

DISTINCTNESS EFFECTS ON VOS ORDER: EVIDENCE FROM YUCATEC MAYA

STAVROS SKOPETEAS
University of Potsdam
ELISABETH VERHOEVEN
University of Bremen

1 Preliminaries

Basic word order in Yucatec Maya has been the source of several controversial accounts. Some researchers claim that the basic order in this language is VOS (see for instance Norman and Campbell 1978: 144, Lehmann 1990: 44, 2003: 28), while other researchers claim that the basic order is SVO (see Gutiérrez-Bravo and Monforte y Madera 2008a). The third logical possibility, also exploited in the literature, is to argue that basic order in this language involves a split between V-initial and S-initial configurations (see Durbin and Ojeda 1978, Gutiérrez-Bravo and Monforte y Madera 2008b).

The root of this controversy lies in two conflicting observations: on the one hand, structural facts suggest that preverbal constituents do not occupy their basic positions but rather positions projected by functional heads that are designated to particular discourse functions (topic and focus); on the other hand, observation of the occurrence of the orders in texts reveals that SVO is the preferred order in a wide range of contexts including the contextual condition in which no presuppositions are involved, while the VOS order appears only very rarely. Thus, the empirical phenomena create a puzzle: structural facts suggest that the basic order is V-initial, behavioral facts suggest that the basic order is S-initial.

The aim of this paper is to make a proposal that unifies the conflicting phenomena. As a starting point, we adopt the view that the reductionist's option to ignore either the structural or the behavioral facts as not crucial for the assumptions of constituent structure simply fails to account for the full range of phenomena at issue. Section 2 presents the basic facts and introduces the empirical puzzle. Section 3 proposes an account based on the idea that particular word order possibilities are banned as the result of a condition on the distinctness of linearizations (Richards 2006). This claim implies some requirements about the constituent structure of Yucatec Maya, which are outlined in Section 4. Section 5 presents empirical

evidence from language comprehension that contributes to our understanding how distinctness effects arise. Finally, the conclusions of the paper are presented in Section 6.

2 Basic facts

2.1 Structural facts

In terms of purely linear permutations, all logically possible orders of a verb, a subject, and an object are grammatical in Yucatec Maya. However, this does not mean that word order in this language is free. There is a structural asymmetry between arguments in the preverbal and the postverbal domain which was already figured out in the earlier grammatical descriptions of the language (see Durbin and Ojeda 1978, Bricker 1979). In the VOS order, which is exemplified in (1), both postverbal arguments occur without any special marking. Reordering to VSO order is licensed by asymmetries in animacy, definiteness or weight, which are not accounted for in this paper (see Skopeteas and Verhoeven 2005, Gutiérrez-Bravo and Monforte y Madera 2008b, Bohnemeyer 2008).

- (1) *k-u* *hàant-ik* *òon* *Pèedróoh.*
 IPFV-A.3 eat:TRR-INCMPL avocado Pedro
 ‘Pedro eats avocado.’

Arguments appear preverbally in two constructions that are exemplified in the following; the crucial point is that preverbal realization of the arguments involves additional morphological marking. The first construction involves a *left dislocated* argument and is exemplified in (2). In comparison to (1), the left dislocated argument involves additional marking, i.e., its right edge is marked by an enclitic *-e* which belongs to a closed set of enclitics that contrast for the encoding of indexical properties (see details on this construction in Lehmann 2003: 28, Bohnemeyer 1998a: 59f., 1998b: 205f., 2008). Following the tradition in Mayan linguistics, this constituent occupies a ‘topic position’ at the left periphery of the clause, however Skopeteas and Verhoeven (2008a) argue that the pragmatic properties of this position are underspecified in syntax and result from the interaction of universal pragmatic preferences with the properties of linearization and prosodic structure in this language.

- (2) *Pèedróoh-e*’ *k-u* *hàant-ik* *òon.*
 Juan-D3 IPFV-A.3 eat:TRR-INCMPL avocado
 ‘Pedro eats avocado.’

A further option in the left periphery is to place a constituent at a position at the left side of the verb and immediately adjacent to it; this construction is frequently used for the expression of narrow focus (see (3)). In contrast to the configuration illustrated in (2), the preverbal constituent in (3) is not enclosed by a right edge clitic and is accompanied by a special inflectional form of the verb under particular structural conditions, i.e. when the constituent at issue is the agent of an active transitive verb. This is the so-called agent focus form of the verb and is characterized by the drop of the A cross-reference marker and the tense/aspect/mood auxiliary. With imperfective reference, the verb appears in the incomplete status (see suffix *-ik* ‘INCMPL’ in (3)). With

perfective reference, it appears in the subjunctive status which is zero in non-clause final position (see details on this construction in Bricker 1979, Bohmeyer 1998b: 189-202, 2008, Lehmann 2003: 28f., Skopeteas and Verhoeven 2008b, Tonhauser 2003, 2007; see also Stiebels 2006 for an outline of the agent focus constructions in Mayan languages). We call this constituent a ‘pre-predicate phrase’ in order to avoid the established term ‘focus position’ that involves further assumptions about the pragmatic properties of this constituent.

- (3) *Pèedróoh hàant-ik òon.*
 Pedro eat:TRR-INCMPL avocado
 ‘PEDRO eats avocado.’

The crucial point in the facts presented in (1) to (3) is the asymmetry between the preverbal and the postverbal realization of the arguments. The option that qualifies as unmarked is certainly the postverbal one. A straightforward account of this observation is to assume that the configuration in (1) is basic while the configurations in (2) and (3) are the result of (probably distinct) syntactic operations that determine the placement of the arguments in the preverbal positions and are accompanied by additional morphological devices. This asymmetry is the main evidence that is used by the accounts that opt for a V-initial basic order in Yucatec Maya.

2.2 Behavioral facts

Things become complicated as soon as we observe the discourse occurrences of the options in (1)-(3). Taking for granted that syntactic operations do not apply just randomly, but they require a particular trigger in order to occur, implies some expectations for the distribution of the alternative orders in discourse. In particular, the unmarked VOS order is expected to be contextually unrestricted, while the derived orders are expected to be triggered by restricted contextual conditions. Following the current assumptions in Mayan linguistics, the configuration in (2) is expected to occur when the contextual requirements for agent topicalization are met, while the configuration in (3) is expected to occur when the context licenses a focused agent. However, this prediction is not borne out, and this is the source of the controversy in word order studies.

Observation in corpora shows that the most frequent order in discourse is SVO. This preference is observed by Durbin and Ojeda (1978: 75) and Bricker (1979: 112)¹, and is supported by corpus measurements in Skopeteas and Verhoeven (2005), Gutiérrez-Bravo and Monforte y Madera (2008b), and Tonhauser (2005, cited in Avelino 2008). Further converging evidence comes from a comprehension experiment reported in Skopeteas and Verhoeven (2005), an experiment on language production discussed in Skopeteas and Verhoeven (2008a) and in Skopeteas and Fanselow (2008) in comparison to further languages, as well as from a forced-choice elicitation task reported in Gutiérrez-Bravo and Monforte y Madera (2008a, 2008b).

The following examples illustrate the facts. Example (4) is elicited through picture description and is part of a study on language production that is reported in Skopeteas and Verhoeven (2008a). The context sentence, given in translation, introduces an inanimate entity. The target sentence presents the following scene, in which an agent performs an action in which

¹ Notably, Bricker (1979: 112) observes that the preference for SVO order is already attested in the *Chilam Balam of Chumayel* (1782).

the given entity is involved as a patient. The important point is that the contextual conditions at the moment that the speaker lexicalizes the target scene do not license a topicalized agent. Nevertheless, the indefinite agent phrase is realized in the preverbal domain and is accompanied by the enclitic *-e'* which encloses left dislocated constituents. This example illustrates that the SVO order with a left-dislocated S is not restricted to contexts that license agent topicalization. Corpus studies report that this configuration occurs very frequently (percentage of SVO out of *n* sentences with two lexically realized arguments: 70% according to Skopeteas and Verhoeven 2005, 50% according to Tonhauser 2005; 75% according to Gutiérrez-Bravo and Monforte y Madera 2008b).

(4) Context: {There is a ball on the table.}

Target: *hun-túul máak-e' túun hats'ik le bòoláah (...)*
 one-CL.AN man-D3 PROG hit-INCMPL DEF ball
 'A man hits the ball (...)' (J 42.271)

The VOS order is also attested in our data, however only rarely, which is again in line with the corpus measurements (percentage of VOS out of *n* sentences with two lexically realized arguments: 10% according to Skopeteas and Verhoeven 2005, 16% according to Tonhauser 2005; 25% according to Gutiérrez-Bravo and Monforte y Madera 2008b). Example (5) illustrates a token elicited as the first sentence of a narrative. Both the agent and the patient constituents are not previously introduced. The agent appears in a definite DP, which reflects the fact that the speaker assumes during the picture description task that this entity is available in the discourse situation. Two points are relevant for the following discussion: (a) the VOS order occurs rarely, contrary to the considerations about the contextual non-restrictedness of basic orders, but also (b) that the VOS order is not categorically excluded.

(5) *táan u hóoyab hun-p'éel pàak'al le máak-o' (...)*
 PROG A.3 water(INCMPL) one-CL.INAN plant DEF man-D2
 'The man is watering a plant (...)' (F 40.1)

These behavioral facts are interpreted by Durbin and Ojeda (1978) and Bricker (1979) as evidence that SVO qualifies as a basic word order in Yucatec Maya. However, this generalization does not take into account the full range of the related phenomena. The examples (4) and (5) and the discussion so far relate to clauses with transitive verbs. If we turn to intransitive clauses, then the preferred pattern is V-initial, while SV orders occur only when the context licenses topicalization of the subject constituent. The sentence in (6) illustrates the V-initial order with intransitive verbs. It is from a fairy tale and occurs at a point of the story where the protagonists (one of them being *nuxib kéeh* 'the old deer' mentioned in (6)) have been introduced and the main story line is developing. This example illustrates the opposite pattern: Though the context involves the necessary conditions for a subject topic, the subject of the intransitive verb is realized postverbally.

- (6) *ma' sáam-e' k-u k'uch-ul nuxib k'éeh (...)*
 NEG some.time.ago-D3 IPFV-A3 arrive-INCMPPL old deer
 'Soon afterwards, the old deer arrived (...)'²

The asymmetry between transitive and intransitive verbs concerning canonical order is reported in Gutiérrez-Bravo and Monforte y Madera (2008b) and in Skopeteas and Verhoeven (2008b). We may conclude that there is a split in the canonical orders of Yucatec Maya, such that the basic word order is {Predicate < Subject} for intransitive verbs and {Subject < Predicate} for transitive verbs. This hypothesis perfectly accounts for the facts presented so far, but it introduces a stipulative trigger such as 'transitivity' that determines the directionality of the specifier of the highest verb projection, i.e. the subject. Note that none of the available studies reports a difference between unergatives and unaccusatives, hence it is not possible to relate the observed word order variation to different projections for internal and external arguments.³

Moreover, evidence from reflexive and reciprocal verbs contradict the transitivity split hypothesis for canonical word order. Reflexive/reciprocal verbs are formally transitive but behave like intransitive verbs as regards the preferred order {Predicate < Subject}, exemplified in (7). This example involves two coordinated lexical DPs in the postverbal domain, which form the subject constituent of the reciprocal verb. The co-indexed object constituent *u báah* 'A.3 self' obligatorily follows the verb. The fact that transitive configurations with co-indexed arguments behave identical to intransitive verbs suggests that transitivity is not the crucial factor for the ordering preferences.

- (7) *te'l-a' túun ts'u'ts'-ik u báah le máak yéetel x-ch'úup-o' (...)*
 there-D1 PROG:A.3 kiss-INCMPPL A.3 self DEF man and F-woman-D2
 'In this (picture), the man and woman kiss each other (...)' (N 26.1)

Moreover, the preference for the {Subject < Predicate} order does not hold true for all instances of transitive verbs in discourse. Crucially, it does not apply when the object occupies the pre-predicate position, as exemplified in (8). In these cases, the subject constituent is more often than not realized in the postverbal position (see experimental data in Skopeteas and Verhoeven 2008a).

- (8) Question: {What is the man kicking?}

hun-p'éel esten k'áanche' k-u kóochek'-t-ik
 one-CL.INAN HESIT chair IPFV-A.3 kick:foot-TRR-INCMPPL
le xib-o'.
 DEF man-D2
 'It is a ehm chair that the man kicks.' (J 41.6)

Finally, the preference for {Subject < Predicate} order does not apply when the patient is not lexically realized. In this configuration, the agent constituent occurs preferably in the postverbal

² Example from Romero Castillo, Moisés 1964, Tres cuentos mayas. *Anales del INAH, Tomo XVII*, 64: 307-309.

³ Furthermore, Gutiérrez Bravo and Monforte y Madera (2008) as well as Skopeteas and Verhoeven (2008a) report that subjects of passives are not preferably left dislocated – just like intransitive verbs.

position (see (9)); it may occur in the preverbal position, but only when the context licenses agent topicalization.

- (9) *ts'u* *y-áant-ik-en* *le* *xòok-a'* (...)

 TERM:A.3 0-help-INCMPL-B.1.SG DEF story-D1

 'This story has helped me (...)'⁴

3 Distinctness condition

We observed in section 2 that though the order {Subject < Predicate} is the dominant pattern in sentences with a transitive verb and two lexically realized arguments, this preference does not hold for intransitive verbs (and passives), transitive reflexive/reciprocal verbs, transitive verbs with one lexically realized argument, and transitive verbs with an object in the preverbal position. Thus, the critical condition that involves the preference for subject-initial order relates to the configuration with an object and a subject that are both realized as lexical DPs. In concluding, structural and behavioral facts suggest the view that the canonical word order in Yucatec Maya is V-initial, with the exception of the configuration with two lexical arguments. Gutiérrez-Bravo and Monforte y Madera (2008b: 9) suggest that the crucial factor is “whether one or two arguments of the verb are overtly expressed”, while Bohnemeyer (2008) notes that there is “a tendency to avoid multiplicity of clause-internal noun phrases”. This section develops a principled account in this direction.

Richards (2006) proposes a syntactic condition that bans configurations involving adjacent syntactic units with identical features. The effects of this condition show up in very different structures, hence the proposal is to establish a fundamental condition on distinctness in syntax, having common cognitive foundations with the ‘obligatory contour principle’ in phonology, which bans sequences of adjacent and identical tonal events. Richards (2006) examines data from English quotative inversion and observes a data pattern that is similar to the Yucatec Mayan data presented above. In particular, while subjects surface postverbally in quotative inversion as exemplified in (10a-b), the configuration with an adjacent subject and object in the postverbal domain is banned, see (10c). The contrast between (10b) and (10c) shows that the constraint at issue does not ban clauses with more than one argument, but clauses with two arguments with identical overt properties.

- (10) a. “*It’s cold,*” *said John.*
 b. “*It’s cold,*” *said John to Mary.*
 c. *“*It’s cold,*” *told John Mary.* (Richards 2006:1)

Richards accounts for the data pattern in (10) in terms of a condition that relates to the output of syntactic rules, i.e., directly to the linearization that evolves after the application of the derivational processes. The crucial condition, which is given in (11), bans linearization statements that contain adjacent syntactic units being in an asymmetric c-command relation.

⁴ Vivas Cámara, Gregorio 1988, *El hijo pródigo*, recorded by Christian Lehmann in Yaxley, Quintana Roo.

Assuming that the linearization proceeds in phases (Chomsky 2001), the condition at issue applies if the suboptimal configuration occurs within one and the same phase.

(11) Distinctness (Richards 2006: 4)

If a linearization statement $\langle \alpha, \alpha \rangle$ is generated, the derivation crashes.

Richards (2006) explores a wide range of phenomena cross-linguistically that may be accounted in terms of the distinctness condition, including stylistic inversion in French, the obligatory A-bar movement of DP subjects in the context of DP predicates in Tagalog, the ban of two PPs with the same prepositional head in nominalization, constraints in sequences of adjacent verbs in several languages, etc. He also discusses data from another Mayan language, namely Chol (based on data reported in Coon 2006), which shows that multiple postverbal DPs are ruled out in this language.

The distinctness condition accounts straightforwardly for the data pattern reported in section 2.2 for Yucatec Maya. Intransitive verbs, transitive verbs with only one lexically realized argument, and transitive verbs with an argument in the focus position, come up in the vP phase with a linearization statement that involves a single postverbal DP, as illustrated in (12a). Agent constituents of passives are realized as PPs, resulting hence in the linearization in (12b). The critical configuration relates to transitive verbs with two lexically realized arguments, which correspond to the linearization statement in (12c). Moreover, the data from reciprocal constructions in (7) shows that only a subset of the linearizations with two adjacent DPs is banned, namely those that involve two lexical DPs, as represented in (12c). Linearizations that involve a pronominal DP, as represented in (12d) are not subject to the distinctness condition.

(12) Linearization statements in the vP phase

- a. $\langle V, DP \rangle$
- b. $\langle V, DP, PP \rangle$
- c. $! \langle V, DP (+lex), DP (+lex) \rangle$
- d. $\langle V, DP (-lex), DP (+lex) \rangle$

Hence, the distinctness condition provides a straightforward account of the data pattern presented so far. The canonical order in Yucatec Maya is V-initial, as suggested by the structural facts as well as by the behavioral evidence apart from the configuration with two adjacent lexical DPs. In this latter case, distinctness applies and the configuration is rendered suboptimal. However, the assumption of distinctness effects has some structural requirements that are not discussed so far. First, it has to be shown that the two postverbal DPs form part of the same phase and this depends on our assumptions about the constituent structure in Yucatec Maya, which will be discussed in section 4. Second, Richards (2006: 4) notes that languages differ with respect to the exact configurations that are banned whenever distinctness is violated. Hence, it is an empirical question which categorical features are subject to the distinctness condition. We address this issue in Section 5.

4 Constituent structure of Yucatec Maya

As basis for the following considerations, we assume the constituent structure in Figure 1 based on previous proposals for Yucatec Maya (Gutiérrez-Bravo and Monforte y Madera 2008a: 10), Tzotzil, Jakaltek, Tojolabal, and Tz'utujil (Aissen 1992: 46f.), Tzotzil (Aissen 1996: 449), Jakaltek (Woolford 1991: 507f.), Kaqchikel (Broadwell 2000: 15), Tz'utujil (Aissen 1999: 172, Duncan 2003: 180), and Chol (Coon 2009). We assume that a lower verb projection (VP) hosts the V and the object constituent and a higher verb projection (vP) hosts the VP and the subject constituent. This implies that:

- a. we assume a subject/object asymmetry (contrary to the assumption of a flat verb projection that is proposed for Kaqchikel in Broadwell 2000: 15, Tz'utujil in Duncan 2003: 180, and Jakaltek in Woolford 1991: 507f.) based on the evidence that is presented in Verhoeven (2007: ch. 4.3) and further facts about the subject/object asymmetry in binding that are reported in Bohnemeyer (2008) and below (see example 16) and
- b. we do not assume that VOS results from predicate fronting in Yucatec Maya (as proposed by Coon 2009 for Chol) for reasons that will be discussed below.

The higher projections, IP (=inflection phrase) and CP (=complementizer phrase), host elements that surface in the left periphery. The evidence for these projections exclusively relates to the structural facts and not to the putative association of the preverbal constituents with pragmatic functions such as topic and focus. Hence, we do not assume a FocP (= focus phrase) and a TopP (=topic phrase) that are used in cartographic approaches to the left periphery (see Rizzi 1997, Kiss 1998 among others); these approaches imply that information structural properties form part of the constituent structure, which is not supported by our data (see Skopeteas and Verhoeven 2008a and 2008b). The crucial structural evidence for the distinction of the left peripheral positions has already been introduced in (2) and (3): evidence from verb morphology and the availability of right edge clitics show that there are two distinct structural operations that underlie the preverbal occurrence of constituents in Yucatec Maya. It is clear that the positions at the left periphery are strictly ordered (see Lehmann 2003: 28): as illustrated in (13a-b), the fronted agent constituent that triggers the particular inflectional properties of the verb cannot precede a left dislocated constituent that is accompanied by the right edge clitic.

- (13) a. *Pèedròoh-e' òon t-u hàant-ah.*
 Pedro- D3 avocado PFV-A.3 eat:TRR-CMPL
 'Pedro, it was avocado that he ate.'
- b. **òon Pèedròoh-e' t-u hàant-ah.*

Following Aissen (1992, 1996), we assume that the pre-predicate constituent (see (3)) occupies spec,IP.⁵ The head I of the IP projection hosts aspect/mood auxiliaries that precede the lexical verb, see *k-* 'IPFV' in example (2), *túun* 'PROG' in example (4), etc. (see Gutiérrez-Bravo and

⁵ There is not yet a detailed account of extraction possibilities and their interaction with islands for movement in Yucatec Maya. Some data are given in Norcliffe (2008) who provides evidence that extraction to spec,IP is sensitive to islands, hence suggesting the view that this operation involves movement.

Monforte y Madera 2008a).⁶ The specifier of the IP is a unique structural position that hosts either constituents in narrow focus or *wh*-constituents (evidence for the uniqueness of this position comes from the complementary distribution of *wh*- and narrow focused constituents, see detailed discussion in Bricker 1979 and Tonhauser 2003).

The head C of the higher functional projection CP hosts (at least a subset of) the complementizers (e.g., *káa* ‘that’), hence left dislocated constituents appear on their left.⁷ The question is whether left dislocated constituents are part of the clause structure or external to it. In the former case, they are projected within the CP layer (spec,CP), while in the latter case they form a separate CP which is adjoined to the clausal material. For this case, we assume a root node termed ‘utterance phrase’ (UP)⁸, i.e. a unit outside the functional layers of the clause, IP and CP. Aissen (1992: 47) shows that both configurations are possible in Mayan languages (but not both are available in all languages of the family). The lower position hosts elements that are moved from their postverbal position and correspond to a gap in situ, which does not hold for the highest position that is base generated. There is evidence that at least a subset of the left dislocated constituents in Yucatec Maya does not correspond to a gap in situ, which implies that they are not the result of movement. As an illustration, see example (14), which presents an instance of hanging topic left dislocation: the left-dislocated constituent is in a hypernymic relation to the object that occupies the pre-predicate position (see further examples involving connectivity violations in Bohnemeyer 2008 and Skopeteas and Verhoeven 2008a). Examples like (14) do not allow for a movement analysis of the left-dislocated material.

- (14) *ch’úich’-o’b-e’ chen x-k’òok’-o’b u k’ahóol.*
 bird-PL-D3 only F-nightingale-PL A.3 know
 ‘As concerns birds, he only knows nightingales.’⁹

A further type of example that very frequently occurs in discourse is the occurrence of a strong pronoun in the pre-predicate position which is co-referent with the left dislocated constituent, as exemplified in (15a). The acceptability of the example (15a) reduces when the pronoun is presented in situ (see 15b), since it is an emphatic pronoun and since it is co-referent with the set A person affix on the verb. However, this does not affect our main argument: the relevant issue is that it is possible to use a pronominal element which is coreferent with the left dislocated material in a CP internal position as the spec,IP, as exemplified in (15a).

- (15) a. *Le ah kòonol-o’ leti’ túun y-áalkab.*
 DEF master seller-D2 that.one PROG:A.3 0-run
 ‘The seller, he is the one who is running.’

⁶ We do not assume a tense phrase (TP) for Yucatec Maya, since this language does not have inflectional categories for the encoding of tense (see Bohnemeyer 1998b). A possible alternative would be the assumption of aspectual and modal phrases (AspP and MoodP respectively), which we do not favour for the reason that the choice of the exact semantic properties that are encoded by the auxiliaries is accidental with respect to constituent structure, and as such irrelevant for syntactic considerations.

⁷ A detailed account on the complementizers in Yucatec Maya is a matter of future research.

⁸ For the same purposes, Aissen (1992) postulates a node E (= Expression).

⁹ This example is part of the Questionnaire of the DFG-project ‘Discontinuous Noun and Prepositional Phrases’ (University of Potsdam, Gisbert Fanselow and Caroline Féry).

- b. ²*Le ah kòonol-o' túun y-áalkab leti'*.
 DEF master seller-D2 PROG:A.3 0-run that.one
 'The seller, that one is running.'

The data in (14) and (15) shows that at least a subset of the left-dislocated constituents is clause external, i.e., it is hosted by a separate CP in terms of the assumptions of Figure 1. The empirical question is whether Yucatec Maya displays a lower position (spec,CP) that hosts left dislocated constituents as well.¹⁰ This possibility has been proposed for topics in Tz'utujil Maya by Aissen (1992: 76); this configuration can be justified if we find evidence that some left-dislocated constituents correspond to a gap in situ. There is no information concerning island constraints on left dislocation in Yucatec Maya (a further desideratum for future research), but binding facts suggest an asymmetry between left dislocation and movement to the pre-predicate position. It has been shown that Yucatec Mayan subjects in VOS order may bind their antecedent objects (see Bohnemeyer 2008) and not vice versa, which provides evidence that subjects asymmetrically c-command objects in the VOS order. The binding possibility of the postcedent subject is exemplified in (16a). (16b) shows that the binding possibilities do not change when the object is placed in the pre-predicate position, which supports the view that this configuration is the result of displacement to a position that is linked to the trace in situ through an A-bar chain. However, these binding possibilities do not apply to left-dislocated constituents as shown in (16c). The reading in which the possessor of the left dislocated constituent is bound by the subject is excluded: this evidence suggests that left dislocated constituents in Yucatec Maya are not linked to a trace in situ which could be bound by the c-commanding subject; further research on the interaction of left dislocation with clausal constraints is needed in order to clarify this issue.

- (16) a. *k-u kol-ik u_{i/j} kòol káadah hun-túul kolnáal_j*.
 IPFV-A.3 cut-INCMPL A.3 milpa every one-CL.AN farmer
 'Every farmer clears his milpa.'
- b. *chen u_{i/j} kòol k-u kol-ik káadah hun-túul kolnáal_j*.
 just A.3 milpa IPFV-A.3 cut-INCMPL every one-CL.AN farmer
 'It is just his milpa that every farmer clears.'
- c. *u_{i/j}*_j kool-e' k-u kol-ik káadah hun-túul kolnáal_j*.
 A.3 milpa-D3 IPFV-A.3 cut-INCMPL every one-CL.AN farmer
 'As concerns his milpa, every farmer clears it.'

¹⁰ This view is advocated by Gutiérrez-Bravo and Monforte y Madera (2008) who make a distinction between preverbal subjects that are enclosed by a right edge clitic and occupy the 'topic' position and preverbal subjects that are not separated from the rest of the clause through an enclitic and presumably occupy a lower position for subjects. We do not adopt this view because we are not aware of any syntactic differences, e.g. in the extraction or binding possibilities, that are determined by the occurrence of the right edge clitics. The optionality in the occurrence of the enclitic *-e* 'D.3' is attested in all environments (i.e., between the left dislocated constituent and the predicate or between main clauses and adverbial subordinate clauses, etc.), hence we assume that the only phenomenon that is informative for syntactic analyses is the *possibility* to use an enclitic at particular boundaries.

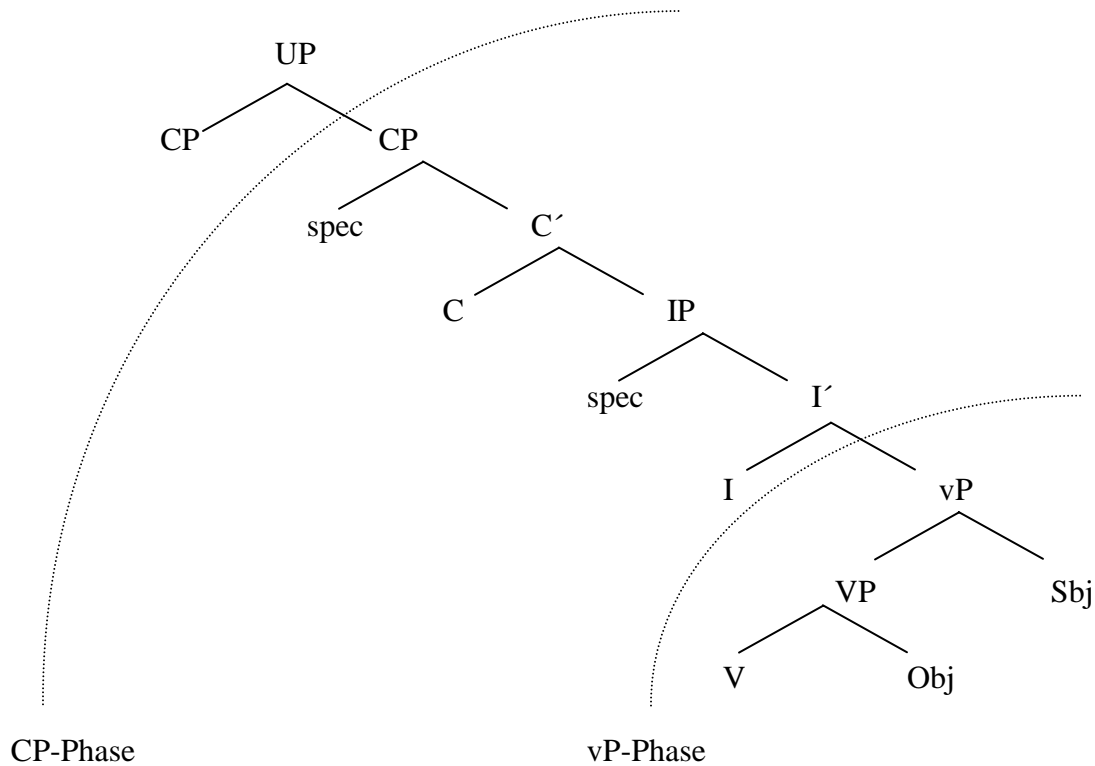


Figure 1. Constituent structure of YM

Phase theoretical accounts assume that the linearization that results from the structure in Figure 1 is derived in phases. Following standard assumptions in this framework, CP and vP constitute strong phases, in contrast to TP/IP and further projections, as indicated in Figure 1 (see Chomsky 2001, 2005). The crucial consequence of these assumptions is that DP subjects and DP objects belong to the same phase, hence they constitute a syntactic domain in which the distinctness condition as formulated in (11) may apply. However, this is not an obvious assumption, in particular with reference to a VOS language. According to antisymmetric accounts on constituent structure (Kayne 1994), word order permutations are universally derived by a basic Specifier-Head-Complement order; languages with VOS order involve an operation that moves the predicate (Head-Complement unit) past the Specifier (see Kayne 1994: 36).

A fundamental difficulty for a predicate-fronting account in Yucatec Maya is that the possible SVO orders in this language cannot be assumed to be basic without stipulative assumptions: the linearization options that are available in this language involve a configuration in which the subject is realized in a lower position (spec,IP) that triggers a special morphological form of the V (under particular circumstances) and a configuration in which the subject constituent surfaces in a higher position that shows the properties of left dislocation. In order to justify the assumption of a basic SVO order, we would like to see the possibility of preverbal subjects to surface in a position lower than the pre-predicate position, a configuration that simply does not exist in this language. Hence, in those syntactic models that require the basic order to be a possible order, the predicate-fronting hypothesis may already be rejected. However, let's assume that this prerequisite does not hold, which implies the theoretical possibility to have *obligatory* predicate fronting, such that the basic configuration does not have any chance to be ever spelled out.

A diagnostic for predicate fronting is proposed in the work of Chung (2005, 2006): fronting a constituent implies that its subconstituents are inaccessible to extraction. Evidence for this claim comes from the fact that non-subjects in predicate-fronting languages, such as Malagasy or Seediq, cannot be relativized and are not accessible for *wh*-movement (see examples and discussion in Chung 2006: 693-697). On this basis, Chung evaluates the possibility to extract objects in other V-initial languages – notably Tzotzil Maya – as evidence against a predicate-fronting account for these languages. Similar facts are available in Yucatec Maya. Hence, there is no restriction to the extraction of VP internal constituents, as exemplified in (17). Positive evidence that sentences such as (17) involve movement comes from the binding possibilities of the pre-predicate position (see discussion above and examples under 16), as well as from the fact that the extracted constituents correspond to a gap in situ (hence they cannot be repeated with a co-referent element as shown in Skopeteas and Verhoeven 2008b).

- (17) *ba'x k-u tul-ik le máak-o'?*
 what IPFV-A.3 push-INCMPPL DEF person-D2
 'What is the man pushing?'

A predicate-fronting account for a Mayan language, namely Chol, is proposed by Coon (2009). The evidence for this account is an asymmetry between bare NP and full DP objects in Chol: while the former type of object occurs in VOS order, the latter type can only occur in VSO order. Coon (2009) accounts for this asymmetry in assuming that VOS order in Chol Maya involves predicate fronting that applies in the case of bare NP objects. Full DP objects obligatorily undergo object shift, moving rightwards to a position outside the vP; the remnant vP is fronted resulting hence in a VSO linearization. These facts are in line with the assumption of predicate fronting, however they do not constitute evidence for predicate-fronting, since the full DP object may be shifted rightwards from a basic VOS order too. However, there is no corresponding constraint on full DP objects in Yucatec Mayan VOS, as exemplified in (18).

- (18) *táan u kóochek'-t-ik le k'áanche' le xib-o'.*
 PROG A.3 kick:foot-TRR-INCMPPL DEF chair DEF man-D2
 'The man is kicking the table.'

Hence, we maintain our original assumption that DP objects and DP subjects are derived within the same phase, which allows for an account of the asymmetries observed in the behavioral data in section 2.2 in terms of the effects of a well established cross-linguistic generalization, namely the distinctness condition on linearizations with two identical syntactic entities that are derived within one and the same phase.

Finally, an alternative account that equally predicts our data pattern is the assumption of a constraint on the Spell-Out of verb projections. Alexiadou and Anagnostopoulou (2001: 193) formulate this constraint as follows: "By Spell-Out VP can contain no more than one argument with an unchecked Case feature". This constraint accounts for the facts of English quotative inversion in (10) and predicts the data pattern summarized in (12) (some ancillary assumptions are needed in order to accommodate 12d). The reason that we do not follow this account is that our evidence shows that the phenomenon in Yucatec Maya cannot be reduced to the consideration of Case properties. The preference against a $\langle \alpha, \alpha \rangle$ configuration proportionally increases, depending on the set of features shared between the postverbal syntactic entities. In

face of this evidence, we assume that the phenomenon at issue in Yucatec Maya relates to performance factors that may be captured through principles for the optimization of linearizations and is not a categorical component of syntax as would be expected for a filter on Case features.

5 The source of distinctness effects

There is a crucial difference between the data from English quotative inversion in (10) and the behavioral facts from Yucatec Maya in section 2.2: in the former case, the $\langle V, DP, DP \rangle$ linearization is categorically banned, in the latter case the same configuration is not ungrammatical, but suboptimal. Hence, derivations resulting into a $\langle \alpha, \alpha \rangle$ linearization statement do not “crash” in Yucatec Maya, as predicted by (11). What the presented data shows, is that native speakers only rarely select such linearizations. All accounts on Yucatec Maya agree that VOS sequences are perfectly grammatical. From this evidence, we conclude that the phenomenon we observe in Yucatec Maya is not a categorical constraint that is part of the syntax, but rather a preference to avoid linearization options that are suboptimal for language performance.

Recall that Yucatec Maya is a head marking language, hence argument DPs do not bear morphological markers of structural case. Note also that the argument order in the postverbal domain is not rigid. The choice of VOS and VSO order may be influenced by asymmetries in animacy, definiteness and weight, as shown in a number of studies (Durbin and Ojeda 1978: 70, Skopeteas and Verhoeven 2005, Gutiérrez-Bravo and Monforte y Madera 2008b, Bohnemeyer 2008).¹¹ Hence, a linearization $\langle V, \alpha, \alpha \rangle$ certainly involves a difficulty in parsing.

Direct evidence for this assumption is provided by an experiment on language comprehension reported in Skopeteas and Verhoeven (2005). Native speakers were auditorily presented sentences of the type illustrated in (19) and were instructed to give a spontaneous translation in Spanish.

- (19) *Pedro-e' t-u y-a'l-ah t-u kiiims-ah lòoxnáal chakmol.*
 Pedro-D3 PFV-A.3 0-say-CMPL PFV-A.3 kill-CMPL boxer puma
 ‘Pedro said that a puma killed a boxer.’

The experimental conditions involved several manipulations of the animacy and definiteness of the postverbal arguments in order to observe the influence of these factors on the choice between a VOS and a VSO reading of the sentence (see Skopeteas and Verhoeven 2005). The crucial point for the issue of this paper is that next to the expected readings, native speakers’ reactions involved a high amount of interpretations in which the two postverbal DPs were concatenated in a single DP object and the subject of the matrix predicate was interpreted as controlling the embedded one. Hence, next to the VOS reading of (19) we also elicited the reading: ‘Pedro said that he killed a boxer (called) puma’.

The overall results of this study show that the proportions of concatenative readings are sensitive to the feature permutations between the two postverbal DPs. Table 1 presents a subset

¹¹ VOS/VSO alternation determined by soft constraints of this kind is very widespread in Mayan languages (see an overview in Norman and Campbell 1978: 144-146).

of the experimental results and summarizes the interpretations of eight native speakers. The number (n) of interpretations that involve a concatenative reading are given in the third column and the percentage (%) with respect to the number of ‘valid’ interpretations is given in the fourth column. In order to meet the normality requirements of parametric tests, we transformed each speaker’s percentages through the *arcsin-square root* transformation which resulted in the values given in the last column (A.S.T.).

Table 1 presents the results of the animacy manipulations that involve sentences with two bare NP postverbal arguments varying in their animacy properties. Condition 1 (C1) involves sentences with two human postverbal NPs (+h), Condition 4 (C4) involves sentences with two non-human postverbal NPs (-h), and Conditions 2 and 3 (C2/C3) involve sentences with a human and a non-human NP in two different orders. A simple comparison of the proportions reveals that speakers are more likely to apply a concatenative reading, when both NPs share the same feature (C1/C4), than when the two postverbal NPs bear different animacy features (C2/C3). Pairwise comparisons of the four conditions in a repeated measures analysis of variance reveal a significant main effect in the comparison C1/C4 vs. C2/C3, $F_{1,8} = 7.79$, $p < .05$, while the probability of the effects in the further two comparisons (C1/C2 vs. C3/C4; C1/C3 vs. C2/C4) is above the chance level (.05).

Table 1. Proportion of concatenative readings in animacy manipulations

	total	valid	concatenative reading		
			n	%	A.S.T.
C1 <DP (+h), DP (+h)>	32	24	15	62.50	52.07
C2 <DP (+h), DP (-h)>	32	26	6	23.08	30.12
C3 <DP (-h), DP (+h)>	32	25	6	24.00	32.76
C4 <DP (-h), DP (-h)>	32	31	16	51.61	45.22

Furthermore, Skopeteas and Verhoeven (2005) report that the concatenative readings almost disappear when a pragmatic cue for the interpretation of the sentences is available (e.g., ‘Pedro said that a dog bit a girl’).

These results give some idea about the ways distinctness works in Yucatec Maya. The phenomenon at issue does not seem to be associated with a closed set of grammatical features that cause a discrete subset of derivations to crash. The necessary condition for the application of the distinctness condition is a configuration of two postverbal arguments without differential marking of their syntactic function, i.e. as a subject and an object. However, this configuration is not a sufficient condition for distinctness to apply. The likelihood of distinctness proportionally increases when the postverbal DPs share more features in common and when the discourse does not supply any cues for their interpretation. The experimental data in Table 1 come from language comprehension; it is an open question for future research whether the same asymmetry applies in language production. It is not possible to check this hypothesis at this moment, since the amount of sentences with two postverbal arguments that is reported in the available corpus studies is considerably small.

6 Conclusions

This paper started from a controversy in the assumptions about the basic word order in Yucatec Maya. We outlined the sources of this controversy in presenting structural facts that support the view that Yucatec Maya is a V-initial language and behavioral facts that show a strong preference for SVO orders in sentences with two lexically realized arguments. Based on this evidence, we hypothesized that the crucial factor is the suboptimality of linearization statements involving two adjacent syntactic units with identical overt features: $\langle\alpha, \alpha\rangle$. Linearization statements of this type are subject to the distinctness condition (Richards 2006) and are banned in several languages. In order to prove the syntactic basis of this explanation, we examined the constituent structure of Yucatec Maya and we concluded that postverbal objects and subjects are derived within the same phase, which is the prerequisite for the distinctness condition to apply. Finally, we discussed the nature of the distinctness condition in our data and we argued that our evidence is against the view that the distinctness effects we observe here are a categorical constraint in the syntactic component. Since our data involve gradience that is affected by the processing difficulty of particular configurations, we concluded that we are rather observing a factor that optimizes language performance in avoiding linearizations that are difficult to process.

Acknowledgments

This paper has been developed within the project D2 Typology of Information Structure, which is part of the research center 632 *Information Structure* at the University of Potsdam/Humboldt University of Berlin (funded by the DFG) and within the project 10/853/05 at the University of Bremen. We are grateful to Christian Lehmann for teaching us the syntax of Yucatec Maya, to Gisbert Fanselow for drawing our attention to the distinctness effects, to Artemis Alexiadou for proposing several theoretical challenges to our account, and to Heriberto Avelino, Rodrigo Gutiérrez-Bravo, and Elisabeth Norcliffe for their comments at the Symposium *New Perspectives in Mayan Linguistics*, held at the 2008 SSILA Annual Meeting, January, 3-4, 2008, Chicago. Grateful thanks are due to our Yucatec Mayan consultants, in particular Amedee Colli Colli, Ramón May Cupul, and Ernesto May Balam.

Glosses

A= person marker, set A; AN= animate; B= person marker, set B; CL= classifier; CMPL= completive; D1= 1st person deixis; D2= 2nd person deixis; D3= 3rd person deixis; DEF= definite; F= feminine; HESIT= hesitative; INAN= inanimate; INCMPL= incompletive; IPFV= imperfective; NEG= negator; PFV= perfective; PL= plural; PROG= progressive; SG= singular; TERM= terminative; TRR= transitivizer; 0= meaningless element; 1= 1st person; 3= 3rd person.

References

- Aissen, Judith L. 1992. Topic and focus in Mayan. *Language* 68.1: 43-80.
- Aissen, Judith L. 1996. Pied-piping, abstract agreement, and functional projections in Tzotzil. *Natural Language and Linguistic Theory* 14: 447-491.
- Aissen, Judith L. 1999. External possessor and logical subject in Tz'utujil. In: Doris L. Payne and Immanuel Barshi (eds.), *External possession*. Amsterdam/Philadelphia: Benjamins, 167-193.
- Alexiadou, Artemis and Anagnostopoulou, Elena 2001. The subject in situ generalization, and the role of Case in driving computations. *Linguistic Inquiry* 32: 193-231.
- Alexiadou, Artemis and Anagnostopoulou, Elena 2007. The subject in situ generalization revisited. In: H.-M. Gärtner and U. Sauerland (eds.), *Interfaces + recursion = language*. Berlin: Mouton de Gruyter, 31-59.
- Avelino, Heriberto 2008. Intonational patterns of topic and focus constructions in Yucatec Maya. Paper presented at the Symposium *New Perspectives in Mayan Linguistics*, held at the 2008 SSILA Annual Meeting, January, 3-4, 2008, Chicago.
- Bohnemeyer, Jürgen 1998a. Sententiale Topics im Yukatekischen. In Dietmar Zaefferer (ed.), *Deskriptive Grammatik und allgemeiner Sprachvergleich*. Tübingen: Niemeyer, 55-85.
- Bohnemeyer, Jürgen 1998b. *Time relations in discourse: Evidence from a comparative approach to Yukatek Maya*. Ph. D. Dissertation, Katholieke Universiteit Brabant.
- Bohnemeyer, Jürgen 2008. Linking without grammatical relations in Yucatec: Alignment, extraction, and control. Ms., University of Buffalo.
- Bricker, Victoria R. 1979. Wh-questions, relativization, and clefting in Yucatec Maya. In Laura Martin (ed.), *Papers in Mayan linguistics*. Columbia, Miss.: Lucas Brothers, 107-136.
- Broadwell, George Aaron 2000. Word order and markedness in Kaqchikel. In: Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG00 Conference*. University of California, Berkeley. CSLI Publications.
- Chomsky, Noam 2001. Derivation by phase. In: Michael Kenstowicz (ed.), *Ken Hale: a life in language*. Cambridge, Mass.: MIT Press, 1-52.
- Chomsky, Noam 2005. Three factors in Language Design. *Linguistic Inquiry* 36.1: 1-22.
- Chung, Sandra 2005. What fronts? On the VP-raising account of verb-initial order. In: Andrew Carnie, Heidi Harley, and Sheila Ann Dooley (eds.), *Verb first: On the syntax of verb initial languages*. Amsterdam, Philadelphia: John Benjamins, 9-29.
- Chung, Sandra 2006. Properties of VOS languages. In: Everaert, Martin and van Riemsdijk, Henk (eds.), *The Blackwell companion to syntax*, vol. 3. Oxford: Blackwell, 685-720.
- Coon, Jessica. 2006. Constituent order and configurationality in Chol (Mayan). Ms., MIT.
- Duncan, Lachlan 2003. The syntactic structure of Tz'utujil Maya. In: Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG03 Conference*, University at Albany, State University of New York. CSLI Publications, 164-183.
- Durbin, Marshall and Fernando Ojeda 1978. Basic word-order in Yucatec Maya. In: Nora C. England (ed.), *Papers in Mayan linguistics*, vol. 2. Columbia: University of Missouri, Department of Anthropology, 69-77.
- Gutiérrez-Bravo, Rodrigo and Monforte y Madera, Jorge 2008a. La alternancia sujeto inicial/verbo inicial y la teoría de la optimidad. In: Rodrigo Gutiérrez Bravo y Esther Herrera Zendejas (eds.), *Teoría de la optimidad: estudios de sintaxis y fonología*. El Colegio de

- México, Centro de Estudios Lingüísticos y Literarios-Laboratorio de Estudios Fónicos, México, 61-90.
- Gutiérrez-Bravo, Rodrigo and Monforte y Madera, Jorge 2008b. On the nature of unmarked word order in Yucatec Maya. Paper presented at the Symposium *New Perspectives in Mayan Linguistics*, held at the 2008 SSILA Annual Meeting, January, 3-4, 2008, Chicago.
- Hoffling, Charles Andrew 1984. On proto-Yucatecan word order. *Journal of Mayan Linguistics* 4.2: 35-64.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge/Mass., London: MIT Press.
- Kiss, Katalin É. 1998. Identificational focus versus information focus. *Language* 74.2: 245-273.
- Lehmann, Christian 1990. Yukatekisch. *Zeitschrift für Sprachwissenschaft* 9.1-2: 28-51.
- Lehmann, Christian 2003. *Possession in Yucatec Maya*, second revised edition. Assidue Nr. 10, Erfurt: Seminar für Sprachwissenschaft der Universität.
- Norcliffe, Elisabeth 2008. Filler-Gap dependencies in Yucatec. Paper presented at the Symposium *New Perspectives in Mayan Linguistics*, held at the 2008 SSILA Annual Meeting, January, 3-4, 2008, Chicago.
- Norman, William M. and Campbell, Lyle 1978. Toward a proto-Mayan Syntax: A comparative perspective on grammar. In: England, Nora C. (ed.), *Papers in Mayan Linguistics*, vol. 2. Columbia: University of Missouri, Department of Anthropology (University of Missouri Miscellaneous Publications in anthropology, 6), 136-155.
- Richards, Norvin 2006. A distinctness condition on linearization. Ms., MIT.
- Rizzi, Luigi 1997. The fine structure of the left periphery. In: Liliane Haegeman (ed.), *Elements of grammar*. Dordrecht: Kluwer, 281-337.
- Skopeteas, Stavros and Gisbert Fanselow 2007. Effects of givenness and constraints on free word order. In: Malte Zimmerman and Caroline Féry (eds.), *Information structure from different perspectives*. Oxford: Oxford University Press.
- Skopeteas, Stavros and Elisabeth Verhoeven 2005. Postverbal argument order in Yucatec Maya. *Sprachtypologie und Universalienforschung* 53.1: 71-79.
- Skopeteas, Stavros and Verhoeven, Elisabeth 2008a. The interaction between topicalization and structural constraints. *Linguistic Review* (in press).
- Skopeteas, Stavros and Verhoeven, Elisabeth 2008b. Licensing focus in Yucatec Maya: an empirical study on the association with focus (manuscript under review).
- Stiebels, Barbara 2006. Agent focus in Mayan languages. *Natural language and linguistic theory* 24: 501-570.
- Tonhauser, Judith 2003. F-constructions in Yucatec Maya. In: Jan Anderssen, Paula Menéndez-Benito and Adam Werle (eds.), *The proceedings of SULA 2*. Vancouver, BC: GLSA, University of Massachusetts, Amherst, 203-223.
- Tonhauser, Judith 2005. Grammar meets discourse: The case of Yucatec Maya. Paper presented at the LSA Meeting, January 6-9, 2005, Oakland.
- Tonhauser, Judith 2007. Agent focus and voice in Yucatec Maya. *Proceedings of the 39th Meeting of the Chicago Linguistic Society*.
- Verhoeven, Elisabeth 2007. *Experiential constructions in Yucatec Maya*. Amsterdam: Benjamins.
- Woolford, Ellen 1991. VP-Internal subjects in VSO and nonconfigurational languages. *Linguistic Inquiry* 22.3: 503-540.