Jürgen Bohnemeyer and Timothy J. Tilbe (Buffalo)

Argument realization and discourse status in Yucatec, a purely head-marking language

Abstract

We examine the impact of givenness on argument realization in a small text corpus of Yucatec folk narratives (660 clause-like units). We find that the default devices for the extension of topic chains are bare cross-reference markers. Weightier expressions – free pronouns, demonstrative pro-forms, clause-mate co-nominals, and left-dislocated pro-forms or nominals – are used for disambiguation among competing referents, resuming previous topic chains, and introducing new referents. Bare cross-reference markers occur exclusively with given referents. Their behavioral properties thus strongly support an anaphoric analysis, despite the existence of evidence suggesting an agreement ('pro-drop') analysis in the presence of clause-mate co-nominals.

1. Introduction

The question addressed in this paper is which factors govern the use of noun phrases (NPs)[1] in purely head-marking languages, given that NPs are by definition syntactically optional in head-marking structures. Based on qualitative observation, Bohnemeyer (2009) suggests that in Yucatec Maya, non-pronominal NPs are used both for the introduction of new referents and for the resumption of previously introduced ones that are not continuing topics in the local context, i.e., for newly returned chain-initial topics in the sense of Givón (1983: 9)). In contrast, bare cross-reference markers not accompanied by NPs are used for extending topic chains. In other words, they are used very much like pronouns in dependent-marking languages such as English. Our primary goal in this article is to test this hypothesis quantitatively. We present the results of a small corpus study. The hypothesis is not without alternative, as Yucatec also has free pronouns as they occur in configurational dependent-marking languages such as English. An important supplementary question is thus what the functional division of labor is between the realization of arguments by bare cross-reference markers vs. by (cross-reference markers augmented by) morphologically unbound pronouns. Our study is rounded out by an additional realization option – left-dislocated NPs – and by a consideration of the role of discourse structure and lexical semantics in reference resolution.

---

[1] Throughout this article, we use the term 'noun phrase' (NP) in the broad, traditional sense of Dixon (2009: 106-108), corresponding to 'determiner phrase' (Abney 1987) in the Minimalist Program and 'reference phrase' (Van Valin 2008) in Role and Reference Grammar.

The article is organized as follows: in Section 2, we define central terms. Section 3 sketches the grammar of Yucatec down to the introduction of the options speakers choose from among for the realization of arguments. In the process, we discuss the evidence for Yucatec cross-reference markers being referential. We also include a summary of the analysis in BOHNEMEYER et al (in press), according to which Yucatec cross-reference markers realize the arguments of the head in the absence of coindexed clause-internal NPs, but serve to express agreement in the presence of such 'co-nominals', much as argued by BRESNAN & MCHOMBO (1987) for the subject markers of Chicheŵa (Bantu; Malawi, Zambia, Mozambique). In Section 4, we present the corpus study. We discuss the findings in Section 5. We include a qualitative comparison with available accounts of realization in other head-marking languages. Section 6 concludes.

2. Argument realization and discourse status

We use the term **argument realization** for the choice among different constructions a given language allows for the expression of a particular semantic argument of a given predicate (similarly ALLEN (2000, 2008), BICKEL (2003), BROWN (2008), DU BOIS (1987, 2003), and others). Argument realization in this sense is one aspect of the larger domain of **argument structure**, the mapping between the lexical meaning of verbs (and other natural-language predicators), their morphosyntactic properties, and the expression of their semantic arguments.[2] For example, as illustrated in (1), the subject of *rebuke* may be ellipsed (a) or expressed by a pronoun (b) or a non-pronominal NP (c).

(1)   a.   *Floyd$_i$ took exception and Ø$_i$ rebuked the officer.*
      b.   *Floyd$_i$ took exception. He$_i$ rebuked the officer.*
      c.   *Floyd took exception. Sally rebuked the officer.*

Which of the available options is (most) appropriate for the realization of a given argument in a particular context depends on a variety of factors, including:

- The semantic role and grammatical relation of the argument in conjunction with the meaning and selectional restrictions of the verb (or more generally, the head) – e.g., some semantic arguments require prepositional phrases, verbal projections, or clauses for their expression;
- The syntactic environment of the construction – e.g., the ellipsis in (1a) is licensed by coordination;
- The status of the referent of the argument in the discourse – e.g., whether the referent has already been mentioned (1a,b) or is newly introduced to the discourse (1c).

[2] Some authors use 'argument realization' in this broader sense, for which we prefer 'argument structure'; e.g., KING (2000), GOLDBERG (2005), and LEVIN & RAPPAPORT-HOVAV (2005).

This study is concerned with the influence of discourse status on argument realization. **Discourse status** can be characterized informally as the *metalinguistic* information about a given referent (most commonly an individual, time, place, or event) that is ideally available to speakers and hearers when processing a particular utterance as a result of the preceding discourse and the general knowledge shared by the members of the linguistic and cultural community. We take discourse status to comprise at least two basic dimensions, which are in first approximation independent of one another. Many different labels for these dimensions have been proposed. We call them 'givenness' and 'pragmatic role'. Both can be conceptualized as variables with a range (possibly a scale) of categorical values.

In the following, we assume an informal, generic dynamic model of discourse, which decomposes discourses into sets of utterances. Utterances are verbal or nonverbal actions that realize tokens of conventional semiotic signals and perform speech acts.[3] Each utterance has a unique position in the discourse vis-à-vis all the other utterances the discourse consists of, which precede it, follow it, or overlap with it. Speakers and hearers keep a record of the discourse as it unfolds, updating with each utterance the Common Ground (STALNAKER 1974; LEWIS 1979), i.e., the information they treat as shared.

By **givenness**, we refer to the relative ease with which a referent can be "picked up" in a given utterance in a manner that is pragmatically felicitous and readily processable for the interactants. This ease can be measured in terms of the relative weightiness or complexity of the expression required for this purpose (CHAFE 1976, 1994; GUNDEL et al 1993). Alternative terms for this property used equally widely in the literature include 'availability' and 'accessibility' (cf. GIVÓN 1983 for both). Speakers treat referents as given or accessible in a given context when they assume a tacit agreement with the hearers to the effect that the referents in question are available in the discourse context for reference by suitable expressions without requiring explicit introduction. This agreement can be licensed by the referents having been mentioned previously in the same conversation. An alternative means by which an expression may be associated with an accessible referent in a given context is that the referent is unique in the 'topic situation' (KRATZER 2014), i.e., the situation around which the conversation revolves at the time of the utterance, which may or may not be the extralinguistic utterance situation. Such uniqueness assumptions may depend on shared cultural (or 'encyclopedic') knowledge. For example, in a conversation about a garden, the definite description *the roses*, used for the first time in this conversation, can

---

[3] We assume that utterances are individuated at the speech act level: a felicitous utterance performs exactly one speech act in the sense of AUSTIN (1962) and SEARLE (1969, 1975). However, we leave open the possible existence of a subatomic level in speech act theory, which might allow for the conflation or amalgamation of "co-speech-acts" in single utterances. Apologies in Western cultures, for example, may be seen as composite acts that involve both the acceptance of blame – a representative or commissive act in SEARLE'S (1969) classification – and the expression of empathy or regret, an expressive act.

generally be expected to pick up a unique referent, namely all and only the roses growing in the garden in question at the 'topic time' (KLEIN 1994; i.e., the time of the topic situation).

Relative givenness is crucially influenced by recency of mention, (presumed) salience in the utterance context, and the availability of competing referents for the same potential expressions (GIVÓN 1983). Additional factors influencing especially the minimum weightiness of the expression required to felicitously introduce a new referent include the availability of a conventional conceptual category under which the referent can be subsumed and the lexicalization of that category in the language in question (GUNDEL et al 1993).

We introduce the term of art **pragmatic role** for the role a referent plays in the *information structure* of the utterance within its discourse context. The roles we have in mind are in particular those of topic and focus. Following BÜRING (1997, 2003), CARLSON (1982), KLEIN & VON STUTTERHEIM (1987, 2002), ROBERTS (1996, 2012), and VAN KUPPEVELT (1995, 1996), we assume that the thematic organization of the utterances that form a coherent discourse can be described in terms of question-answer relations. Every utterance responds to one or more implicit or explicit questions. These **questions under discussion** (QUDs) ultimately derive from the communicative goals of the interaction, which may of course change during the interaction. To make this more concrete, let us cite an actual Yucatec example. This example will also serve us in §3 to introduce some of the structural devices involved in argument realization in Yucatec. The discourse fragment is reproduced in (2). What we identify as topics in the following discussion are underlined; foci are bolded.

(2) a. U láak'e chan k'anche'a',
   U=láak'   le=chan   k'an+che'=a'
   A3=other   DEF=DIM   four+wood=D1
   'Another one of those little chairs (lit. four-sticks) here,'

   b. le pàarte tu'x ku kutal máako',
   le=pàarte   tu'x   k-u=kutal   máak=o'
   DEF=part   where   IMPF-A3=sit:INCH.DIS   person=D2
   'the seat (lit. the part where people sit),'

   c. **chik'in** yàan.   O sea - bèeyo'.[4]
   chik'in   yàan   o sea   bèey=o'
   west(B3SG)   EXIST(B3SG) that.is   thus(B3SG)=D2
   '**in the west** is where it is. Or rather – (yes,) it's like that.'

   d. E k'anche'o', **tseleb ti' to'n** yàan.
   le=k'an+che'=o',   tsel-eb   ti'   to'n yàan.
   DEF=four+wood=D2   side-?(B3SG)   PREP   us   EXIST(B3SG)

---

[4] The director appears to be starting a correction here, aborting it upon realizing that he was in fact correct in the first place.

'The chair (lit. four-stick), **sideways with respect to us** is how it is.'

e. <u>Le tu'x ku naktal máako'</u>, **lak'in** súutul.
   le=tu'x   k-u=nak-tal   máak=o', lak'in   súut-ul
   DEF=where IMPF-A3=lean-INCH.DIS person=D2 east(B3SG) turn/ACAUS-INC(B3SG)
   'The backrest (lit. where one leans (against)) **east** is where it's turned.'

f. **Te – te chik'ino', náats' te lu'mo'**,
   te'l   chik'in=o'   náats'   te=lu'm=o'
   there west=D2   near(B3SG) PREP:DEF=earth=D2
   '**There in the west, close by on the ground**,'

g. **ti' pekekbal** <u>ump'ée chan bòolai'</u>.
   ti'   pek-ekbal   hun-p'éel chan bòola=i'.
   PREP(B3SG) supported.as.if.fallen.down-DIS(B3SG) one-CL.IN DIM   ball=D4
   '**that's** where <u>a little ball is lying</u>.'

h. Ta na'tah bixih?
   t-a=na't-ah   bix-ih
   PRV-A2=intuit-CMP(B3SG) how(B3SG)-?
   'Did you understand how it is?'

This fragment is part of a conversation between two Yucatec speakers that was recorded by the first author in the course of a referential communication study (CLARK & WILKES-GIBBS 1990) on the use of spatial reference frames in Yucatec. The participants – there are two in each trial, a 'director' and a 'matcher' – match photos by describing them, while a screen between them prevents them from sharing a visual field. They are looking at identical copies of photos, which are however placed on the table in front of them in different orders. A trial involves a set of 12 pictures (there are four trials/sets altogether). Each picture features a ball and a chair. The pictures differ from one another in terms of the orientation and disposition (standing up, turned over, lying on one side) of the chair and the location of the ball with respect to it.[5]

In order to enable the matcher to identify a given picture, the director must seek to communicate to the matcher the orientation of the chair and the location of the

---

[5] The objective of this task is to study the use of so called 'spatial frames of reference' in discourse. Reference frames are systems of axes (not unlike coordinate systems in analytical geometry) used to individuate regions of space. The axes may be derived from an observer's body (egocentric frames), a reference entity (intrinsic frames), or the environment (geocentric frames). Which of these options are available and which are preferred for a particular type of task varies with language and culture. See BOHNEMEYER (2011) for detailed Yucatec results and background references and Bohnemeyer et al (in press) for a comparison of data collected with the *Ball & Chair* task from speakers of 11 linguistic varieties.

ball with respect to the chair. Each of these goals can be framed as a question: 'How is the chair oriented?' and 'Where is the ball located with respect to it?' To answer each of these questions, the director has a variety of different strategies at his or her disposal. For example, (s)he may orient the chair by referring to the facing direction of the seat or the back of the chair or the chair in its entirety. Each of these strategies can be framed as a question in its own right, as illustrated in Figure 1.[6] The diagram references the lines of the example (2) that represent the utterance responding to the particular QUD.
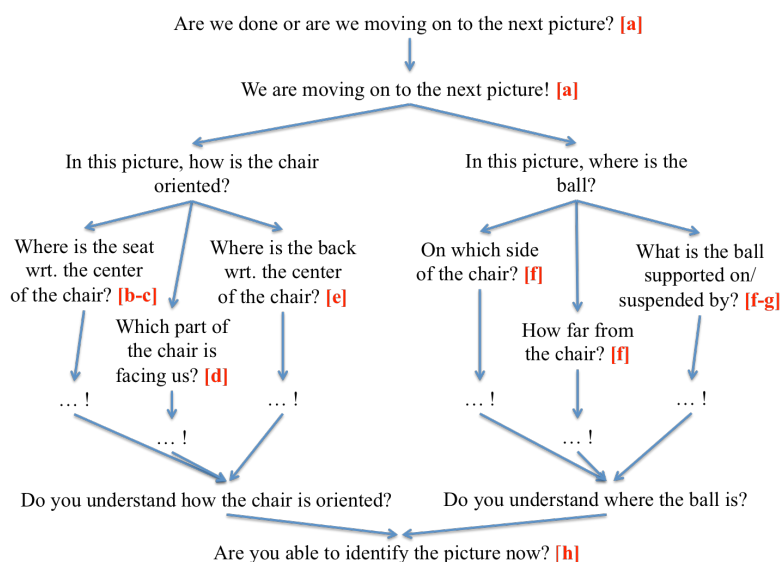
Are we done or are we moving on to the next picture? **[a]**

We are moving on to the next picture! **[a]**

In this picture, how is the chair oriented?

In this picture, where is the ball?

Where is the seat wrt. the center of the chair? **[b-c]**

Where is the back wrt. the center of the chair? **[e]**

On which side of the chair? **[f]**

What is the ball supported on/ suspended by? **[f-g]**

Which part of the chair is facing us? **[d]**

How far from the chair? **[f]**

… !

… !

… !

… !

… !

… !

Do you understand how the chair is oriented?

Do you understand where the ball is?

Are you able to identify the picture now? **[h]**

Figure 1. *Communicative goals, topical organization, and* QUDs *in example (2)*

The rationale behind framing these strategic moves as questions is that each of them opens up a number of discrete alternatives, just like questions do. For example, the QUD 'On which side of the chair is the ball?' opens up the alternatives 'in front of it', 'behind it', 'left of it', 'right of it', 'below it', and 'above it'.[7] We

---

[6] Similar diagrams are employed in BÜRING (1997, 2003) and VAN KUPPEVELT (1995, 1996). BÜRING calls them 'd-trees'.
[7] These are the choices in egocentric or intrinsic terms. If geocentric frames are used, the appropriate labeling will of course depend on the orientation of the set up vis-à-vis the environment. Furthermore, the alternatives listed here are those available in the universe of the Ball & Chair pictures. Outside this set, the ball

informally define the **focus** of an utterance as that part of its meaning – if any – that eliminates some of the available alternatives (ideally, all but one) of the immediate QUD to which the utterance responds (Rooth 1985, 1992). 'Eliminates' here means that if the utterance is accepted by the participants, the proposition entered into the CG will be the open proposition underlying the QUD with the "open" (technically, lambda-abstracted) variable replaced by the meaning of the focus of the response. We define the **topic** of an utterance as comprising the set of discourse referents mentioned in its immediate QUD but not in focus. This approach to defining 'topic' is less psychological, but otherwise broadly compatible with REINHART's (1982) proposal, according to which topics are an organizing principle that helps structure mental representations of the CG. Propositions that are entered into the CG are "stored" and made accessible via the topics to which they relate. However, REINHART's approach can offer at most an indirect explanation for the complementarity of topic and focus, via a preferred association between topics and 'old', previously established information on the givenness dimension motivated by processing considerations. Note also that our definition of 'topic' entails that an utterance can involve multiple topical referents. Our view of topic and focus as complementary notions aligns closely with the Prague School view of theme and rheme as being differentiated in terms of a continuum of 'communicative dynamism', i.e., that which "pushes the communication forward" (FIRBAS 1971: 135-136).[8]

We defer a detailed discussion of the structural resources involved in argument realization in Yucatec to §3. A few comments will suffice to tie (2) back into the current discussion of background assumptions. It will be noticed that up until line f, every sentence starts out with a representation of the topic, followed by a representation of the focus. The representation of the topic (in lines (a), (b), (d), and (e)) always terminates in the particle =*a'* (a) or =*o'* ((b), (d), (e)). This construction is a left-dislocation (LD). The left-dislocated (LDed) NP is always marked as definite, except in the first line, where a possessed nominal refers to the next picture to be taken up by the participants (the next picture is etymologically represented as the previous picture's "other" or "fellow/kin"). The participants were on their third trial at the time; the fact that every trial involves 12 pictures and every picture features exactly one chair and exactly one ball was thus well established. Definiteness was thus licensed by unique identifiability. What motivates the choice of LD – a relatively "weighty" strategy – for reference to these topics is their implicit *contrastiveness*. The ball and the various parts of the chair are about equally likely to be taken up as topics at this stage of the task. This is represent by the top-most branching in Figure 1.

Contrastive topics have attracted considerable attention in the recent literature on information structure. BÜRING (2003), in an approach based like the one

---

could of course also be, say, at a 45° angle with respect to the front of the chair and thus neither clearly in front of nor clearly beside it.

[8] An important potential challenge for the view that topic and focus derive from complementary parts of the QUD is the phenomenon of contrastive topics. We address this issue momentarily.

sketched here on ROBERTS' (1996, 2012), describes them as foci within topics. However, BÜRING does not offer a proposal for the analysis of topics *per se*, only for that of contrastive topics. We believe that it is possible to adapt the tentative definition of 'topic' we proposed above to reconcile it with BÜRING's analysis of contrastive topics, via a hierarchic ordering of QUDs as sketched in Figure 1.

The focus is expressed by a focus construction in lines (c), (d), and (e). Like all Mayan languages, Yucatec is strictly head-initial. The appearance of a spatial relator – *chik'in* 'west' in (c), *tseleb* 'sideways' in (d), and *lakin* 'east' in (e) – in a position between the LDed position and the verb is an expression of focus. (It is in fact arguably licensed by a biclausal, cleft-like construction; cf. §3.) All of these utterances have in fact narrow focus, as Figure 1 predicts.

However, in lines (f) and (g), a peculiar reversal seems to occur: LDs are now used to express part of what according to Figure 1 should be the focus, and the structure of the clause in line h is that normally used for broad focus. And the putative topic, the ball, is referred to using a post-verbal indefinite NP. The form of the utterance suggests that the speaker is responding, not to the ball-related QUDs in Figure 1 (see (3)), but rather to the QUD in (4):

(3)   On which side of the chair is the ball?
        How far from the chair is it?
        What is it supported on or suspended from?

(4)   What is featured in this photo?

The Yucatec speakers this task was conducted with used both of the strategy represented in Figure 1 and the alternative general QUD in (4) for framing the task, and it was not unusual for a speaker to switch between the two in the description of a single picture. However, interestingly, the framing in (4) was almost never used for utterances involving the chair, only for utterances involving the ball. This might be a reflection of the difference in size and featuredness between the two objects.

Given (4), what is the function of the phrases in (2f), which have the same structural properties as the LDs in (2a), (b), (d), and (e)? Since we define the topic and focus of an utterance in terms of the immediate QUD it responds to, it follows that an utterance can include referential constituents that refer neither to its topic nor to its focus, but rather to a discourse referent not mentioned in the immediate QUD. This suggests that there are pragmatic roles beside topic and focus. One candidate for such a role is that of **frame setter** (JACOBS 2001), which corresponds to CHAFE's (1976) notion of 'topic' as a framework that serves to restrict a predication to a particular situation (namely, what KRATZER (2014) calls the 'topic situation').

The examples discussed above make it clear that both givenness and pragmatic role influence argument realization. However, they influence it in different ways. By hypothesis, givenness is the primary semantic factor governing pronominalization, whereas pragmatic role governs the use of information perspective devices such as focus constructions, topicalizations, and left-

dislocations. Since we are primarily interested in the effects of head-marking on realization, and specifically in the role of morphologically bound expressions that semantically behave much like reference tracking devices (pronouns) and morphosyntactically behave as if they saturated argument positions of a head – that is, in cross-reference markers – the corpus study we present in §4 focuses on givenness rather than pragmatic role. However, in analyzing the results of the corpus study, we discuss the apparent semantic conditions of the use of other realization constructions, such as LDs, as well. Therefore, we have tried to sketch a more comprehensive picture of discourse status in this section.

The last question to be addressed here is whether (or to what extent) these two variables are indeed independent of one another, as we assumed above. GUNDEL et al (1993), for example, treat focus as a givenness property. However, in the framework assumed here, it is topicality where the two dimensions meet, since by the definition proposed above, topics are necessarily given. Foci and frame setters, in contrast, can be both given and newly introduced.

3. The grammar of argument realization in Yucatec

The language scientifically known as 'Yucatec' is called *Maya* by its speakers. It gave its name to the Mayan language family. It is spoken by approximately 780,000 first-language speakers in the three Mexican states of the Yucatan peninsula and a few thousands more in neighboring Belize and the Mexican state of Tabasco (LEWIS et al 2015).[9] It forms the Yucatecan branch of the Mayan language family together with its much smaller and lesser-documented sisters Itza', Mopán, and Lacandon.

Typologically, Yucatec is a head-initial and, with the exception mostly of certain operators/functional categories,[10] exclusively right-branching language. It has a split argument marking (or 'alignment') system, which treats the single argument of intransitive predicates sometimes on a par with the actor of a transitive verb and sometimes on a par with the undergoer. What makes this system typologically unusual is that the split is governed neither lexically nor pragmatically, but by an inflectional category that conflates viewpoint aspect and mood (called 'status' in BOHNEMEYER 2002, following a proposal by KAUFMAN (1990) for the entire Mayan language family); cf. BOHNEMEYER (2004), KRÄMER & WUNDERLICH (1999), and references therein.[11]

---

[9] Published data from the Mexican government's 2010 census show 786,113 speakers age five and older (http://cuentame.inegi.org.mx/hipertexto/todas_lenguas.htm; last consulted 3/19/2015).

[10] There is also a small, closed class of around 40 true adjectives, which occur pre-nominally.

[11] Alignment splits governed by viewpoint aspect are well known, especially from Indo-Iranian languages. What distinguishes the system found in Yucatec and, to varying degree, other languages of the Yucatecan and Cholan branches of the

Crucially for present purposes, Yucatec is an entirely head-marking language. There is no case marking of any kind on dependents. We use the term 'head-marking', not in the broad sense of NICHOLS (1986), but in the narrower sense of VAN VALIN (1985). In this usage, it is equivalent to 'cross-reference' (BLOOMFIELD 1933: 191-194) and the 'concentric' syntactic type of MILEWSKI 1950. A construction is **head-marking** in this sense if, and only if, it involves a head that carries one or more bound morphemes indexing the fillers of its argument positions, and that saturate them in the absence of clause-mate coindexed NPs or **co-nominals**. We use the traditional term **cross-reference markers** for the bound indices; other common terms include 'bound pronominal arguments' and 'argument affixes'. A construction is **dependent-marking** in the narrow sense if, and only if, it involves a nominal dependent and its head such that the dependent bears a morphological reflex of the (semantic and/or syntactic) relation to the head – in other words, a case-marker, irrespective of whether the nominal is in fact a syntactic argument of the head or merely coindexed with an argument. A construction is **double-marking** if, and only if, it is both head-marking and dependent-marking. A language is purely head-marking if, and only if, its grammar licenses no dependent-marking (or double-marking) constructions.

Examples (5) and (6) illustrate head-marking in Yucatec. In (5), two possessed nominals are used as heads of nonverbal predicates. A first-person 'Set-A' clitic cross-references the possessor and a second-person 'Set-B' suffix the theme, the semantic subject of the predication. The following example (6) features the same two markers cross-referencing the actor and undergoer of a transitive verb. This distribution is found across the Mayan language family. Up to this point, it instantiates ergative alignment.

(5) Síi   **in**=ìiho-**ech**,   **in**=pàal-**ech**,   ko'x!
    yes   A1SG=son-B2SG A1SG=child-B2SG   HORT
    '**You** ARE **my** son alright, **you** ARE **my** child; let's go!' (Lehmann ms.a)

(6) T-**inw**=il-ah-**ech**   te=ha'ts+kab+kìin=a'
    PRV-A1SG=see-CMP-B2SG   PREP:DEF=divide:PASS+earth+sun=D1
    '**I** saw **you** this morning.'

Like most Mayan languages, Yucatec has two paradigms of cross-reference markers. Mayanists have become accustomed to labeling these 'Set A' and 'Set B.' Table 1 summarizes the distribution of the two paradigms across syntactic contexts. Table 2 lists the forms.

Table 1. Distribution and functions of the two paradigms of Yucatec cross-reference markers.

Environment	Set A	Set B

Mayan language family from this better known system is that the split occurs in intransitive clauses in the Mayan case, but in transitive clauses in the Indo-Iranian case.

Transitive verbs (active voice)	A(ctor)	U(ndergoer)
Intransitive verbs; transitive verbs in non-active voice	S (the single argument of intransitive clauses in incomplete 'status')	S (the single argument of intransitive clauses in completive, subjunctive, extrafocal 'status')
Other lexical categories	Possessor of nominal	S of non-verbal predicates

Table 2. The morphological forms of the two paradigms of cross-reference markers.

Number	Person	Set A	Set B
SG	1	*in(w)=*	*-en*
	2	*a(w)=*	*-ech*
	3	*u(y)=*	*-Ø (/-ih)*
PL	1	*(a)k=…(-o'n)*	*-o'n*
	1 INCL	*(a)k=…-o'ne'x*	*-o'ne'x*
	2	*a(w)=…e'x*	*-e'x*
	3	*u(y)=…o'b*	*-o'b*

As illustrated in (7), the co-nominals are syntactically optional. In their absence, the cross-reference markers are interpreted like pronouns.

(7) a. T-**u**_i=nes-ah-∅_j   [hun-túul   pàal]_j [le=xoh]_i=o'
  PRV-A3=gnaw-CMP(B3SG) one-CL.AN child DEF=cockroach=D2
  'The cockroach bit a child.' [elicited]

  b. T-**u**=nes-ah-∅
  PRV-A3=gnaw-CMP(B3SG)
  'It bit it.' [constructed]

EVANS (1999) and HASPELMATH (2013) challenge the traditional assumption that cross-reference markers are bound pronominal arguments. Evans shows that some head-marking languages use cross-reference markers indiscriminately in contexts where they pick up a given referent (i.e., are interpreted definitely, insofar as they behave similarly to definite descriptions), introduce a new one (i.e., are interpreted indefinitely), or are used non-referentially (e.g., in impersonal contexts such as those in which English speakers might use *one* or generic *you*, French speakers use *on*, and German speakers use *man*). He cites primary data from Bininj Gun-wok (Gunwinygu; Northern Territory, Australia), but also draws on secondary data from a range of other languages. He argues that this versatility does not match the behavior of free pronouns in European languages and suggests that cross-reference markers are inherently non-referential and should really be treated as agreement markers. In response, MITHUN (2003) shows that Yup'ik (Eskaleut, Alaska) and Navajo (Athabaskan, Arizona and New Mexico) reserve **bare cross-**

reference markers – cross-reference markers unaccompanied by co-nominals – for picking up given referents ('definite' uses) and use other expressive devices in indefinite and non-referential contexts. This is true for Yucatec as well. Indefinite and non-referential uses require a cross-reference marker to be accompanied by a co-nominal. Example (2e) above, replicated here for convenience, illustrates impersonal reference. To express this, the A3 marker *a=* combines with the bare nominal *máak* 'person'. 'Where one leans against' is thus literally rendered as 'where (a) person leans against':

(8) Le=tu'x   k-**u**=nak-tal   **máak**=o', lak'in   súut-ul
     DEF=where IMPF-A3=lean-INCH.DIS person=D2 east(B3SG) turn/ACAUS-INC(B3SG)
     'The backrest (lit. where one leans (against)), east is where it's turned.'

Without the nominal *máak* 'person', *u=* would have to be interpreted as picking up a given referent.

Example (2g), repeated in (9), illustrates a referential indefinite expression used to introduce a new referent. It combines the phonologically unpronounced B3SG suffix with a co-nominal constituted by the numeral *hun-* 'one', the 'generic' classifier for inanimate referents *p'éel*, an optional diminutive marker, and the head *bòola* 'ball':

(9) (…) ti'   pek-ekbal   **hun-p'éel chan   bòol**a=i'.
     PREP(B3SG) supported.as.if.fallen.down-DIS(B3SG) one-CL.IN   DIM   ball=D4
     '(…) that's where a little ball is lying.'

Without the co-nominal, (9) would be understood to refer to some contextually given individual. The utterance would be infelicitous in a context in which no obvious candidate is available.

Lastly, whereas the indefinite description in (9) introduces a referent as an instance of the category lexicalized by the head (in this case, the category 'ball'), (10) illustrates a strategy used for introducing a referent without specifying more than the broadest, most general category possible. This strategy is used for example in order to suggest that the identity of the referent is unknown. To this effect, the cross-reference marker is coindexed with the pro-forms *máax* for humans or *ba'x* for non-human referents (both animate and inanimate ones), optionally augmented by the interrogative particle *wáah*. In this use, (*wáah*)*máax* and (*wáah*)*ba'x* translate as 'someone' and 'something', respectively. Outside the kind of context illustrated in (10), *máax* and *ba'x* are primarily used as interrogative pro-forms.

(10) Yàan   **wáah+máax** ti'   k-a=ch'a'-ik   ts'àak?
      EXIST(B3SG) ALT+who   PREP(B3SG) IMPF-A2=take-INC(B3SG) cure\ATP
      'Is there someone from whom you get medicine?' (Blair & Vermont-Salas 1967)

Without the co-nominal, (10) would be interpreted as inquiring about the existence of a specific person given in context from whom the addressee receives medicine.

It sum, it appears that bare cross-reference markers are exclusively used for given referents in Yucatec. The findings of the corpus study we present in §4 confirm this. To this extent, they behave like pronouns in European languages.

HASPELMATH (2013) suggests that head-marking may be analyzable as a combination of agreement and argument ellipsis. However, what matters for our purposes is that heads in combination with bare cross-reference markers *behave* semantically *as if* their argument positions were saturated by pronouns: they are morphosyntactically complete, and they occur with given referents, but not by themselves with new ones or in non-referential contexts. It is not clear to us how meaningful the question is as to whether this behavior is attributable to a property of the cross-reference markers alone (in particular, the putative property of being bound pronominal indices) or only to the combination of cross-reference marker, head, and the absence of a co-nominal. If there is no empirical way of adjudicating between these alternatives, then this question is in our view meaningless. Notice, however, that this does not mean that it is not possible to empirically test the hypothesis that head-marking is a combination of agreement and argument ellipsis. A systematic crosslinguistic difference in the possible or typical discourse status of bare cross-reference markers and argument ellipsis would argue against the hypothesis. The present study can be seen as contributing to laying the groundwork for such a test.

The relation between cross-reference markers and co-nominals has been a notorious headache for syntactic theory. The traditional view is that this relation is one of apposition (HUMBOLDT 1836, BOAS 1911, BLOOMFIELD 1933, MILEWSKI 1950, NICHOLS 1986). However, as LEHMANN (1985: 92) observes, this view rests on the otherwise quite non-traditional assumption that there are syntactic relations between phrases and bound morphemes. The co-nominals have also been treated as clausal adjuncts (VAN VALIN 1985), as detached from the clause (JELINEK 1984, PENSALFINI 2004), and most recently as occupying a special 'extra-core slot' position immediately dominated by the clause (VAN VALIN 2013).

BRESNAN & MCHOMBO (1987) propose an ambiguous or 'PRO-drop' analysis (PERLMUTTER 1971) for the subject markers of Chichewa (Bantu; Malawi, Zambia, Mozambique). According to this analysis, the subject markers satisfy the head's argument positions in the absence of co-nominals, but express agreement with the co-nominals in their presence. AUSTIN & BRESNAN (1996) extend this analysis to Warlpiri. BOHNEMEYER et al. (ms.) propose a BRESNAN & MCHOMBO-style PRO-drop analysis for the cross-reference markers of Yucatec. They draw on two sources of evidence. First, distributional evidence suggests that co-nominals can occur as constituents of the same projection that immediately dominates the head, called the **(verbal) core** in BOHNEMEYER et al (ms.), following FOLEY & VAN VALIN (1984), VAN VALIN (ed.) (1993), VAN VALIN (2005), and VAN VALIN

& LAPOLLA (1997), *inter alia*.[12] Consider the examples in (11), which feature instances of the so-called 'motion-cum-purpose' construction, which combines a motion and a non-motion core, the latter describing an event that stands in a purposive relation to the motion event, but is also understood to (be intended to) occur at the endpoint of the motion event.[13]

(11) a. Pablo=e'   t-u=túuxt-ah   **le=pàal**
      Pablo=TOP   PRV-A3=send-CMP(B3SG) DEF=child
      u=ch'a'   le=ta'kin=o'
      A3=take(SUBJ)(B3SG)   DEF=money=D2
      'Pablo, he sent **the child** to take the money'

   b. Pablo=e'   t-u=túuxt-ah   u=ch'a'
      Pablo=TOP   PRV-A3=send-CMP(B3SG) A3=take(SUBJ)(B3SG)
      le=ta'kin   **le=pàal**=o'
      DEF=money   DEF=child=D2
      'Pablo, he sent **the child** to take the money'

The referent of the NP *le=pàal* 'the child' receives the undergoer role (in this case, more specifically a theme role) from the first core and simultaneously the actor (agent) role from the second. This is a control construction: the arguments are coindexed, and maximally one of them can be realized by a co-nominal in combination with the (always obligatory) cross-reference marker. Crucially, the co-nominal can occur both between the verbs (as in (a)) and in sentence-final position. Given the head-initial syntax of the language, the most straightforward explanation for the position of the co-nominal in (a) is that it is a constituent of the first core. Other multi-core constructions show similar behaviors. If and when co-nominals occur as core constituents, it is unclear what their syntactic function could be other than that of arguments of the head. However, a co-nominal and its coindexed cross-reference marker cannot both satisfy the same argument position of the head. An argument position can have at most one filler. This can be (i) the co-nominal, (ii) the cross-reference marker, or (iii) the combination of the two, but not both of them independently of one another. The 'PRO-drop' analysis opts for option (iii), by treating the cross-reference markers as agreement markers when they are accompanied by core-internal co-nominals.

Additional evidence for the PRO-drop analysis comes from a series of two production experiments. Yucatec is a language with optional plural marking. Even if morphological plural marking is absent on both the verb and the co-nominal, this does not prevent a plural interpretation, as illustrated in (12a). The plural marker can also occur on the verb alone, as in (12b), where it instantiates a discontinuous

---

[12] The core is a projection that immediately dominates a head and all of its syntactic arguments, as well as a 'periphery' of modifiers. Verbal cores can be considered subject-internal verb phrases.
[13] Cf. AISSEN 1987 for Tsotsil and ZAVALA MALDONADO 1993 for an overview including other members of the Mayan language family.

constituent of the Set-A cross-reference marker (cf. Table 2 above). This of course forces a plural interpretation. The same holds if the plural marker appears solely on the co-nominal, as (12c). And lastly, the plural marker may appear in both positions, as illustrated in (12d).

(12) a. Táan   u=k'àay   le=x-ch'úupal=o'
        PROG   A3=sing\ATP   DEF=F-female:child=D2
        'The girl(s) is/are singing'

    b. Táan   u=k'àay-**o'b**   le=x-ch'úupal=o'
        PROG   A3=sing\ATP-3PL   DEF=F-female:child=D2
        'The girls are singing'

    c. Táan   u=k'àay   le=x-ch'úupal-**o'b**=o'
        PROG   A3=sing\ATP   DEF=F-female:child-PL=D2
        'The girls are singing'

    d. Táan   u=k'àay-**o'b**   le=x-ch'úupal-**o'b**=o'
        PROG   A3=sing\ATP-3PL   DEF=F-female:child-PL=D2
        'The girls are singing'

If the cross-reference markers are the true syntactic arguments, double marking as in (12d) should be redundant and dispreferred. However, BOHNEMEYER et al (in press) show that the participants in their production studies strongly preferred the pattern in (12d) over all others, and the pattern in (12a) over those in (12b) and (12c). BOHNEMEYER et al interpret these findings as indicating weakly grammaticalized syntactic agreement between the verb and co-nominals, in line with the PRO-drop analysis.

On the PRO-drop analysis, co-nominals can occupy argument positions inside the core. In addition, they occur in adjoined or detached positions, such as the left-dislocations illustrated in (2b), (2d), and (2e) above. Here is (2d) again for convenience:

(13) **Le=k'an+che'=o'**, tsel-eb   ti'   to'n yàan.
     DEF=four+wood=D2   side-?(B3SG)   PREP   us   EXIST(B3SG)
     '**The chair (lit. four-stick)**, sideways with respect to us is how it is.'

The syntactic position of the **left-dislocation** (LD) is identified by the occurrence on the left edge of the sentence, combined with a falling "comma" intonation and a clitic particle such as 'D2' =o' in (13). In the dialect of Yucatec this study draws on (the variety spoken in Quintana Roo and northeastern parts of the state of Yucatán), there are four particles that appear in this position.[14] The

---

[14] HANKS (1990: 18-19) mentions a fifth particle, =be', used with referents of whose presence the speaker has auditory evidence, but no visual evidence.

same four particles also occur at the right edge of a matrix clause, with their realization there being subject to the same set of constraints.

These particles are typologically quite unusual. Similar phenomena occur in other Mayan languages, but we are unaware of close parallels outside the Mayan language family. The function of these particles is to indicate that the clause or left-dislocated (LDed) expression they attach to contains a constituent of certain referential properties. The use of these particles thus directly reflects the discourse status of the referent. The particle D1 =*a'* indicates the presence of an expression with a referent given exophorically in the spatiotemporal proximity of the speech situation. In (2a) above (repeated as (15) below), this is the picture the speaker is about to describe. In (6) above, it is a time interval that forms part of the day of utterance ('this morning').

The particle D2 =*o'* indicates the presence of an expression with a referent that is given in the speech situation or treated as uniquely identifiable in the topic situation (cf. §2), for example because it was previously mentioned. If the referent is understood to be given in the speech situation, a scalar implicature licensed by GRICE'S (1975) first Quantity maxim, "Make your contribution as informative as is required", suggests that it is non-proximal, for otherwise the speaker should chose the more informative =*a'* (cf. BOHNEMEYER 2012). An example can be seen in (13).

In fictional narratives – but apparently only in that genre – D3 =*e'* often replaces D2 =*o'* with referents that are uniquely identifiable in the topic situation. This is illustrated in (14):

(14)   (…) **le=kéeh=e'**, t-uy=a'l-ah   ti' **le=chan=áak=e'**: (…)
       DEF=deer=D3   PRV-A3=say-CMP(B3SG)   PREP DEF=DIM=turtle=D3
       '(…) **the deer**, he said to **the little turtle**: (…) (ROMERO CASTILLO 1964: 307)

Both protagonists, the deer and the turtle, have already been introduced at this point in the story and thus are treated as definite. In a conversation or in non-fictional narratives, D2 =*o'* would be expected in such contexts.[15]

In addition, and across genres, D3 =*e'* occurs with phrases that lack a trigger for D1 =*a'* or D2 =*o'*. The following examples illustrate some of the possibilities. In (15) and (16), the phrase in the LDed position is an indefinite NP. In (15), it refers to a time span that serves as a frame setter (cf. §2). In (16), the LDed phrase is an indefinite co-nominal. This is possible because the sentence has generic reference.

(15)   **Hun-p'éel k'ìin=e'**, chéen táan u=máanhun-túul chàan=áak ichk'áax, (…)

However, this particle appears to be restricted to the variety HANKS' study is based on, which is spoken in the northwest of the state of Yucatán.

[15] In the excerpt of a *Frog Story* narrative in (20) below, both particles occur in the relevant contexts, but =*o'* is more prevalent. It is tempting to speculate that this is the result of the narrative, though fictional, having been elicited as a retelling of a picture book.

one-CL.IN sun=D3   only   PROG A3=pass one-CL.AN DIM=turtle in jungle
'**One day**, a little turtle was just walking about in the jungle, (…)' (ROMERO CASTILLO 1964: 307)

(16)   Pos   **hun-túul   chàan=pèek'=e'**,   **kìinseh díias=e'**,
        well   one-CL.AN   DIM=dog=D3   fifteen   day:PL=D3
        k-u=p'íil-il   uy=ich.
        IMPF-A3=open\ACAUS-INC   A3=eye/face
        'Well, **a puppy**, [within] **fifteen days** [after its birth], its eyes open.'
        (Lehmann ms.b)

Example (16) also illustrates a second left-peripheral phrase terminating likewise in =*e'*. Like the one in (15), this phrase expresses a frame setter, and like the one in (15), it is indefinite.

AISSEN (1992) distinguishes between two 'topic' positions in Mayan languages, a clause-internal one and a clause-external one. Both are distinct from focus positions. Some languages (e.g., Ts'utujul) realize both topic positions; others (e.g., Jakaltek and Tsotsil) only realize the external one. On Aissen's analysis, the internal topic position involves extraction and must bind a coindexed gap, like topicalizations in English, whereas the external one is not subject to such a constraint and can, but need not, be coindexed with a coreferential pronoun inside the clause, as is the case with English LDs. As (16) shows, the Yucatec LD construction involves an external topic position in AISSEN's sense. Further evidence for this analysis is presented in BOHNEMEYER (2009: 189-190). The grammar of Yucatec does not license a clause-internal topic position.

Another very common use of =*e'* at the right edge of LDs is with subordinate clauses (cf. BOHNEMEYER 1998a; 2002: 129-143 for details). If such a clause contains no trigger of =*a'* or =*o'*, it will be marked by =*e'* on its right edge. In (17), this is the case for the antecedent of a conditional:

(17)   **Wáah a=k'áat=e'**,   ko'x chan=áalkab (…)!
        ALT   A2=wish(B3SG)=D3 HORT DIM=run
        '**If you want**, let's go for a run (…)!' (ROMERO CASTILLO 1964: 307)

D3 =*e'* also commonly accompanies independent pronouns in the position of LDed co-nominals, as illustrated in (18):

(18)   **Tèen=e'**, k-in=bin   kutal   tu=xùul   le=bèeh=o'.
        me=D3   IMPF-A1SG=go sit:DIS.INCH PREP:A3=end\ATP DEF=way=D2
        '**As for me**, I go sit down at the end of the road.' (ROMERO CASTILLO 1964: 308)

The independent pronouns of Yucatec are of special interest in this study due to their presumed functional overlap with the bare cross-reference markers. More on them below.

D3 =*e'* also occurs at the right edge of matrix clauses, but (with the exception of fictional narratives; cf. (14)) only when a lexical trigger inside the clause selects for it. The D4 particle =*i'* is exclusively licensed by such a lexical trigger – negation (though only in realis mood)[16] – and by locative focus constructions, as illustrated in (9) above, repeated for convenience in (19):

(19) (…)**ti'**   pek-ekbal   hun-p'eel   chan bòola=**i'**.
      PREP(B3SG) supported.as.if.fallen.down-DIS(B3SG) one-CL.IN   DIM ball=D4
      '(…) **that's** where a little ball is lying.'

Since the particles occur only on the right edge of matrix clauses and on the right edge of LDs, and the latter can be constituted by subordinate clauses (as in (17)), the term **clause-final particles**, understood as a conventional technical term, may not be too far off the mark for them. Only one of the particles can be realized at the right edge of a LDed expression or clause. Should the expression contain multiple distinct triggers, the particle is chosen according to a hierarchy that is reflected by the numbering of the glosses (D1 > D2 > D3 > D4). Cf. BOHNEMEYER (2012) and the literature cited there.

As the above examples illustrate, LDS are quite pervasive in Yucatec discourse. Moreover, Yucatec LDS occur with cross-reference markers whose linking properties strikingly resemble those of subjects in European languages. Let us illustrate this with another extended example, in this case one excerpted from a *Frog Story* narrative collected by CHRISTEL STOLZ in 1992 with the picture book by MERCER MAYER (1969). The three characters, a boy, his pet dog, and a deer, are all given at the topic time of the excerpt. Co-nominals are bolded and LDs are underlined in (20).

(20) a. Pwes, <u>**le=kéeh=o'**</u>,   túun   bin,
          well   DEF=deer=D2   PROG:A3   go
          'Well, <u>**the deer**</u>, it is going,...'

      b. u=kuch-mah   **le=pàal**   y=éetel
          A3=carry.on.back-PERF(B3SG)   DEF=child   A3=COM
          '...it has **the child** shouldered ("backed") with …'

---

[16] In Yucatec, realis mood occurs with factual past- and present-time reference to individual states of affairs, whereas irrealis mood occurs with future-time reference, habitual and generic reference, and counter-factual past- and present-time reference. The contrast is overtly expressed in subordinate clauses, but remains largely covert in matrix clauses except for the distribution of =*i'* under negation. This mood contrast has not so far been described comprehensively in publication, although Bohnemeyer (2002) contains references to it throughout. Combined with the distribution of =*e'* in fictional vs. non-fictional narratives mentioned above and the observations in Hanks (1984, 1990) regarding aspects of evidentiality in the use of the particles, it suggests a possible broader veridical orientation of the particle system that remains to be studied.

c. u=ho'l=o',   táan   u=bin.
    A3=head=D2   PROG   A3=go
    '…its head, as it is going.'

d. Pwes, káa=h [new start] **le=pèek'** xan=e'   te'l
    well CON=PRV   DEF=dog also=D3   there
    'Well, (when/and then) [new start] **the dog** as well, there...'

e. ts'ay-a'n   tu'x   yàan   t-u=pàach
    hit-RES(B3SG) where   EXIST(B3SG) PREP-A3=back
    '…it was hit where it was behind...'

f. u=yùumil=o',   táan xan   u=tohol-t-ik
    A3=master=D2   PROG also A3=bark-APP-INC(B3SG)
    '...its master, the dog was also barking at…'

g. **le=kéeh**   xan=o'; pwes, **le=kéeh=o'**,
    DEF=deer   also=D2 well DEF=deer=D2
    '…the deer; well, as for the deer, …'

h. chich   u=bin   túun=e'.
    hard(B3SG)   A3=go   so.then=D3
    '...fast was how it went.'

i. Le=káa=t-u=pik+ch'iin-t-ah
    DEF=CON=PRV-A3=fling+pelt\ATP-APP-CMP(B3SG)
    '(When/and then) it threw off…'

j. **le=pàal=o'**,   káa=h-lúub   **le=pàal=e'**,
    DEF=child=D2   CON=PRV=fall(B3SG) DEF=child=D3
    '...**the child**, (when/and then) **the child** fell, …'

k. tak   **le=pèek'** túun=o',   h-lúub-ih.
    as.far.as DEF=dog so.then=D2 PRV-fall-CMP(B3SG)
    '…and even **the dog**, it fell.'

The three LDs in this excerpt are coindexed with the single argument of an intransitive verb in lines (a) and (g)-(h) and with the undergoer of a passivized transitive verb form in lines (d)-(e). Moreover, the intransitive clause in line (a) is followed by a clause projected from a transitive active verb form in line (b), with the actor co-referential with the LD. An example in which a LD is coindexed with the actor of an active transitive verb form that immediately follows it can be seen in (14) above. What does not occur in this textlet and in Yucatec discourse in general is LD of an undergoer nominal of a transitive active verb form. The results of a large-scale production experiment confirm this (BUTLER et al ms.). It is

conceivable that this pattern is the result of extension and regularization of the LD construction under contact influence from Spanish clause structure, with Yucatec speakers using LDs to emulate Spanish subjects. However, contrary to DURBIN & OJEDA (1978) and GUTIÉRREZ-BRAVO & MONFORTE (2010), LDs cannot synchronically be analyzed as verb-initial subjects. At least four properties rule out a subject analysis:

- The contested position can be filled by nominals not co-indexed with cross-reference markers, as in (15)-(17) above;
- A single sentence can carry multiple instances of the contested construction, as illustrated in (16) above;
- When a sentence does contain multiple such constructions, one constituted by an adverbial can intervene between one filled by a co-nominal and the matrix clause, as in (16);
- Predicate-initial sentences without a filler of the contested position do occur in conversation; an example is (10) above.[17]

So far, we have introduced three structural devices for argument realization in Yucatec: bare cross-reference markers, clause-internal co-nominals, and LDed co-nominals. BOHNEMEYER (2009) formulates hypotheses regarding the use of these in discourse that are summarized in Table 3, recast in the framework introduced in §2:[18]

Table 3. *Yucatec resources for argument realization (adapted from* BOHNEMEYER *2009:195)*

Referent Realization	Given	New
Bare cross-reference marker	Extending 'topic chains'	N/A
Cross-reference marker plus clause-mate co-nominal	Resuming discontinued topics; maintaining multiple competing topic chains	Introducing new referents
Cross-reference marker plus LDed co-nominal	Contrastive topics and frame setters	Introducing new referents in thetic utterances and as topics in generic categorical utterances

The notion of the 'topic chain' was introduced in GIVÓN 1983 without receiving a clear definition there. A narrow and somewhat simpleminded definition would be

---

[17] GUTIÉRREZ-BRAVO & MONFORTE (2010) treat such sentences as evidence of a constituent order split.
[18] The notion of the 'topic chain' and that of 'thetic utterances' are introduced in the discussion immediately following Table 3.

the following: a **topic chain** is a sequence of references to the same discourse referent, distributed across contiguous utterances. One question this definition raises is whether the utterances in the topic chain have to be adjacent. In many if not all languages, there are contexts in which a referent is picked up from an antecedent in a non-adjacent utterance using a structural device that is also prototypically used for picking up referents from antecedents in immediately preceding utterances. Consider (21):

(21)   *Sally waved to Floyd. A car passed by. She motioned for him to approach.*

Clearly, the intervening sentence does not prevent the last sentence from picking up the referents from the antecedents in the first. It is as if the second sentence has no impact at all on the topic chain. But if topic chains can leap utterances, then how are we to distinguish between such leaping chains and the resumption of a topic that was previously discontinued? Table 3 predicts that such discontinued topics require more structural effort for their resumption. But how are we to test this hypothesis without an independent semantic diagnostic of topic chains? In the corpus study we present in §4, we circumvented this problem by restricting the coding of topic chains to adjacent utterances.[19]
　　Despite the label, a referential expression that extends a topic chain does not have to be 'topical' in the sense introduced in §2. It can also be in focus, as illustrated by the pronoun in the exchange in (22):

(22)   *Who filed the complaint, Sally or Floyd? – HE did!*

Topic chains in GIVÓN's sense are a givenness phenomenon, not a topicality phenomenon.
　　Examples (2) and (20) provide ample illustration of the use of LDs for the expression of contrastive topics and frame setters. But according to Table 3, LDs also occur with new referents. One type of context in which this happens is under generic reference, as in (16) above. Another case in point is thetic utterances. A **thetic** utterance is one in which all discourse referents are newly introduced (KURODA 1972; LADUSAW 1994; SASSE 1987). As a result, thetic utterances also have no topic in the framework introduced in §2. An example of a LD in a thetic utterance is (23), the introductory sentence of a fictional narrative:

(23)   Hun-túul   xib=e',
         one-CL.AN   male=D3

         h-ts'o'k   u=bèel   y=éetel   hun-túul   x-ch'úupal
         PRV-end(B3SG) A3=way   A3=COM   one-CL.AN   F-female:child

---

[19] A residual problem is whether referring expressions in subordinate clauses and projections below the clause level can initiate or continue topic chains. We address this question in §4.

ma'   t-uy=ohel-t-ah   wáah x-wàay=i'.
NEG(B3SG) PRV-A3=knowledge-APP-CMP(B3SG) ALT   F-sorcerer(B3SG)=D4

'A man, he married a girl not knowing that she was a witch.' (ROMERO CASTILLO 1964: 305)

Two further options for argument realization in Yucatec not listed in Table 3 because they are not considered in Bohnemeyer 2009, but which we would like to add here, are focus positions and independent pronouns.[20] Several focus constructions are features in (2) above; but in all of these, the referent of the focus constituent stands in a spatial relation to the background clause. The following examples illustrate focus constructions in which the focus constituent is coindexed with an argument of the background clause:

(24) a. Tèech=wáah   túun **le=**k-a=k'ay=a',   mùuch? - Pos tèen!
         you(B3SG)=ALT CON DEF=IMPF-A2=sing=D1 frog   well me(B3SG)
         'So are you the one who is singing here, frog?' – 'Well I am!' (Lehmann ms.c)

      b. Tèech=wáah   túun k-a=k'ay=a',   mùuch?
         you(B3SG)=ALT CON IMPF-A2=sing=D1   frog
         'So are YOU singing here, frog?' [constructed]

The construction in (24a) involves a nominalized subordinate background clause and is thus readily identifiable as a kind of cleft. The question is whether the same string, but without the nominalization, as in (24b), is likewise a cleft or whether it instead instantiates a mono-clausal focus construction. Proponents of the cleft analysis include BOHNEMEYER (2002, 2009); BRICKER (1979); TONHAUSER (2003, ms.); and VAPNARSKY (2013). The mono-clausal analysis has been advocated by GUTIERREZ-BRAVO & MONFORTE (2009) AND VERHOEVEN & SKOPETEAS (2015). The correct choice between these competing analyses hinges on a complex set of properties, several of which are contested, with different authors citing conflicting data. A discussion that does the issues justice would take us too far afield here.

Lastly, we have now seen the independent pronouns of Yucatec in LDed (cf. (18) above) and focus positions (in (24)). Clause-internally, they mostly occur as oblique arguments, as in (25):

(25)   A'l   **tèen**, José, ba'x   le=he'l=o', ba'x   u=k'àaba'.
       say(B3SG) me   José what(B3SG) DEF=PRSV=D2 what(B3SG) A3=name

---

[20] In addition to left-dislocations, there are right-dislocations in Yucatec (called 'antitopics' in BOHNEMEYER (2002: 133-135), following LAMBRECHT 1994). However, it is not clear that simple NPs ever occur in this position. If they do, they do so exceedingly rarely.

'Tell **me**, José, what that there is, what is it's name.' (BLAIR & VERMONT-SALAS 1965)

Expressions of recipients and similar semantic roles are not cross-referenced on the verb and are not targeted by any voice operations. We therefore treat them as obliques.

The full paradigm of independent pronouns is represented in Table 4:

Table 4. *The paradigm of independent pronouns*

Number	Person	Form
SG	1	*tèen*
	2	*tèech*
	3	*leti'*
PL	1	*to'n*
	2	*te'x*
	3	*leti'o'b*

The first- and second-person forms can be analyzed as portmanteaus formed out of the general purpose preposition[21] *ti'* and the appropriate Set-B suffix (cf. Table 2 above). In contrast, the third-person forms are analyzable as nominalizations of *ti'*. One piece of evidence in support of these analyses comes from the fact that *ti'* rarely co-occurs with the first- and second-person pronouns in the same environments in which it is obligatory with the third-person pronouns:

(26) a. Hay-p'éel àanyos yàan   **tèech**?
         how-CL.IN year:PL EXIST(B3SG)   you
         'How old are you (lit. how many years are **with you**)?'

     b. Hay-p'éel ha'b yàan   **ti'**   **leti'-o'b**?
         how-CL.IN year EXIST(B3SG)   PREP   it-PL
         'How old are they (lit. how many years are **with them**)?' (BLAIR & VERMONT-SALAS 1965)

This concludes our survey of the grammar of argument realization in Yucatec. As we move on to the presentation of the corpus study in §4, we have three objectives: First, we would like to conduct a quantitative test of the generalizations in Table 3. Secondly, we hope to elucidate the functional division of labor between bare cross-reference markers and independent pronouns. The latter are not represented in Table 3, and since they overlap with the bare cross-reference markers semantically and with co-nominals in terms of their syntactic distribution (as in (18)), their use is of obvious interest for the study of realization in head-

---

[21] *Ti'* is a default marker for adjuncts and obliques across the board, with the exception of instruments and comitatives. Having such a single semantically (nearly) empty general-purpose adposition is not unusual for Mesoamerican languages, especially for Mayan languages.

marking languages. And lastly, we aim to prepare the ground for a comparison of argument realization across head-marking languages.

4. Realization in Yucatec narratives: A small corpus study

In order to quantify the distribution of the various realization strategies in Yucatec discourse, we conducted a small corpus study involving four fictional narrative texts. Our decision to focus on narratives was motivated in part by utilitarian considerations: most of the literature on argument realization has been based on narrative discourses (including, e.g., BICKEL 2003; DU BOIS 1987; and the contributions in GIVÓN 1983), and narrative is also the genre in which by far the greatest number of transcribed, analyzed, and glossed Yucatec texts is available. Furthermore, from a theoretical perspective, there is a well-motivated expectation of the properties of argument realization being maximally simple in fictional narrative discourse. In particular, the narrator can generally expect to be the sole participant with active control over the Common Ground; the role of exophoric reference, and with it that of gestural and gaze cues and other nonlinguistic information in reference resolution, is minimized; and interference effects from other interactions (such as givenness of a referent due to mention in a prior conversation) likewise tend to be minimal. All of these factors should ensure relatively ideal conditions for the study of argument realization. By the same token, however, generalizations from narrative texts should not be assumed to hold for other genres of discourse without modification.

The corpus our study draws on consists of four texts. Information about these is listed in Table 5 below. The amount of linguistic material each text comprises is measured in 'predication units' in Table 5, a term of art for a structural unit that has proven useful in previous corpus studies on Yucatec (BOHNEMEYER 1998b; 2003).[22] A **predication unit** comprises no more than one clause, regardless of whether it is a matrix clause or a subordinate one. Furthermore, any expression that is dependent on a clause and a co-constituent of the same sentence, but is not a constituent of the clause itself, belongs to the same predication unit as the clause in question. This covers adjoined/detached material such as LDs and topicalizations. A predication unit comprises maximally a single conversational turn. However, we treated small (usually single-word) turn-constituting units that form an intonation unit with a sentence and have a conventionalized sequential position with respect to it as part of an appropriate predication unit – usually the one that contains the matrix clause. This applies to vocatives[23] and interjections such as in ***Yes, Virginia, there is a Santa Claus***, which we would thus code as a single predication unit.

---

[22] BOHNEMEYER (1998b, 2003) uses the term 'utterance unit' for what we call 'predication unit.'

[23] LAUNEY (2004: 64-65) suggests that vocatives are predications in 'omnipredicative' languages such as Classical Nahuatl, i.e., languages in which the privilege of heading a syntactic predicate is not restricted in terms of lexical category. And as BOHNEMEYER (2002: 108-129) and VAPNARSKY (2013) show, Yucatec qualifies as an omnipredicative language at least to some extent. Indeed,

To illustrate, here is how we would segment (23) and (24a) above into predication units, using braces for tagging to avoid unwanted phrase structure interpretations:

(23') {Hun-túul   xib=e',
        one-CL.AN   male=D3

        h-ts'o'k   u=bèel y=éetel   hun-túul   x-ch'úupal}
        PRV-end(B3SG) A3=way A3=COM   one-CL.AN   F-female:child

        {ma'   t-uy=ohel-t-ah}   {wáah x-wàay=i'.}
        NEG(B3SG)   PRV-A3=knowledge-APP-CMP(B3SG)   ALT F-sorcerer(B3SG)=D4

        'A man, he married a girl not knowing that she was a witch.' (ROMERO
        CASTILLO 1964: 305)

(24) a.'{Tèech=wáah túun} {le=k-a=k'ay=a',   mùuch?} - {Pos tèen}!
        you(B3SG)=ALT CON   DEF=IMPF-A2=sing=D1 frog   well me(B3SG)
         'So are you the one who is singing here, frog?' – 'Well I am!' (LEHMANN
        ms.c)

Note that whereas the LD in (23') forms a predication until with the following matrix clause, the focus constituent on the left edge of (24a') constitutes its own predication unit on the cleft analysis of the focus construction, which we assume here. Subordinate clauses such as the right-most two clauses in (23') form their own predication units regardless of their syntactic position in the sentence, which is why tagging in terms of predication units does not quite align with phrase structure.[24]

Title	Narrator	Recorded	Published	Sentences	Predication units	Predication units included in the analysis
Bix kahnal le nukuch máako'b úuch	VICENTE EK CATZIN	By JÜRGEN BOHNE-	N/A	357	621	318

there is preliminary morphological evidence supporting a predicative analysis of Yucatec vocatives; but the issue awaits further study. We did not code vocatives as predication units.

[24] We used orthography as the criterion for identifying sentences. Orthography in turn reflects above all intonation. That sentences comprise on average multiple predication units – i.e., *mutatis mutandis*, clauses – in Yucatec narratives may seem surprising. It should be borne in mind, though, that clauses often consist of a single morphological (let alone phonological) word in Yucatec and that they are commonly integrated into sentences, not by hypotaxis, but by structural devices situated in a grey area between hypotaxis and parataxis.

wayo' 'How the ancestors used to live here in the old days' *(Kahnal)*		MEYER in Yaxley, Quintana Roo, in 1999				
Huntúul kòolkab 'A campesino' *(Kòolkab)*	DOMINGO DZUL POOT	N/A	DZUL POOT (1986: 15-23)	65	152	114
Huntúul óotsil koonchúuk 'A poor charcoal seller' *(Koonchúuk)*	AMBRO-SIO DZIB	By Manuel J. Andrade in Chichen Itzá, Yucatán, in 1930	ANDRADE & MÁAS COLLÍ (1991: 426-433)	52	145	66
T'u'l yéetel hkoh 'Rabbit and Puma'[25] *(T'u'l & Hkoh)*	BERNARD-INO TÚN		ANDRADE & MÁAS COLLÍ (1990: 502-517)	95	279	162
Total				569	1197	660

As is apparent from the table, we excluded some material from the analysis. There were two types of exclusions. First, we excluded all character utterances represented as direct speech. This amounts to *all* character utterances in these texts; indirect speech is extremely unusual in Yucatec stories. Two factors motivated the exclusion of *oratio recta* from the analysis. The first of these factors is the genre difference, as direct speech is conversational (notwithstanding the conversation being fictional when embedded in fictional narratives). The second reason for the exclusion is that character speech is anchored to a Common Ground that is distinct from, though not entirely independent of, the Common Ground of the narrative text.

The second exclusion applies uniquely to the *Kahnal* text (the first text listed in Table 5). This text has actually a composite genre structure. The first 87 sentences, or 177 predication units, are descriptive rather than narrative. The speaker describes the conditions of life in remote settlements in the jungle "in the old days", i.e., before the advent of modern infrastructure. We only included the second part of the *Kahnal* text in our analysis, which is a demon story. The two parts are tied together by the descriptive part providing background for the demon story and the narrative at the same time providing an illustration of the difficult conditions described in the first part. We excluded the descriptive part from the analysis primarily due to the pervasive occurrence of generic and habitual reference (cf. BOHNEMEYER 2003 for details).

---

[25] ANDRADE & MÁAS COLLÍ translate this title into Spanish as 'El conejo y el coyote'. However, the Yucatán peninsula is not part of the traditional range of coyotes. Yucatec speakers customarily translate *koh* as 'león', which in Yucatecan Spanish is used for any large cat.

We coded the referents associated with the arguments positions of both matrix and subordinate verbal cores and stative predicates in the texts for four givenness levels: 'new'; 'old chain-medial/final'; 'old chain-initial'; and 'other'. 'Other' here mainly refers to argument positions coindexed with utterances, clauses, and verbal cores. We treated a referent as chain-medial/final if, and only if, it was mentioned in the immediately preceding clause. A chain-initial old referent is an old referent that is resumed after having been discontinued as a topic. We did not code the referents for pragmatic role, as our main concern in this study is with pronominalization; above all, with the use of bare cross-reference markers and free pronouns.

Among realization strategies, we coded for the following options: 'LDed non-pronominal NPs'; 'focalized non-pronominal NPs'; 'clause-mate non-pronominal NPs'; 'morphologically unbound pro-forms' in any syntactic function coindexed with an argument position of a verbal core; 'bare cross-reference markers'; and 'other'. LDed, focalized, and clause-internal co-nominals are of course accompanied by cross-reference markers on the verb or stative predicate. Morphologically unbound pro-forms include the emphatic pronouns discussed at the end of §3, demonstrative pro-forms such as (27) below (or pronominal demonstratives; cf. BOHNEMEYER 2012), and indefinite pro-forms used in content questions (e.g., (25) above) and as pronominal indefinite NPs (e.g., in (10) above), including as relative clause constructions with pronominal heads. This 'other' category comprises verbal cores, clauses, and utterances coindexed with argument positions of a verbal core and reflexive constructions.

(27)   A=ti'a'l   **lel=a'**?
         A2=property(B3SG)   DEM=D1
         'Is **this** [pointing] yours?'

Figures 2-5 break down the distribution of the realization strategies over the givenness levels:
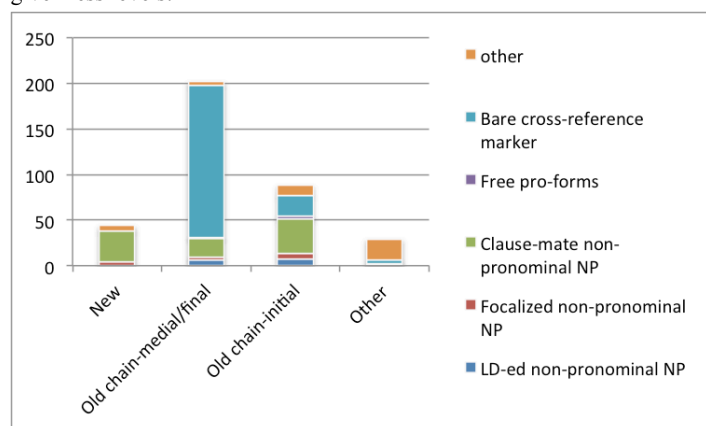
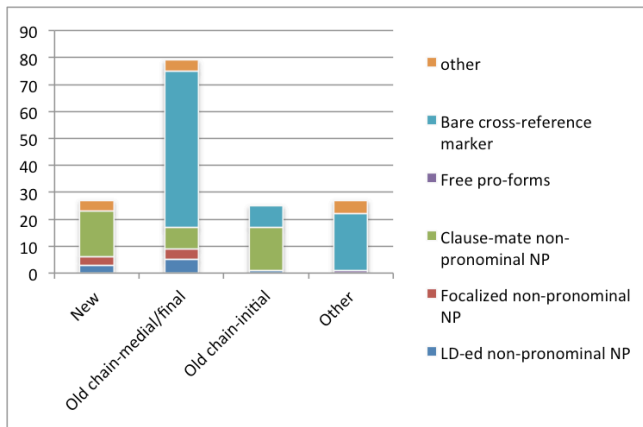**Figure 2.** *Frequency of realization strategies by givenness levels in the* Kahnal *narrative text*

Legend:
- other
- Bare cross-reference marker
- Free pro-forms
- Clause-mate non-pronominal NP
- Focalized non-pronominal NP
- LD-ed non-pronominal NP

Figure 3. *Frequency of realization strategies over givenness levels in the* Kòolkab *narrative text*

Legend:
- other
- Bare cross-reference marker
- Free pro-forms
- Clause-mate non-pronominal NP
- Focalized non-pronominal NP
- LD-ed non-pronominal NP
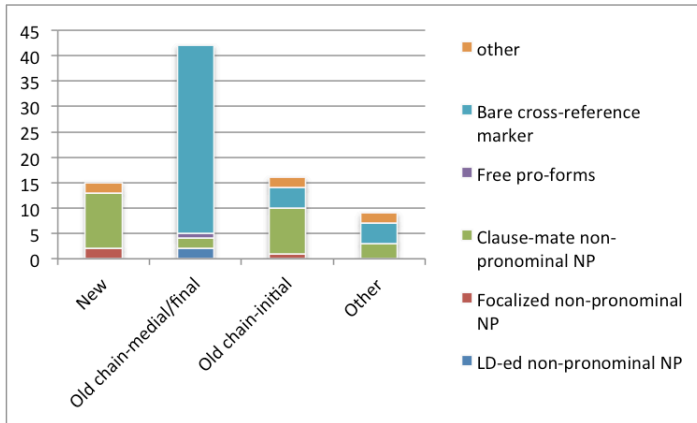
Figure 4. *Frequency of realization strategies by givenness levels in the* Koonchúuk *narrative text*
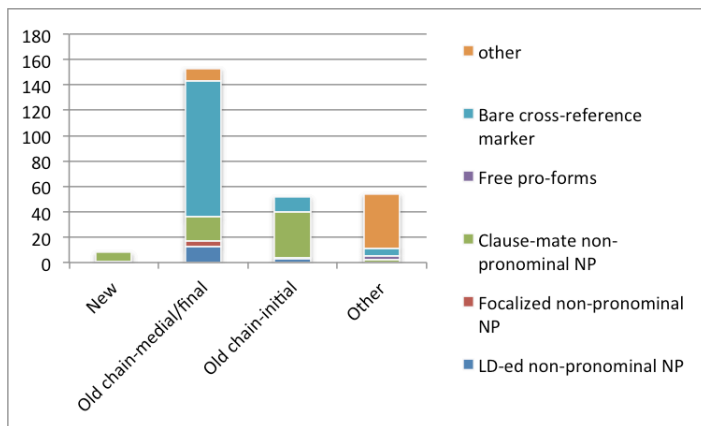
Figure 5. *Frequency of realization strategies by givenness levels in the* T'u'l & Hkoh *narrative text*

The distribution of the strategies seems overall remarkably similar across the four texts. The differences are mainly the following:

- The *T'u'l & Hkoh* text stands out for its surprisingly small number of new discourse referents introduced in the course of the story. This appears to be an artifact of referents often being introduced in this text in places where we did not count as them as such: direct speech and locative adverbials.
- The *Kahnal* and *T'u'l & Hkoh* texts show a much higher incident of cross-reference markers coindexed with utterances or dependent cores/clauses than the other two texts. This is due to differences in how direct speech is marked in these texts: in *Kahnal* and *T'u'l & Hkoh*, the narrator used primarily lexical speech act verbs, whereas a formulaic quotative construction dominates in the case of *Kahnal* and in *Koonchúuk*, direct speech is often not marked at all.

Figure 6 summarizes the distribution across the four texts. For the purposes of inferential statistics, we lumped LDs and clause-mate non-pronominal NPs and distributed pro-forms across focus constructions and 'Other' strategies. The resulting breakdown, visualized in Figure 7, is highly significant (df = 9; χ^2 = 577.3; p < 0.001).

The findings can be summarized as follows:

- New referents are overwhelmingly introduced by clause-mate non-pronominal NPs. Marginally, LDs are also used for this purpose, but never bare cross-reference markers.

- For the extension of topic chains, bare cross-reference markers appear to be the default strategy, although LDed and clause-internal co-nominals are likewise used for this purpose. However, free pronouns play no more than a marginal role in Yucatec narratives.
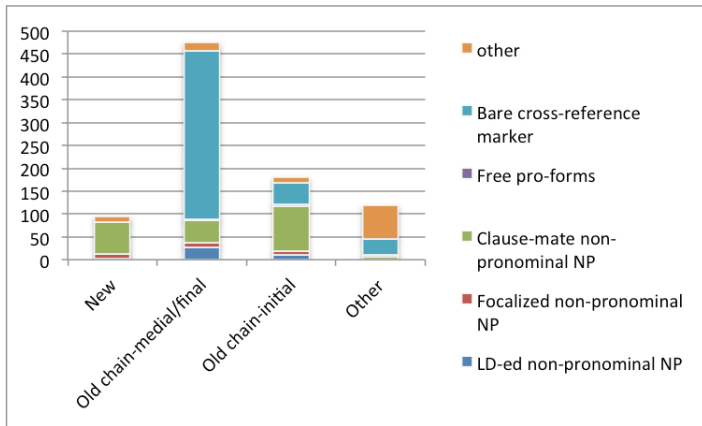

Figure 6. *Frequency of realization strategies by givenness levels across the narrative corpus*
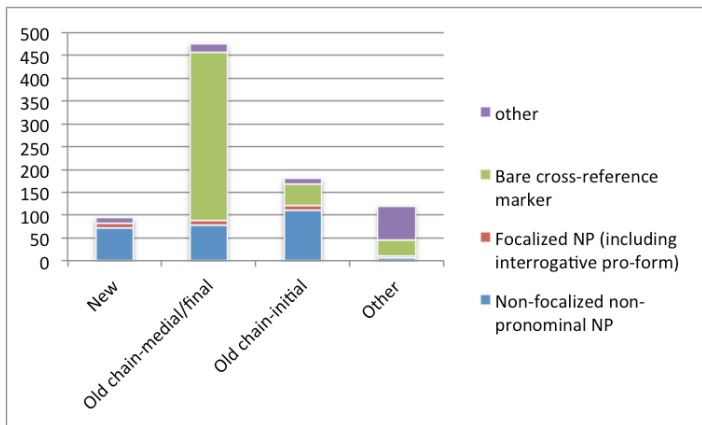

Figure 7. *Simplified distribution of realization strategies over givenness levels across the corpus*

- Bare cross-reference markers and clause-internal co-nominals compete for the function of resuming a temporarily discontinued topic (in the Givónian sense), with co-nominals having a slight (and not necessarily significant) advantage of numbers.

Thus, despite the availability of free pronouns in Yucatec, it is not them, but bare cross-reference markers that are used to continue topic chains, pragmatically resembling the free pronouns of configurational dependent-marking languages such as English. In the next section, we offer a discussion of these findings and attempt to draw some preliminary comparison to available data on argument realization in other head-marking languages.

5. Discussion and comparison

There are two types of morphologically unbound pro-forms in Yucatec that are semantically definite and thus could in theory be used to extend topic chains: the free personal pronouns listed in Table 4 above and the demonstrative pro-forms exemplified in (27) above. However, our corpus does not contain a single token of either device in the function of extending a topic chain. This is by no means a surprising finding. It is essentially a confirmation, based on corpus data, of the observation that bare cross-reference markers are semantically anaphoric in this language. Combined with Grice's (1975) third Manner Maxim ("Be brief (avoid unnecessary prolixity)", the anaphoricity (or pronominal force) of bare cross-reference markers ensures that free pro-forms are not used to extend topic chains unless their use for this purpose in a given context is motivated by additional factors.

In our corpus, demonstratives were used anaphorically only in reference to previously mentioned states of affairs, not to individuals.[26] We found just two tokens of free personal pronouns. They were used for disambiguation when an interrupted topic chain was resumed, as illustrated in (28):

(28) (…) ba'l=e',   k-uy=il-ik   t-u=k'uch-ul
      thing=D3   IMPF-A3=see-INC(B3SG) PREP-A3=arrive-INC

      u=k'ìin-il   u=bis-a'l   tuméen   le=k'àas-il+ba'l=o',
      A3=sun-REL   A3=go:CAUS-PASS.INC   CAUSE   DEF=bad-REL+thing=D2

      leti'=e',   ma'   **u**=k'áat   h-bin=i'.
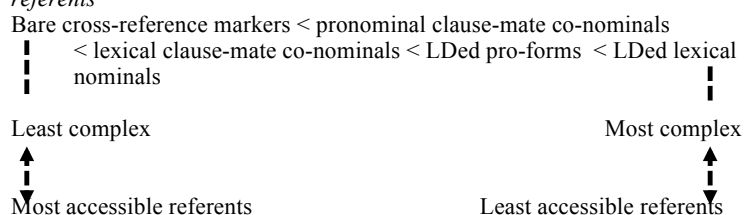   it=D3   NEG(B3SG)   A3=wish(B3SG)   NMLZR-go=D4

      '(…) but, (when) he saw the day he was to be taken by the Devil arriving, as for **him**, **he** didn't want to go.' (ANDRADE & MÁAS COLLÍ 1991: 426-433)

---

[26] Exophoric uses occurred in character speech. As mentioned, we did not include these utterances in the analysis.

The third-person pronoun *leti'* appears LDed in the last line. It is coreferential with the following cross-reference marker *u=*, which indexes the possessor of the nominal predicate *k'áat* 'wish'. Without the intervention of *leti'*, the closest potential nominal antecedent for the cross-reference marker in the last line is *k'àasilba'l* 'Devil' at the end of the second line. The presence of *leti'* triggers a Manner implicature that discourages the interpretation of *k'àasilba'l* as the antecedent of the cross-reference marker.

Example (28) illustrates the function of resuming a previous topic chain. The structural devices that meet this function include, aside from free pro-forms, clause-mate co-nominals and LDed NPs. An eminently plausible hypothesis that remains to be tested is that the preference among these types of expressions is iconically correlated with the distance of the previous mention, with the heaviest and most complex device, LD, being used for the most distant antecedents and the least weighty device, free pro-forms, being used for the most proximate antecedents, such as in the case of (28):

(29) *Complexity scale of Yucatec referring expressions and givenness of referents*
Bare cross-reference markers < pronominal clause-mate co-nominals
< lexical clause-mate co-nominals < LDed pro-forms < LDed lexical nominals

Least complex   Most complex

Most accessible referents   Least accessible referents

However, more surprisingly, as Figures 6 and 7 show, there is also a sizeable number of cases in which a bare cross-reference marker is used to resume a discontinued topic chain. Resumed topics occur with bare cross-reference markers under two licensing conditions: disambiguation by discourse structure and disambiguation by lexical semantics and world knowledge. The following excerpt from the *Kahnal* text illustrates both types of effects. The story involves a demon,[27] who at the onset of (30) has just tracked down, killed, and eaten a man who was out hunting alone in the bush. In this excerpt, we use boldface for bare cross-reference markers, simple underlining for cross-reference markers accompanied by clause-mate co-nominals, and double underlining for cross-reference markers accompanied by LDed nominals.

(30) a.   (…) káa,   bin, t-**u**=ki'=k'ax-ah-**Ø**
          CON   HS   PRV-A3=nicely-tie-CMP-B3SG

---

[27] The demon is referred to throughout as *k'àasilba'l*, which etymologically means 'thing of evil', just like the Devil in (28).

‘(…) and, they say, **he** [the demon] nicely tied **them** [lit. ‘it’; the bones] together,’

b. káa   t-**u**=k'uch-ah-**Ø**,   bin,
    CON   PRV-A3=load/carry.on.back-CMP-B3SG   HS
    ‘and **he** [the demon] loaded **them** [lit. ‘it’; the bones] on his back.’

c. káa t-**u**=ch'a'-ah-**Ø**,   bin, u=ts'òon   (…) le=òotsil   máak
    CON PRV-A3=take-CMP-B3SG HS A3=shoot\ATP   DEF=poor person
    ‘and **he** [the demon] grabbed, they say, the gun of (…) the poor man’

d. ts'-**u**=hàan-t-ik=o',
    TERM-A3=eat-APP-INC(B3SG)=D2
    ‘**he** had eaten,’

e. káa=h-bin-**ih**.
    CON=PRV-go-B3SG
    ‘and **he** [the demon] took off.’

f. Pwes, le=òotsil   nohoch   máak=o',
    well   DEF=poor   big   person=D2
    ‘Well, the poor old man,’

g. ohel-a'n=e'   biha'n-**Ø**   h-ts'òon,   tuméen **u**=pamiilya=o'.
    knowledge-RES(B3SG) go:RES-B3SG NMLZR-shoot\ATP CAUSE A3=family=D2
    ‘it was known (that) he was gone hunting by his family.’

h. Káa, bin,   h-k'uch-**Ø**   te=hòol+nah,   bin=o',
    CON HS   PRV-arrive-B3SG PREP:DEF=aperture+house HS=D2
    ‘When, they say, **he** [the demon] arrived at the door, they say,’

i. *hehten!,*   káa,   bin,   t-**u**=pul-ah-**Ø**.
    IDEO   CON   HS   PRV-A3=throw-CMP-B3SG
    ‘*hehten!,* and, they say, he [the demon] threw **them** [lit. ‘it’; the bones] down.’

In lines a and b, two topic chains can be seen extended simultaneously by bare cross-reference markers, one referring to the demon, the other to the bones of his[28] victim. Even though Yucatec cross-reference markers and pro-forms are not gendered, no ambiguity arises thanks to the animacy difference of the two referents and the two transitive verbs involved selecting for human-like agents.

Furthermore, twice in this textlet is a bare cross-reference marker used to resume a previously interrupted topic chain. In both cases, the cross-reference marker in question resumes reference to the demon. The first instance occurs in

---

[28] It becomes clear later in the story that this particular demon happens to be male.

line e. There are two potentially competing referents at this point both of which have been mentioned more recently by non-pronominal NPs: the dead man and his gun. But neither is a viable referent for the bare cross-reference marker in line e, since both at the point fail to meet the selectional restrictions of *bin* 'go'.

The second instance of resuming reference to the demon with just a bare cross-reference marker occurs in line h. This follows a sequence of two lines, f and g, in which the topical referent is again the dead man. We hypothesize that in this case it is discourse structure that helps disambiguate the reference of the cross-reference marker in h. Lines f and g provide background information that is not part of the main story line. They are recognizable in this capacity by the use of stative result-state verb forms in line g (cf. BOHNEMEYER 2003 for details). We tentatively conclude that speakers turn to the more "weighty" devices for resuming an old topic chain – clause-internal and LDed pronominal and non-pronominal NPs – predominantly when lexical semantics and discourse structure are not sufficient for supporting unambiguous reference resolution.

Although the finding that bare cross-reference markers are the primary means for extending topic chains in Yucatec and that morphologically free pro-forms are only used for this purpose when their use is motivated by additional factors is by no means surprising, it nevertheless in our view represents an important empirical contribution. This is, first of all, due to the special circumstances involved in the Yucatec case – such as (cf. §3 and references therein for details):

- The co-presence of head-marking and free personal pronouns;
- The PRO-drop-like, ambiguous nature of cross-reference markers in this language;
- The absence of gender and noun class marking on cross-reference markers and pro-forms;
- The obviative alignment system of this language;
- The frequent use of left-dislocations in Yucatec discourse.

All of these properties make Yucatec an interesting language for inclusion in future crosslinguistic and typological studies of argument realization. However, secondly, and more generally, we currently still lack much reliable empirical work on argument realization in head-marking languages altogether. The languages studied in detail in the contributions to GIVÓN ed. (1983) are all either purely dependent-marking (English, Japanese) or show at least some degree of double-marking (in the sense of the co-presence of two features: (i) person-marked heads licensing argument ellipsis (PRO-drop) and (ii) some form of nominal case marking; Amharic, Chamorro, Hausa, Spanish, Ute). Similarly, the three Himalayan languages studied in BICKEL (2003) – Belhare (Sino-Tibetan), Maithili, and Nepali (the latter two both Indo-Iranian) are all double-marking.[29]

---

[29] Several of the languages we consider double-marking are classified as dependent-marking in NICHOLS & BICKEL (2013). What matters most for present purposes, however, is that there is agreement that none of the languages in question is (purely) head-marking.

The first study of argument realization in a purely head-marking language was DuBois' (1987) well-known investigation of another Mayan language, Sakapultek (Quichean; Guatemala). This study was based on a corpus of 18 Pear Story narratives elicited with the film and protocol described in CHAFE (1980). The corpus comprised 443 clauses, roughly comparable in size to ours. DuBois's primary interest was the distribution of givenness levels over the argument position classes 'A' (transitive actor), 'O' (transitive undergoer), and 'S' (intransitive argument). Hence, he provided only a limited amount of information about the distribution of the structural devices involved in realization. However, his findings appear to be broadly compatible with ours. Thus, he notes:

"In Sacapultec discourse, independent pronouns are rare; mentions that would be realized pronominally in a language like English are realized affixally." (P814)

Lastly, our interest in argument realization in Yucatec was sparked (or rekindled) in part by recent work by KOENIG & MICHELSON (2012, 2013, 2014, in press). KOENIG and MICHELSON tackle argument realization in another purely head-marking language, Oneida (a Northern Iroquoian language of Ontario, New York, and Wisconsin), based on a much larger corpus than ours (approximately 8230 clauses; approximate 31000 words). Like DuBois', their focus is again different from ours. KOENIG & MICHELSON's work on realization in Oneida has been motivated by a confluence of two features of this language: first, KOENIG & MICHELSON (2012, in press) claim that selectional restrictions play no role in Oneida grammar; and secondly, Oneida happens to be a language with a small nominal lexicon, in which most individual concepts are expressed by (co-)predicative expressions and deverbal expressions rather than by lexical nominals. As a result, the use of referring expressions at the phrase level – which overwhelmingly do not involve lexical nouns – is largely restricted to introducing new referents and to meeting the demands of information perspective.

KOENIG & MICHELSON did not study the distribution of realization strategies by givenness levels. However, it is apparent from their data that just as in Yucatec, free pronouns are rare in Oneida discourse: the authors report 490 tokens in their ca. 31000-words corpus. Another noteworthy feature is the ratio of referring phrasal expressions to clauses in KOENIG & MICHELSON's corpus. This is about 0.4. We counted 303 phrasal referring expressions over 660 predication units, a ratio of approximately 0.5 (just over one per sentence). Given the different composition of the two corpora and differences in which kinds of referring expressions are included in the analysis and in the criteria the two teams use for identifying clauses/predication units, it is difficult to assess how meaningful the proximity of these two figures is. However, both figures are in line with DuBois' (1987: 829) 'One Lexical Argument Constraint', which states the proposed generalization that universally, clauses with multiple non-pronominal arguments are dispreferred. This hypothetical generalization entails that the average number of phrasal referring expressions per clause in head-marking languages should not be greater than one. Obviously, both Oneida and Yucatec discourses seem to

actually operate at a significantly lower density of referring expressions than that. The actual number will be influenced by a variety of language-specific factors, such as:

- The extent to which clauses are combined into sentences;
- The extent to which argument sharing of various kinds occurs across clauses that are thus combined;
- The extent to which the presence of gender or noun-class marking on cross-reference markers supports both longer topic chains and more frequent resumption of discontinued topics with bare cross-reference markers.

At least the last factor is bound to be responsible for differences in realization patterns between Oneida and Yucatec, as Oneida has a complex pronominal gender system, whereas gender marking in Yucatec is restricted to nouns, where it is optional.

Clearly, such potential differences call for cross-linguistic comparisons of realization across different head-marking languages such as Oneida and Yucatec. We hope to have made a small contribution in this paper toward paving the way for such future comparison.

6. Conclusions

We have presented one of the first corpus studies of the impact of givenness on argument realization in a purely head-marking language. We found that in narrative discourse, Yucatec speakers use bare cross-reference markers to extend topic chains, much like speakers of dependent-marking languages use independent pronouns. Combined with the property of entirely eschewing new referents, this behavior strongly supports an anaphoric analysis of the bare cross-reference markers, contra EVANS (1999). This finding is all the more significant given the evidence for an ambiguous (or PRO-drop) analysis of the cross-reference markers, which in the presence of clause-mate co-nominals seem to express agreement, as argued in BOHNEMEYER et al (in press).

Although morphologically unbound pronouns and demonstrative pro-forms are available in Yucatec as well, they are not used for the purpose of tracking highly accessible referents unless additional factors motivate their use. Noun phrases are used both for the introduction of new topics and for the resumption of discontinued old ones. Bare cross-reference markers occur with discontinued topics where discourse structure and lexical semantics provide cues that help disambiguate their reference. A hypothesis that remains to be tested is that for the purposes of resuming an interrupted topic chain, there is a complexity scale of devices, the choice among which iconically reflects the relative accessibility of the referent, as illustrated in (29) above.

Abbreviations

1/2/3	First/Second/Third person	HORT	Hortative
A	Cross-reference Set-A	HS	Hearsay evidential
ALT	Alternative worlds (conditional, disjunction, question focus)	IDEO	Ideophone
ACAUS	Anticausative	IMPF	Imperfective
AN	Animate	IN	Inanimate
APP	Applicative	INC	Incompletive status
ATP	Antipassive	INCH	Inchoative
B	Cross-reference Set-B	LD	Left-dislocation
CAUS	Causative	NEG	Negation
CAUSE	Causal/agentive preposition	NMLZR	Nominalizer
CG	Common ground	NP	Noun phrase
CL	Classifier	PASS	Passive
CMP	Completive status	PERF	Perfect
COM	Comitative preposition	PL	Plural
CON	Continuative connective	PREP	Generic preposition
D1	Proximal-deictic particle	PROG	Progressive
D2	Distal/anaphoric particle	PRSV	Presentative
D3	Locative focus particle	PRV	Perfective
D4		QuD	Question under discussion
DEF	Definite article	REL	Relational
DIM	Diminutive	SG	Singular
DIS	Dispositional stative form	SUBJ	Subjunctive status
F	Feminine		

References

ALLEN, SHANLEY E.M. (2000): A discourse-pragmatic explanation for argument representation in child Inuktitut, in: *Linguistics* 38(3), 483-521.

ALLEN, SHANLEY E.M. (2008): Interacting pragmatic influences on children's argument realization, in: BOWERMAN, MELISSA & BROWN, PENELOPE (eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability*. Mahwah, NJ: Erlbaum, 191-210.

AUSTIN, JOHN L. (1962*). How To Do Things With Words,* 2nd Edition, ed. J.O. URMSON and M. SBISÁ. Cambridge, MA: Harvard University Press.

BICKEL, BALTHASAR (2003). Referential density in discourse and syntactic typology. *Language* 79(4), 708-736.

BLAIR, ROBERT W. & VERMONT SALAS, REFUGIO (1965). *Spoken (Yucatec) Maya. Book 1: Lessons 1-12*. Chicago: University of Chicago, Dept. of Anthropology (Reprint: Columbia, Miss.: Lucas Brothers, 1979).

BLAIR, ROBERT W. & VERMONT SALAS, REFUGIO (1967). *Spoken (Yucatec) Maya. Book 2: Lessons 13-18*. Chicago: University, Dept. of Anthropology.

BLOOMFIELD, LEONARD (1933). *Language*. New York, NY: Henry Holt.

BOHNEMEYER, JÜRGEN (1998). Die Stellung sententialer Topics im Yukatekischen [On the status of sentential topics in Yukatek], in ZAEFFERER, DIETMAR (ed.), *Deskriptive Grammatik und allgemeiner Sprachvergleich* [descriptive grammar and general language comparison]. Tübingen: Niemeyer, 55-85.

BOHNEMEYER, JÜRGEN (2002). *The grammar of time reference in Yukatek Maya*. Munich: LINCOM.

BOHNEMEYER, JÜRGEN (2003). Invisible time lines in the fabric of events: Temporal coherence in Yucatec narratives. *Journal of Linguistic Anthropology* 13(2): 139-162.

BOHNEMEYER, JÜRGEN (2004). Split intransitivity, linking, and lexical representation: the case of Yukatek Maya. *Linguistics 42*(1): 67-107.

BOHNEMEYER, JÜRGEN (2009). Linking without grammatical relations in Yucatec: Alignment, extraction, and control, in NISHINA, YOKO; SHIN, YONG-MIN; SKOPETEAS, STAVROS; VERHOEVEN, ELISABETH & HELMBRECHT, JOHANNES (eds.), *Issues in functional-typological linguistics and language theory: A Festschrift for CHRISTIAN LEHMANN on the occasion of his 60th birthday*. Berlin: Mouton de Gruyter, 185-214.

BOHNEMEYER, JÜRGEN (2011). Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences* 33, 892-914.

BOHNEMEYER, JÜRGEN (2012). Yucatec demonstratives in interaction: spontaneous vs. elicited data, in SCHALLEY, ANDREA C. (ed.), *Practical theories and empirical practice*. Amsterdam/New York: John Benjamins, 99-128.

BOHNEMEYER, JÜRGEN; BUTLER, LINDSAY K. & JAEGER, T. FLORIAN (In press). Head-marking and agreement: Evidence from Yucatec Maya, in LATRUITE, ANJA & OSSWALD, RAINER (eds.), *Explorations of the syntax-semantics interface*. [tba.]

BOHNEMEYER, JÜRGEN; DONELSON, KATHARINE T.; MOORE, RANDI E.; BENEDICTO, ELENA; CAPISTRÁN GARZA, ALEJANDRA; EGGLESTON, ALYSON; HERNÁNDEZ GREEN, NESTOR; HERNÁNDEZ GÓMEZ, MARÍA DE JESÚS S.; HERRERA CASTRO, SAMUEL; O'MEARA, CAROLYN K.; PÉREZ BÁEZ, GABRIELA; PALANCAR, ENRIQUE; POLIAN, GILLES & ROMERO MÉNDEZ, RODRIGO (In press). The contact diffusion of linguistic practices: Reference frames in Mesoamerica. *Language Dynamics and Change* (2015).

BRICKER, VICTORIA R. (1979). Wh-questions, relativization, and clefting in Yucatec Maya, in MARTIN, LAURA (ed.), *Papers in Mayan linguistics, vol. 3*. Columbia, MO: Lucas Brothers, 107-136.

BROWN, PENELOPE (2008). Verb specificity and argument realization in Tzeltal child language, in BOWERMAN, MELISSA & BROWN, PENELOPE (eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability*. New York, NY: Lawrence Erlbaum, 167-189.

BÜRING, DANIEL (1997). *The 59th Street Bridge Accent*. London: Routledge.

BÜRING, DANIEL (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26: 511-545.

BUTLER, LINDSAY. K., T. FLORIAN JAEGER, & JÜRGEN BOHNEMEYER (Ms.). Effects of animacy on sentence production: Ease of retrieval or topicality? Manuscript, University at Buffalo.

BUTT, MIRIAM & HOLLOWAY KING, TRACY (eds) (2000). *Argument realization*. Stanford, CA: CSLI Publications.

CHAFE, WALLACE L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view, in LI, CHARLES N. (ed.), *Subject and topic*. New York, NY: Academic Press, 27-55.

CHAFE, WALLACE. L. (ed.) (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

CHAFE, WALLACE L. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.

CLARK, HERBERT H. & WILKES-GIBBS, DEANNA (1990). Referring as a collaborative process, in COHEN, PHILIP; MORGAN, JERRY & POLLACK, MARTHA (eds.), *Intentions in communication*. Cambridge, MA: MIT Press, 463–493.

DU BOIS, JOHN W. (1987). The discourse basis of ergativity. *Language* 63(4), 805-855.

DU BOIS, JOHN W. (2003). Argument structure: Grammar in use, in DU BOIS, JOHN W.; KUMPF, LORRAINE E. & ASHBY, WILLIAM J. (eds.), *Preferred argument structure: Grammar as architecture for function*. Amsterdam: Benjamins, 11-60.

DURBIN, MARSHALL, & OJEDA, FERNANDO (1978). Basic word-order in Yucatec Maya, in ENGLAND, NORA C. (ed.), *Papers in Mayan linguistics. Vol. 2*. Columbia, MO: University of Missouri, Department of Anthropology, 69-77.

FIRBAS, JAN (1971), On the concept of communicative dynamism in the theory of functional sentence perspective. *Brno Studies in English* 7,12–47.

FOLEY, WILLIAM A., & ROBERT D. VAN VALIN, JR. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

GIVÓN, TALMY (1983). Introduction, in GIVÓN, TALMY (ed.), *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins, 1-42.

GIVÓN, TALMY (ed.) (1983). *Topic continuity in discourse: A quantitative cross-language study*.

Amsterdam: John Benjamins.

GOLDBERG, ADELE (2005). Argument Realization: the role of constructions, lexical semantics and discourse factors, in ÖSTMAN, JAN-OLA & FRIED, MIRJAM (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*. Amsterdam: John Benjamins, 17-43.

GRICE, PAUL (1975). Logic and conversation, in PETER COLE & JERRY MORGAN (eds.), *Syntax and semantics 3: Speech acts*. New York, NY: Academic Press, 41–58.

GUNDEL, JEANETTE. K., NANCY HEDBERG, & RON ZACHARSKI (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.

GUTIÉRREZ-BRAVO, RODRIGO & MONFORTE, JORGE (2009). Focus, agent focus and relative clauses in Yucatec Maya, in AVELINO, HERIBERTO; COON, JESSICA & NORCLIFFE, ELISABETH (eds.), *New Perspectives on Mayan Linguistics*. Cambridge, MA: MIT (Massachusetts Institute of Technology Working Papers in Linguistics), 83-95.

GUTIÉRREZ-BRAVO, RODRIGO & MONFORTE, JORGE (2010). On the nature of word order in Yucatec Maya, in CAMACHO, JOSÉ; GUTIÉRREZ-BRAVO, RODRIGO & SÁNCHEZ, LILIANA (eds.), *Information Structure in Indigenous Languages of the Americas*. Berlin: Mouton de Gruyter, 139-170.

HANKS, WILLIAM F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago, IL: University of Chicago Press.

JACOBS, JOACHIM (2001). The dimensions of topic-comment. *Linguistics* 39, 641-681.

KAUFMAN, TERRENCE (1990). Algunos rasgos estructurales de los idiomas mayences con referencia especial al k'iche' [Some structural traits of the Mayan languages, with special reference to K'iche'], in ENGLAND, NORA C. & ELLIOTT, STEPHEN R. (eds.), *Lecturas sobre la lingüística maya* [Readings on Mayan linguistics]. Guatemala: CIRMA, 59-114.

KOENIG, JEAN-PIERRE & MICHELSON, KARIN (2012). The (non)-universality of syntactic selection and functional application, in PIÑÓN, CHRISTOPHER (ed.), *Empirical issues in syntax and semantics, 9*. Paris: CNRS (http://www.cssp.cnrs.fr/eiss9/). 185-205.

KOENIG, JEAN-PIERRE & MICHELSON, KARIN (2013). Counting nouns is not always the right question. Departmental colloquium, University at Buffalo.

KOENIG, JEAN-PIERRE & MICHELSON, KARIN (2014). Direct syntax meets information structure: The case of Iroquoian. Workshop *Information structure in head-marking languages*, Max Planck Institute for Psycholinguistics.

KOENIG, JEAN-PIERRE & MICHELSON, KARIN (In press). Invariance in argument realization: The case of Iroquoian. *Language*.

KRÄMER, MARTIN & WUNDERLICH, DIETER (1999). Transitivity alternations in Yucatec, and the correlation between aspect and argument roles. *Linguistics* 37, 431 480.

KRATZER, ANGELIKA (2014). Situations in Natural Language Semantics, in ZALTA, EDWARD N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). (http://plato.stanford.edu/archives/spr2014/entries/situations-semantics, last consulted 06/11/2015).

LAUNEY, MICHEL (2004). The features of omnipredicativity in Classical Nahuatl. *STUF: Language Typology and Universals* 57(1): 59-69.

LAMBRECHT, KNUT (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

LEHMANN, CHRISTIAN (Ms.a). El hijo prodigo [The prodigal son]. By GREGORIO VIVAS CÁMARA. Recorded by CHRISTIAN LEHMANN and transcribed with the aid of JULIO EK MAY. Manuscript, University of Erfurt (no page numbers).

LEHMANN, CHRISTIAN (Ms.b). U tsikbalil u kuxtal don FELIPE CARRILLO PUERTO [The story of the life of don FELIPE CARRILLO PUERTO]. By GREGORIO VIVAS CÁMARA. Recorded by CHRISTIAN LEHMANN and transcribed with the aid of RAMÓN MAY CUPUL. Manuscript, University of Erfurt (no page numbers).

LEHMANN, CHRISTIAN (Ms.c). Xùunáan much [The frog princess]. By ESTÉBAN PPOL KAAW. Recorded by CHRISTIAN LEHMANN and transcribed with the aid of RAMÓN MAY CUPUL. Manuscript, University of Erfurt (no page numbers).

LEVIN, BETH & MALKA RAPPAPORT HOVAV (2005). *Argument Realization*. Cambridge: Cambridge University Press.

LEWIS, M. PAUL; SIMONS, GARY F. & FENNIG, CHARLES D. (eds.) (2015). *Ethnologue: Languages of the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

MAYER, MERCER (1969). *Frog, where are you?* New York, NY: Dial Books.

MILEWSKI, TADEUSZ (1950). La structure de la phrase dans les langues indigenes de l' Amérique du Nord, in MILEWSKI, TADEUSZ (ed.), *Études typologiques sur les langues indigenes de l' Amérique*. Kraków, Poland: Polska Akademia Nauk, 7-101.

NICHOLS, JOHANNA (1986). Head-marking and dependent-marking grammar. *Language* 62, 56-119.

NICHOLS, JOHANNA & BICKEL, BALTHASAR (2013). Locus of marking in the clause. In DRYER, MATTHEW S. & HASPELMATH, MARTIN (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/23, Accessed on 2015-06-09.)

ROOTH, MATS (1985). *Association with focus*. Doctoral dissertation, University of Massachusetts Amherst.

ROOTH, MATS (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75-116.

SEARLE, JOHN (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SEARLE, JOHN (1975). Indirect speech acts, in P. COLE & J. L. MORGAN (eds.), *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press, 59–82.

STALNAKER, ROBERT (1974). Pragmatic presuppositions, in MUNITZ, MILTON K. & UNGER, PETER K. (eds.), *Semantics and philosophy*. New York, NY: University Press, 197-214.

TONHAUSER, JUDITH (2003). F-constructions in Yucatec Maya, in ANDERSSEN, JAN; MENÉNDEZ BENITO, PAULA & WERLE, ADAM (eds.), *Proceedings of Semantics of Under-Represented Languages in the Americas II*. Amherst, MA: Graduate Linguistics Student Association, 203–223.

TONHAUSER, JUDITH (Ms). The syntax and semantics of Yucatec Mayan focus constructions. Manuscript, The Ohio State University.

VAN KUPPEVELT, JAN (1995). Discourse structure, topicality and questioning. *Linguistics* 31, 109–147.

VAN KUPPEVELT, JAN (1996). Inferring from topics. *Linguistics and Philosophy* 19, 393–443.

VAN VALIN, ROBERT D., JR (1985). Case marking and the structure of the Lakhota clause, in NICHOLS, JOHANNA & WOODBURY, ANTONY (eds.), *Grammar inside and outside the clause*. Cambridge: Cambridge University Press, 363-413.

VAN VALIN, ROBERT D., JR (ed.) (1993). *Advances in Role and Reference Grammar*. Amsterdam: Benjamins.

VAN VALIN, ROBERT D., JR (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

VAN VALIN, ROBERT D., JR (2008). RPs and the nature of lexical and syntactic categories in Role and Reference Grammar, in VAN VALIN, ROBERT D. (ed.), *Investigations of the syntax-semantics-pragmatics interface*. Amsterdam: Benjamins, 161-178.

VAN VALIN, ROBERT D., JR (2013). Head-marking languages and linguistic theory, in BICKEL, BALTHASAR, GRENOBLE, LENORE A., PETERSON, DAVID A., & TIMBERLAKE, ALAN (eds.), *Language typology and historical contingency: In honor of JOHANNA NICHOLS*. Amsterdam: John Benjamins, 91-123.

VAN VALIN, ROBERT D., JR. & LAPOLLA, RANDI J (1997). *Syntax*. Cambridge: Cambridge University Press.

VAPNARSKY, VALENTINA (2013). Is Yucatec an omnipredicative language? Predication, the copula, and focus constructions. *STUF: Language Typology and Universals* 66(1), 40-86.

VERHOVEN, ELISABETH & SKOPETEAS, STAVROS (2015). Licensing focus constructions in Yucatec Maya: An empirical study on the association with focus. *International Journal of American Linguistics* 81(1), 1-40.

JÜRGEN BOHNEMEYER

Department of Linguistics
University at Buffalo – SUNY
609 Baldy Hall
Buffalo, NY 14260
U.S.A.
jb77@buffalo.edu