed to induce expressions with varying information structures. The speakers described the pictures to the instructor who was a further native speaker participating in the field session.⁵

The dataset consists of 309 clauses in total ($n_F=79$, $n_L=86$, $n_N=77$, $n_R=67$). Excluding subordinate clauses (38 tokens, 12.2%), utterances that do not form a clause (17 tokens, 5.5%) and some incomprehensible utterances which are classified as “other” (6 tokens, 1.9%), a set of 248 main clauses (80.3%) remains. Since the aim of this Section is to assess the pure effect of contextually established givenness, we exclude further 16 clauses that involve a local (1st or 2nd person) argument: these referents are not given through the context but through the discourse situation (but see Section 4 on configurations with local persons). Furthermore, we excluded 8 clauses that involve a

⁴ All native speakers are inhabitants of Felipe Carrillo Puerto in Quintana Roo, use Yucatec Maya in their everyday communication, and are bilingual in Spanish.

⁵ The stimuli and the experimental procedure are part of the *Questionnaire on Information Structure* (see Skopeteas et al. 2007: 36-38, 77-82, 122-137, 167-169).

constituent in the focus position which is contrasted with other contextually available referents and 6 clauses with reflexive and reciprocal verbs. These decisions restrict our corpus to 218 clauses (70.6% of the entire dataset).

Subjects of intransitive and transitive verbs in our corpus are annotated for discourse status: a referent is classified as “given”, (a) if it is aforementioned in discourse, or (b) if its availability is inferable from the presence of a referent which is aforementioned in discourse.⁶ Otherwise, the referent is annotated as non-given, [-g]. Furthermore, a given referent is annotated as prominent, [+p], if it is co-referent with the highest non-local (i.e., 3rd person) referent in the immediately preceding utterance. Otherwise, the referent is annotated as non-prominent, [-p]. The concept of “highest referent” refers to the hierarchy of syntactic functions: subject > object > oblique complement > adjunct (see examples in Section 3.1.2). The subject of the target clause is co-referent to the highest argument of the preceding clause if it has exactly the same referential content with it (a criterion that applies equally to singular, plural or quantified NPs).

3.1.2. Results. A summary of the word order patterns that occur in our corpus is given in Table 1. Prominent given referents are not realized through lexical NPs in the majority of cases, see example (10) (see previous observation in Tonhauser 2003: 206-207). This result is expected for a language that obligatorily encodes subjects and objects through cross-reference markers and is independent of the transitivity of the verb (see “Pred” in Table 1: 92% with intransitives, 95.9% with transitives).

⁶ Some further types of information that is accessible without requiring a mention in the preceding text (such as nouns with unique reference, e.g., ‘the moon’, or referents that are part of the shared knowledge of the interlocutors, e.g., ‘the chief’) are relevant for our research question but do not occur in our corpus.

- (10) Context: ‘A cat is in the water ...’

Túun b àab.

PROG:A.3 swim

‘It is swimming.’ (L 26.2)

Prominent referents are sometimes re-introduced in discourse (6 examples in our corpus, see Table 1), typically when the speaker wishes to signal a new text section through resetting the involved referents, as illustrated in (11).

- (11) Context: ‘In this [picture], there is a dog that is jumping and he sees a boy and begins to follow him ...’

He'l-o' le pèek'-o' káa t-u pul-ah u báah

PRSV-D2 DEF dog-D2 CNJ PFV-A.3 throw-CMPL(B.3.SG) A.3 self

u chi'-eh-e'.

A.3 bite-SUBJ(B.3.SG)-D3

‘Here, the dog jumped down to bite him.’ (R 22.1)

The crucial part of Table 1 for the understanding of the factors that influence left dislocation in Yucatec Maya is the subset of the data in which the speaker lexicalizes both the predicate and the subject (“Pred | S” in Table 1). In this case, the speaker has a choice between the predicate-initial order (“Pred S” in Table 1) and the order with a left dislocated subject (“S_{LD} Pred” in Table 1). The data from intransitive verbs suggest that the left dislocated vs. postverbal realization of the subject is influenced by information structure.⁷ When the subject is new information, the predominant order is verb-subject (60 out of 68 clauses; 88.2%, see “Pred S” in Table 1), as exemplified in (12).

⁷ ‘Intransitive verbs’ include all classes of intransitives in this language: active, inactive, inchoative, positional, passives, and non-verbal predicates (see Bohemeyer 2004). The size of the corpus does not

- (12) *H lúub hun-p'éel che'[-g],*
 PFV fall one-CL.INAN branch
káa h tàal hun-túul x-ch'úup-e'[-g].
 CNJ PFV come one-CL.AN F-woman-D3
 'A branch fell down, then a girl came.' (L 37.2)

Given non-prominent subjects are illustrated in (13). The subjects of the second and third clause are already introduced at the very beginning of the narrative, hence they both represent given information. The subject of the second clause denotes a member of the set of referents denoted by the highest non-local constituent of the first clause, hence it does not qualify as co-referent to it and is annotated as [-p]. Similarly, the referent of the subject of the third clause is distinct from the referent of the subject of the immediately preceding clause and qualifies as [-p] too. Example (13) illustrates the most frequent pattern: non-prominent given subjects of intransitive verbs occur preferably left dislocated, see “S_{LD} Pred” in Table 1 (17 out of 28 clauses with a lexically realized subject; 60.7%).⁸ The difference to the frequency of left dislocated new subjects is

allow for generalizations concerning differences between these classes or between different aspects. The crucial question is whether unergative verbs display a preference for left dislocation of the subject constituent that is independent of information structure and hence pattern with transitive verbs. Our data shows that 4 out of 9 new subject constituents (44.4%) of unergative verbs are left dislocated. The comparison of this result to the overall result of intransitives in the condition new subject (S[-g]) in Table 1 suggests that agentive subjects of intransitives are more likely to be left dislocated than non-agentive subjects of intransitives (11.8%), but less likely than agents of transitive verbs (88.9%).

⁸ Surprisingly, non-prominent given referents are sometimes not lexically realized (see intransitive and transitive verbs for S[-p] in Table 1: 2+4=6 out of 30+12=42 cases, 14.3%). One of those examples is the first sentence of (13), in which the plural person affix on the verb refers to the entire set of introduced

statistically significant (two-tailed paired-sample *T*-test carried out on the proportions of left dislocation per speaker, $t_3 = 6.8$, $p < .01$).

- (13) Context: ‘Near the water, there is a woman. The door of the house is open and a man is going there ...’ (further 12 clauses containing actions of the two individuals)

Káa t-u sut-ah u báah-o'b-e',
 CNJ PFV-A.3 turn-CMPL(B.3.SG) A.3 self-3.PL-D3
le máak-o'[-p] káa h wa'l-lah yéetel u cubo,
 DEF person-D2 CNJ PFV stand-CMPL(B.3.SG) with A.3 pot
le x-ch'úuppal-o'[-p] káa káa h wa'l-lah tu'x
 DEF F-woman:child-D2 CNJ CNJ PFV stand-CMPL(B.3.SG) where
yàan le p'óok-o'.
 exist(B.3.SG) DEF hat-D2
 ‘And they turned around, the man stopped with his pot, the girl stopped where the hat is.’ (L 40.1)

If we compare the results from intransitive verbs to the results from transitive verbs in Table 1 (excluding 6 instances of reflexives and reciprocals), we observe the following difference: while new subject constituents of intransitive verbs preferably occur in situ (88.2%), new subject constituents of transitive verbs are preferably left dislocated (88.9%), as illustrated in (14) (see similar findings in Gutiérrez Bravo and Monforte y Madera 2008a). The same preference holds for subjects of transitive verbs in the condition $S[-p]$. The three “Pred S” tokens with transitive verbs correspond to utterances with a subject in situ in the VOS order, see (15).

individuals. The occurrence of such examples shows that the antecedent of a cross-reference marker may be identified through pragmatic inferences.

- (14) *Te'l-a' hun-túul máak u mach-mah hun-p'éel k'áanche'.*
 there-D2 one-CL.AN person A.3 seize-PART.PF(B.3.SG) one-CL.INAN chair
 'There, a man holds a chair.' (L 31.4)
- (15) *Táan u hóoyabt-ik hun-p'éel pàak'al le máak-o'.*
 PROG A.3 water:TRR-INCMPL(B.3.SG) one-CL.INAN plant DEF person-D2
 'The man is watering a plant.' (F 40.1)

Table 1: Corpus data⁹

	S[-g]		S[-p]				S[+p]					
	intr.		trans.		intr.		trans.		intr.		trans.	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
TOTAL	68		9		30		12		50		49	
Pred	0		0		2	6.7	4	33.3	46	92	47	95.9
Pred S	68	100	9	100	28	93.3	8	66.7	4	8	2	4.1
Pred S	60	88.2	1	11.1	11	38.3	1	12.5	2	50	1	50
S _{LD} Pred	8	11.8	8	88.9	17	60.7	7	87.5	2	50	1	50

The crucial issue for the function of left dislocation is the interaction between the discourse status of the subject constituent and the transitivity of the verb. The critical

⁹ Abbreviations: Pred = predicate (the subject constituent is not lexically realized); Pred | S = predicate and subject in either order; Pred S = predicate-subject in this order; S_{LD} Pred = left dislocated subject and predicate; S[-g] = new subject; S[-p] = non-prominent given subject; S[+p] = prominent given subject. The notation "Pred" conventionally refers to the unit containing the aspectual auxiliary, the crossreference markers, the lexical verb, and its complements. We use this term descriptively, through it may correspond to different constituents in the "Pred S" and in the "S_{LD} Pred" tokens (predicate in the former case, clause in the latter).

subset of the data is presented in Figure 1 and involves the utterances with a lexically realized subject that occur whenever the subject is either new or non-prominent given information. An analysis of variance on the percentages of left dislocation per speaker (transformed through the *arcsin-root* transformation in order to meet the normality requirement of parametric tests) reveals a significant main effect of transitivity (intransitive vs. transitive), $F_{1,3} = 80.1, p < .01$, a significant main effect of information structure (S[-g] vs. S[-p]), $F_{1,3} = 37.9, p < .01$, and a significant interaction between the two factors, $F_{1,3} = 22.3, p < .05$.

<Insert Figure 1 here>

3.1.3. Discussion. An important question is the relation of the text counts in Table 1 with assumptions about topics. The heuristics we used for the annotation of the discourse status are similar to the measurements presented in Givón (1994, ed.). The main difference is that referential distance to co-referent antecedents is treated as a continuum in Givón's framework. In our study, we make a sharp distinction between co-reference with the highest non-local referent of the preceding clause and further given referents. This distinction is crucial, since it allows for predictions concerning the lexical realization of the subject constituent and since it leads to a more restrictive description of the discourse contexts in which left dislocation occurs (these phenomena are not relevant for the studies on voice alternations presented in Givón 1994). These heuristics allow for generalizations concerning the relation between the left-dislocated constituent and the context, in particular in order to examine whether the constituent at issue is a link to the

common ground (in terms of Vallduví 1992), but they are not informative for hypotheses concerning the relation between the left-dislocated constituent and the clause (“aboutness” relation in terms of Reinhardt 1981).

Elements annotated as prominent given subjects, (S[+p]), correspond to continuing topics. Aissen (1992: 74) shows that topics in Tz’utujil Maya occur in this discourse configuration, while topics in Tzotzil and Jakaltek do not. She argues that the pragmatic function of coding continuing topics correlates with the structural property that Tz’utujil topics are projected in a lower layer of the clause structure than topics in Tzotzil and Jakaltek, which are primarily used for new topics or topic shift. Our data in Table 1 shows that Yucatec Maya patterns with Tzotzil and Jakaltek: apart from a few instances in which the speakers reset the discourse referents, as illustrated in (11), generally continuing topics are not lexically realized.

Elements that are annotated as non-prominent given, (S[-p]), are given referents that are not the current topic at the critical point in discourse in which the speaker plans the target utterance. This is the ideal discourse condition for the choice of an expression which introduces a link to the common ground. In this regard, the results from intransitive verbs in Table 1 show that exactly this discourse condition licenses left-dislocation in Yucatec Maya. This finding supports the view that Yucatec Maya patterns with Tzotzil and Jakaltek in which constituents in the topic position are used to signal a new topic (following Aissen 1992).

Elements that are annotated as non-given, (S[-g]), are new referents. The major finding in the data from intransitive verbs is that the frequency of left dislocation is significantly lower in the condition (S[-g]) than in the condition (S[-p]). However, the

(S[-g]) results involve some instances of left dislocation, see (16). The possibility of left dislocation in this condition is due to the fact that givenness is not a necessary condition for topicalization: an indefinite NP may be used as topic in order to create a new “address” in the common ground (see Krifka 2007: 29).

- (16) *Ka'-túul máak-o'b-e' túun xòok-o'b (...)*
 two-CL.AN person-PL-D3 PROG:A.3 read\INTRV:INCMPL-3.PL
 ‘Two men are reading.’ (L 9.1)

Apart from the effect of information structure in clauses with intransitive verbs, we identified an effect of transitivity such that subjects of transitive verbs are preferably left dislocated independently of information structure. We will come back to the structural basis of this difference in Section 4.

3.2. Agents vs. patients of transitive verbs

The corpus data in Section 3.1. led to some observational generalizations concerning the left-dislocation of subject constituents. In order to observe the properties of patient topics, we carried out a controlled elicitation task.

3.2.1. Method. The task consists in the semi-spontaneous description of two subsequent scenes.¹⁰ The first scene (=context scene) presents an entity. The second scene (=target scene) presents an event in which an agent and a patient are involved. The speakers were

¹⁰ A full documentation of the conditions and materials used for this elicitation task can be found in Skopeteas et al. (2007: 39-73).

presented the pictures by the instructor and were asked to describe the scenes as if they were part of the same story. Crucially, speakers were shown the target scene after completing the description of the context scene.

In the following, we discuss two different conditions of this task: (a) condition A[+p]: the agent of the target scene is introduced in the context while the patient is new information, (b) condition P[+p]: the patient of the target scene is introduced in the context while the agent is new information. These two conditions are implemented in 8 items presenting diverse events with an agent and a patient, resulting in $8 \times 2 = 16$ picture pairs (see material in Skopeteas et al. 2007). These picture pairs were distributed over 4 field sessions such that each session contained four different picture pairs (each condition twice).

Two native speakers, both inhabitants of Yaxley (Quintana Roo), participated in this experiment. They have been presented every field session twice (in different appointments), which resulted in a dataset of 2 (times) $\times 4$ (sessions) $\times 4$ (picture pairs) $\times 2$ (speakers) = 64 descriptions.

3.2.2. Results. Some descriptions (15 descriptions, 23.4%) were classified as non-valid, either because the speaker did not produce a transitive verb or because he described the scene with more than one clauses, such that the givenness of the arguments in the target clause does not correspond to the intended discourse condition.

The valid descriptions differ with respect to the voice of the verb and the word order, as presented in Table 2. When the prominent referent is the agent constituent, the most frequent structure in the obtained descriptions is $S_{LD}VO$, as exemplified in (17). The explicit mention of the prominent agent in this context is due to the fact that the task does

not elicit a continuous narrative. Since speakers are shown the target scene after completing the description of the context scene, they often reset the involved referents, which results in a lower occurrence of clauses without lexical realization of the agent (see 4 VO tokens out of 25 valid in the condition A[+p] in Table 2).

- (17) Context: ‘Here is a dog ...’
Be’òoráah le pèek’-o’ ts’u chuk-ik le kàan-o’.
 now DEF dog-D2 TERM:A.3 catch-INCMP(LB.3.SG) DEF snake-D2
 ‘Now, the dog has caught the snake.’ (E 42.G7)

When the prominent referent is the patient (P[+p]), then it is often left dislocated. However, sentences with active verbs and left dislocated objects do not occur in the data. Instead, patient preposing is always accompanied by passivization, i.e., the left dislocated patient is a subject, see (18). The impact of patient prominence on the selection of passive voice is shown by the difference in the frequency of passives in the conditions A[+p] (4%) and P[+p] (12.5+41.7=54.2%), see Table 2. This difference is statistically significant (two-tailed paired samples *T*-test, performed on the items, $t_7 = 3.9$, $p < .05$).

- (18) Context: ‘Here, a small table is standing on the chairs ...’
Be’òoráah le chàan mesa-o’ túun léench’int-a’l
 now DEF little table-D2 PROG:A.3 push:TRR-PASS.INCMP(LB.3.SG)
téen le x-ch’úuppal-o’.
 by DEF F-woman:child-D2
 ‘Now, the small table is being pushed by the girl.’ (E 42.C6)

Passivization does not imply left dislocation of the subject patient, as exemplified in (19), in which the subject constituent is realized postverbally. The patient subjects of passive verbs preferably precede the oblique agents.

- (19) Context: ‘Here, a little girl is jumping on the table ...’
Be’òoráah ts’u léench’int-a’l le chàan x-ch’úuppal
 now TERM:A.3 push:TRR-PASS.INCMPL DEF little F-girl
téen le máak-o’.
 by DEF person-D2
 ‘Now, the little girl is pushed by the man.’ (E 42.A2)

Furthermore, in the same condition, we encountered two clauses with V-initial order. Example (20) is an instance of the canonical VOS order. Example (21) illustrates the VSO order. It is known that the order of postverbal arguments may be influenced by asymmetries in animacy (animate-first) or weight (heavy-last) (see references in Section 1). Both principles license a VSO order in (21).

- (20) Context: ‘Here is a chair ...’
Lela’ muka’h u koh le k’áanche’-o’ máak-o’.
 it:D1 IMM.FUT A.3 push(SUBJ)(B.3.SG) DEF chair-D2 person-D2
 ‘[On] this [picture], a man is going to push the chair.’ (R 42.C3)

- (21) Context: ‘Here is a traffic sign ...’
Lela’ túun hok-ik le máak le senyal de transito.
 it:D1 PROG:A.3 pull.up-INCMPL(B.3.SG) DEF person DEF traffic.sign
 ‘[On] this [picture], the man is pulling up the traffic sign.’ (E 42.C4)

Table 2: Elicitation task on prominent agents and patients¹¹

	A[+p]		P[+p]		TOTAL	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
TOTAL	32		32		64	
non-valid	7	21.8	8	25	15	23.4
valid	25	78.1	24	75	49	76.6
V _{ACT} S	4	16	-		4	8.2
V _{ACT} OS/SO	-		2	8.3	2	4.1
S _{LD} VO	20	80	9	37.5	29	59.2
V _{PASS} SObl	-		3	12.5	3	6.1
SV _{PASS} Obl	1	4	10	41.7	11	22.4

3.2.3. Discussion. The data in Table 2 reveals an asymmetry between active and passive clauses. We encountered a single instance of passive voice in the context A[+p], but a substantial amount of active clauses in the context that licenses passives (P[+p]). We assume that the occurrence of an instance of passive in A[+p] is accidental (probably due to priming from previous trials of this experiment). The occurrence of the two voices is asymmetric, because active is the only option in the A[+p] context, while both voices are possible in the P[+p] context. This asymmetry reflects the markedness of the passive voice: the marked pattern occurs only in the licensing context, while the unmarked pattern is contextually unrestricted (see Skopeteas and Fanselow 2009).

¹¹ Abbreviations: S=subject; O=object; V=verb; ACT=active; PASS=passive; LD=left dislocation; A=agent; P=patient; [+p]=prominent referent.

Furthermore, the results reveal a subject/object asymmetry in left dislocation. Left dislocation of the subject is the default structure for the encoding of a transitive verb with two lexically realized arguments, whereas left dislocation of the object does not occur at all, though it is a grammatical possibility in Yucatec Maya, see (4). A preference for passivization instead of object preposing can also be found in other languages that allow for both options (Skopeteas and Fanselow 2009), both in languages in which object preposing is only possible through movement to positions in the left periphery (e.g., English, Canadian French, and Dutch) as well as in languages with free word order (e.g., German).

4. The licensing conditions of left dislocation

The empirical data presented so far suggests that there are two independent licensing conditions for left dislocation. In the corpus data presented in Section 3.1., left dislocation of subjects of intransitive verbs is more likely to occur when the subject is part of the given information, but not co-referent with the highest non-local constituent of the preceding clause, in which case it is a continuing topic that is most likely not realized as a lexical NP. Hence, the dominant patterns that are summarized in (22) lead to the conclusion that left dislocation of subjects of intransitive verbs is sensitive to information structure.

- (22) a. $\langle V, S[-g] \rangle \rightarrow VS$
 b. $\langle V, S[-p] \rangle \rightarrow S_{LD}V$

Additional support that givenness may induce left dislocation comes from the elicitation task presented in Section 3.2. This data set differs from the corpus data in that the target utterances are not part of a continuous narrative to the effect that prominent referents are lexically realized. The most frequent pattern when the prominent information is the agent, is an active clause with a left dislocated agent, while the dominant pattern when the prominent information is the patient, is a passive clause with a left dislocated patient, as shown in (23).

- (23) a. $\langle V, P[-g], A[+p] \rangle \rightarrow A_{LD} V_{ACT} P$
 b. $\langle V, P[+p], A[-g] \rangle \rightarrow P_{LD} V_{PASS} A$

The data presented so far could lead to the conclusion that left dislocation is inherently associated with particular information structure, i.e., it is a topic position. The puzzling evidence comes from the distribution of word order patterns in configurations that involve a transitive verb with two lexically realized arguments. In contrast to the dominant patterns of intransitive verbs in (22), subjects of transitive verbs are preferably left dislocated independently of information structure (see Table 1):

- (24) a. $\langle V, P[-g], A[-g] \rangle \rightarrow A_{LD} VP$
 b. $\langle V, P[-g], A[-p] \rangle \rightarrow A_{LD} VP$

There are two possible explanations for this phenomenon. Assuming that the behavioral data directly reflects configurational properties of the constituent structure would lead to the conclusion that Yucatec Maya displays a subject position that has the same properties as left dislocation, i.e., it precedes focus and is enclosed by a right

boundary enclitic; this view has been advocated by Durbin and Ojeda (1978) and Gutiérrez Bravo and Monforte y Madera (2008a). It involves a complication of the constituent structure, since it assumes a subject position for transitive verbs that differs from the subject position for intransitive verbs. A further problem for this account is that VOS is a grammatical possibility in Yucatec Maya, even if it occurs rarely in discourse (see “Pred S” results with transitive verbs in Table 1 and V_{ACT}OS results in Table 2). The assumption of S_{LD}VO sentences as canonical creates a markedness paradox, since the allegedly basic configuration (S_{LD}VO) involves more morphological marking than the allegedly marked configuration (VOS); see the occurrence of the right edge clitic in (2) vs. (1).

These complications may be avoided if we assume that the behavioral facts reflect the impact of a constraint that leads to the preference for left-dislocation of subjects of transitive verbs (see detailed discussion in Skopeteas and Verhoeven 2009 and a similar view in Bohnemeyer 2008). The crucial constraint is a distinctness condition that renders configurations with two adjacent argument NPs suboptimal (see Richards 2006; Neeleman and Van de Koot 2005).¹² The intuition behind this assumption is that adjacent syntactic units of the same category that have to be interpreted as functionally distinct (i.e., they bear different thematic roles), but are not marked for their function (i.e., do not bear case marking) are difficult to parse.

The critical evidence that empirically proves that the preference for left dislocation is not related to the subjects of transitive verbs but to configurations with two lexically realized arguments comes from clauses involving a transitive verb and a sole lexically

¹² We are grateful to Gisbert Fanselow for drawing our attention to the effects of distinctness.

realized argument. In a corpus of 20 spontaneously elicited narratives, we encountered 19 clauses with a lexically realized agent constituent and a local (1st or 2nd person) patient that is not realized through a lexical pronoun. In this data set of transitive verbs, the V-initial order exemplified in (25) is more frequent when the agent constituent is new information, A[-g], (3 out of 5 clauses, 60%), while the order with a left dislocated subject in (26) is more frequent when the agent constituent is given and not co-referent with the highest non-local constituent of the preceding clause, A[-p], (9 out of 14 clauses, 64.3%). Hence, this data suggests that transitive verbs with a single lexically realized argument pattern with intransitive verbs as concerns the interaction of information structure with left dislocation.

(25) *Ts'u* *y-áant-ik-en* *le* *xòok-a'* ...
 TERM:A.3 0-help-INCMP-L-B.1.SG DEF story-D1
 'This story has helped me ...' (HIJO_043)

(26) *Wa'pach'-e'* *ta'itak* *u* *chukpacht-ik-o'n*.
 giant-D3 almost A.3 reach-INCMP-L-B.1.SG
 'The giant is almost reaching us.' (HK'AN_134)

Additional support for the application of the distinctness condition comes from focus constructions. When the focused patient is realized preverbally, then the distinctness condition does not apply to postverbal arguments. In a dataset of spontaneously elicited answers to constituent questions (16 native speakers; 2 answers per speaker; total: 32 answers), 7 answers contain the patient constituent in the focus position and a lexically realized agent. In all seven answers, the agent is realized postverbally as exemplified in (27) (see discussion in Skopeteas and Verhoeven 2007).

- (27) Question: What is the man kicking?
Hun-p'éel esten k'áanche' k-u kóochek'-t-ik
 one-CL.INAN HESIT chair IPFV-A.3 kick:foot-TRR-INCMPL(B.3.SG)
le xib-o'.
 DEF man-D2
 'It is a... ehm... chair that the man kicks.' (J 41.6)

On the basis of this evidence, we conclude that the preference for left dislocation of subjects is the result of the distinctness condition and is independent of information structural properties. The distinctness condition bans two adjacent syntactic entities of the same type (NP) and is related to the avoidance of structural ambiguity. Two adjacent postverbal NPs involve two sources of ambiguity: (a) the ambiguity between a VOS and a VSO reading (given that rightwards object shift is possible in Mayan languages), and (b) the ambiguity between a reading with two arguments, [[V NP] NP], and an appositive reading involving a single postverbal NP, [V [_{NP} NP NP]].¹³ Left dislocation resolves the ambiguity of the latter type. However, it is not obvious how left dislocation contributes to the resolution of the thematic ambiguity in (a), since left dislocated constituents are not thematically specified.¹⁴ The data presented in this article suggests that the thematic properties of a left dislocated argument are specified through the choice of voice. The experimental results in Section 3.2 show that whenever a patient constituent is left

¹³ Evidence from language comprehension shows that the appositive reading is a very frequent interpretation of V NP NP sequences in Yucatec Maya (see Skopeteas and Verhoeven 2005: 355).

¹⁴ See an overview of the disambiguation strategies employed by several head-marking languages in De Swart (2007: 99-105). See also Stiebels (2006) concerning the use of agent focus constructions for the resolution of role ambiguities in several Mayan languages.

dislocated, the speakers consistently choose passive voice which is thematically unambiguous (see previous observation in Tonhauser 2003: 206). Hence, we may conclude that the choice of the unmarked voice (active) leads to the implication that the licensing condition for passivization does not apply, i.e., the left dislocated argument is an agent constituent.

The assumption that this choice is a disambiguation strategy is supported by corpus evidence from configurations involving a local agent constituent that is not lexically realized. In the corpus of spontaneous narratives mentioned above, we encountered 48 clauses with a non-lexically realized local agent and a non-prominent given patient, P[–p]. The majority of these clauses (87.5%) involve a postverbal patient; the crucial finding is that the 6 clauses (12.5%) that involve a left dislocated patient do not involve a passive verb.

- (28) *Tak u y-o'ch le k'éek'en-o'b-a' táan in hàant-ik.*
 as.far.as POSS.3 0-food DEF pig-PL-D1 PROG A.1.SG eat:TRR-INCMPL
 'I am eating even the food of these pigs.' (HIJO_103)

This evidence shows that left dislocation is not associated with a particular syntactic function (i.e., it is not a subject position). The preference for the passive option applies exactly when ambiguity is involved.

5. Conclusions

This article presented behavioral data on the occurrence of left dislocation in discourse. Our data shows that the properties of left dislocation are not uniform. For a subset of the

data, left dislocation is licensed by the context, i.e., the left dislocated constituent qualifies as a pragmatic topic. This applies to the single arguments of intransitive verbs in Section 3.1, to the left dislocated agents and patients in Section 3.2., as well as to the single lexically realized arguments of transitive verbs in Section 4. For another subset of left dislocated constituents, our data suggests that left dislocation is not sensitive to information structure. We claimed that these instances of left dislocation are accounted for if we assume a purely structural constraint that renders configurations of two postverbal arguments suboptimal.

These generalizations lead to the conclusion that left dislocation in Yucatec Maya is not inherently associated with a unique information structural concept, i.e., it is not a topic position. The preference for topics to be realized early in the sentence is a cross-linguistic phenomenon that has its roots in functional aspects of human communication that are independent of the constituent structure. The structural possibility that is available in order to fulfill this grammar-independent preference in Yucatec Maya is left dislocation. Independently of discourse-related properties, structural constraints such as the distinctness condition render particular linearizations suboptimal. In order to express a propositional content that involves a transitive verb and two arguments that have to be realized through lexical NPs, speakers frequently select the structural possibility of left dislocation. Thus, the phenomena at issue suggest that left dislocation is a structural possibility that is used for an array of functions that are independent of each other.

References

- Aissen, Judith L. (1992). Topic and focus in Mayan. *Language* 68.1:43-80.
- Bohnmeyer, Jürgen (1998a). Sententiale Topics im Yukatekischen. In *Deskriptive Grammatik und allgemeiner Sprachvergleich*, Dietmar Zaefferer (ed.), 55-85. Tübingen: Niemeyer.
- Bohnmeyer, Jürgen (1998b). Time relations in discourse: evidence from a comparative approach to Yukatek Maya. Ph.D. Dissertation, Katholieke Universiteit Brabant.
- Bohnmeyer, Jürgen (2004). Split intransitivity, linking, and lexical representation: the case of Yukatek Maya. *Linguistics* 42:67-107.
- Bohnmeyer, Jürgen (2008). Linking without grammatical relations in Yucatec: alignment, extraction, and control. Ms., University of Buffalo.
- Bricker, Victoria R. (1979). Wh-questions, relativization, and clefting in Yucatec Maya. In *Papers in Mayan Linguistics*, Laura Martin (ed.), 107-136. Columbia, Miss.: Lucas Brothers.
- Clark, Herbert and Eve Clark (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Clark, Herbert and Susan Haviland (1977). Comprehension and the given-new contrast. In *Discourse Production and Comprehension*, Roy O. Freedle (ed.), 1-40. Hillsdale, NJ: Lawrence Erlbaum Associates.
- De Swart, Peter (2007). Cross-linguistic variation in object marking. Ph.D. Dissertation, Radboud University Nijmegen (LOT publications).
- Durbin, Marshall and Fernando Ojeda (1978). Basic word order in Yucatec Maya. In *Papers in Mayan Linguistics, Vol. 2*, Nora C. England (ed.), 69-77. Columbia: University of Missouri, Department of Anthropology.

- É. Kiss, Katalin (1998). Identificational vs. information focus. *Language* 74.2:245-273.
- Givón, Talmy (ed.) (1994). *Voice and Inversion*. Amsterdam/Philadelphia: Benjamins.
- Gutiérrez Bravo, Rodrigo and Jorge Monforte y Madera (2008a). La alternancia sujeto inicial/verbo inicial y la teoría de la optimidad. In *Teoría de la optimidad: estudios de sintaxis y fonología*, Rodrigo Gutiérrez Bravo y Esther Herrera Zendejas (eds.), 61-90. México: El Colegio de México, Centro de Estudios Lingüísticos y Literarios-Laboratorio de Estudios Fónicos.
- Gutiérrez Bravo, Rodrigo and Jorge Monforte y Madera (2008b). On the nature of unmarked word order in Yucatec Maya. Paper presented at the Symposium *New Perspectives in Mayan Linguistics*, held at the 2008 SSILA Annual Meeting, January, 3-4, 2008, Chicago.
- Krifka, Manfred (2007). Basic notions of information structure. In *The Notions of Information Structure*, Caroline Féry, Gisbert Fanselow, and Manfred Krifka (eds.), 13-55. Potsdam: University Press (Interdisciplinary Studies on Information Structure 6).
- Lambrecht, Knud (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Lehmann, Christian (1990). Yukatekisch. *Zeitschrift für Sprachwissenschaft* 9.1-2:28-51.
- Lehmann, Christian (2003). *Possession in Yucatec Maya*. Second revised edition. Erfurt: Seminar für Sprachwissenschaft der Universität (Assidue 10).
- Neeleman, Ad and Hans Van De Koot (2005). Syntactic haplology. In *The Blackwell Companion to Syntax: Volume IV*, Martin Everaert, Henk van Riemsdijk, Rob Goedemans, and Bart Hollebrandse (eds.), 684-710. London: Blackwell.

- Neeleman, Ad and Hans Van De Koot (2008). Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Linguistics* 11.2:137-189.
- Norman, William M. and Lyle Campbell (1978). Toward a proto-Mayan Syntax: A comparative perspective on grammar. In *Papers in Mayan Linguistics*, vol. 2., Nora C. England (ed.), 136-155. Columbia: University of Missouri, Department of Anthropology.
- Reinhart, Tanya (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27:53-94.
- Richards, Norvin (2006). A distinctness condition on linearization. Ms., Cambridge, MA: MIT.
- Skopeteas, Stavros and Gisbert Fanselow (2009). Effects of givenness and constraints on free word order. In *Information Structure from Different Perspectives*, Malte Zimmermann and Caroline Féry (eds.), in press. Oxford: Oxford University Press.
- Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry, and Manfred Krifka (2006). *Questionnaire on Information Structure: Reference Manual*. Potsdam: University Press (Interdisciplinary Studies on Information Structure 4).
- Skopeteas, Stavros and Elisabeth Verhoeven (2005). Postverbal argument order in Yucatec Maya. *Sprachtypologie und Universalienforschung* 58.4:347-373.
- Skopeteas, Stavros and Elisabeth Verhoeven (2007). Licensing focus constructions in Yucatec Maya: an empirical study on the association with focus. Ms., University of Potsdam and University of Bremen.

- Skopeteas, Stavros and Elisabeth Verhoeven (2009). Distinctness effects on VOS order: evidence from Yucatec Maya. In *New Perspectives in Mayan Linguistics*, Heriberto Avelino, Jessica Coon, and Elisabeth Norcliffe (eds.), in press. Cambridge, MA: MIT (MIT working papers in Linguistics).
- Stiebels, Barbara (2006). Agent focus in Mayan languages. *Natural Language and Linguistic Theory* 24:501-570.
- Tonhauser, Judith (2003). F-constructions in Yucatec Maya. In *The Proceedings of SULA 2*, Jan Anderssen, Paula Menéndez-Benito, and Adam Werle (eds.), 203-223. Vancouver, BC: GLSA, University of Massachusetts, Amherst.
- Tonhauser, Judith (2007). Agent focus and voice in Yucatec Maya. *Proceedings of the 39th Meeting of the Chicago Linguistic Society*.
- Vallduví, Enric (1992). *The Informational Component*. New York: Garland.
- Verhoeven, Elisabeth (2007). *Experiential Constructions in Yucatec Maya*. Amsterdam/Philadelphia: Benjamins.