HEAD MARKING IN USAGE AND GRAMMAR:

A STUDY OF VARIATION AND CHANGE IN YUCATEC MAYA

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Elisabeth Norcliffe

September 2009

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Paul Kiparsky)   Principal Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Ivan Sag)   Principal Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Beth Levin)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Tom Wasow)

Approved for the University Committee on Graduate Studies.

iii

Abstract

Many Mayan languages make use of a special dependent verb form (the Agent Focus, or AF verb form), which alternates with the normal transitive verb form (the synthetic verb form) of main clauses when the subject of a transitive verb is focused, questioned or relativized. It has been a centerpiece of research in Mayan morphosyntax over the last forty years, due to its puzzling formal and distributional properties. In this dissertation I show how a usage-oriented approach to the phenomenon can provide important insights into this area of grammar which resists any categorical explanation.

I use Yucatec Maya as the lens through which to examine the phenomenon. I first show that the AF alternation is a subtype of resumptive pronoun (RP)/gap alternation. This sets the AF alternation in typological context, providing a basis for analyzing the variation, both within and across Mayan languages, as well as from a cross-linguistic perspective. The difference between Yucatec, and other types of RP/gap alternating languages lies in the locus of the alternation. Rather than involve the presence or absence of an independent pronoun, the alternation in Yucatec relates to two verb forms: one which carries a morphologically dependent subject pronoun (the synthetic verb), the other (the AF verb) which does not.

I extend this analysis to a range of head-marking languages, where similar phenomena have been documented. Just like Mayan languages, such languages make use of special verb forms in dependent clauses in A-bar dependencies (wh-questions, relative clauses and clefts), which are distinguished (minimally) from their main clause counterparts by the absence of person marking inflection. And just as in Mayan, the distributions of the verbal alternants are conditioned by the same factors that are implicated in RP/gap distributions cross linguistically, providing support for the notion that they are all exemplars of the same basic phenomenon.

I propose that the historical origins of these special verb forms can be traced to the

emergence of head marking. Drawing on a wide range of cross-linguistic and historical data, I argue that the special verbs that occur in A-bar dependencies are byproducts of the frequency-sensitive gramaticalization process by which independent pronouns become pronominal inflection on verbs. I show that the relatively low frequency of adjacent pronoun-verb combinations in extraction contexts (where gaps are more frequent than resumptive pronouns) can give rise to asymmetric patterns of pronoun grammaticalization, and thus lead to the emergence of these morphological alternations.

The asymmetric frequency distributions of gaps and RPs (within and across languages) in turn can be explained by processing preferences. I present three experiments which show that Yucatec speakers are more likely to use the resumptive verb form in embedded environments, and where the antecedent is indefinite. Specifically, these studies indicate the need to bring discourse-level processing principles into the account of what have often been taken to be autonomously sentence-internal phenomena: factors such as distance and the referential salience of the antecedent have been shown to influence referential form choice in discourse, suggesting that the same cognitive principles lie behind both types of variation.

More generally, the Yucatec studies demonstrate that production preferences in Yucatec relative clauses reflect patterns of RP/gap distributions that have been attested *across grammars*. The Highest Subject Restriction (the ban on subject RPs in local dependencies), which is apparently a categorical constraint in many languages, is reflected probabilistically in Yucatec in terms of production preferences. The definiteness restriction (RPs are obligatory with indefinite antecedents), which has been reported categorically in other languages, is also visible probabilistically in Yucatec production. This lends some statistically robust support to the view that that typological patterns can arise via the conventionalization of processing preferences.

Acknowledgements

My first and greatest thanks are to my parents, Jim and Joan, and my brother, Tom. Despite the large body of water that lies between California and Christchurch, they have been at my side for this entire enterprise. I thank them for their love and support throughout.

To my wonderful committee: Paul Kiparsky, Ivan Sag, Beth Levin and Tom Wasow. I have benefited enormously from Paul's input ever since he brought champagne to class after the Red Sox won the World Series. Paul has unwaveringly given me his time and support, always with great enthusiasm and good humor, and has done much to shape my thinking with the acuity and depth of his. He has been my intellectual mentor and sounding board for six years. I am deeply indebted to him for all of this. Above all, I thank him for his friendship, which I will always hold dear. Ivan Sag has been an invaluable support and role model, both intellectually and musically. His passion for language is infectious and I have always come away from meetings with him buoyed up and head abuzz with new ideas and directions. He has been an excellent critic, constantly keeping me on my toes with his sharp questions and eye for detail. I have valued not just his direct input into my work, but also our off-topic philosophical conversations about grammar, music, life and everything, as a result of which he has become a firm friend. Without Beth Levin, I doubt this dissertation would ever have happened. Since day one she has helped me in countless ways – intellectually and practically. She has read over every abstract, proposal, paper and presentation I have ever written, always with a meticulous eye for detail and insightful commentary. With her encyclopedic knowledge of linguistic scholarship, she has pointed me to innumerable references from which my work has benefitted considerably. She has reminded me of every upcoming deadline, helped me with all sorts of paperwork, spurred me on with words of encouragement. Not only is she a brilliant scholar, she is one of the kindest people I have ever had the good fortune to meet. She also proof read this entire

experiments. I also had the benefit of his collaboration and company during one of the field trips to Valladolid. But above and beyond that, I want to thank him for the close friendship we have shared over many years. For the laughs, the jaunts, the molecular powers.

My California experience would have been greatly diluted were it not for the wonderful people who helped to make it my home. To Philip Hofmeister, my steadfast friend, my sanity check, cook, colleague. I thank him for his linguistic insights, the clarity of his thinking, the countless ways he has helped me with my work. For all the laughter, for cheering me on in rough times. For his Mexico survival kit. That head lamp was a lifesaver. I don't think I could have done all this without you. To my dear family at 29th Street, Inbal Arnon and Harry Tily. Inbal, thank you for all the good times over these years (and to many more to come!). To our Chiapan adventures, always beware of scorpions falling from the roof and check the bed for spiders! To our beach in Tulum. To those countless caltrain rides, la camisa negra. To your voice of reason. To Hal, for lepodoptery and its impact on the rise of function words, hot rice and avocado and other important discoveries. Thank you for all the Japanese curries and the synchronicities. Here's to better luck with public transportation. To Uriel Cohen, for all those coffees and long chats. For your loyalty and friendship. For your passion about language. For gitar gitar gitar. To my friend and office mate, Doug, in accordance with prophecy. Lauren Hall-Lew, cahoot, May child, we did it! Start to finish! To John Beavers, for the friendship, the music and the Honda. To a number of others who have each played an important part in my life here: Alexei, Graham, Itamar, Na'ama, Liz, Tatiana, Jason, Abby, Neal, Natalie, David, Nathan, Fabian, Florian, Héctor, Ximena, Felipe y Elif. To Dead Tongues, for laying down the funk.

And on farther shores, but no less closer to my heart, Heidi, Adam, Andrew, Judith, Brian, Nic, Ben A, Ben W, Teena, Potch. To T.B.A.L.C and our collective lack of vision. To my whanau.

Finally, and above all, to Alejandro, my greatest gift, my purpose in life. For dropping in, and all that came afterwards.

Abbreviations

A	Set A person marker
ABS	Absolutive
AUX	Auxiliary
ACAUS	Anticausative
ACC	Accusative
ACT	Active
ALT	Alternative (disjunction, interrogative focus, conditional protasis)
AN	Animate
APP	Applicative
ART	Article
ASP	Aspect
ATP	Antipassive
B	Set B person marker
CAUS	Causative
CL	Classifier
COLL	Collective
CMP	Completive
CON	Consequential
DEF	Definite
DET	Determiner
D1	Proximal deixis
DAT	Dative
D2	Distal deixis
D3	Textual deixis

D4	Location/Negation deixis
ENC	Enclitic
ERG	Ergative
F	Feminine
FUT	Future
GEN	Genitive
HAB	Habitual
HUM	Human
INAL	Inalienable
IRR	Irrealis
IMPF	Imperfective
INC	Incompletive
ITR	Intransitive
LOC	Locative
LV	Linking vowel
M	Mode
N	Nominative
NC	Noun classifier
NEG	Negation
NR	Near distal deixis
OBL	Oblique
OBJ	Object
OST	Ostensive
PART	Participle
PASS	Passive
PRED	Predictive
PREP	Preposition
PRSV	Prospective
PRV	Perfective
PL	Plural
PROG	Progressive
PST	Past

R	Realis
RDP	Reduplication
REC	Recent past
REP	Repetitive
SUBJ	Subjunctive
SG	Singular
SR	Subordinator
T	Tense
TERM	Terminative
TOP	Topic
TRANS	Transitive
VO	Voice
1	First person
2	Second person
3	Third person

Contents

xiv

List of Tables

List of Figures

Chapter 1

Introduction

"Everyone knows that language is variable" Sapir once remarked (1921, 147). Indeed, it takes little inspection to see that variation in human language manifests itself at every structural level. In English a speaker may sometimes pronounce the participle *doing* and sometimes *doin*. She may sometimes say *can't* and at other times *cannot*. She may announce on one occasion that she is going to *pick a friend up*, and on another that she is going to *pick up a friend*.

The existence of linguistic variation is an obvious fact. But how a student of language chooses to engage with this fact varies considerably across theoretical orientations. In the dominant schools of linguistic thought of the last century –Saussurean theory, American and Prague School structuralism, and, most lastingly, Chomskyan generative theory– linguistic variation has been outside the scope of inquiry. Sapir's contemporaries, and his immediate successors adopted instead what Chambers (1995) refers to as the *axiom of categoricity*, the assumption that linguistic data is to be regularized for analysis, to eliminate real-world variability. This axiom is made clearest perhaps, in Chomsky's oft-quoted programmatic statement:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky, 1965, 3).

In classical generative theory, questions about the nature of grammar are separated from questions about how the grammar is employed. Understanding internalized grammars consists in establishing the innate, universal constraints which govern their structure, together with the 'parameter settings' which account for cross-linguistic differences. External variability in performance, though analyzable in its own right (Chomsky, 1980, 24-25), offers no explanatory potential for grammatical structure.

While the generative linguist searches for top-down principles and parametric constraints derived from an innate universal grammar, an alternative approach understands language structures as emergent phenomena, influenced by the regularities of language use. This usage-based approach to language builds on the fact that variation in performance is not typically 'free' in any meaningful sense of the word, but rather can be shown to be systematic and probabilistic.[1] It may be conditioned, for example, by socio-pragmatic factors, by perceptuo-motor biases and by cognitive limitations (see Beckner et al., 2008).

The connection between usage and grammar follows from the observation that the statistical preferences exhibited by some languages may emerge as categorical grammatical constraints in other languages.[2] Diachronic change is understood to be the key mediating mechanism that allows performance factors to shape grammars (Bybee, 1988; Keller, 1994; Kirby, 1999; Nettle, 1999; Haspelmath, 1999, 2004; Croft, 2000a; Hawkins, 2004). Over time, frequent patterns may become grammaticalized, via mechanisms of sequential processing, categorization and conventionalization (Bybee, 1998; Heine and Kuteva, 2002).

In this dissertation, I show how a usage-oriented approach to language can throw light on an area of grammar which resists any categorical and parameterized explanation. The phenomenon in question is a morpho-syntactic alternation which occurs in many Mayan languages. These languages possess a verb form, known as the *Agent Focus* (AF) verb, which occurs in a restricted set of environments, namely in dependent clauses where the

---

[1]This was first demonstrated systematically by Labov (1963) and subsequent scholars in the nascent field of sociolinguistics in the 1960s.

[2]For instance, Givón (1979, 26ff) observes that many languages prohibit referential indefinite NPs in subject position, while others allow them but still show a clear preference for definite subject NPs. Bresnan et al. (2001) observe that some languages (such as Lummi) do not allow passives with first or second person agents, while other languages (such as English) show a significantly lower frequency of passives with first or second person agents. And Hawkins (1994, 2004) derives the typologically observed preference for SVO languages from the same principles of efficient processing that predict the short-before-long preference in head-initial VP languages like English.

subject of a transitive verb is focused, questioned or relativized. Across the language family, the AF verb differs (minimally) from its main clause counterpart by the absence of otherwise obligatory morphologically bound subject person marking.

The AF phenomenon has been a centerpiece of research in Mayan morphosyntax over the last thirty years, due to its puzzling formal and distributional properties (i.a. Day, 1973; Craig, 1977; Ayres, 1983; England, 1983; Dayley, 1981; Aissen, 1999, 2003; Campbell, 2000; Tonhauser, 2003; Stiebels, 2006; Bohnemeyer, 2009). It has often been thought to represent a type of Mayan-specific voice category, triggered by an underlyingly ergatively organized syntax (it only targets subjects of transitive verbs). This view is undermined by the fact that in some Mayan languages, Yucatec included, the Agent Focus verb is not obligatory in subject extraction contexts: it may alternate with the fully inflected verb form (referred to here as 'synthetic') which is found in main clauses. The inter- and intra-language variation that exists in this domain has proved troublesome for categorical accounts of the phenomenon, preventing any solution that can extend adequately across the whole family. From a usage-based perspective, by contrast, the variation provides a window into the origin and evolution of these verb forms.

In this dissertation I use Yucatec Maya, an AF-variable language, as the lens through which to examine the phenomenon. The account I develop is structured rather like a Russian doll: over the course of the dissertation I uncover increasingly deeper levels of explanation in order to arrive at an overall emergentist picture of the phenomenon. I begin with what might be described as a *taxonomic explanation*, whose objective is to show that a phenomenon is part of a more widespread cross-linguistic phenomenon. I will show that the AF alternation is a specialized instance of a resumptive pronoun (RP)/gap alternation. Establishing this is crucial, not only for descriptive adequacy, but because it provides the departure point for a deeper level of explanation, a *generalizing-historical explanation*,[3] which aims to show that a phenomenon is a consequence of general constraints on historical change. I will show that AF verbs are a byproduct of frequency-sensitive gramaticalization processes which occurred elsewhere in the grammar: the morphologization of person marking. It is the asymmetric frequency distributions of pronouns and gaps intrasententially which accounts for the exemption of verbs in certain subordinate environments from this process. Finally, I argue that the asymmetric frequency distributions of RPs and gaps

---

[3]I borrow this term from Haspelmath (2002).

within and across languages derive in turn from processing preferences. This constitutes the final *processing-based explanation* which aims to explain the deeper cognitive factors underlying the existence of distributional patterns and emergent linguistic structure.

Below, I briefly outline the main hypotheses and findings of this *matroyshka* approach to the Mayan AF alternation.

Taxonomy: AF alternations and resumption/gap typology (Chapters 2-4)

The peculiar distributional and formal properties of the AF alternation have typically encouraged Mayanists to adopt the Boasian approach of positing some kind of special language particular category when describing and analyzing the phenomenon. While I agree that there exists a need to define certain categories in language-particular terms (see, e.g., Dryer, 1997; Croft, 2000b),[4] one must also be careful not to do so at the expense of overlooking parallels across other languages which can help to shed light on the nature of a phenomenon and its historical-functional origins. In the case of Mayan AF, I suggest that the tendency to label the alternation a type of Mayan-specific voice alternation has obscured Mayan's place in a larger cross-linguistic typology.

A cross-linguistic perspective is crucial in arriving at an understanding of the phenomenon of AF. My departure point is the observation that the variation between the AF verb and the synthetic verb in A-bar dependencies parallels typologically attested patterns of *RP/gap* distributions.

The typological record reveals that within and across languages, gaps and RPs exhibit asymmetric frequency distributions. Gaps are more frequent than RPs, for example, in direct argument positions (especially subject position); resumptives are more likely to be realized as obliques (Keenan and Comrie, 1977). Gaps are more frequent in shorter dependencies; RPs are more frequent at the foot of longer, more complex dependencies (Hawkins, 1999). Gaps tend to occur more often in wh-questions; RPs appear more in relative clause constructions Hawkins (1999); Boeckx (2003). Gaps are more likely with definite antecedents; RPs with indefinite antecedents (Ariel, 1999).

---

[4]Especially for example, for syntactic functions such as 'subject'.

Inspection of the patterns of variation in Yucatec reveals that the AF verb is more frequent and is preferred in contexts where gaps are more frequent within and across languages. The fully inflected (synthetic) verb form occurs more and is preferred where resumptives are more frequent, cross-linguistically. On the basis of such parallels, together with close inspection of Yucatec's person marking system, I propose a new classification of the Mayan AF alternation as a type of RP/gap alternation. The difference between a language like Yucatec, and other types of RP/gap alternating languages lies in the locus of the alternation. Rather than involve the presence or absence of an independent pronoun, the alternation in a language like Yucatec relates to two verb forms: one which carries a morphologically dependent subject pronoun (the synthetic verb), the other (the AF verb) which does not.

Identification of these typological parallels shifts the Mayan AF alternation from a phenomenon *sui generis* which is not cross-linguistically comparable, to a subtype of a larger cross-linguistic typology against which it now can be compared. This both enriches the typology of resumption in interesting ways and in the process, allows us to reevaluate assumptions about the organization of Mayan syntax – Mayan syntax has variously been argued to be, i.a., ergative (Larsen and Norman, 1979; England, 1983, 1988; Campana, 1992; Campbell, 2000), obviative (Aissen, 1999), harmonically aligned (Aissen, 2003) or information structurally based (Tonhauser, 2003) and the existence of AF verbs in Mayan languages has typically been invoked as the central piece of evidence in support of these theories.

Classifying the Mayan AF alternation as a morphological subtype of RP/gap alternation is directly supported by the important observation that this subtype is cross-linguistically widespread: I show that alternations equivalent to the AF alternation are in fact exhibited in an impressive range of head-marking languages genetically unrelated to Mayan, including, but not limited to Hausa, Kinande, Somali, Fiorentino and Trentino, Halkomelem Salish, Chamorro, Palauan, Tukang Besi, Moore, Konjo, Chalcatongo Mixtec, Selayerese, Makaresse, Irish, Yimas, Yagua, Apurina and Bare. Just like Mayan languages, such languages make use of special verb forms in dependent clauses in A-bar dependencies (wh-questions, relative clauses and clefts), which are distinguished (minimally) from their main clause counterparts by the absence of person marking inflection.

The language specific facts relating to such verbal alternations have been reported separately for individual languages, and analyzed in a variety of different ways. In the generative literature the phenomenon has been discussed for (subsets of) these languages under the term *wh-agreement* or *anti-agreement*. These approaches attempt to isolate categorical parameterized constraints to account for the form and distribution of these verb types. I argue that these analyses fail to account for the amount of cross-linguistic variation in this domain. Perhaps most strikingly, they also fail to identify any logical connection between the phenomenon and the language types in which it occurs. The deeper question is: why do RP/gap alternations in some languages manifest themselves in the verbal domain, and why are these languages all head-marking languages?

To answer these questions we must look to a deeper level of explanation, a generalizing-historical explanation. I argue that the origins of analytic verbs can be traced diachronically to the emergence of head-marking.

Diachrony: analytic verbs as the by-product of the rise of head-marking (Chapter 5)

A rich tradition of research on grammaticalization has established that the higher the frequency of use of co-occurring linguistic expressions the more likely that a fused unit comprising these expressions will be formed and reanalyzed by speakers. (Bybee, 1988; Keller, 1994; Kirby, 1999; Nettle, 1999; Haspelmath, 1999, 2004; Croft, 2000a; Heine and Kuteva, 2002). As Bybee (2002b, 112) puts it: 'items that are used together fuse together'. Higher frequency of use also leads to greater predictability, allowing the speaker to reduce her articulatory effort. Thus, frequent expressions exhibit a greater tendency to be phonologically shortened than rare expressions (Zipf, 1935). In this way, the main aspects of grammaticalization, tighter structures and attenuated forms, are linked to frequency.

Applied to the domain of person marking, a necessary precondition for the development of synthetic (pronominally inflected) verbs is a high frequency of pronoun-verb (or verb-pronoun) adjacency. While numerous grammaticalization paths have been identified by which the appropriate conditions emerge to give rise to the morphologization of pronouns onto verbs in *main clauses* (see e.g. Givón, 1976; Ariel, 1998, 2000; Fuß, 2005), I argue that it is precisely in subordinate clauses where gaps are most frequent that this grammaticalization process is less likely to take place, because of the relatively lower frequency of pronoun-verb co-occurrence in these environments. This can result in the retention of

conservative verb forms in these contexts, which remained exempt from the grammaticalization of person marking. In short, analytic verbs are the byproduct of grammaticalization processes occurring elsewhere in the grammar: the morphologization of person marking.

A number of case studies supports this conclusion. I examine a cluster of French and Northern Italian varieties in which subject person clitics are in the process of developing into affixes. This study allows us to see the very beginnings of the emergence of verbal alternations in nascent head-marking languages. I show that while, in a number of respects, there is evidence in these languages of increasing fusion between the subject person marker and the verb in main clauses, significantly, dependent verbs in subject extraction contexts remain exempt from this process.

I also show how grammaticalization processes in other domains can result in further morphological differentiation between main and dependent verbs, and hence the appearance of apparently exceptional morphology on dependent verbs in A-bar dependencies in certain languages. For example, a study of Classical Yucatec Mayan texts and Colonial grammars reveals that the rise of imperfective aspect marking in main clauses has led to the increased morphological differentiation of main and AF verb forms in the modern language. In Western Malayo-Polynesian, so-called 'wh-agreement' morphology has been shown to be the residue of the old Philippine type voice system which has eroded in main clauses in Chamorro and Tukang Besi through the development of head marking in main clauses (Donohue and Maclachlan, 1999; Dukes, 2000). I point to a similar phenomenon in Central Mayan languages, where the special AF suffix on dependent verbs in extraction contexts is identifiable as an old Proto Mayan suffix found on finite main verbs which has been lost in many modern languages in main clauses.

Processing: asymmetric RP/gap distributions are driven by processing (Chapter 6)

The result then, is that it is the asymmetric frequency distribution of gaps and pronouns which can give rise to verbal alternations in A-bar dependencies in head-marking languages via the grammaticalization of frequently co-occurrent sequences. A deeper question remains, however: why do these asymmetric frequency distributions exist? What governs the patterns of distribution that we find, within and across languages? Fully understanding the verbal alternations in head-marking languages is part of the larger problem of explaining the principles underlying resumptive/gap distributions more generally.

Various researchers have converged on the notion that the distributions are ultimately tied to processing preferences. This is based on the observation that resumptive pronouns tend to occur more frequently in more complex environments. This is visible both within languages which have productive RP/gap alternations in terms of the patterns and preferences found in performance and is also mirrored by the patterns one finds across languages in the fixed conventions of grammars.

Why do resumptives tend to occur in more complex environments? What role do they play in facilitating processing difficulty in these contexts? While various storage based processing hypotheses that have been put forward in the literature (Keenan and Comrie, 1977; Hawkins, 1999, 2004; Ariel, 1999; Alexopoulou and Keller, 2007), to date, very little experimental data on production exists to directly confirm the predictions of any or all of these theories. This is an obvious consequence of the fact that English, the target language of most processing research (see Jaeger and Norcliffe 2009), does not have productive resumption. For productive RP/gap alternations, the empirical evidence on which processing theories are based has come mainly from small corpus analyses, observations of written and spoken discourse, magnitude estimation studies, impressionistic generalizations, or through indirect extrapolating from the results of the much more numerous set of filler-gap experiments on English. Given that asymmetric RP/gap distributions have constituted a signature phenomenon evoked in support of the performance-grammar hypothesis (see e.g., Hawkins, 2004), there is an obvious need to increase our understanding of the factors driving variation in production in this domain.

To this end, I present a series of three experiments on Yucatec (two production studies and an interpretation task) which constitute a modest first step at procuring more controlled, quantitative production data on a productive RP alternating language. As far as I am aware, these represent the first controlled production studies to be undertaken on a productive RP/gap alternating language. They are also among the first controlled experimental studies on morphosyntactic variation in production to be undertaken on a Mayan language.[5]

Three important findings emerge from these studies. Most broadly, they demonstrate that Yucatec variation in production mirrors some of the patterns of asymmetrical RP/gap distributions that have been attested *across* grammars, providing some of the first robust quantitative support for the performance-grammar hypothesis in this domain. We find that

---

[5]Yucatec has also been the focus of recent experimental research on syntactic (word order) variation (Skopeteas and Verhoeven, 2009, 2005; Gutiérrez-Bravo and Monforte, 2008).

in Yucatec, dependency complexity (as measured by embedding depth) and the salience of the antecedent (as measured by definiteness) are significant predictors of the choice between the two structures: the resumptive structure tends to occur in more complex dependencies, and where the antecedent is indefinite.

A more specific finding is that in complex dependencies, Yucatec speakers, in addition to having the option of either using the AF structure or the synthetic (resumptive) structure, also make use of a third strategy, in which the relativized argument is expressed as an independent resumptive pronoun. Analysis of the variation between the three response types (gap vs. synthetic resumptive vs. independent resumptive) reveals that the level of formal attenuation of the referential expression concluding the dependency is inversely correlated with the complexity of the dependency (as measured by embedding depth), and with the salience of the antecedent (as measured by definiteness).

(1.1) gap   < bound resumptive < independent resumptive

       less complex < more complex

This result mirrors results reported for referential form choice in discourse, where specificity of referential form is inversely correlated with the accessibility of its antecedent and the length between the antecedent and subsequent referent (Givón, 1976; Ariel, 1990; Gundel et al., 1993). This correlation suggests that processes shown to operate at the intersentential level are also detectable at the intrasential level, indicating the need to bring discourse-level processing principles into the account of what have often been taken to be autonomously sentence-internal phenomena (cf. Ariel, 1999).

Finally, the Yucatec studies speak directly to the issue of ambiguity avoidance, and its relevance for choices in production. The notion of ambiguity avoidance has played a strong role in processing theories of resumption. On many accounts, the presence of a resumptive is argued to facilitate the identification of the extraction site (Givón, 1973, 1975; Keenan, 1972, 1975; Keenan and Comrie, 1977; Tarallo, 1986; Silva-Corvalán, 1996; Hawkins, 1994, 1999).

In Yucatec, along with a subset of morphological RP languages, the presence of the resumptive element can create globally ambiguous strings. While the gap-alternant (the AF structure) only has one available interpretation, that of a subject relative reading, the resumptive-alternant (the synthetic structure) has two interpretations: it can (in appropriate

contexts) be interpreted as either an object or a subject relative clause. If ambiguity considerations are a lead factor in determining the choice between pronouns and gaps, then, interestingly, this leads to the inverse prediction for a language like Yucatec: Yucatec's gapping structure would be expected to be preferred where resumptives are preferred in other languages, because it is only the gapping structure that unambiguously identifies the relativized argument.

The Yucatec studies presented in this dissertation collectively show that Yucatec speakers do not exhibit any preference for producing the disambiguating AF verb more in contexts where the intended interpretation is pragmatically biased against. Despite the fact that in these contexts, it is experimentally demonstrated that hearers are more likely *interpret* strings with the full verb form as object relatives rather than subject relatives, this has no influence on the choice between verb forms in production.

<p style="text-align:center">∞∞∞∞∞∞∞∞</p>

The remainder of this introductory chapter is devoted to an overview of Yucatec Maya, its geographic distribution and genetic affiliation, and aspects of its grammar which are relevant for the present study.

1.1 Introduction to Yucatec Maya

1.1.1 Geographic distribution and genetic affiliation

Mayan languages are spoken in Guatemala, Mexico and Belize, by around 3 500 000 speakers. There are about thirty different languages in the family, divided into five major branches (Campbell and Kaufman, 1990). The language this dissertation is chiefly concerned with is Yucatec, from the Yucatecan branch of the family. It is a language spoken by around 800 000 people in the Yucatán peninsula in the southeast of Mexico, and in parts of Belize and Guatemala.

Yucatec is one of the largest indigenous languages of the Mesoamerican area. Bohnemeyer (2002) calls attention to the fact that despite Yucatec's populousness, dialectally the language is remarkably homogeneous. Native speakers from any part of the peninsular can communicate with each other without difficulty. He surmises that the lack of obvious diversity probably is due to the uniform nature of the topography of the Yucatán peninsula,

and also due to the fact that large parts of the center and east of the present day Yucatec area were largely uninhabited for the first three centuries following the conquest.

On the basis of certain regular lexical and morpho-phonological differences (see Pfeiler, 1995), it is possible to differentiate between a north-western variety of Yucatec, which includes the urban areas around Merida and Campeche, and another variety, spoken principally in Valladolid and surrounding villages, and in the rural areas to the east and south of Valladolid, including those parts of Belize where the language is spoken. Following Bohnemeyer (2002), I will refer to these dialects as western Yucatec and eastern Yucatec. The western dialect displays a greater influence from Spanish, but is also more conservative, lexically and morpho-syntactically. Edmonson (1986) suggests that these dialects may have already been differentiated by pre-Hispanic times. The present study is based on the eastern variety of Yucatec.

Classical Yucatec (also referred to as Colonial Yucatec) is ancestral to modern Yucatec. It is considered to have been spoken between the middle of the 15th century and the middle of the 17th century (McQuown, 1967, 202). A small number of texts have been collected from this period, written in an orthography developed by Spanish missionaries (based on the Spanish script). A number of contemporary Colonial grammars exist, including Beltrán (1746); Coronel (1620); de San Buenaventura (1684). There are two 20th century descriptions of Classical Yucatec: McQuown (1967) and Smailus (1989).

1.1.2 Data collection and sources

The Yucatec data used in this dissertation are drawn from a mixture of direct elicitation, controlled spoken production experiments (involving computer based translation tasks), as well from secondary sources (naturally occurring textual examples and elicited utterances).

Direct elicitation sessions were undertaken in the field with six Yucatec Maya speakers over the course of three field trips, the first in the village of Chan Chen, Quintana Roo, during March 2007, the second and third in Valladolid, Yucatán, during April 2008, and between September and October 2008. In addition, regular consultation sessions were undertaken with a native Yucatec Maya speaker living in San Francisco throughout 2007 and 2008. All speakers were bilingual (Yucatec-Spanish), who had grown up speaking Yucatec in the home as their first language. All consulting sessions were conducted in Spanish.

The experimental studies described in chapter 6 were undertaken with a pool of around 40 student participants from the Universidad de Oriente, a state university in the Yucatán. This university is located just outside of the town of Valladolid, and is attended by students living throughout the state of Yucatan, with the bulk of the population concentrated in Valladolid and surrounding villages. The students who participated in the studies were all bilingual (Spanish/Yucatec) speakers, who spoke Yucatec in the home.

Secondary sources of data for Modern Yucatec are taken from Bohnemeyer (2002), Bohnemeyer (2009), Verhoeven (2007), Lois and Vapnarsky (2003) and Lehmann (1998). Classical Yucatec data are taken from two primary texts, *The Morley Manuscript* (Author unknown, c. 1576; annotated and translated by Whalen (2003)), and from excerpts of the *The Book of Chilam Balam of Chumayel* (1782), translated and transcribed by Tozzer (1921), as well as two Colonial grammars, Coronel (1620) and Beltrán (1746).[6] I also draw from two early twentieth century grammars, Lopez Otero (1914) and especially Tozzer (1921).

1.1.3 Some background on Yucatec grammar

The grammatical overview presented here is based on insights from Bohnemeyer (2002); Verhoeven (2007); Lois and Vapnarsky (2003) and Lehmann (1998) which in turn have built on earlier twentieth century work on Yucatec by Andrade (1955); Blair (1964); Blair and Vermont Salas (1967); Owen (1968); Bricker (1978, 1979) and Bricker et al. (1998). I restrict this sketch to those aspects of Yucatec grammar which are particularly relevant for the discussions of the AF alternation presented in the following chapters. For fuller accounts, the reader is directed to the above-mentioned sources.

Person marking

Mayan languages are head-marking (in the sense of Nichols, 1986), in that they mark core grammatical relations on the predicate head (verbal, nominal, adjectival or locative). This is done by means of two sets of person markers, typically referred to in the Mayan structuralist tradition as Set A and Set B. These markers may indexically represent a referent introduced

---

[6]Coronel's grammar, *Arte en lengua maya* represents the oldest extant Colonial grammar; Beltran's later grammar is taken to reflect the transitional stage between Classical and Modern Yucatec (Andrade, 1955, iv)

in discourse or given in the speech situation. Noun phrases co-indexed by the set A and B markers are syntactically optional. The status of these markers is an important question, and one which bears directly on the analysis of the Agent Focus alternation. In chapter 3 I will motivate their classification as pronominal arguments. Until this point I will use the neutral terms 'person marker' or 'set A/B marker' when referring to these elements.

In Yucatec the set A series are preverbal elements which are generally classified as clitics (Bohnemeyer, 2002; Verhoeven, 2007).[7] Before a vowel initial verb they are typically preceded by a glide (the bracketed consonants in the table below). The set B series are suffixes, which attach to the verb stem.

	Set A		Set B	
	Sg	Pl	Sg	Pl
1	in(w)	k(...-o'n)	-en	-o'n
2	a(w)	a(w)...-e'x	-ech	-e'x
3	u(y)/y	u(y)...-o'b	-ø/-ih	-o'b

Table 1.1: Yucatec's bound pronominal system

Set A cross references transitive subjects of verbs (the 'A' argument in the sense of Dixon, 1979), and set B cross references direct objects, and intransitive subjects (Dixon's O and S arguments, respectively). This basic ergatively aligned head marking pattern is common across Mayan languages, though Yucatec distinguishes itself by additionally exhibiting a split intransitive system of argument marking, which depends on the overt aspect/mood category of the verb. In the imperfective aspect, the sole participant of an intransitive verb is marked like the agent of a transitive (i.e, with a set A marker), while in the perfective and subjunctive, it is marked like the patient of a transitive verb (i.e with a set B marker).

The following examples of simple clauses illustrate these person marking patterns. In (1.2), the transitive verb *il* 'to see', takes the third person set A clitic *uy*, which references the agent of the verb, and the second person set B suffix *en*, which references the undergoer. In (1.3), the agent of the intransitive verb *sùut* is referenced with the set A marker (first person), because the verb is in the imperfective. In (1.4), because the verb is in the perfective, the agent of the verb is referenced with the set B suffix:

---

[7]Set A markers in fact exhibit certain mixed properties of clitics and affixes; see chapter 3 for discussion.

(1.2) *K-**uy**=il-ik-**en***
IMPF-A3=see-INC-B1
'He/she watches me.'

(1.3) *K-**in**-sùut*
IMPF-A1-return
'I return/am returning.'

(1.4) *H-sùut-nah-**en***
PRV-return-CMP-B1
'I returned.'

In addition to the dependent person markers, Yucatec also possesses a set of independent personal pronouns, which may bear emphatic stress. These occur in topic or focus positions, or otherwise in indirect object/oblique functions, or as the complement of a preposition (see §3.2.2).

Word order

When lexical NPs are present in the clause, word order is canonically VOS.[8] Topical subject NPs are frequently left dislocated (clause external), giving the appearance of SVO ordering.

(1.5) VOS

T-u=kin-s-ah   *le*   *sina'an*   *le*   *máak-o'*
PRV-A3-die-CAUS-CMP   DEF   scorpion   DEF   man-D2
The man killed the scorpion.'

(1.6) S$_{topic}$VO

Le   *máak-o'*   *t-u=kin-s-ah*   *le*   *sina'an-o'*
DEF   man-D2   PRV-A3-die-CAUS-CMP   DEF   scorpion-D2
'The man, he killed the scorpion.'

---

[8] See §3.2.2 for a discussion of postverbal word order constraints.

Deictic enclitics

Yucatec possesses a set of four deictic enclitics which occur at the right edge of an in-
dependent clause. No more than one clitic may occur per clause. The choice between
clitics is determined by a dominance hierarchy (cf Hanks, 1990, chs 6,10): a clause con-
taining a noun phrase which takes the definite determiner *le* must be followed by either
the 'proximal' enclitic *-a'* or the distal enclitic *-o'* (1.7). If the clause contains no definite
noun phrase, then the neutral/textual deictic marker *-e'* may occur clause finally. The topic
marker (TOP) is one instance of this function (1.8). Finally, the particle *-i'* is triggered by
the negation maker *ma'* (1.8) and also occurs in locational focus constructions.

	D1/D2	*-a'/-o'*	proximal/distal deixis (in space and time)
>	D3	*-e'*	neutral/textual deixis
>	D4	*-i'*	locational focus; negation

Figure 1.1: Yucatec clause final deictic enclitics

(1.7) *T-inw=il-ah*   *x-maria*   *yéetel u*   *mama ti'*   *le*   *mercardo-o'*
      PRV-A1-see-CMP FEM-maria and   A3 mother LOC DEF market-D2
      'I saw Maria and her mother at the market.'

(1.8) *Ko'lel-o'b-e'* *ma'*   *táan*   *u*   *bin-i'*
      lady-PL-TOP NEG PROG A3 go-D4
      'Women don't go there.'   (Verhoeven, 2007, 105)

Internal clause structure

In describing the structure of the Yucatec clause I follow the standard terminology first
introduced by Lehmann (1993a). The basic structure of the verbal clause is illustrated in
(1.1.3):

The verbal clause is made up of a pre-verbal AM MARKER (Aspect/Mood marker),
here the terminative *ts'o'k*), plus the VERBAL CORE. The verbal core consists of the set
A marker (here the third person singular *u*), an optional sentence adverbial clitic (here the
repetitive *ka'*), the VERB ROOT (*wèen*), optionally a derivational suffix (the causative suffix
-s), the postverbal STATUS SUFFIX (the incompletive *-ik*), and a set B suffix (second person
singular *-ech*).

A/M marker ≺ Verbal core

	A.clitic ≺	(SAdv) ≺	Root	Deriv.	Status	B.suffix	
ts'o'k	*u=*	*ka'=*	*w*	*-s*	*-ik*		*-ech*
TERM	A3	REP	sleep	CAUS	INC	B2	

'He made you sleep again'

Figure 1.2: Structure of the Yucatec independent verbal clause

As can be seen from the above example, aspect/mood are marked simultaneously in two different positions in Yucatec, by the preverbal AM marking, and at the same time by a post-verbal status suffix. Each independent verbal clause contains exactly one out of fifteen different preverbal AM markers, which express different aspects, modalities, or degrees of temporal distance (e.g. remote vs. recent past) (see Bohnemeyer, 2002). Each preverbal AM marker in turn triggers one of three possible postverbal status suffixes: incompletive, completive, or subjunctive. The terminative AM marker *ts'o'k* in the example above triggers the incompletive status suffix *-ik*.

The morphosyntactic relationship between the preverbal AM and the rest of the verbal complex (the 'verbal core') is not entirely agreed upon in the literature, and the situation is made more complex by the fact that the AM markers themselves do not form an entirely uniform class. Most AM markers, as in the example above, function as independent predicates. Bohnemeyer (2002) offers two diagnostics which may be used to determine the predicative status of an element in Yucatec, both of which involve the use of polar questions in discourse. Polar questions are typically formed with a special question focus marker *wáah* which cliticizes to the main predicate of the clause. In clauses with independent AM markers, it is the independent aspect marker which attracts the question focus marker:

(1.9) *Táan   wáah   u   hats'-ik-ech?*
    PROG   ALT   A3   beat-INC-B1
    'Is he beating you?'

The second piece of evidence comes from affirmative answers to polar questions, which in Yucatec require either the entire proposition to be repeated, or just the main predicate. (1.9)

is corroborated by means of the aspect marker alone, indicating again that the independent
aspect marker functions predicatively:

(1.10) *Táan*
     PROG
    'He is.'

Authors are in agreement that in clauses with independent AM markers, the AM marker
functions as the matrix predicate, selecting the rest of the clause as its argument. There
are differences of opinion as to what category these predicative aspect markers belong to,
however. Bricker (1981) argues that they are grammaticalized verbs, Lehmann (1993a)
proposes that they are auxiliaries. Bohnemeyer (2002) argues that they are best analyzed as
stative predicates, because they bear none of the properties of verbs in Yucatec (and hence
are not auxiliaries either): they don't themselves inflect with status suffixes for example
(see the section below on verbal categories), and they don't function like auxiliaries in
the Indo-European sense - they are not themselves *carriers* of verbal inflection. On this
analysis the AM marker functions as a stative predicate that takes the rest of the verbal
complex as an argument.

    In addition to the set of AM stative predicates, there is also a set of two morphologically
bound AM markers (the perfective and the imperfective), which occur prefixed to the set
A marker. In verbal clauses with bound AM markers, the AM marker in combination with
the verbal constituent functions as the main predicate. This is made evident by the polar
question diagnostics: *waáh* attaches to the verbal predicate. The question is corroborated
via the repetition of the verbal predicate:

(1.11)   a. *T-u=sat-ah*   *wáah?*
            PRV-A3=lose-CMP   ALT
            'Did he lose it?'

        b. *T-u=sat-ah*
            PRV-A3=lose-CMP
            'He did.'

The table below presents Yucatec's set of bound and independent aspect markers:[9]

---

[9]For a complete list, including modal and distance predicates, see Bohnemeyer (2002, 103).

Type	Category	Form	Status triggered
Bound	Perfective	*t-/h-*	Completive
	Imperfective	*k-*	Incompletive
Aspectual predicates	Progressive	*táan*	Incompletive
	Terminative	*ts'o'k*	Incompletive
	Prospective	*mukah / mikah / bikah*	Incompletive (V_{itr}) / Subjunctive (V_{tr})

Table 1.2: Preverbal aspect markers (adapted from Bohnemeyer, 2002, 103)

Valency alternations

Yucatec, like many Mayan languages, exhibits a rich system of voice derivations, including an antipassive, a passive and an anticausative (often referred to as a 'middle voice' by Mayanists). Passivization renders the verb intransitive, with the undergoer of the verb appearing as the subject in the passive clause. The agent is optionally expressed by means of the preposition *tuméen* 'by, because of'. The anti-causative also functions to detransitivize, but unlike the passive excludes any implication of a controlling agent. In the antipassive the undergoer argument of the transitive base verb is not realized.

(1.12) gives the three voices for the root transitive *hek'* 'break':

(1.12)   a. Active transitive

        K-in   hek'-ik
        IMPF-A1 break-INC
        'I break it.'

   b. Passive

        K-u   he'k'-el
        IMPF-A3 break.PASS-INC
        'It is broken.'

   c. Anticausative

        K-u   ***héek'-el***
        IMPF-A3 break.ACAUS-INC
        'It is breaking.'

  d. Antipassive

        K-in   ***hèek'***
        IMPF-A1 break.ATP
        'I break.'

While the passive applies regularly to all transitive verbs, and is therefore considered an inflectional category (Verhoeven, 2007; Bohnemeyer, 2002), the antipassive and the anticausative only apply to certain subsets of transitive verbs and are therefore considered derivational. The anticausative applies only to underived transitives, while the antipassive is formed from basic and causativized transitives only.

Predicate classes

Like most Mayan languages, Yucatec distinguishes morphologically between transitive and intransitive verb classes on the basis of person marking (as described above) as well as status inflection. Status marking in Yucatec additionally distinguishes between different classes of intransitive verbs (inactive, active, inchoative and positional).[10] Each verb class has its own inflectional paradigm for the three status categories (incompletive, completive and subjunctive).[11] [12]

    Yucatec possesses a rich array of predicate-class changing derivations. These include causativization with *-s* (and occasionally *-bes*), which applies to verbs of the inactive intransitive class and applicative transitivization (with *-t*), which applies to active intransitive verbs.

---

[10]See Lehmann (1993b); Lucy (1994); Krämer and Wunderlich (1999) and Bohnemeyer (2004) for studies on the semantic extension of these classes.

[11]There is an additional fourth inflectional status category, the extrafocal, which is related to the completive. It marks the out-of-focus predicate in manner focus constructions. See Bohnemeyer (2002) for discussion.

[12]The passive may be formed in two distinct ways, which are in complementary distribution. In most root transitives, a glottal stop is inserted into the root and the status suffixes (the same as those of the intransitive inactives) are added. Derived transitives take a suffix, but do not undergo a stem change, as shown in the second line of the passive in table 1.3

	Incompletive	Completive	Subjunctive
tr. active	-ik	-ah	-ø/-eh
passive	...'...-Vl	...'...-ab	...'...-Vk
	/-a'l	/ -a'b	/ -a'k
intr. inactive	Vl	-ø	-Vk
intr. active	-ø	-nah	-nak
intr. inchoative	-tal	-chah	-chahak
intr. positional	-tal	-lah	-l(ah)ak

Table 1.3: Yucatec status inflection paradigms according to verb stem classes

1.1.4   Relativization

I conclude this chapter with an overview of the suite of three constructions which are the focus of this dissertation: wh-questions, focus-clefts and relative clause constructions. The Agent Focus alternation, to be introduced in chapter 2, is restricted to a subset of these construction types. In the remainder of this section I provide some of the necessary details of their structure, which will be called upon in the remaining chapters.

It will be noted that I have chosen to label this section 'relativization'. This reveals at the outset the particular view of these constructions, espoused by Bricker (1979); Bohnemeyer (2002) and built on here, that wh-questions, focus-clefts and relative clause constructions in Yucatec share a structural affinity: all three are composed of a higher matrix predicate and an embedded relative clause. I will begin by introducing the relative clause construction proper, and then will turn to the properties of wh questions and clefts.

Relative clause constructions proper

Relative clause constructions in Yucatec are postnominal, relativizerless and head-external. An example of a subject-extracted relative clause is presented in (1.13), and of an object-extracted relative clause in (1.14). It may be observed that the placement of the deictic clitic provides evidence that the relative clause is embedded under the head noun. In relative clause constructions which feature a definite head, the deictic clitic attaches to the right edge of the NP *containing* the relative clause.

(1.13) *le   x-ch'upaal [k-uy=ok'ol]=o'   inw=íitsin*
     DEF FEM-girl   IMPF-A3=cry-D3   A1=sister
     'The girl who is crying is my sister'

(1.14) *In   k'ahóol le   k'oha'an [t-u=ts'ak-ah   le   doctor]=o'*
     A1   know   DEF patient   PRV-A3=cure-CMP DEF doctor-D3
     'I know the patient that the doctor cured.'

It is typical to describe relativization (of core arguments) in Yucatec (and Mayan languages generally) as involving a 'gapping' strategy; that is, that the syntactic-semantic role of the head noun in the relative clause is expressed by a 'gap', an empty element in the syntax (as opposed, e.g. to a resumptive pronoun), which is co-indexed with the head-noun (e.g. Gutiérrez-Bravo and Monforte, 2009). Schematically, this may be represented as such:

(1.15) *In   k'ahóol le   k'oha'anᵢ [t-u=ts'ak-ah   ___ᵢ le   doctor]=o'*
     A1   know   DEF patient   PRV-A3=cure-CMP   DEF doctor-D3
     'I know the patient that the doctor cured.'

Describing the termination of Yucatec unbounded dependencies in terms of a *syntactic* gap vs. an overt syntactic element is not something adopted in this dissertation. As I will argue in chapters 3 and 4, how dependencies are terminated in a language like Yucatec is not in fact determined by the presence/absence of independent syntactic elements, but rather by an alternation relating to two verb forms: one which carries a resumptive person marker, the other (the so-called Agent Focus verb) which does not. This point is developed in some detail in chapters 3 and 4. The remainder of the discussion in this section outlines attendant aspects of Yucatec relative clause structure.

In realis contexts, relative clauses in Yucatec take neither relative pronouns, nor complementizers. In certain irrealis contexts however, the subordinate clause is introduced by the 'irrealis subordinator' *kéen*. *kéen* occurs exclusively at the left edge of subordinate clauses, and only with future, habitual and generic reference. *Kéen* is in complementary distribution with the preverbal AM markers, which in relative clauses (and focus and wh-questions) are restricted to realis reference.

(1.16) *Le   kàarta [kéen   a=ts'íib-t   bèey]=o', hay=p'éel   tyèempo*
     DEF letter   SR.IRR A2=write-APP thus=D2   how.many=CL.IN time
     k-a=tukul-ik   u=xàan-tal?
     IMPF-A2=think-INC   A3=take.time-INCH.INC

'The letter you are going to write thus, how much time do you think it will take?'
(Bohnemeyer, p.c.)

Headless relatives are productive in Yucatec; they are particularly frequent in topic position. In such cases the head of the relative clause may contain only the definite determiner:

(1.17) *Le   [kéen   k=ts'a'   túun   he'l]=a'   u=k'àaba'-e',*
     DEF SR.IRR A1pl=put so.then PRSV=D1 A3=name=TOP
ka'anal+pàach+nah
  high+back+house
'So then the (one) we put here, as for its name, (it is) ka'anal pàach nah'

A range of positions may be relativized on, including subject, object, and oblique prepositional phrases. In the latter case, the preposition may either remain in situ postverbally within the relative clause (1.18a), or it may follow the head of the relative clause construction (1.18b):

(1.18) a. *K'àas le   máak [t-a=ts'a'-ah   le   ta'kin ti']=o'*
       bad   DEF person PRV-A2=put-CMP DEF money PREP=D2
      'The person you gave the money to is bad'

    b. *K'àas le   máak ti'   [t-a=ts'a'-ah   le   ta'kin]=o'*
      bad   DEF person PREP PRV-A2=put-CMP DEF money=D2
      'The person to (whom) you gave the money is bad'   (Bohnemeyer, 2009, 200)

Relative clause dependencies in Yucatec are potentially unbounded. Complement clauses may intervene between the head noun and the dependent clause:

(1.19) ... *le   peek t-uy=aal-ah   x-maria   chi le   paal-o'*
    ... DEF dog PRV-A3=say-CMP FEM-maria bite DEF boy-D3
    '...the dog that Maria said bit the boy'

Wh-questions and focus constructions

While some Mayanists have adopted a monoclausal displacement analysis of wh-questions and focus constructions, in which the wh-word or focussed NP is an argument which is

moved to a preverbal position within the clause,[13] it is generally agreed for Yucatec that wh-questions and focus constructions are types of clefts: the focussed/questioned element functions as the matrix predicate in a biclausal (relative clause) construction (Bricker, 1979; Bohnemeyer, 2002, 2009; Tonhauser, 2003).[14]

Some examples of focus constructions and wh-questions are given below. In (1.20) the second person independent pronoun *teech* is focused, in (1.21) the bare nominal *wàah* 'tortilla' is focussed, and in (1.22), the question pro-form *máax* 'who' is focussed:

(1.20) *Tèech   ximbalt   Juan*
      you   visit   Juan
      'It's you that visited Juan'

(1.21) *Wàah   k-u=haànt-ik   Juan*
      tortilla   IMPF-A3=eat-INC   Juan
      'Tortillas is what Juan eats.'

(1.22) *Máax   ximbalt   Juan?*
      who   visit   Juan
      'Who visited Juan?'

Bohnemeyer and Bricker specifically analyze focus and wh-constructions in Yucatec as pseudo-clefts: the remainder of the construction (the extra-focal material) is subordinate to the focussed predicate in the form of a headless relative clause, which functions as the subject of the main predicate. This yields the following basic constructions:

---

[13]Norman (1977) for example, argues that focus constructions in Mayan languages are monoclausal, with the focused NP extracted from a postverbal to a preverbal position. Aissen (1992) builds on this proposal, providing a more articulated analysis of focus (and topic) in Tzotzil, Jakaltek and Tzutujil. On her GB analysis, the focused NP moves to spec, I' from its original postverbal position, leaving behind a co-indexed trace. While Aissen (1992) has provided evidence that different expressions of topic are detectable across various Mayan languages, cross-family comparison of focus structures remains to be undertaken, in order to determine whether they exhibit uniform structures across the family. The classification of Mayan AF as an RP/gap alternation that I propose in Chapter 3 does not hinge on any particular structural analysis of wh and focus constructions (displacement vs. cleft), only on the fact that they are A-bar dependencies of some sort.

[14]Parallel analyses, based on similar argumentation, have also been proposed for Jakalteko (Norcliffe, 2006) and Mam (England, 1983), inter alia.

(1.23) Wh-question

[WH]*predicate* [HRC]*subject*

Báax *[NP ø [RC t-a=hàant-ah?]]*
what   PRV-A2=eat-CMP
'What is (what) he ate?'

(1.24) Focus construction

[FOCUS]*predicate* [HRC]*subject*

Wàah   *[NP ø [RC t-a=hàant-ah]]*
tortillas   PRV-A2=eat-CMP
'Tortillas is (what) he ate'

The predicativity of the focused/questioned constituent is made evident by Bohnemeyer's *waah*-test (introduced above). In focus and wh constructions, the focus marker attaches to the focussed constituent, indicating that it is the main predicate:

(1.25) *Tèech **wáah** he' le bèentana-o'?*
     you   ALT   open   DEF   window-D2
     'Was it you who opened the window?'   (Bohnemeyer, 2002, 119)

(1.26) *Séeban **wáah** túun k'a'náan a ts'o'k-s-ik?*
     fast   ALT   CON   NEC   A2   end-CAUS-INC
     'Is it fast that you have to finish it?'

The predicative nature of these forms is supported further by the simple fact that elements which are focused or questioned belong to the range of expressions that may function predicatively in Yucatec. The examples below demonstrate the predicativity of question pro forms (1.27a) and independent pronominals (1.27b)

(1.27)   a. *Aw-ohel   máax-en?*
          A2-knowledge   who-B1SG
         'Do you know who I am?'

     b. *Tèech*
       you
       'It is you'

In addition to the predicativity of the focused element, the biclausality of focus and wh question constructions is also by the fact that in irrealis contexts, the irrealis subordinator (introduced above) occurs to the right of the verb. This is not anticipated on a monoclausal displacement analysis:

(1.28) *Káa=t-u=núuk-ah*   *Pedro=e,*   *"Téen*   *[kéen*
      CON=PRV-A3=answer-CMP(B3SG)   Pedro=D3   me(B3SG)   SR.IRR
      hats   *hit]."*
      hits(B3SG)   hit
      Pedro answered, It's ME who is going to bat a hit.   (Bohnemeyer, 2002)

1.2   Overview of the dissertation

In chapter 2 I present an overview of the AF alternation in Yucatec and throughout the Mayan family, and critically review previous analyses of the phenomenon. In chapter 3 I present a new classification of Yucatec's AF alternation as a type of RP/gap alternation, and show how the variation exhibited in this domain mirrors typologically attested asymmetric frequency distributions of RPs and gaps. In chapter 4 I look beyond Mayan and show how alternations equivalent to the AF alternation are exhibited in a range of other head-marking languages. Just as in Yucatec, the choice between verbal alternants is conditioned by the same factors that are implicated in RP/gap distributions cross-linguistically, providing support for the notion that they are all exemplars of the same basic phenomenon. In chapter 5 I present evidence that the special verb forms that show up in extraction contexts in head marking languages are conservative verb forms which remained exempt from the grammaticalization of head marking. It is the relatively low frequency of adjacent pronoun-verb combinations in extraction contexts, by comparison with main clause verbs, which gives rise to asymmetric patterns of pronoun grammaticalization, and thus leads to the emergence of these morphological alternations. I present a number of case studies which support this basic diachronic trajectory. In chapter 6 I explore possible processing explanations for the asymmetric distributions of gaps and RPs and present three experiments on Yucatec which provide support for the Performance-Grammar Correspondence Hypothesis. These experiments also offer some of the first production data against which processing theories of resumption can be tested. In chapter 7 I summarize the findings of the dissertation and discuss some implications and open questions that they raise.

Chapter 2

Agent Focus in Mayan Linguistics

In Mayan languages, when a subject is extracted from a transitive clause, the relative clause verb may be realized in a special form, the so-called *Agent Focus*, or AF form. In Yucatec, this verb is distinguished from transitive verbs of main finite clauses in that it does not occur with the set A person marker, nor preverbal aspect marking, both of which are obligatory elements of all finite main clauses. In this dissertation I will refer to the person-marked transitive verbs of main clauses as *synthetic*. I will refer to subject extractions from transitive clauses as A-extractions (i.e., in the sense of Dixon's A argument).

(2.1) illustrates the contrast. In (2.1a) the transitive verb formed with the root *xok* 'to read', occurs in a main clause and thus is realized in the synthetic form. In the wh-question in (2.1b), where it occurs in an A-extracted relative clause, the verb is realized in the AF form, carrying neither the subject person marking, nor the preverbal aspect marking associated with the synthetic form:

(2.1)   a. *Le   máak=o' **k-u=xok-ik**   le   libro=o'*
               DEF man=D2 IMPF-A3=read-INC DEF book=D2
               'The man reads the book'

         b. *Máax t-aw=a'l-ah   **xok-ik**?*
               who   PRV-A2=say-CMP read-INC
               'Who did you say reads it?'

AF verb forms occur across the triad of relative clause construction types introduced in §1.1.4: wh-questions, as in (2.1) above, focus clefts (2.3a), and relative clause constructions proper (2.3b):

(2.2)   a. *Juan **xok-ik***
         Juan   read-INC
         'It is Juan who reads it.'

    b. *...le   máak **xok-ik=o'***
        ...DEF   man   read-INC
        'the man who reads it'

The distribution of the AF verb is restricted to subject extractions out of transitive clauses
(A-extractions). Where either the O or S argument of the relative clause is relativized on,
the AF verb does not occur. In such cases the verb form realized in the relative clause is
the same verb form of main clauses. Compare the object-extracted relative clause verb in
(2.3) with its identical main clause counterpart in (2.1a):

(2.3) *le   libro **k-u=xok-ik**   le   máak=o'*
     DET   book   IMPF-A3=read-INC   DET   man=D2
     'the book that the man reads'

(2.4) compares a main intransitive verb form (2.4a) with its subordinated counterpart in a
subject-extracted relative clause (2.4b), and in (2.3). Again, the subordinate intransitive
verb form is identical to that of main clauses:

(2.4)   a. *Le   máak=o' **h-sùut***
         DEF   man=D2   PRV-returned
         'The man returned.'

    b. *T-inw=il-ah   [NP le   máak [RC **h-sùut]]=o'***
        PRV-A1-see-CMP   DEF   man   PRV-returned=D2
        'I saw the man who returned.'

AF verbal alternations have been identified for around two-thirds of Mayan languages in
some or all of wh-questions, focus clefts and relative clause constructions. Apart from the
Huastecan branch, all major branches of the family include at least one language which
possesses AF alternations (Stiebels, 2006). Across languages, the phenomenon is identi-
fied strictly by its distribution: it is restricted to relative clause constructions which target
subjects of transitive verbs; and, loosely, by its formal properties: it involves the use of a
verb form which is distinct from the finite transitive verb form of matrix clauses.

AF verbs across the Mayan family share a common property: they do not carry set A (subject) person marking. Formally however, the morphosyntactic manifestation of AF is not uniform across languages. In Yucatec, as shown above, the AF verb is distinguished from synthetic verbs by not bearing preverbal aspect marking and set A person marking. In other Mayan languages, the AF verb, though similarly distinguished by lack of set A person marking, does not undergo any change in aspectual marking. Additionally, in languages throughout the Mayan family (but not in Yucatec), AF verbs take a suffix. Across the family, there are two (apparently etymologically unrelated) suffixes which occur in these contexts. In greater Mamean, and in the Western branches of the family (Cholan and Kanjobalan), the suffix is -(*V*)*n*. In greater Quichean, -(*V*)*n* also occurs, but only on (polysyllabic) derived transitive verbs, while *-ow* ($\sim$ *-o* $\sim$ *-w*) occurs on (monosyllabic) root (= underived) transitives.

Historically and synchronically, and within and across languages, the distribution of these suffixes is rather complicated. In some languages, they also productively derive antipassive verbs (e.g. Greater Quichean, Mamean), in others (e.g. Tzutujil) the suffix appears to be formally identical to an antipassive, though the verb does not exhibit all the behavioral properties of an antipassive. In yet other languages the suffix appears, at least historically, to have had a nominalizing or dependent marking function, and is also found on verbs in a range of non-finite contexts (e.g. Kanjobalan).

Beyond the differences in verb types, we also find a lack of consistency across languages in terms of the agreement relations involving the verb's person marking, the realization of the verb's object, and whether the AF verb is obligatory in A-extracted relative clauses, or may alternate with the finite verb form of main clauses.

In this chapter I provide an overview of the phenomenon of AF, both in Yucatec and across the family and discuss the various accounts that have been proposed to explain its distributional and formal properties.

2.1   Agent Focus in Yucatec

Yucatec is alone among the Yucatecan languages in possessing AF in A-extraction contexts. In closely related languages such as Itzaj, the subordinate verb is realized in the synthetic form of main clause verbs:

(2.5) Itzaj (Lois and Vapnarsky, 2003, 57-58)

    a. *La'ayti' k-uy-il-ik-ech*
       PR3   HAB-A3-see-INC-B2
       'He is the one who sees you.'

    b. *Max t-u-kin-s-ah*   *a'*   *balum he'l-o'*
       who CP-A3-kill-CAUS-CP DET jaguar OST-D2
       'Who killed that jaguar?'

In Yucatec, aspect-mood marking in the AF form exhibits a contrast between bare incompletive status for imperfective reference (2.6b) and the bare 'subjunctive' status for past perfective reference (2.7b).[1] That is, the completive status suffix which is required on main verbs in past perfective contexts (2.7a) does not occur on AF verbs in the past perfective, instead the AF verb inflects for the subjunctive status suffix. The subjunctive suffix is only ever overtly realized in clause-final position, and so is not visible if there is any other lexical material following the verb (compare 2.7b and 2.7c):

(2.6)   a. *Le*   *máak* **k-u=xok-ik**   *le*   *libro=o'*
         DEF man   IMPF-A3=read-INC DEF book=D2
         'The man read the book'

        b. *Máax* **xok-ik***?*
         who   read-INC
         'Who reads it?'

(2.7)   a. *Le*   *máak* **t-u=xok-ah**   *le*   *libro=o'*
         DEF man   PRV-A3=read-CMP DEF book=D2
         'The man read the book'

        b. *Máax* **xok-eh***?*
         who   read-SUBJ
         'Who read it?'

        c. *Máax* **xok** *le*   *libro=o'?*
         who   read DEF book=D2
         'Who read the book?'

---

[1]For future-time, habitual and generic reference, the irrealis subordinator *kéen* is required, plus the subjunctive form of the subordinate verb (see §1.1.4). I don't deal with these constructions in this dissertation.

The imperfective AF verb (verb stem + *-ik*) occurs outside of subject extraction contexts in certain subordinate environments. For example, it may occur in the subordinate clause of object control structures (2.8), and after motion verbs in extrafocal clauses with prospective aspect/mood marking (2.9). In such cases, the transitive subject of the subordinate verb is not expressed by any person marking, and is obligatorily bound by a higher NP.

(2.8) *Pedro-e'*   *k-u*   *túucht-ik* *Maria* ***man-ik*** *hun-p'éel*   *báaxal*
     Pedro-TOP IMPF-A3 send-INC Maria buy-INC one-CL.INAN toy
     'Pedro sends Maria to buy a toy'   (Verhoeven, 2007, 139)

(2.9) *tèen* *kin*   *bin* ***il-ik***   *u*   *kin-s-a'l*
     me   SR.IRR:SBJ.1.SG go see-INC A3 die-CAUS-PASS.INC
     '(...)it is me who will go (and) see him being killed'   (Verhoeven, 2007, 129)

Subjunctive inflected transitive verbs are found in a range of dependent contexts outside of AF constructions, including as embedded predicates of higher non-verbal aspectual or modal predicates, such as the predictive predicate *bíin*.

(2.10) *Bíin*   *u*   *ts'íib+óol-t*   *x-ch'úup-tal*   *xib-o'b,* *(...)*
      PRED A3 write+soul-APP(SUBJ) F-female-PROC.INC male-PL
      'The men shall wish to become women, (...)'   (Vapnarksy 1995:89, cited in Bohnemeyer 2002)

In time focus constructions with perfective reference:

(2.11) *Tèen-e'* *domìngo-ak* ***in*** ***hats'***   *hun-p'éel* *hit*
      me-TOP Sunday-ak A1 beat(SUBJ) one-CL.IN hit
      'Me, it was last Sunday that I batted (lit. beat) a hit'   (Bohnemeyer, 2002, 124)

And in irrealis contexts following the subordinator *káa*:

(2.12) *In* *k'áat* *káa* ***u*** ***bis-en***   *Cancun* *in* *tàatah*
      A3 wish that A3 take(SUBJ)-B3 Cancun A1 father
      'I want my father to take me to Cancun'   (Verhoeven, 2007, 132)

However, note that in all these cases the subjunctive-inflected verb is preceded by the set A person marker, unlike in A-extraction contexts.

Yucatec has often been cited in the literature as obligatorily requiring the AF verb in A-extractions. (Tonhauser, 2003; Lois and Vapnarsky, 2003; Bohnemeyer, 2002), yet it has been confirmed (at least for the eastern variety of Yucatec), that the AF verb form is not in fact obligatory under agent extraction in certain conditions (Gutiérrez-Bravo and Monforte, 2009; Norcliffe, 2009). Rather, it may alternate with the synthetic verb form. The following is an example of a minimal pair, (2.13a) shows the AF verb alternant, (2.13b) shows the synthetic alternant:

(2.13)   a. *Le   chàan xibpàal   **ts'éent-ik** le   mono=o'   yàan u   ya'ax*
            DEF little   male.child feed-INC   DEF monkey=D2 exist A3   green
         p'ok
            hat
            'The little boy who is feeding the monkey has a green hat.'

         b. *Le   chàan xibpàal   **k-u=ts'éent-ik**   le   mono=o'   yàan*
            DEF little   male.child IMPF-A3=feed-INC   DEF monkey=D2 exist
         u   ya'ax p'ok
            A3   green hat
            'The little boy who is feeding the monkey has a green hat.'

It is important to note that subject extractions involving the synthetic verb form are ambiguous between a subject relative and an object relative interpretation. Thus, (2.13b) also has the possible meaning: 'The little boy who the monkey is feeding has a green hat'. The only interpretation available to the AF alternant, by contrast, is that of a subject relative.

The choice between alternants in A-extracted relative clause constructions can be shown to be conditioned by a number of different factors. For example, among my consultants the AF form is preferred over the synthetic form, and is spontaneously produced more in wh-questions (2.14). In relative clause constructions, by contrast, both the synthetic and the AF form are acceptable, and both occur in spontaneous speech (2.15):

(2.14)   a. *??Máax **t-uy=il-ah**   x-maria=o'*
            who   PRV-A3=see-CMP FEM-maria=D2
            'Who saw Maria'?

    b. *Máax **il**   x-maria=o'*
      who   see FEM-maria=D2
      'Who saw Maria'?

(2.15)   a. *le   máak **t-uy=il-ah**   x-maria=o'*
       the man   PRV-A3-see-CMP   FEM-maria=D2
       'the man who saw Maria'

      b. *le   máak **il**   x-maria=o'*
        the man   see FEM-maria=D2
        'the man who saw Maria'

Finally, there also seems to be strong a preference among my consultants to use the AF form when the reference of the relative clause object is pronominal (especially third person):

(2.16)   a. *??le máak t-uy=isíin-t-ah=o'*
        DEF man   PRV-A3-wash-APP-CMP
        'The man who washed it/him/her'

      b. *le   máak isíin-t=o' pèek'?*
        DEF man   wash-APP A3   dog
        'The man who washed it/him/her'

Examples such as (2.16a), with the synthetic form, are preferentially interpreted as object relatives ('the man who she/he washed'), and the AF version is strongly favored in production (2.16b), at least in simple relative clauses.

    The factors conditioning the variation in Yucatec AF will be discussed in more detail in chapters 3 and 6. For now, I will note a couple of environments where no variation is possible. The first is in contexts where the relative clause object is a reflexive, and is bound by the relativized NP. In such cases, only the synthetic verb is possible in the subordinate clause. This is illustrated in (2.17) for a wh-question (the same generalization holds for focus clefts and relative clause constructions). The AF alternant in (2.17b) is ungrammatical:

(2.17)   a. *Máax t-uy=il=u=bah   ti'   le   néen-o'*
        who   PRV-A3=see=A3=self LOC DEF mirror-D2
        'Who saw himself in the mirror?'

    b. **Máax il=u=bah   ti'   le   néen-o'*
       who   see=A3=self LOC DEF mirror-D2
       Intended: 'Who saw himself in the mirror?'

The reflexive object is always interpreted as bound by the extracted NP. Thus in (2.18), *Juan* cannot be the antecedent of the reflexive:

(2.18) *Máax t-uy=a'al-ah   Juan t-u=p'uch'=u=bah?*
      who   PRV-A3-say-CMP Juan PRV-A3-hit-A3-self
      'Who$_i$ did Juan$_j$ say bruised himself$_i$'

A similar pattern obtains when the possessor of an object is relativized. A bound reading is possible with the synthetic verb (2.19). This structure also has a disjoint reading. The construction with the AF verb, by contrast, can only receive a disjoint reading (2.19b):

(2.19)   a. *Máax t-uy=isíin-t-ah   u   pèek'?*
         who   PRV-A3-wash-APP-CMP A3 dog
         'Who$_i$ washed his$_{i/j}$ dog?'

    b. *Máax isíin-t   u   pèek'?*
       who   wash-APP A3 dog
       'Who$_i$ washed his$_j$ dog?'

These reflexive constraints appear to be common across the Mayan family. They have been reported, for example, for Jakalteko (Craig, 1977) and Tzotzil (Aissen, 1999). The source of these constraints remains unclear and deserves further research.

2.2 Agent Focus across the Mayan family

Beginning with the eastern division of languages and then traveling westwards, I will briefly survey in the next few subsections the range of subtypes of Agent Focus found in the Mayan family beyond Yucatec. This is not intended as an exhaustive survey, but is aimed rather at drawing out the range of differences exhibited across the family in the expression of AF alternations.

2.2.1 Eastern Mayan languages

Quichean

AF alternations are attested in all of the languages of the Quichean family: K'iche', Tz'utujil, Sipakepense, Sakapultek, Kaqchikel, Poqomam, Poqomchi', Uspantek and Q'eqchi'.

In Q'eqchi', Poqomam and Poqomchi', the verb forms used in AF contexts are, formally, inflected like antipassive verbs. Antipassivization in these languages serves to intransitivize a transitive verb, such that the object is either syntactically unrealized (inducing an indefinite/generic interpretation of the object), or occurs as an oblique, prepositional phrase, or is incorporated. Accordingly, the subject of the verb, the sole direct argument, is referenced with a person marker from the set B series, like other intransitive verbs.

Q'eqchi' will serve as an example for illustrating the formal identity of antipassive and AF verbs in this subset of languages. In Q'eqchi', there are two different suffixes which derive antipassives, depending on the verb type. For root transitive verbs, which are underived, and monosyllabic, the antipassive suffix is *-o*. For derived transitive verbs (polysyllabic), the suffix is *-n*. (2.20a) gives an example of a (derived) transitive verb; (2.20b) its antipassive counterpart inflected with *-n*; and (2.20c) an antipassive verb derived from a root transitive with *-o*:[2]

(2.20) Q'eqchi' (Dayley, 1981, 19–20)

        a. *X-at-ch'iila*
           T-B2-A1-scold
           'I scolded you.'

        b. *X-in-ch'iila-n*
           T-B1-scold-ATP
           'I was scolding.'

        c. *X-at-b'is-o-k*
           T-B2-measure-ATP-M
           'You were measuring.'

---

[2]Here and throughout the dissertation I generally conserve the original glosses of the source text in citing language examples, with the following exceptions. In all Mayan examples I employ the conventional A and B glosses for the two sets of dependent pronouns, in order to maintain consistency. I have also changed the abbreviation for the antipassive (AP) in many texts, to ATP in order to differentiate this more clearly from the applicative abbreviation (APP). All abbreviations are listed on the abbreviations page at the beginning of the document.

In AF contexts (wh-questions, focus clefts, or relative clause constructions), the same verb forms, inflected with the same suffix as the antipassive verb, occur. (2.21a) shows the *-o* suffix of root transitives; (2.21b) shows the *-n* suffix of derived transitives:

(2.21) Q'eqchi' (Dayley, 1981, 20)

      a. *Laa'at x-at-sak'-o-k*   *w-e*
         you   ASP-B2-hit-ATP-M A1-to
         'It was you who hit me.'

      b. *li   winq li   x-kamsi-n r-e*
         the man who T-kill-ATP A3-to
         'the man who killed him'

The object of the relative clauses in (2.21a) and (2.21b) occurs in a relational noun phrase. Relational nouns function like prepositions, but are formally nouns and are usually followed by their head noun and possessed by it (with the set A series, which function as possessors). The oblique realization of the object indicates that the antipassive inflection preserves its function in these AF contexts in Q'eqchi': the relative clause verb is intransitive. This holds both for pronouns, and for lexical NPs, as shown in (2.22):

(2.22) Q'eqchi' (Berinstein, 1985, 150)

    Li   c'anti' x-lop-o-c   *r-e   li   cuink*
    the snake   REC.PAST-bite-ATP-M A3-to DET man
    'It was the snake that bit the man.'

Tzutujil, Kaqchikel and K'iche' present a similar situation, though in these cases the suffixes used on AF verbs are only partially identical to those used in antipassive constructions. Specifically, while antipassives and AF verbs are formally identical for derived transitive verbs (-*Vn* for K'iche', *-n* for Kaqchikel, and *-Vn* or *-n* for Tzutujil), for root transitives they are distinct. In K'iche', for root transitives, the suffix found in AF contexts is *-ow*, while for antipassives it is *-oon*. For Kaqchikel the AF suffix is *-o* in these contexts, but *-on* for the antipassive. For Tzutujil it is *-o(w)* for AF and *-oon* for the antipassive.

(2.23) Tzutujil (Dayley, 1981, 25–26)

a. *X-in-aa-ch'ey*
   T-B1-A2-hit
   'You hit me.'

b. *X-at-ch'ey-oon-i*
   T-B2-hit-ATP-M
   'You were hitting.'

c. *Jar iixoq   x-ø-ch'ey-ow-i jar aachi*
   the woman T-B3-hit-AF-M the man
   'The woman was the one that hit the man.'

Note that unlike Q'eqchi', in Tzutujil, the NP object of the relative clause is realized as a direct argument, not an oblique.

Quichean languages display an interesting pattern of person marking in these constructions, which is only detectable in focus clefts where first and second persons can be focused. In these contexts, the AF verb form is inflected for a set B person marker, which agrees with the focused element:

(2.24) Q'eqchi' (Berinstein, 1985, 152)

a. *T-at-cu-a'bi*
   FUT-B2-A1-listen
   'I will listen to you.'

b. *Lain t-in-a'bi-n-k   acu-e*
   I   FUT-B1-listen-AF-FUT.ITR A2-DAT
   'I'm the one who will listen to you.'

Mamean

In Mamean languages, A-extraction involves the use of a verb form inflected with *-n* (∼ *-Vn*). This is formally identical to the suffix used to derive antipassives in the languages, as the following examples from Mam demonstrate:

(2.25) Mam (Dayley, 1981, 29–30)

a. *Ma ø-ky-tx'aj*
   T   B3-A3p-wash
   'They washed it.'

b. *Ma cyi-txj-on   mes*
   T   B3p-wash-ATP   table
   'They washed tables.' [Incorporating antipassive]

c. *Na'yan e ∅-kub' b'y-on   t-e   n-man*
   I   T B3-D   hit-ATP A3-to   A1-father
   'I was the one who hit my father..' [AF verb]

(2.25c) shows that the object of the relative clause verb, when it occurs as a lexical noun, is realized as an oblique relational noun, as in Q'eqchi' in (2.22). However, the Mamean languages pattern like Tzutujil in that object pronouns, realized solely by the set B person markers, may be realized directly on the verb, as the following example from Ixil shows:

(2.26) Ixil (Dayley, 1981, 34)

In   kat q'os-on axh
B1   T   hit-ATP B2
'I was the one who hit you.'

2.2.2 Western Mayan languages

Kanjobalan

The Kanjobalan languages differ from the eastern Mayan languages described above in that the verbal suffix found in A-extraction contexts (*-n*) is distinct from that which is used to derive antipassives (*-w*) in these languages. Thus, the two suffixes, whose reflexes in the greater Quichean languages are associated with different classes of transitive verbs, with both occurring in antipassive and A-extraction contexts alike, in Kanjobalan languages exhibit a distributional split, with *-w(a)* restricted to antipassive contexts, and *-n* to A-extraction contexts. The *-n* inflected agent focus verb, like the antipassive verbs we have already seen in eastern Mayan, bears no set A person marking.

The following examples illustrate this for Jakaltek (Poptí). (2.27a) gives a main transitive clause; (2.27b) an A-focus cleft, with the *-n* inflected relative clause verb.

(2.27) Jakaltek (Craig, 1977, 11)

a. *X-s-lok naj pel no' cheh c'ej-iñ*
ASP-A3-buy NC Peter NC*animal* horse black
'Peter bought the black horse.'

b. *(Ha') naj pel x-lok-**n**-i no' cheh c'ej-iñ*
DEM NC*male* peter ASP-buy-n-i NC*animal* horse black
'It's Peter who bought the black horse.'

(2.28) shows the contrasting antipassive verb form, with *-wa*:

(2.28) *Ch-ach-il-wa-yi*
T-B2-see-ATP-M
'You watch.'

While the inflectional morphology associated with AF in Kanjobalan is distinct from that of the antipassive, it does occur productively in other areas of the grammar, chiefly in (historically) nominalized or subordinated non-finite environments.

In Akatek, for example, *-on* is found in embedded clauses that follow three types of higher predicates: following perception verbs (2.29a), following some types of adverbial predicates (2.29b), and following some grammaticalized auxiliaries.

(2.29) Akatek (Zavala Maldonado, 1997, 446)

a. *Sa'al-ø ø-ey-i-on-aa-tej*
good-B3 B3-A2p-carry-NOM-DIR:up-DIR:hither A2p-salt
'It's good that you pick up your salt' (Lit., 'It's good, your picking up your salt')

b. *X-ø-y-il ix in-aw-ante-on an*
COM-B3-A3-see she B1-A2-cure-NOM CL1s
'She saw that you cured me' (Lit., 'She saw it, your curing me')

Zavala Maldonado (1997) observes that historically, the embedded clause in these cases is either the grammatical subject or object of the higher predicate, and thus (historically) nominalized. The set A marking on these forms is thus possessor marking (set A markers function as possessors, as well as subjects of transitive verbs).

Similar uses of *-n* are documented for Kanjobal, Chuj and Jakaltek (Craig, 1977). In Jacaltek, we also find *-n* marking temporal sequence in conjoined sentences or in the context of other sentences:

(2.30) Jakalteko (Dayley, 1981, 41)

> *X-ø-ul*   *naj*   *xqu-in*   *y-w-il-**ni***   *naj*
> PST-B3-arrive   NC*he*   PST-B1   A3-see-suff   NC*he*
> 'He arrived, then he saw me'

Greater Tzeltalan

Tzotzil is the only language in the greater Tzeltalan branch that has been documented as possessing an AF alternation. As in the case of Kanjobalan languages, the verb form used in AF contexts is inflected for *-on* (2.31b). This is formally distinct from the antipassive suffix *-wan* (2.31c):

(2.31) Tzotzil (Dayley, 1981)

> a. *I-ø-s-maj*   *Maryan*   *li*   *nlo'*   *e*
>   T-B3-A3-hit   Mariano   the   Chamulan   dem
>   'The Chamulan hit Mariano'

> b. *Ulo'*   *i-ø-maj-on*   *li*   *Maryan*
>   Chamulan   T-B3-hit-ATP   the   Mariano
>   'The Chamulan was the one who hit Mariano'

> c. *N-i-mil-wan*
>   T-B1-kill-AP
>   'I killed' (Tzotzil)

As in the Kanjobalan languages, the relative clause object in Tzotzil AF constructions remains a direct argument of the verb. I am unaware of instances of the AF morphology in Tzotzil occurring outside of AF contexts.

2.2.3 Summary

The above survey reveals a wide range of verb types and patterns of behavior in AF across the family. The table below summarizes the morphological form of the AF verb for each language, and notes the contexts in which this morphology is productively found in other

areas of the grammar. The question mark signifies that I have been unable to find descriptions of AF suffixes for those languages which refer to their appearing in other environments outside of A-extractions. "Root trans" in parentheses indicates that the AF suffix is identical to that of the antipassive for root transitive verbs only.

Branch	Language	AF verb form	Wider Distribution
Yucatecan	Yucatec	verb stem + -ik/-eh	subordinations
Greater Tzeltalan	Tzotzil	-on	?
Greater Kanjobalan	Jakaltek	-n	embedded nominalizations
	Akatek	-on	embedded nominalizations
	Chuj	an	embedded nominalizations
Mamean	Ixil	-on/-n	antipassive
	Awakatek	-oon/-Vn	antipassive
	Mam	-(VV)n	antipassive
Greater Quichean	Sipakapense	-w	?
	Sakapultek	-Vw/-n	antipassive (root trans)
	Tzutujil	-o(w)/-Vn/-n	antipassive (root trans)
	Kaqchikel	-o/-n	antipassive (root trans)
	K'iche'	-ow/-Vn	antipassive (root trans)
	Poqomam	-w/-in	antipassive
	Poqomchi'	-w/-Vn	antipassive
	Q'eqchi'	-o/-n	antipassive

Table 2.1: Form and wider distribution of Mayan AF suffixes

2.2.4   Language-internal variability

Yucatec is not the only language in the family which exhibits variation between the synthetic and AF verb in A-extraction contexts. To date, variation has also been reported for K'iche, Tzotzil, Mam, Ixil, Poqomam, Poqomchi'. It remains to be seen whether, with more extensive documentation, pockets of variation will be detected for supposed obligatory-AF languages.

Language vary as to which of three construction types (wh-questions, clefts and relative clause constructions) exhibit variability. In K'iche', AF appears to be optional even in

questions. As illustrated in (2.32), the dependent transitive verb in subject extractions may be realized with the synthetic form:

(2.32) *Xacin   s-ø-u:-kuna-x   ri: ci:x*
　　　　who   CMP-B3-A3-cure-ACT.VO   the   sheep
　　　　'Who cured the sheep?'

Tzotzil has received the most thorough documentation of AF variation to date (Aissen, 1999, 2003). In Tzotzil, as in K'iche', the AF verb form alternates with the synthetic verb form across all three construction types.

(2.33) *Buchu'   i-s-mil   li   Xune?*
　　　　who   CP-A3-kill   the   Juan
　　　　'Who killed Juan?'

According to Aissen, across all three construction types, the choice between the AF and the synthetic structure is determined by the referential and discourse properties of the participants of the relative clause: the AF form is preferred, or in some cases required where the relative clause object outranks the relativized subject in terms of topicality or definiteness or animacy. Thus (2.34a) is realized with the AF form because the object is inanimate than the subject is animate. And (2.34) is realized with the synthetic form, because the subject is animate and the object is inanimate:

(2.34) Tzotzil (Aissen, 2003)

　　　　a. *K'usi   i-subtas-on   li   antzetike?*
　　　　　　what   CP-frighten-AF   the   women
　　　　　　'What frightened the women?'

　　　　b. *li   vinik   ta   x-chon   paxak-e*
　　　　　　the   man   CP   A3-sell   pineapple-ENC
　　　　　　'The man who's selling the pineapple'

I will return to the Tzotzil data, and Aissen's analysis in §2.3.2

　　　Table 2.2 summarizes the existence of variability of AF exhibited across the family. For a given language, each construction type is labeled as *Var* 'Variable', *Syn* 'synthetic' or *AF* 'Agent Focus. 'Variable' in this context signifies that variability has been documented

in descriptions of the language for that construction type; 'Synthetic' means that it has been explicitly observed in the literature that only the synthetic form is possible for that construction type; 'AF' signifies that it has been explicitly observed that only the 'AF' form is possible for that construction type. Languages in the family for which such descriptions do not exist are not included in the table.

Language	Wh-question	Focus-cleft	Relativization
Yucatec	Var	Var	Var
Lacandon	Syn	Syn	Syn
Itzaj	Syn	Syn	Syn
Mopan	Syn	Syn	Syn
Chorti	Syn	Syn	Syn
Cholti	Syn	Syn	Syn
Chontal	Syn	Syn	Syn
Chol	Syn	Syn	Syn
Tzotzil	Var	Var	Var
Tzeltal	Syn	Syn	Syn
Tojolabal	Syn	Syn	Syn
Motozintlec	Syn	Syn	Syn
Jakalteko	AF	AF	AF
Akatek	AF	AF	AF
Ixil	AF	AF	Var
Mam	AF	Var	Var
Sipakapense	AF	AF	Syn
Sakapultek	AF	AF	AF
Tz'utujil	AF	AF	AF
K'iche'	Var	Var	Var
Poqomam	Var	Var	Var
Poqomchi	Var	Var	Var
Q'eqchi	AF	AF	AF

Table 2.2: Variability in Agent Focus across the Mayan family

2.3   Agent Focus in Mayan Linguistics

Considered from a cross-Mayan perspective, the phenomenon of AF represents an interesting descriptive and analytic challenge, being, distributionally, strikingly homogeneous, while formally displaying a remarkable degree of heterogeneity across the family. In this section I review the major approaches to AF in the literature.

2.3.1   Voice-based accounts

The most prevalent, and longstanding approach to AF has been to locate the alternation within the internal typology of voice alternations exhibited by these languages. This has been motivated by the obvious formal resemblance between the AF verb form and the antipassive in a number of eastern Mayan languages, as we saw in §2.2. While some early accounts treated the AF verb form specifically as an antipassive, others have argued that, in the bulk of languages at least, it represents not an antipassive, but instead a different kind of voice. Various descriptive labels have been offered for the voice type, including 'agent focus voice' (Tonhauser, 2003), 'actor focus voice' (Aissen, 1999) , 'focus antipassive' (Dayley, 1981), 'agentive voice' (Smith-Stark, 1978). The variation in terminology indicates, at the outset, the rather sui generis nature of the proposed voice, from a typological standpoint.

The explanandum for voice-based approaches (antipassive or otherwise) to AF is to account for why the voice should be required in the contexts in which it is found: subject extractions out of transitive clauses. In the following two sections I review two of the major approaches to this question. The first treats the AF voice as a reflex of syntactic ergativity; the second treats it as a special voice type with a particular discourse-pragmatic function.

Syntactic ergativity

A language can be said to exhibit ergative traits if it treats subjects of intransitives the same as objects of transitives, and differently from subjects of transitives. This treatment may be morphological in nature, or (more rarely), syntactic.[3] Mayan languages clearly exhibit

---

[3]Syntactic and morphological ergativity are logically independent of each other, though it is probably significant that all languages which have been argued to exhibit syntactic ergativity are also morphologically ergative (Dixon, 1979).

morphological ergativity as I have shown: S and O are cross-referenced on the verb with the same paradigm of person markers, while there is a separate paradigm for marking the A argument.

The treatment of S and O as a natural class may also be syntactic in nature. In Mayan languages, the restriction of AF to subjects of transitive relative clauses has been presented as evidence that AF is a reflex of underlying syntactic ergativity (Larsen and Norman, 1979; Larsen, 1981; Van Valin, 1981; England, 1983, 1988; Campana, 1992; Campbell, 2000). According to this view, the syntax of extraction in Mayan languages operates on an 'absolutive' pivot, such that only absolutive arguments (S and O) are privileged to undergo these operations. On early antipassive analyses of the AF verb, the function of the antipassive was understood to convert ergative NPs into absolutives, so that they could be questioned, relativized or clefted (Larsen and Norman, 1979; Dayley, 1981).

It was soon recognized however, that the antipassive analysis is not feasible for many Mayan languages. Even among those languages in which the suffix on the agent focus verb is identical (at least for certain morphological classes of verbs) to antipassive morphology, for a subset of these languages (e.g. Tzutujil and Ixil), the patient argument is not realized as an oblique in AF contexts, and thus exhibits a distinct behavior to that of antipassivization. The extension of an antipassive analysis of the AF verb is even more problematic for languages which possess a morphologically distinct antipassive verb (e.g. Kanjobalan languages and Yucatec).

The fact that in many languages the AF verb does not behave exactly like an antipassive (i.e., it does not demote the object), and, in some languages, co-exists with a true antipassive in the language, has resulted in the AF verb being ascribed the status of a special type of voice, e.g. a 'focus antipassive' (Dayley, 1981). On syntactic ergativity accounts, this voice is assumed to perform the function of licensing the extraction of subjects out of transitive clauses (Campbell, 2000), though here it is rather less clear what exactly the voice is doing, given that it does not involve any change in syntactic valency.

Despite its acceptance from many quarters, the connection between the AF alternation and syntactic ergativity is problematic in several respects. Why should wh-questions, clefts and relative clause constructions in particular be sensitive to underlyingly ergative syntax? The rest of the syntax in Mayan languages has an accusative alignment —syntactic operations such as control, binding and imperatives, for example, treat S and A arguments

alike. Without a theory of *why* ergative syntax should target extraction contexts in particular, this restriction remains stipulative. Indeed, the only evidence for ergative syntax in the language is (potentially) the behavior of the very constructions that the ergative syntax hypothesis is attempting to explain, with the result that syntactic ergativity, rather providing an explanation of the distribution, serves merely to describe it.

A second, and probably more insurmountable difficulty for syntactic ergativity based approaches is the fact that in many Mayan languages the AF verb is only optionally realized in these contexts. As I have shown, in a sizable number of languages, from four different branches of the family, the AF verb form may alternate with the synthetic verb form in A-extraction environments. Voice analyses framed in terms of syntactic ergativity cannot account for this sort of variability, unless one has a very elastic theory of ergative syntax.

Discourse-pragmatic approaches

Discourse-pragmatic approaches to voice (Cooreman, 1982; Rude, 1985; Thompson, 1989; Givón, 1990, 1994), view voice alternations as means of encoding different degrees of topicality of the two main participants of a semantically transitive event, agent and patient. A set of voice-based accounts of Mayan AF has sought to understand the AF verb in these terms, by charting the internal typology of voice contrasts in Mayan languages and their discourse-pragmatic functions, and locating the AF verb type within this typology. On these accounts, the AF voice is argued to represent a 'marked' voice of some type, in that it occurs in contexts where the agent is low in topicality, in contrast to the active transitive voice.

Zavala Maldonado (1997) situates his analysis of Akatek voices in terms of functional prototypes of the four major voice alternations found across languages: active-direct, inverse, passive and antipassive, which, following Cooreman (1987), are the following:

1. Active-Direct: both agent and patient are topical, but the agent is more topical than the patient

2. Inverse: Both agent and patient are topical, but the patient is more topical than the agent

3. Passive: the patient is topical and the patient is completely non-topical

4. Antipassive: The agent is topical and the patient is completely non-topical

Using quantitative text-based methods, Zavala provides a typology of Akatek voice constructions in which the active transitive voice, the passive and the antipassive (morphologically distinct from the AF verb in Akatek) all correspond to Cooreman's functional prototypes, according to various measures of topicality. Against this typology of voices he compares agent focus clefts. His analysis shows that in these constructions the agent is lower in topicality than active transitive clauses, though the patient remains at the level of topicality of the active transitive clause. He suggests on the basis of this that the agent focus construction bears similarities to the functional prototype of an inverse voice, where the patient is more topical than the agent.

It must be noted that Zavala's voice typology does not speak directly to the status of the AF verb itself, but rather to one of the constructions in which it occurs. It is the construction as a whole (the focussing of an agent) is to be understood serving a particular pragmatic function, that of encoding eventualities where the agent is low in topicality: "structure aside, the agent-in-focus clause in Akatek fulfills the discourse-pragmatic function of neither the passive nor the antipassive. If anything, it seems to be somewhat reminiscent, functionally, of an inverse..."(Zavala Maldonado, 1997, p.3).[4] Thus, this approach does not attempt to provide an account of the special morphology associated with focusing of agents in Akatek (on his analysis the AF suffix *-on* is simply, descriptively, an 'agent-in-focus marker'), or why this verb form is also found in two other constructions in Akatek and other Mayan languages (relative clause constructions and wh-questions).

The basic observation that focus constructions differ discourse-pragmatically from active transitive constructions forms a building block for Tonhauser's (2003) discourse-based analysis of the function of the AF verb. She connects the distribution of the Yucatec AF verb to contexts where the agent is low in topicality. Like Zavala for Akatek, she argues that voice in Yucatec is sensitive to the discourse status of the agent, and the normal active transitive voice requires the agent to be the current discourse topic. The AF verb form represents a distinct voice category which expresses a transitive predication where the agent is not the current discourse topic. Tonhauser argues that this explains the use of the verb not just in focus constructions, but also in wh-questions and relative clause constructions, because in all of these cases the agent is focused, and thus not the discourse topic. The Agent

---

[4]This may be contrasted with Aissen (1999) who treats the Agent Focus verb itself as an inverse verb.

Focus voice, on this analysis, is the more 'marked' transitive voice due to its restricted occurrence and because it marks eventualities as realizing a non-discourse topic agent.[5]

We can trace this type of approach to the AF verb at least as far back as Dayley (1981), who argues for Tzutujil that any voice category which is not the transitive active voice is a 'marked' category because it indicates the relationship of the arguments to a transitive predicate in a situation that deviates from the simple basic transitive relationship. In this context, the passive is marked in that it removes the agent from being in a direct relationship with the verb, the antipassive is marked in that it removes the patient. In AF contexts (in his terms, the 'focus antipassive voice'), the basic transitive relationship is not fundamentally altered, but the agent is brought into focus. It is thus marked compared to the active transitive voice, where agent and patient are more or less equally topical.

The variation between the AF verb and the synthetic verb which is attested in various Mayan language proves problematic for discourse-pragmatic accounts. Tonhauser's analysis is developed for Yucatec, which she assumed obligatorily requires the AF verb in A-extraction contexts. This may well be the case in the variety of Yucatec on which her analysis was based.[6] But the type of categorical account that Tonhauser or Dayley offer cannot extend to variable AF-languages, such as the variety of Yucatec studied in this dissertation, or any of the other Mayan languages for which variation has been documented.

For variable-AF languages, a categorical account like Tonhauser's which posits that the synthetic verb is restricted to environments where the agent is highly topical cannot predict its occurrence in 'low topicality' environments like agent focusing or questioning. Likewise, given that the non-discourse topic status of focused or questioned agents remains the same regardless of whether the AF verb or the synthetic verb is realized in these constructions, such an analysis cannot fully predict the occurrence of the AF verb. It can only offer the weaker claim that the AF verb, while not required, is licensed only in these conditions, and thus cannot occur elsewhere. But this stops short of providing any account of the factors driving the variation between the two verb forms.

Furthermore, while a discourse function-based account like Tonhauser's or Dayley's may provide an adequate characterization of the synchronic licensing conditions for AF in a

---

[5]If Tonhauser's use of 'marked' also implies morphological markedness, then this doesn't quite capture the Yucatec case, at least in imperfective, where the AF verb form in fact involves *less* morphology than the synthetic verb form.

[6]No dialect studies have been undertaken of Yucatec, to determine how widespread the existence variability is in this language.

non-variable language, neither address the connection between the formal properties of the AF verb, and the conditions in which it occurs. Why should the AF verb be characterized by an absence of subject person marking? How (if at all?) does this relate to low topicality? The only connection between form and function relates to the notion of 'markedness' - the AF verb, on these analyses, is a morphologically marked verb form, which occurs in 'marked' transitive environments. Thus, the absence of person marking, being essentially nothing more than a diacritic for morphological markedness, remains an arbitrary fact. I will return to this point in more detail below.

Some general problems for voice-based approaches to AF

Regardless of what one postulates as the conditioning factors governing its distribution, the formal classification of Mayan AF verb as a type of voice is beset by a number of difficulties. In many Mayan languages, the AF verb does not fit any standard definition of voice. On the classical generative view voice alternations involve a rearrangement of the basic syntactic organization of a clause, by valency increasing or decreasing operations. On this definition, the AF verb in the majority of Mayan languages does not represent a voice, because it exhibits the same bivalency as its synthetic counterpart.

As Tonhauser (2003) points out, this can be clearly shown in a language like Yucatec, where the AF verbs in imperfective contexts inflect for the same status suffix *-ik*, which belongs to the transitive verb paradigm (cf. Table 1.3, §1.1.3).

It can also be seen in a range of languages by virtue of the fact that these verbs select for a direct NP object argument. This is the case in Jakaltek for example, where the NP *no' cheh c'ej-iñ* 'the black horse" in the example below is not realized as any sort of prepositional oblique. This is also straightforwardly the case for Yucatec, Tzotzil, Chuj, Akatek, Awakatek Ixil and K'iche'.

(2.35) Jakaltek (Craig, 1977, 11)

(Ha')	*naj*	*pel*	*x-lok-**n**-i*	*no'*	*cheh*	*c'ej-iñ*
DEM	NC$_{male}$	peter	ASP-buy-n-i	NC$_{animal}$	horse	black

'It's Peter who bought the black horse'

In such languages, the AF verb does not represent a voice category in the classical generative sense of the term.

Of course, not all definitions of voice are based on syntactic valency alternations. According to a more semantic or discourse-pragmatic oriented conception, as we have seen, voice alternations can affect the relative topicality of the participants (Givón, 1994), or the control (agentivity) and affectedness of the subject (Klaiman, 1988), and include, among others, inverse systems. But even then, voice alternations are *clause-bound*: they encode changes in the discourse salience of participants in finite clauses, which includes, most obviously, main clauses.

However, in most Mayan languages, the AF verb is *restricted* to subordinated environments: in Tzotzil, for example, the AF verb, (which is morphologically and behaviorally distinct from the antipassive) never occurs in main finite clauses. Indeed, I believe this holds for every language with the exception of Q'eqchi' (see below). Categorizing the AF verb form as a voice type in most Mayan languages thus forces the concomitant assumption that voice categories can be restricted to dependent clauses in A-bar dependencies. As a voice, this would be typologically unique.

Even if one is willing to ascribe a sui generis voice category to the AF verb, inexplicably restricted to subordinate environments, there is the additional problem that for some languages there is not any obvious unique morphological exponent of such specialized agent focus voice category. In a language like Akatek or Jakaltek, for example the morphology found on agent focus verbs has a nominalizing function in other areas of the grammar. Similarly, in Yucatec, the verb forms that occur in A-extraction contexts occur in a range of other subordinate contexts. If we choose to analyze these forms in A-extraction contexts as representing a special voice, then we are either forced to treat these verb forms as accidentally homophonous with the identical subordinate/nominalized verb forms, or, alternatively to assume that in these contexts too, the verb forms represent the same agent focus voice, at which point the term 'voice' becomes so generally applicable as to be nothing more than a synonym for 'suffix'.

It appears to me that attempts to categorize AF as a voice category in Mayan languages have been influenced unduly by its obvious similarities with the antipassive. However, there is probably only one language in the family where a voice categorization of the verb form is justified, and this is because it is identical, both behaviorally and morphologically to the language's antipassive, and thus, by any measure, simply *is* an antipassive. As we saw in 2.2, Q'eqchi', AF verbs are inflected identically to antipassives in the language and, like

the antipassive they take oblique objects. I suggest (tentatively) that the complete formal and behavioral identity between the antipassive and AF in Q'eqchi' represents a recent innovation.

(2.36) Q'eqchi' (Berinstein, 1985, 150)

> | *Li* | *c'anti'* | *x-lop-o-c* | | *r-e* | *li* | *cuink* |
> |------|-----------|-------------|--|-------|------|---------|
> | the | snake | REC.PAST-bite-AP-M | | A3-to | DET | man |
>
> 'It was the snake that bit the man.'

In no other language in the family is the AF verb both morphologically and behaviorally identical to that language's antipassive. And in no other language, I submit, can the AF verb be categorized as a voice type, because it does not occur in main clauses, and thus does not form a paradigmatic relationship with other morphological voice categories in the language.

2.3.2 Disambiguation oriented accounts

In the second dominant set of approaches to AF in Mayan languages, the AF alternation is connected to the functional role of recoverability/disambiguation (Craig 1977 for Jakaltek, Mondloch 1978 for K'iche', Gutiérrez-Bravo and Monforte 2009 for Yucatec, Aissen 2003 for Tzotzil, Stiebels 2006 pan-Mayan). This is based on the observation that the extraction of one of the transitive verb's arguments from a postverbal position to a preverbal position leads to ambiguities in the case where the two arguments are third person: as we have seen, NPs themselves are not case marked in Mayan languages, and generally in a main clause with two postverbal NPs, it is word order that will distinguish the thematic role of each argument. However, an extracted third person argument leaves the string potentially ambiguous between a subject and an object extraction. Thus, in the Yucatec example in (2.37), where the relative clause verb is realized in the synthetic form, both a subject or an object relative reading is possible:

(2.37) | *le* | *chàan* | *xibpàal* | **_k-u=ts'uts'-ik_** | *le* | *x=ko'olel=o'* |
|------|---------|-----------|------------------|------|----------------|
| DEF | little | male.child | IMPF-A3=kiss-INC | DEF | FEM=woman=D2 |

'the little boy who is kissing the woman' or 'the little boy who the woman is kissing'

The extraction of an argument from a relative clause featuring an AF verb is, by contrast, unambiguous: it can only receive a subject relative reading:

(2.38) *le chàan xibpàal **ts'uts'-ik** le x=ko'olel=o'*
   DEF little male.child kiss-INC DEF FEM=woman=D2
   'the little boy who is kissing the woman'

The disambiguating function of the AF form has been interpreted in a variety of ways in the literature. One early account is Craig (1977) who argues for Jakalteko that the morphological alternations associated with AF, specifically the loss of subject person marking and the presence of the suffix *-n*, function jointly as a 'disambiguation mechanism'. This is obligatory in Jakalteko in order to avoid the ambiguity that would otherwise arise by extraction out of a main transitive clause.

 Mondloch (1978) suggests a similar function for the AF verb in K'iche' (which he classifies formally as an 'agent focus voice'). However, in contrast to Jakalteko, where AF is obligatory, AF in optional in K'iche'. Mondloch argues that speakers choose the AF verb form precisely in cases where ambiguities would otherwise arise. He suggests that typically, the semantic features of participants, combined with the lexical semantic features of verbs will resolve ambiguities in extraction contexts. Thus (2.39) is unambiguous even with the synthetic verb form, because typically only humans can be agents of curing events.

(2.39) K'iche' (Mondloch, 1978)

  Xacin s-ø-u:-kuna-x   ri: ci:x
   who CMP-B3-A3-cure-ACT.VO the sheep
   'Who cured the sheep?'

On this account, the primary function of the AF verb form is to disambiguate where context or semantic features will not. For example, (2.40) is ambiguous because both participants are human and either can be plausibly understood as the agent of the event.

(2.40) *Xacin s-ø-u:-kuna-x   ri: acih*
   who CMP-B3-A3-cure-ACT.VO the man
   'Who cured the man?' / 'Who did the man cure?'

The AF verb form is used to disambiguate in precisely these contexts:

(2.41) *Xacin s-ø-kuna-n   ri: acih*
    who   CMP-B3-cure-AF the man
    'Who cured the man?'

In the following two sections I review the two most spelled out disambiguation-based ac-
counts of AF. The first is Aissen's (2003) bidirectional OT analysis of Agent Focus in
Tzotzil, in which AF is taken to be an instance of differential coding, required by the gram-
mar to disambiguate in contexts in which the intended interpretation is not the predictable
one. The second is Stiebels' (2006) pan-Mayan Lexical Decomposition Grammar analy-
sis of AF, which aims to account for variation in AF across the Mayan family in terms of
grammaticalized disambiguation strategies.

Harmonic Alignment and Differential Coding in Tzotzil

In Tzotzil, as in K'iche', as described above, and in many other Mayan languages, the
AF may alternate with the synthetic verb in extraction contexts. (Aissen, 1999, 2003)
argues that in Tzotzil, the AF form is preferred, or in some cases required where the object
outranks the subject in terms of topicality or definiteness or animacy. Thus, in the examples
in (2.42) the AF verb form is used because in both cases the object outranks the subject, in
topicality in (2.42a), and in animacy in (2.42b):

(2.42)   a. *Pero buch'u i-mil-on?*
         but   who   CP-kill-AF
         'But who killed her?'

       b. *K'usi i-subtas-on   li   antzetike?*
         what   CP-frighten-AF the   women
         'What frightened the women?'   (Aissen, 2003)

By contrast, the finite transitive verb form is used in (2.43) because in this instance the
animate subject outranks the inanimate object.

(2.43) *li   vinik [ta x-chon   paxak]-e*
     the man   CP A3-sell pineapple-ENC
     'the man who's selling the pineapple'   (Aissen, 2003)

Aissen (2003) treats the AF verb form as an instance of *differential coding*.[7] Differential coding systems mark subjects and/or objects which diverge from the prototype; those which are more prototypical are left unmarked. She observes that certain semantic and pragmatic properties are prototypical (statistically more frequent) for objects: they tend to be low in animacy, definiteness and topicality. Transitive subjects by contrast, typically fall at the high end of these scales. The AF verb is the more marked verb form which is used just in case the intended interpretation of the string is the non-canonical one, that is, where the subject is ranked lower than the object on the animacy/topicality/definiteness scale.

The functional motivation of differential coding is recoverability of grammatical function. Aissen argues that the use of AF is exactly determined by recoverability: overt marking is required exactly when its absence would lead to an unintended interpretation. The unmarked form is always preferred as long as it can be associated with the target interpretation. Thus, the AF verb serves a disambiguating function in non-prototypical situations, which is enforced by grammatical constraints.

She suggests however, that there is no general proscription against ambiguity in the Tzotzil grammar. In (2.44), both arguments are human and neither (she assumes) are topical. Because in such a case neither argument outranks the other, both the AF verb form and the synthetic verb form are permitted by the grammar.

(2.44) *Buchu' i-s-mil   li   Xune?*
      who   CMP-A3-kill the Juan
    'Who killed Juan?' and 'Who did Juan kill?'

It is interesting to note that this is in fact precisely the type of context which Mondloch suggests for K'iche' is most likely to induce the use of the AF verb form (cf (2.40) above).

---

[7]In earlier work she proposes an analysis of the AF phenomenon in Tzotzil in terms of abstract obviation (Aissen, 1999), influenced by Algonquian morphosyntax, where there is an overt morphological category of obviation which is only relevant to 3rd persons, and is only significant when there are multiple 3rd persons in a discourse segment. In these obviation systems, 3rd persons are ranked according to properties like animacy and topicality, and this ranking determines various aspects of morphosyntax. Unlike the later work, this work does not relate these phenomena directly to the problem of recoverability.

Grammaticalized disambiguation across the family

Stiebels (2006) extends the disambiguation accounts which have been proposed for various individual languages, to the entire Mayan family. She proposes that the marker found in A-extraction contexts emerged initially as a means of disambiguation, and was subsequently grammaticalized in various languages to environments where it was not needed to recover grammatical function, and thus, synchronically, does not serve any disambiguating role. For example, in argument extraction contexts, ambiguity will only arise in cases where both participants are third person singular (3:3 environments), because person markers only distinguish number and person, but no other features (such as gender). If the arguments differ in their person features, however, the person marking will guarantee a unique reading with respect to the extracted argument. Thus, in Yucatec, (2.45) is unambiguous with the synthetic form, because the object is second person:

(2.45) *le   chàan xibpàal   **t-u=ts'uts'-ah-ex***
     DEF   little   male.child   PRV-A3-kiss-CMP-B2
    'the little boy who is kissing you'

Across the family, languages differ as to whether AF occurs in environments involving first or second person objects. In Tzotzil, for example, the AF form only occurs in 3:3 environments, while in Ixil, AF occurs also where the object is a first or second person pronoun. Stiebel's suggestion is that in cases like Ixil, the use of the AF form has been grammaticalized to non-ambiguous environments, where it was not functionally required:

(2.46) Ixil (Dayley, 1981, 34)

    in   kat q'os-on axh
    B1   T   hit-AP   B2
    'I was the one who hit you.'

Against disambiguation

While it is descriptively true that the AF verb in Mayan languages is unambiguous, whereas the use of the synthetic verb form can lead to ambiguities between subject and object extraction, this must be distinguished from the proposition that the function of the AF verb form is to disambiguate, that is, that the special morphology associated with A-extraction

should be classified as some kind of disambiguation marker, and that its very historical genesis owes to this functional motivation. I would like to argue for a number of reasons that this is not an appropriate characterization of Mayan AF.

The first problem is that teleological statements of the type "structure X emerged in order to do Y" or "structure X emerged in order to avoid Z" do not directly explain the emergence of linguistic structures. Haspelmath (1999, 188) offers the following biological analogy: "Antifreeze proteins are surely useful for polar fish, indeed necessary for their survival, but this does not suffice as an explanation for their presence". In both biology and language, useful or needed things are not sufficiently explained by their usefulness or their need for them. As Haider (1998:98) remarks "the fact that the design is good for a function is not the driving force that led to the design".

Stiebel's teleological statement 'It is proposed that the agent focus marker emerged as a means of disambiguation' (2006, 501) could perhaps be converted into a causal explanation. We could posit that at some earlier time, there was structural variation between the AF verb and the synthetic verb in A-extraction contexts. Use of the AF verb avoided ambiguity and thus speakers chose it increasingly often, such that in some languages it became obligatory in these contexts.

This proposal is still not adequate, however. First, it doesn't in itself shed any light on the existence of AF verbs in extraction contexts to begin with, or their formal properties. Second, it is not clear that ambiguity avoidance would be the driving force governing the choice between verb forms to begin with. While it is true that some theories have posited that ambiguity avoidance constrains the shape of grammars (Hankamer, 1973; Frazier, 1985), recent research on language production has shown that production decisions are not necessarily directly influenced by concern for ambiguity avoidance – even in cases where it might have a large effect on the success of communication. It has been shown in production studies of English that speakers don't consistently avoid syntactic ambiguity with word order (Arnold et al., 2004) , or with optional 'that' complementizers (Ferreira and Dell (2000). Given that grammaticalization of structures is taken to arise through frequency of usage (Hopper and Traugott, 1993; Bybee et al., 1994), and given that the production data, for English at least, indicates that speakers do not generally employ disambiguation devices when they could, it is unclear that ambiguity avoidance would account for the spread and conventionalization of the AF structure.

Finally, even if, for argument's sake, we were comfortable with postulating disambiguation as a functional motivation for the emergence and grammaticalization of AF verbs, it is logically possible that either a marker of agent extraction or of patient extraction could have arisen historically, as a means of disambiguating between the two. Why then did marking of *agent* extraction arise? Stiebels' answer is that objects are more natural foci than subjects, which in turn often surface as topics. In a similar vein to Aissen's differential coding account for Tzotzil, Stiebels argues that the various structures for non focus, agent focus and object focus in Mayan languages relate meaning and form in an iconic way: the less marked interpretation is assigned to a less marked morphosyntactic structure.

Stiebels' appeal to iconicity is problematic, however, because both the types of markedness being proposed are dubious. First, with regard to interpretative markedness, while it might be the case that objects occur more frequently than agents as the focused argument in focus constructions as Stiebels claims (though this would need to be corroborated by corpus data), it is less evident that object relativization or object wh-questions are more frequent than subject relativization or subject wh-questions. While again, this would need to be corroborated by Mayan corpus data, there is at least robust cross-linguistic evidence that subject relativization is both more frequent, both within and across languages, than object relativization (Keenan and Comrie, 1977; Hawkins, 1999, 2004). On Stiebels account, this would predict that disambiguating marking should have emerged on object extracted relative clause verbs, if anywhere. Second, with regard to morphosyntactic markedness, it is unclear what 'markedness' means in this context. If it is about additional morphological form, then the Yucatec case is problematic, for AF in Yucatec involves *less* form (the loss of subject person marking only).

Aissen's differential coding account for Tzotzil is no more explanatory in this respect: it remains arbitrary on her account why subject relatives where the object outranks the subject should be formally marked. Why not instead mark object relatives where the object outranks the subject? This, according to Aissen, is equally non-prototypical, and would equally well serve to disambiguate such cases from subject relative interpretations.

Of course, it is possible that it *is* simply arbitrary that the markedness should arise in this direction. There is, after all, plenty of arbitrariness in grammars. Yet it then remains rather perplexing that across the entire language family, we find AF verbs systematically occurring in subject extraction environments. This is despite the fact that the morphology

associated with the AF verb is not clearly historically related across all members of the family. Thus, if certain branches of Mayan languages independently of each other grammaticalized special morphology for the functional purpose of disambiguation, it is striking that they should all have converged upon precisely the same contexts in which to do so.

2.3.3 Some remarks on form and function

A common, if often implicit component of many accounts of Mayan AF is that the actual morphology associated with the AF verb form is arbitrary. For disambiguation accounts, for example, the morphology is a marker of disambiguation, a 'morphosyntactic quirk' (Gutiérrez-Bravo and Monforte, 2009), or simply the formal manifestation of markedness (Aissen, 2003). For voice accounts similarly, on many accounts, the verb morphology simply reflects a category of voice which is more 'marked' by contrast with the active transitive voice (e.g. Tonhauser, 2003; Mondloch, 1978).

Especially for those that are interested in developing accounts of the phenomenon that will reach beyond individual languages to the family at large, this is a hand that is somewhat forced on analysts, given the seeming mismatch between the formal heterogeneity of the phenomenon across languages, combined with the distributional consistency. Pan-Mayan approaches thus have tended to focus on explaining the distributional patterns at the expense of the formal realization of AF, which, in cross-linguistic perspective, does appear to be somewhat arbitrary.

However, as we have seen there is one consistent property of the verb forms that show up in AF contexts: in all cases, the verb form that occurs in A-extraction contexts does not bear set A person marking. The antipassive verbs of Q'eqchi' (2.47), the near-antipassives of Tzutujil (2.48), the -n suffixed forms of Jakaltek (2.49), and subjunctive and incompletive status suffixed verbs of Yucatec (2.50) all exhibit a notable absence of set A person marking.

(2.47) Q'eqchi' (Dayley, 1981, 20)

> *li   winq   li   **x-kamsi-n**   r-e...*
> the   man   who   T-kill-ATP-   A3-to
> 'The man who killed him...'

(2.48) Tzutujil (Dayley, 1981, 25-26)

>   *Jar iixoq   **x-ch'ey-ow-i** jar aachi*
>   the woman T-hit-ATP-M the man
>   'The woman was the one that hit the man...'

(2.49) Jakaltek (Craig, 1977, 11)

>   *(Ha') naj   pel   **x-lok-n-i**   no'   cheh   c'ej-iň*
>   DEM NC*male* peter ASP-buy-n-i NC*animal* horse black
>   'It's Peter who bought the black horse.'

(2.50) *Máax t-aw=a'l-ah   **xok** le   libro=o'?*
>   who   PRV-A2=say-CMP read DET book=D2
>   'Who did you say read the book?'

The question we want to ask then is why do the verbal alternations in these contexts systematically involve the presence/absence of subject person marking? This is the question I turn to in the following chapter.

Chapter 3

RP/gap alternations in Yucatec Maya

The objective of this chapter is to demonstrate that the Mayan AF alternation, rather than representing some special language particular category, is an instance of a more widespread cross-linguistic phenomenon: it is a type of resumptive pronoun/gap alternation. This classification not only solves the riddles surrounding its formal properties (presence/absence of person marking) and its distribution (its restriction to A-extraction contexts); it also provides the basis for typological comparison (chapter 4), which in turn will help to unmask its historical origins (chapter 5).

3.1   RP/gap alternations

Across languages, we can identify at least three different types of morphosyntactic mechanisms by which the syntactic-semantic role of the head noun in a relative clause (NP$_{rel}$) is expressed. The first, which I will have little more to say about in the remainder of this chapter, involves the use of a relative pronoun. In (3.1), for instance, the NP$_{rel}$ in English subject relative clauses may occur as the relative pronoun *who*:

(3.1) The knight **who** stormed the castle

In the second, NP$_{rel}$ appears as a resumptive pronoun (RP). This is an especially common strategy for embedded postnominal relative clauses (de Vries, 2001). Irish and Hebrew are both languages which exhibit pronominal resumption:

(3.2) Irish (McCloskey, 2007, 2)

> *an ghirseach ar   ghoid na síogaí í*
> the girl   COMP stole the fairies her
> 'The girl who the fairies stole'

(3.3) Hebrew (Borer, 1984, 256)

> *Raʔiti ʔet ha-yeled she   rina ʔohevet ʔoto*
> saw.1S ACC the-boy COMP Rina loves   him
> 'I saw the boy that Rina loves.'

The phenomenon of resumption extends (at least in certain languages) to the entire range of unbounded dependency constructions, including constituent questions (3.4a) and clefts (3.4b), as the following examples from Irish show:

(3.4) Irish (McCloskey, 2007, 3)

    a. *Céacu   fear ar   labhair tú   leis*
       which.of.two man COMP spoke   you with.him
       'Which man did you talk to?'

    b. *Tigh beag caol   gur   mhaireamar ann.*
       house little narrow COMP we.lived   in.it
       'It was a narrow little house that we lived in.'

The third mechanism by which NP_{rel} may be relativized is by not using any morphosyntactic element at all, i.e. by means of a gap. This is exemplified in (3.5) and (3.6) for English and Hebrew:

(3.5) This is the knight$_i$ that ___$_i$ stormed the castle

(3.6) Hebrew (Borer, 1984, 244)

> *ha-ʔarie$_i$ she-   ___$_i$ taraf   ʔet ha-yeled barax*
> the-lion   COMP   devoured ACC the-boy escaped
> 'The lion that devoured the boy escaped.'

3.1.1 The morpho-phonological realization of resumptive pronouns

McCloskey (2002) identifies an important morphological property of resumptive pronouns: resumptive pronouns are the ordinary pronouns of the language. He observes:

> A remarkable but little commented on property of resumptive pronouns is that they simply *are* pronouns [emphasis in original]. McCloskey (2002:192)

Irish may serve to illustrate this point. Compare the resumptive pronouns in (3.7a) and (3.8a) with the ordinary pronoun examples in (3.7b) and (3.8b). The Irish resumptive pronouns are simply the pronominal forms that would occur in the same positions outside of long distance dependency environments:

(3.7) Irish (McCloskey, 1979, 6)

      a. *an fear ar dhúirt mé go dtiocfadh sé*
         the man COMP said I COMP would.come he
         the man that I said (he) would come

      b. *dúirt mé go dtiocfadh sé*
         said I COMP would.come he
         I said he would come

(3.8) Irish (McCloskey, 1979, 6)

      a. *an scríbhneoir a molann na mic léinn é*
         the writer COMP praise the students him
         the writer whom the students praise (him)

      b. *molann na mic léinn é*
         praise the students him
         The students praise him

An important corollary of this observation is that pronominal information contributed by any source can function resumptively (Asudeh, 2004, 114). As will be discussed in more depth below, across and often within languages, pronominal elements do not exhibit a uniform structure.[1] What universally characterizes a pronoun are its referential role and

---

[1] As argued across a wide variety of frameworks, among others typologically oriented functional syntax (Givón, 1976; Nichols, 1986; Siewierska, 2004), Lexical Functional Grammar (Bresnan and Mchombo, 1987; Andrews, 1990), optimality theoretic syntax (Grimshaw and Samek-Lodovici, 1998; Bresnan, 2001), Head Driven Phrase Structure Grammar (Miller and Sag, 1997) and GB syntax (Jelinek, 1984).

functions, not its phrase structure category. Cross-linguistically, the range of forms includes phonologically null structures, affixal structures on a head, clitics, weak pronouns, and full independent pronouns. Given that resumptive pronouns simply are the ordinary pronouns of the language, accordingly, pronominal elements of any form can function resumptively. Cross-linguistically, we find the whole range of possibilities. Resumptives may be fully independent pronouns, as in English:

(3.9) This is the knight that we were all wondering why he stormed the castle

They may be clitics, as in Bantu or Lebanese Arabic:

(3.10) Lebanese Arabic, (Aoun, 2000, 15)

> *l-kteeb   yalli 'tarayt-o   mbeerifi Daaʕ*
> the-book that   bought.1S-it yesterday is-lost.3SM
> 'The book that I bought yesterday is lost.'

Or they may be inflectional elements, for example, on verbs, as in Chicheŵa incorporated object pronouns, or on prepositions, as in Hebrew:

(3.11) Chicheŵa (Bresnan and Mchombo, 1987, 763)

> *munthu   a-méné ndí-n´-mú-yéndêra*
> person(1) 1=rel   I-PAST-OM(1)-visit
> 'the person that I visited'

(3.12) Hebrew (Borer, 1984, 257)

> *raʔit-i ʔet   ha-yeled she- ʔRina xashva   ʔalav*
> saw.I   ACC   the-boy   REL   Rina   thought about.him
> 'I saw the boy that Rina thought about'

By contrast with resumptive pronouns, characterizing the morphophonological form of gaps would on the face of it, seem to be a rather trivial exercise: they are simply defined negatively by the absence of any overt morphophonological form. However, I would like to submit that in a typologically well-defined subset of cases, the alternation between gaps and resumptives does not merely involve presence or absence of a pronoun. In head-marking languages, it may involve an alternation between distinct paradigms of verb forms. The Mayan Agent Focus alternation, I argue, is an instance of exactly this.

3.2   The AF alternation as a RP/gap alternation

The recognition that pronominal resumption can be formally instantiated in a variety of ways is crucial for understanding the AF alternation. In Mayan languages we find two distinct sets of verbal paradigms which may occur in the dependent clause of unbounded dependency constructions. The first set carries subject dependent person marking; the second set does not. I argue that subject (set A) person marking in Mayan languages is pronominal. The consequence of this is that the alternation between these two verb forms is a type of resumptive/gap alternation. Its distinguishing characteristic is that it targets the *verbal* head of the relative clause. That is, gapping and resumptive strategies in Mayan languages relate to two verb forms.

I will present evidence in chapter 5 that diachronically, AF verbs are relics of older stages of the Mayan finite verbal system, which remained exempt from the grammaticalization of head marking. Due to subsequent shifts in aspectual marking in main clauses, these verbs are now morphologically differentiated from their pronominally inflected counterparts in additional ways. In this chapter I focus on their synchronic properties.

Synchronically, they can be typed as instances of RP/gap alternations. In some languages, only one of the alternants is available in long distance dependencies. For example, the Cholean branch of the family has lost AF verbs entirely, such that all subject extracted relatives are concluded with resumptive verb forms. In Tz'utujil and Q'eqchi', by contrast, AF verbs are obligatory in subject extracted dependencies. In a third set of languages, both verb forms are possible in these contexts. Inspection of the patterns of distribution between the two verb forms in this latter type of language is revealing: we find that these verbal alternations exhibit the same distributional tendencies that we find typologically attested for independent RP/gap alternations. Yucatec is a variable language of this type. In the remainder of this chapter I focus on the AF alternation in Yucatec, and show how it should be classified as a type of RP/gap alternation.

3.2.1   Dependent person marking

It is generally accepted that clitic and affixal forms of person marking evolve from independent person markers. This is not always evident synchronically within a language, due to the fact that person affixes in a language can be of a considerable age, and the original

dependent pronouns from which they developed may no longer be present in the language. Person affixes may also undergo substantial subsequent grammaticalization so as to no longer be recognizably related to independent forms.

However, there are cases where the diachronic relation is detectable. Indeed, it is a relation whose existence was posited already by the 19th century, when grammarians such as Bopp (1816) and Müller (1875) working on the reconstruction of Proto Indo-European identified resemblances between the reconstructed forms of independent pronouns and of person/number agreement on verbs.[2] In living languages we also find clear examples of the relation. Siewierska (2004, 252) mentions Asumboa (Oceanic), Mao Naga (Tibeto-Burman) and Mupun (West Chadic) as cases where the dependent markers are phonologically identical to their independent counterparts. She notes however, that more commonly, dependent forms are not identical to, but are obviously derived from independent forms. An example of this type is Wambaya (non-Pama-Nyungan Australian language):

(3.13) Wambaya (Nordlinger, 1998, 86)

	Independent	S/A
1SG	ngawurniji	ngi-
2SG	nyamirnji	nyi-
1PL INCL	ngurruwani	ngurru-
1PL EXCL	ngirriyani	ngirri-
2PL	girriyani	girri-
3PL	irriyani	irri-
1DU INCL	mirndiyani	mirndi-
1DU EXCL	ngurluwani	ngurlu-
2DU	gurluwani	gurlu-
3DU	wurluwani	wurlu-

Other examples where the diachronic relation between independent and dependent person markers is transparent include the Nilotic language So (Carlin, 1993, 79), Bantu languages such as Swahili (Givón, 1976) or Chicheŵa (Bresnan and Mchombo, 1987), various Mongolian languages (Comrie, 1981), various Papuan languages (Donohue, 2002) and native

---

[2]These initial insights were built on in the work of twentieth century Indo-Europeanists such as Meillet (1912); Kuryłowicz (1964) and Watkins (1969). I owe these references to Fuß (2005).

American languages (Mithun, 1991). The number and genetic diversity of these cases has led to the consensus that typically, pronouns are the primary source of dependent person markers (cf. e.g. Greenberg, 1978; Givón, 1976; Lehmann, 1988, 2002; Hopper and Traugott, 1993; Corbett, 1995; Siewierska, 1999, 2004; Fuß, 2005).[3]

In some cases, the changes are still in process, or have taken place during the recorded history of a language. In non-standard French, for example, what were originally subjects clitics have developed into a new form of agreement marking (e.g. Lambrecht, 1981; Auger, 1993, 1994; Gerlach, 2002). Similar processes have also been observed in various Northern Italian and Rhaeto-Romance dialects (Rizzi, 1986b; Linder, 1987; Brandi and Cordin, 1989; Haiman, 1991; Haiman and Benincà, 1992; Poletto, 1995, 1997, 1999; Roberts and Roussou, 2003; Tortora, 2003) , and in some Germanic dialects in inversion contexts (e.g. Weiß (1998) for Bavarian).

The formal grammaticalization process is generally assumed to proceed as follows (cf. e.g. Lehmann 1988, 2002; Corbett 1995; Hopper and Traugott 1993; Siewierska 2004 and Fuß 2005). Phonological erosion results in a formerly independent pronoun becoming a phonologically weak clitic, which requires a lexical host to attach to (typically the finite verb or auxiliary). The pronominal element then continues to erode, and is reanalyzed as an obligatory part of the verbal inflection.[4]

(3.14) strong pronoun > weak pronoun > clitic > affix

The grammaticalization of form often proceeds hand-in-hand with semantic and functional changes. As formal erosion progresses, so too can the function of the person marker shift from that of a referential pronoun to that of a syntactic agreement marker. Semantically, there may be a reduction or loss of information about the referential identity of person markers, for example, number or gender features. As Siewierska notes "The endpoint of the historical evolution of agreement markers from anaphoric personal pronouns is the loss of referentiality on the part of the person marker and the obligatory presence of the nominal with which it agrees" (Siewierska, 1999, 225).

---

[3]There are other historical sources of dependent pronouns, deriving for example, from the reanalysis of tense/aspect inflection. See Siewierska (2004).

[4]The resulting agglutinative marker may subsequently reduce even further, and undergo fusion with other inflection markers because it eventually disappears, resulting in a return to the first stage of the grammaticalization process.

The different dimensions of grammaticalization (formal, semantic and functional) often occur in parallel, though they do not always coincide. In particular, clitics and affixes, though no longer morphologically independent, may continue to function as referential pronouns. It has long been recognized for a variety of languages that such elements may function as more than simply redundant markers of agreement (among others Boas 1911; Bloomfield 1933 and more recently Van Valin 1977 on Lakhota, Jelinek 1984 on Warlpiri (but see Simpson 1991), Bresnan and Mchombo 1987 on Chicheŵa, Baker 1991 on Mohawk and Mithun 2003 on Yup'ik and Navajo). As Bresnan (2001) argues, what defines a pronoun is not its formal structure, but its semantics (reference) and its morphology (it encodes features such as number and person).

The pronominal function of dependent person markers is made evident by the fact that they can occur in clauses without any other overt NP, such that a verb with its affixes forms a complete sentence, as the following example from Central Alaskan Yup'ik indicates:

(3.15) Central Alaskan Yup'ik (Mithun, 2003, 236)

> *Nayurtu-q*
> watch-3SG
> 'He watches'

The (im)possibility of a free pronoun or an NP occurring in the same clause as the marker also gives an indication as to the pronominal argument status of a verbal marker. Free pronouns are typically in complementary distribution with full NPs having the same function within the clause. Indeed, theories of pronominals often trade on the assumption that a pronoun will not be coreferential with an NP within the clause. This would be ensured, for example, by the Binding Theory, or by the LFG principle of 'functional uniqueness'. If a dependent person marker functions as a pronominal argument, the same reasoning should apply. Co-occurrence with an NP or independent pronoun within the clause should be impossible.

We can apply this heuristic in both directions — the pronoun/nominal may block the marker and the marker may block the pronoun/nominal. For example, in the Carib language Macushi, the third person subject marker *aa-* only occurs in the absence of an overt pronoun or NP subject (data are taken from Abbott 1991 and discussed in Siewierska 2004, 123):

(3.16) Macushi (Abbott, 1991, 84)

    a. *Aa-ko'mamî-'pî asakîne wei kaisarî*
       3-remain-PST   two   day up:to
       'He remained two days'

    b. *U-yonpa-kon   Joao ko'mamî-'pî miarî*
       1-relative-COLL John remain-PST there
       'Our relative John stayed there'

    c. **Joao aa-ko'mamî-'pî*
       'Joao   3-remain-PST'

    d. **Môîkîrî aa-ko'mamî-'pî*
       'He   3-remain-PST'

Cases such as these, where bound pronominal markers and external arguments are in complementary distribution are reasonably clear cut. More complicated are cases such as Gumawana, where the person marker on the verb appears in different environments, specifically both together with a full NP, as well as without:

(3.17) Gumawana (Olson, 1992, 326, 308)

    a. *I-situ   vada sinae-na*
       3SG-enter house inside-3SG(INAL)
       'He entered the inside of the house'

    b. *Kalitoni i-paisewa*
       Kalitoni 3SG-work
       'Kalitoni worked'

Some take a syntactic view of such cases, and suggest different functions for a single marker (it is a pronoun when there is no full NP argument, and an agreement marker otherwise). This preserves functional uniqueness or its equivalent, at the cost of giving particular markers a dual function. This is an approach taken by Bresnan (2001) to account for the apparently dual behavior of subject person markers in Chicheŵa, which can appear both with and without a subject NP. She contrasts this with the behavior of Chicheŵa's object marker, which can also appear with co-indexed nominals. In such cases the nominals are analyzed as dislocated topics, however, not clause internal arguments. Siewierska (2004),

building on Bresnan (2001), proposes a typology which is based on the notion of the possibility of dual roles for person markers. She distinguishes three types of person markers cross-linguistically:

a. *Syntactic*: cannot occur without an overt controller in the clause, as in the case of English *-s*.

b. *Ambiguous*: can occur in the presence of an overt nominal as agreement markers, or in the absence of such a controller, as in the case of Chicheŵa's subject person markers.

c. *Pronominal*: always function pronominally, even in the presence of an overt co-indexed NP, as in the case of Chicheŵa's object markers.

Others take a more morphological view, and suggest that if a form cannot be distinguished across environments, then it should be assigned to a single category. (e.g. Georgopoulos 1991 for Palauan). This preserves a single function for the marker, but means that the theory of pronominal binding must be made more complex, or loosened. Indeed, distinguishing between Siewierska's types b and c is only made necessary if one ascribes in the first place to a theory that universally prohibits the co-occurrence of elements within the same clause indexed to the same referent. Mithun (2003) argues that this is not a universal, rather, languages which possess promominal affixes can permit clause internal co-indexation. Comparing the situation in German and English with Yup'ik, a language with pronominal affixes, she writes that while in English and German, there is a restriction against the establishment of pronominal reference for core arguments by material within the clause, "Yup'ik simply lacks this restriction. In Yup'ik, reference can be established in all of the same places as in English and German, by extralinguistic context, inference, the speech act itself and linguistic context beyond the clause, and within the clause as well". (2003, 245).

In the following discussion of Yucatec's dependent person marking system, I will show that Yucatec's set A person markers exhibit pronominal characteristics, both in the absence of co-indexed nominal expressions within the clause, and in their presence. I shall remain agnostic as to what this ultimately means for the status of these nominals in Yucatec; this is a decision that must be made on theory internal grounds.

3.2.2 The Yucatec person marking system

As I have already shown, Yucatec, like all Mayan languages, possess two paradigms of bound person markers, set A and set B. The set A series references the actor of transitive verbs, possessors, and the sole actor of intransitive verbs in the imperfective. The set B series references the undergoer of a transitive verb, the sole actant in stative clauses and in intransitive clauses with completive or subjunctive status. Yucatec's full paradigm of dependent person markers is given in Table 3.1 (cf. Lehmann 1998; Bohnemeyer 2002; Verhoeven 2007):

	Set A		Set B	
	Sg	Pl	Sg	Pl
1	in(w)	k(...-o'n)	-en	-o'n
2	a(w)	a(w)...-e'x	-ech	-e'x
3	u(y)/y	u(y)...-o'b	-ø/-ih	-o'b

Table 3.1: Yucatec's dependent person markers

As can be seen from the above table, the set A markers trigger an onglide if they immediately precede a vowel-initial verb or noun.[5] The third person *u* may actually be dropped in front of the glide *y*, in which case the *y* itself carries the indexical information. The plural paradigm for the set A series is discontinuous; the plural marker suffixes to the verb/noun. Third person set B marking is typically zero. The alternative suffix *-ih* only attaches to intransitive verbs in clause final position.[6]

Set B markers are suffixal, and attach to nouns, adjectives, and verb stems, following all other elements of the verb. The Set A markers are sometimes described as prefixes (see e.g. Hanks 1990), sometimes as clitics (Lehmann, 1998; Verhoeven, 2007; Bohnemeyer, 2002); this is indicative of the fact that they exhibit certain mixed properties of both classes.

In many respects, they exhibit the behavior of enclitics. First, they don't attach to what

---

[5]These onglides are originally are the original set A pronouns which date back to proto Mayan in the case of first and second person (Kaufman, 1990, 71), and to classical Yucatec, in the case of third person *u(y)* (Smailus, 1989, 12).

[6]This suffix is more appropriately characterized as a portmanteau affix which carries information about person as well as aspect. It occurs exclusively with intransitive verbs, where it has been reanalyzed from an older completive status suffix (Bohnemeyer, 2002).

they syntactically belong to, but rather to material immediately preceding them.[7] Lehmann (1998) observes that the natural pronunciation of (3.18) is [miʔn wohlu k'à:bʔiʔ]

(3.18) *Ma'   inw ohel   u   k'àaba'-i'*
       NEG A1   know A3   name-D4
       'I don't know his name.'

This is also evident in their behavior when preceded by either of the preverbal aspect markers *t-* (perfective) or *k-* (imperfective). In such cases the set A marker forms a phonological word with the preceding aspect marker.[8] Second, in disfluent speech, pauses may intervene between the clitic cluster (aspect marker + person marker) and the verb stem.

    However, they do not exhibit all of the expected behavior of clitics. A common argument put forward in support of a clitic analysis of the set A marker is the fact that adverbials may intervene between the verb stem and the set A marker. (3.19) gives an example with *háan* 'quickly', (3.20) with *chen* 'only, just'.

(3.19) ***T-in-háan**-bin*   *yiknal   in-padim*
       PROG-A1-quickly-go   by   A1-godfather
       'I'm quickly going to my godfather's house.'   (Lois and Vapnarsky, 2003, 117)

(3.20) *Le'   t'èel-o' (...) ken   uy-u-('u)b a-t'àan-e'   k-u=**chen**-t'àan*
       DEF   hen-D2   PROS   A3-hear   A2-speak-D4 HAB-A3=just-speak
       'The rooster (...) when he hears you speaking, he just speaks' (Lois and Vapnarsky, 2003, 117)

However, it should be noted that this type of adverbial placement is not entirely productive: the adverbials that may occur in this position belong to a closed class of particles. Complex adverbials occur postverbally. Kaufman (2002) argues that this type of adverbial modification is in fact incorporation; it is a process which occurs in Ch'olan, Huastec and

---

    [7]Enclisis is not particular to the set A marker: it is an interesting feature of Yucatec quite generally that function words are often encliticized in speech.

    [8]Synchronically, it is not clear whether the aspectual suffix and the person marker should be treated as a portmanteau clitic cluster of both aspect and person or as person markers inflected for tense/aspect. Siewierska (2004) observes that such portmanteau forms combining tense/aspect and person are also common in Africa, e.g. Mande languages, also Chadic, and in Austronesia. Typically they are the result of fusion of a subject person marker and an auxiliary verb.

Yucatecan languages, and may be an old pattern in Mayan: across the family many languages, including Yucatec, have reflexes of **kaʔ* 'two' in this position with the meaning 'again'.

Miller (1992) offers a diagnostic for affix vs. clitic status based on coordination facts. He suggests that affixes cannot take wide scope over a coordination of hosts, and thus that coordination facts can help to determine the lexically-attached status of an element. In Yucatec (and across the family) dependent person markers are an *obligatory* part of the finite transitive verbal complex. They cannot, for example, take wide scope over coordinated verb phrases (3.21a). Instead, the preverbal marker must be realized in each conjunct (3.21b):

(3.21)   a. **T-u=pax-ah   gitaara yéetel pax-ah   marimba*
            PRV-A3-play-CMP guitar   and   play-CMP marimba
            'He played the guitar and played the marimba.'

         b. *T-u=pax-ah   gitaara yéetel t-u=pax-ah   marimba*
            PRV-A3-play-CMP guitar   and   PRV-A3-play-CMP marimba
            'He played the guitar and he played the marimba.'

In addition to the dependent pronominal markers, Yucatec has a set of independent personal pronouns (3.2.2). Historically, this set grew out of the contraction of the preposition *ti'* (locative) inflected for the set B markers.

	Sg	Pl
1	tèen	to'n
2	t'eech	te'x
3	(le-)ti'	(le-)ti'o'b

Table 3.2: Yucatec's independent person markers

Independent pronouns occur in emphatic contexts (topic or focus positions), or otherwise in indirect object/oblique function, or as the complement of a preposition. Lehmann (1998) distinguishes between two forms of the third person independent forms, a strong form, which includes the element *le-*, and which occurs in topic or focus position, or when governed by a preposition, and a weak form, *ti'*, which occurs postverbally, referencing indirect objects. (3.22) illustrates the strong form:

(3.22) *Le'eti-e   t-u=tàas-ah   le   pelota-o'*
     he-TOP   PRV-A3-bring-CMP   DET   ball-D2
     'He, he brought the ball.'

The pronominal status of the set A markers

That Yucatec's bound person markers *can* serve a pronominal function is a fairly uncontro-
versial proposal (Bohnemeyer, 2002; Lois and Vapnarsky, 2003; Verhoeven, 2007). Lexical
NPs are optional, and the verb, together with the bound person markers, may exhaust the
clause. In such cases the person markers function as definite pronouns which serve to
indexically represent a referent introduced in discourse or given in the speech situation:

(3.23) *t-u=tàas-ah   u   hàan-t   u   yùumil*
     PRV-A3=bring-CMP   A3   eat-APP   A3   owner
     'and he brought it for his master to eat'   (Verhoeven, 2007, 147)

The morphologically dependent person markers are the only elements which may serve
as pronominal arguments of verbs: as I have shown, while Yucatec does possess a set of
independent pronouns, these have a highly restricted function, occurring only in topic or
focus position, or following prepositions. Syntactically, these cannot function as direct
verbal arguments in Yucatec. Thus, the independent pronoun *le'eti* is ungrammatical when
in the postverbal subject position (3.24b):

(3.24)   a. *T-u=tàas-ah   le   pelota=o'*
         PRV-A3-bring-CMP   DET   ball=D2
         'He brought the ball'

       b. **T-u=tàas-ah   le   pelota   le'eti=o'*
          PRV-A3-bring-CMP   DET   ball   he=D2
          'He brought the ball'

The functional and referential restrictions of Yucatec's independent pronouns rule out a
'pro-drop' analysis of Yucatec's pronominal system (i.e., where the head markers are agree-
ment markers, agreeing with unexpressed independent pronouns), at least from a deriva-
tional perspective: the language does not possess a series of independent pronouns in the
lexical inventory that may be 'dropped'. One could, of course, technically maintain an

analysis in which the head markers agree with obligatorily covert pro (cf. Aissen 1987 for Tzotzil). However, instead of invoking entire paradigms of null pronouns, the more parsimonious approach, which I follow here, is to assume that Yucatec morphologically incorporates referential pronominal arguments into the verbs that subcategorize for them.

Whether Yucatec's person markers *always* function as morphologically incorporated pronouns is, however, a separate issue. It is standard to treat these markers as serving an agreement function in the presence of independent NPs (Bohnemeyer, 2002; Lehmann, 1998), i.e, they belong to group B (ambiguous person markers) in Siewierska's typology. In cases where postverbal independent NPs occur in the clause, the bound person markers are thus assumed to do nothing more than locally index the person and number information of the verb's nominal arguments.

I suggest that it is possible to detect the effects of the anaphoric properties of the set A person markers in the presence of co-indexed NPs within the sentence, and that this argues against these forms functioning as redundant agreement markers in such contexts.

The argumentation I will develop to support this view draws from the functional and psycholinguistic literature on pronoun resolution. Research in this area has demonstrated that various factors contribute to how the reference of a pronoun gets resolved in discourse. At the most general level, pronouns are preferentially interpreted as being co-referential with antecedents which are highly accessible (or salient) (Ariel, 1990; Arnold, 1998). Accessibility is determined both by inherent properties of the referent, as well as the nature and the direction of the dependency between the referent and the anaphor. Factors contributing to accessibility include, among others: the direction of the dependency relation (anaphoric rather than cataphoric dependencies) (Hawkins, 2004), the recency of mention of the referent (Clark and Sengul, 1979; Gernsbacher, 1990), the semantic plausibility of the referent (Garvey and Caramazza, 1974; McDonald and MacWhinney, 1990), competing referents (Ariel, 1990; Givón, 1983) and referent 'topicality' (Siewierska, 2004; Givón, 1983; Gernsbacher, 1990; Ariel, 1990).[9]

On the assumption that Yucatec's set A marker is pronominal, then we might expect to see a similar sensitivity to accessibility effects in the resolution of its reference. I suggest that this is indeed the case.

---

[9]Topicality itself is a complex notion, which may have to do with any or all of: subjecthood/first mention, protagonist identity, repeated mention, prior pronominal reference; see Arnold (1998) for review.

First, there is a tendency in Yucatec for word order to be arranged such that the set A marker stands in an anaphoric, rather than a cataphoric relationship with its antecedent, that is, where the antecedent occurs before the set A marker. The effects of this play out in both production and comprehension. Despite VOS often being described as Yucatec's canonical word order, the postverbal placement of the subject is subject to referential and discourse restrictions, especially in transitive contexts where there are two lexical NPs in the clause. Typically, the transitive agent is topicalized to a preverbal position. In such a structure, as in (3.25a) below, the topical element antecedes the co-referential set A marker:

(3.25) *Le   pèek'-o'   t-u=chi-ah   le   paal-o'*
      DEF   dog-D2   PRV-A3-bit-CMP   DEF   child-D2
      'The dog, it bit the child.'

Indeed, it has been suggested that if frequency alone is considered, the topicalized SVO structure should be regarded as the unmarked word order of the language (Gutiérrez-Bravo and Monforte, 2008).[10] The frequency of topicalized NP V structures follows on the assumption that the set A marker is a pronominal argument and is preferentially interpreted anaphorically.

The preference for anaphoric dependencies is also detectable in comprehension. In a comprehension experiment designed to test post-verbal argument order preferences in Yucatec Maya, Skopeteas and Verhoeven (2005) presented Yucatec Mayan subjects with complex sentences in which properties of definiteness and animacy of postverbal arguments was manipulated in the subordinate clause. An example stimulus is given in (3.26)

(3.26) *Roberto-e'   t-u=ya'l-ah   t-u=chi'-ah   x-ch'úupal pèek'*
      Roberto-TOP   PRV-A3=say-CMP   PRV-A3=bite-CMP   F-girl   dog
      'Roberto said that a dog bit a girl.'

They were interested in whether subjects would interpret the second post-verbal argument or the first postverbal argument in the complement clause as the subject (i.e, (a) a dog bit a girl, or (b) a girl bit a dog). Significantly, they found a high frequency of an interpretation they weren't directly testing for: a "one argument" interpretation, in which the two

---

[10]However, as Bohnemeyer (2009) argues, despite the fact that topicalized (preverbal) subject structures are more common than post verbal structures, the topic position can be shown to be external to the clause on the basis of the distribution of deictic clitics. Moreover, the topic position is not a dedicated subject position – other grammatical functions and adjuncts may also be topicalized.

post-verbal arguments were actually interpreted as one argument, either conjoined or compounded, *with the set A marker interpreted as co-referential with the agent of the matrix clause*. I.e, in (3.26) above, subjects frequently interpreted the string as 'Roberto$_i$ said that he$_i$ bit a dog-girl' or 'Roberto$_i$ said that he$_i$ bit a dog and a girl'. This behavior is not anticipated on the assumption that the set A marker indexes nothing but agreement features of the local arguments in the clause, but is immediately understandable if the set A marker is a pronominal element which is preferentially interpreted anaphorically, and thus obtains its reference from the immediate available antecedent in the higher clause.[11]

This of course, is not meant to imply that Verb NP NP constructions are not possible. However, it is significant, that in verb initial environments, where a cataphoric relation is established between the set A marker, and a post-verbal NP, the set A marker tends to be co-indexed with the most definite, topical NP in the clause. Indefinite NPs, which introduce new discourse referents, are dispreferred as post-verbal subjects.

(3.27) *??t-u-chi'-ah*   *le*   *pàal*   *hun-túul*   *x-chìiwol-o'*
      PRV-A3-mouth-CMP   DEF   child   one-CL.AN   FEM-tarantula-D2
      Intended: 'A tarantula bit the child.'   (Bohnemeyer, 2009, 192)

Instead, strings such as (3.27) in out-of-the-blue contexts typically receive an interpretation where the definite, topical NP is the subject, i.e, that the child bit a tarantula. To convey the intended message, that a tarantula bit a child, speakers will use other structures, such as passives or a left dislocated topic construction.

This tendency to avoid post-verbal transitive subjects on the basis of discourse-related and inherent properties of the participants has been understood in terms of the theory of harmonic alignment, which postulates that thematic roles (e.g. agent and patient) are canonically aligned with a saliency hierarchy (determined by e.g. animacy, person, definiteness, and, above all else, topicality (Aissen, 2003). These accounts attribute the choice between different structures to harmonic alignment. In Yucatec, VOS word order is argued to be restricted to canonical alignment, i.e, the patient must not outrank the agent in VOS structures (Bohnemeyer, 2009). Non-canonical alignment, where the patient outranks the agent, is assumed to trigger other structures, such as focus, topic or passive structures. Example (3.27) is therefore infelicitous because the patient argument, being definite, is more prominent than the agent, which is indefinite.

---

[11]A similar effect has also been reported by England for K'iche'.

Rather than appeal to an abstract notion of harmonic alignment, I suggest these restrictions on the positioning and interpretation of NPs can be attributed instead to the presence of the pronominal argument (the set A marker) procliticized to the verbal head. The postverbal NP word order patterns we find in Yucatec can be understood as arising from the interaction of the factors which influence the referential resolution of this pronoun: where there is a choice between competing antecedents (or rather, procedents), the reference of the set A pronoun will be resolved towards the more accessible (= topical) of the two.

The preference for co-reference between the set A marker and topical/definite external NPs is a tendency, rather than a hard constraint. And in certain contexts in particular, it is less relevant. For example, postverbal indefinite NPs co-indexed with the pronominal subject are more acceptable in cases where the object is a first or second person pronoun:

(3.28) *T-u=chi'-ah-en*   *hun-túul*   *pèek'*
      PRV-A3=bite-CMP-B1   one-CL.AN   dog
     'A dog bit me'

The absence of competing referents increases the accessibility of an antecedent (Ariel, 1990; Givón, 1983). In (3.28) there is one unique available referent for the subject pronoun, the single postverbal NP, because the first person pronoun can only refer to the object. The indefinite postverbal NP is thus more acceptable because there is no other competing referent for the subject pronoun.

In sum, the existence of referential and ordering restrictions on NPs aligned with the subject function in transitive clauses support a pronominal argument analysis of the set A subject marker. The preference for NPs co-indexed with the set A marker to be accessible/topical discourse participants point to the pronominal status of this element. On an agreement analysis, by contrast, these referential effects are unexpected.

3.2.3   Evidence in favor of the AF alternation as a RP/gap alternation

Previous analyses of long distance dependencies and AF in Yucatec either implicitly or explicitly (Gutiérrez-Bravo and Monforte, 2009) assume that in extraction contexts, the set A markers are agreement markers. This obviously has a direct bearing on how to characterize the termination of these long distance dependencies. On an agreement analysis of the set A marker, the variable position in the dependency is defined by a gap (the zero realization of

an argument), regardless of whether the subordinate verb is realized as the full verb form or the AF form, because in both cases, the foot of the dependency is occupied by a missing argument, and the set A marking, if present, is simply functioning to locally index the agreement features of the gapped argument. Thus both the AF and the synthetic versions in (3.29) are, on this analysis, characterized as gap structures:

(3.29) a. *T-in=kíin-s-ah*   *le*   *x-chìiwol$_i$*   *t-u=chi'-ah-en*
           PRV-A1=kill-CAUS-CMP DEF FEM-spider PRV-A3=bite-CMP-B1
           ____$_i$ *=o'*
                 =D2
           'I killed the tarantula that bit me.'

       b. *T-in=kíin-s-ah*   *le*   *x-chìiwol$_i$*   *chi'-en*   ____$_i$ *=o'*
           PRV-A1=kill-CAUS-CMP DEF FEM-spider bite-B1   =D2
           'I killed the tarantula that bit me.'

The pronominal analysis of the set A marker has a different consequence. On the pronominal analysis, the alternation between the AF and the full verb form results in distinct terminations of the dependency. The full verb bears a subject pronominal. Dependencies that feature the full dependent verb form are thus terminated with a resumptive pronoun:

(3.30) *T-in=kíin-s-ah*   *le*   *x-chìiwol$_i$*   *t-u$_i$=chi'-ah-en=o'*
        PRV-A1=kill-CAUS-CMP DEF FEM-spider PRV-A3-bite-CMP-B1=D2
        'I killed the tarantula that bit me.'

The AF verb form does not bear the set A subject pronominal. Thus, in extractions featuring the AF verb form, the dependency is concluded with a gapped argument:

(3.31) *T-in=kíin-s-ah*   *le*   *x-chìiwol$_i$*   ____$_i$=*chi'-en=o'*
        PRV-A1=kill-CAUS-CMP DEF FEM-spider   =bite-B1=D2
        'I killed the tarantula that bit me.'

As I have shown, there is good reason to assume that the pronominal analysis of the set A marker is correct. And as I show in the following sections, there is, independently, a good deal of evidence to support the view that the alternation between the two verb forms in Yucatec does indeed involve the alternation between the presence or absence of a resumptive pronoun.

3.2.4 The distribution of gaps and resumptive pronouns

When both resumptive pronouns and gaps are available in a language, they generally have overlapping, but often non-identical distribution with respect to syntactic positions/ grammatical functions. This distribution tends to conform to certain well-attested patterns. In this section I will discuss how these patterns of distribution are detectable in the Yucatec AF alternation. Many of these patterns of variation, and their sources, will be discussed again in greater detail in chapter 6. Here, the purpose is to use these cross-linguistically attested patterns as diagnostics for the RP/gap classification of the AF alternation.

The 'Highest Subject Restriction'

It has been reported for many languages, that resumptive pronouns are not permitted in the 'Highest Subject position', that is in a subject position immediately subjacent to their binder. This is the case for Hebrew. While (3.32a) is unacceptable with the resumptive, the resumptive is preferred over a gap structure in more embedded positions (3.32b).[12] Similar tendencies have been noted by McCloskey (1990) for Irish and Sells (1984) for Welsh and Swahili.

(3.32) Hebrew (Ariel, 1999, 223)

      a. Ha+makhela she+ ∅/ *hi hirshima oti beyoter ...
         'The chorus that (it) impressed me most ...'

      b. Ha+makhela she+ Maya shamaa she+ Iddo amar she+ ?∅ / hi kibla pras
         'The chorus that Maya heard that Iddo said that it got a prize'

The ungrammaticality of such local subject resumptives has often been understood as an aspect of the kind of antilocality property typical of pronouns (Borer, 1984; McCloskey, 1990; Aoun and Choueiri, 1996): it is generally assumed that the binding of pronouns is subject to the requirement that a pronoun and its antecedent cannot occupy the same local domain. In the domain of A-binding, this requirement is captured by Condition B of the Binding Theory. To capture the similar constraints on resumptives in highest subject position, it has been proposed that the disjointness requirement defined by Principle B of

---

[12]The orthography used in (31) follows that of the cited source.

the Binding Theory should be extended to the domain of A-bar binding. (McCloskey, 1990; Aoun and Li, 1989). This rules out Highest Subject resumptives, but allows them in embedded contexts.

A difficulty for a universal and categorical Principle B approach to the Hightest Subject Restriction (HSR) is, however, that not all languages which exhibit resumptive pronouns exhibit a categorical HSR effect. We have already seen this for Yucatec, which allows subject resumptives. Colloquial Spanish (Suñer, 1998) and Yiddish (Prince, 1990), are also languages which allow resumptives in the Highest Subject position. Keenan and Comrie (1977) also mention Aoban and Urhobo. What is generalizable cross-linguistically, however, is an implicational pattern with respect to embedding and resumptive pronoun distribution: if a language allows Highest Subject resumptives, it will also allow them in more embedded positions, but not the reverse. Thus, Irish bans Highest Subject resumptives, but allows subjects of embedded clauses to serve the resumptive function (McCloskey, 1990). But in Lebanese Arabic (Aoun and Choueiri, 1996) subject resumptives can occur neither in highest position, nor in more embedded positions.

In Yucatec we find similar patterns of distribution. Acceptability judgments with three Yucatec consultants reveal that in the case of relative clause constructions, the AF verb is marginally preferred over the resumptive verb form in Highest Subject extractions:[13]

(3.33)   a. *In   k'ahóol   le   máak **t-u=ts'uts'-ah**   x-maria=o'*
         A1   know   DEF   man   PRV-A3=kiss-CMP   FEM-maria=D2
         'I know the man who kissed Maria.'

         b. *In   k'ahóol   le   máak **ts'uts'**   x-maria=o'*
         A1   know   DEF   man   kiss   FEM-maria=D2
         'I know the man who kissed Maria.'

With one level of embedding, however, the resumptive form is preferred over the AF form:

(3.34)   a. *In   k'ahóol   le   máak t-aw=a'l-ah   **t-u=ts'uts'-ah**
         A1   know   DEF   man   PRV-A2=say-CMP   PRV-A3=kiss-CMP
         x-maria=o'*
         F-maria=D2
         'I know the man that you said kissed Maria.'

---

[13]The experiments described in chapter 6 provide statistically significant evidence for these patterns of distribution in production, in keeping with the impressionistic acceptability judgement data reported here.

    b. *In k'ahóol le   máak t-aw=a'l-ah   **ts'uts'** x-maria=o'*
      A1 know DEF man PERF-A2=say-CMP kiss   F-maria=D2
      'I know the man that you said kissed Maria.'

And in even more complex constructions, with a further level of embedding, the resumptive verb form is strongly preferred over the AF form:

(3.35) *In k'ahóol le   máak t-aw=a'l-ah   t-uy=a'l-ah   juan*
      A1 know DEF man PRV-A2-say-CMP PRV-A3-say-CMP juan
      t-u-ts'uts'-ah   *x-maria=o'*
      PRV-A3-kiss-CMP F-maria=D2
      ' I know the man that you said that Juan said kissed Maria.'

    b. *??In k'ahóol le   máak t-aw=a'l-ah   t-uy=a'l-ah   juan*
      A1   know DEF man PRV-A2-say-CMP PRV-A3-say-CMP juan
      ***ts'uts'** x-maria=o'*
      see   F-maria=D2
      'I know the man that you said that Juan said kissed Maria.'

There is virtually no data available on other Mayan languages (that I am aware of), that pertains to embedding effects on the choice of the AF verb. One exception is Craig (1977) who discusses embedding in Jakalteko. She observes that while the AF verb is obligatory in subject relative clauses in 'Highest Subject' contexts (3.36a), it is optional in embedded contexts, where it may alternate with the full verb form (3.36b):

(3.36) Jakalteko (Craig, 1977)

    a. *W-ohtaj   naj x-lok-n-oj   no' cheh*
      A1-know him ASP-buy-SUFF-FUT the horse
      'I know the man who will buy the horse'

    b. *W-ohtaj   naj x-y-al   ix   ta   x-s-lok-o'   no' cheh*
      A1-know him ASP-A3-say she that ASP-A3-buy-FUT the horse
      'I know the man that she said will buy the horse'

These distributional patterns support the view that the RP/gap analysis of the AF alternation is applicable to languages in the family beyond Yucatec.

Islands

Ever since Ross (1967) it has been recognized that there are constraints operable on languages which limit the region within the relative clause in which NP$_{rel}$ can appear. For example, Ross's Complex NP Constraint (CNPC) says that it is impossible to relativize an NP contained within an S that modifies another NP. The ungrammatical English example of (50) below is a CNPC violation:

(3.37) The people who John denied the claim that Mary had insulted got angry.

The constraints apply particularly to gaps. Resumptive pronouns are often acceptable in island environments, as the following examples from English, Welsh, and Egyptian Arabic show:

(3.38) The people who John denied the claim that Mary had insulted them got angry.

(3.39) Welsh (Keenan, 1985, 156)

...	*'r*	*het*	*y*	*gwn*	*y*	*dyn*	*a'*	*i*	*gadewodd*	*ar*	*y*	*ford*
...the	hat	the	I	know	the	man	that	it	left		on	the	table

'the hat that I know the man who left it on the table'

(3.40) Egyptian Arabic (Keenan, 1985, 156)

al-rajul	*allathi*	*hua*	*wa*	*ibna-hu*	*thahabu*	*ille*	*New*	*York*
the-man	who	he	and	son-his	went	to	New	York

'the man who he and his son went to New York'

In the GB tradition such constraints have been formulated in terms of subjacency (Chomsky, 1981), preventing movements out of all environments that involve crossing more than one bounding node, and with the selection of bounding nodes being subject to parametric variation (Rizzi, 1982). As we will see in chapter 6, the gradience and cross-linguistic variation in this domain proves a challenge for analyses framed in terms of innate parametrized constraints, and point rather to a processing-based account of the preference for resumptive pronouns over gaps, at least in certain types of island environments (Hawkins, 2004; Ariel, 1990).

In this chapter the aim is to simply use these typologically attested tendencies to formulate predictions for the patters of variation exhibited in the AF alternation. If the AF alternation represents an alternation between a resumptive pronoun and a gap, we might expect to find that the synthetic (resumptive) verb form is available in island contexts, while the gapped (AF) construction is not, or at least, the gap should be highly disfavored. This turns out to be correct. The synthetic verb form is permitted in wh-islands, as the following two examples show:

(3.41)   a. *Lela' leti' le   artista hach uts   ti' inw ich báax*
             this   it   DEF artist   very good in A1 eye what
          k-u=pintart-ik
          IMPF-A3=paint-INC
          'This is the painter$_i$ that I like what he$_i$ paints.'

         b. *Máax t-u=tukl-ah   Juan wáah **t-u=ts'u'uts'-ah**   x-maria-o'*
             who   PRV-A3=think-CMP Juan ALT   PRV-A3=kiss-CMP F-Maria-D2
          'Who$_i$ did John wonder whether he$_i$ kissed Maria?'

It is also available in complex NP islands:

(3.42)   a. *Lela' leti' le   x-ko'olel le   máak-o'ob **t-u=kaxt-ah-o'ob***
             this   it   DEF F-woman DEF person-PL PRV-A3=meet-PL
          t-uy=a'l-o'ob   hach ya'ab uy ohel
          PRV-A3=say-PL   very much A3 knowledge
          'This is the woman$_i$ that the people that she$_i$ met said that she was very intelligent.'

         b. *Teen t'aan-ah-en   yéetel le   policia   le   òokol*
             I   speak-CMP-B1 with   DEF policeman DEF thief
          t-u=mach-ah   *puts'   ti'*
          PRV-A3-catch-CMP escape to
          'I talked with the policeman$_i$ that the thief that he$_i$ caught escaped from him.'

Significantly, the AF versions of these sentences are, by contrast, ungrammatical:

(3.43)   a. **Lela' leti' le   artista hach uts   ti' inw ich báax **pintart-ik***
             this   it   DEF artist   very good in A1 eye what paint-INC
          'This is the painter$_i$ that I like what he$_i$ paints.'

    b. **Máax   t-u=tukl-ah   Juan wáah **ts'u'uts'** x-maria-o'*
       who   PRV-A3=think-CMP Juan ALT   kiss   F-Maria-D2
       'Who$_i$ did John wonder whether he$_i$ kissed Maria?'

(3.44) a. **Lela' leti' le   x-ko'olel le   máak-o'ob **kaxt-o'ob***
        this   it   DEF F-woman DEF person-PL meet-PL
      t-uy=a'l-o'ob   hach ya'ab uy ohel
      PRV-A3=say-PL very   much   A3   knowledge
      'This is the woman$_i$ that the people that she$_i$ met said that she was very intelligent.'

    b. **Teen t'aan-ah-en   yéetel le   policia   le   òokol **mach** puts'*
       I   speak-CMP-B1 with   DEF policeman DEF thief   catch escape
      ti'
      to
      'I talked with the policeman$_i$ that the thief that he$_i$ caught escaped from him.'

Construction type

A less often cited, but typologically attested generalization regarding the distribution of gaps and resumptions involves the specific construction type. It has been reported for a variety of languages that the unbounded dependencies which allow resumptives may be restricted: while they may occur in relative clause constructions, they are disallowed or dispreferred in wh-questions (Boeckx, 2003; Alexopoulou, 2009). This is attested for Hebrew, as the following ungrammatical examples show (also discussed in Asudeh 2004), as well as Levantine Arabic (Boeckx, 2003).

(3.45) Hebrew (Sells, 1984, 63)

    *Mi raʔiti   ʔoto?
    who saw.1S him?
    'Who did I see?'

(3.46) Hebrew (Sharvit, 1999, 591)

    *Mi nifgašta ito
    who you.met with.him
    'Who did you meet with?'

Strikingly, we find a very similar situation in Yucatec. In bare wh-questions, the full verb form is strongly dispreferred by my consultants in simple contexts, while in relative clauses, it is readily accepted:

(3.47)   a. *??máax **t-uy=il-ah**   x-maria=o'*
             who   PRV-A3=see-CMP   F-maria=D2
             'Who saw Maria'?

         b. *le   máak   t-uy=il-ah   x-maria=o'*
             The   man   PRV-A3-see-CMP   F-maria-D2
             'the man who saw Maria'

The acceptability of wh questions with the resumptive verb interacts with embedding depth. With one level of embedding, both the resumptive form and the AF form are acceptable in questions:

(3.48)   a. *Máax   t-aw=a'l-ah   **t-uy=il-ah**   le   x-maria=o'*
             who   PRV-A2=say-CMP   PRV-A3=see-CMP   DEF   F-maria=D2
             'Who did you say saw Maria?'

         b. *Máax   t-aw=a'l-ah   **il**   le   x-maria=o'*
             who   PRV-A2=say-CMP   see   DEF   F-maria=D2
             'Who did you say saw Maria?'

And in even more complex constructions, with a further level of embedding, the resumptive verb form is preferred over the AF form in bare wh-questions:

(3.49)   *Máax   t-aw=a'l-ah   t-uy=a'l-ah   juan **t-uy-il-ah***
          who   PRV-A2-say-CMP   PRV-A3-say-CMP   juan   PRV-A3-see-CMP
          x-maria=o'
          F-maria=D2
          'Who did you say that Juan said saw Maria?'

      b. *??Máax   t-aw=a'l-ah   t-uy=a'l-ah   juan **il**   x-maria=o'*
          who   PRV-A2-say-CMP   PRV-A3-say-CMP   juan   see   F-maria=D2
          'Who did you say that Juan said saw Maria?'

The asymmetries across construction types are also evident in other languages in the family. In Mam, for example, AF is obligatory with questions, but optional with focus clefts and relative clauses (England, 1983). In Ixil AF is obligatory with both wh questions and focus clefts, but optional with relative clauses (Ayres, 1983). In Sipakapense, AF is obligatory with wh questions and focus clefts, but the synthetic form is obligatory with relative clauses. Table 3.3 illustrates the patterns. '+' signifies obligatory use of the synthetic form, '-' signifies obligatory use of AF form; '±' signifies variability. At the time of writing I was unaware of data for any language pertaining to non-local extractions (i.e., how embedding might interact with construction type in affecting the choice between verb forms).

Language	Wh-question	Focus-cleft	Relativization
Sipakapense	-	-	+
Ixil	-	-	±
Mam	-	±	±

Table 3.3: Distribution of synthetic and AF verb forms across construction types

Differences have also been reported across languages between bare wh-questions and *which*-questions, with respect to resumption. Sharvit (1999) observes that in simple contexts, *which*-questions in colloquial Hebrew allow resumptives, though bare wh-questions are not possible:

(3.50) Hebrew (Sharvit, 1999, 591)

    Eyze   student nifgašta   ito?
    Which student you.meet with.him
    'Which student did you meet with?'

The same is found in Yucatec. In simple environments, *which*-questions are acceptable both with the resumptive verb form, and the AF verb form:

(3.51)   a. *Máax wàach-il okol-t   le   libro=o'*
         who   soldier-of steal-APP the book-D2
         'Which soldier stole the book?'

b. *Máax wàach-il t-uy-okol-t-ah le libro=o'*
who soldier-of PRV-A3-steal-APP-CMP the book-D2
'Which soldier stole the book?'

Summary

The patterns of distribution exhibited by the synthetic and the AF verb form in Yucatec are consistent with cross-linguistically documented patterns of resumptive/gap distributions: the AF verb form is barred from island environments, the synthetic verb isn't. The synthetic form is preferred in complex dependencies, the AF form is preferred in local environments. The AF form is preferred in questions, the synthetic form is more acceptable in relative clause constructions.

	Syntactic RP/gaps	AF alternation
Islands	√	√
Embedding	√	√
Construction type	√	√

Table 3.4: Conditioning factors for choice between alternants

The limited data available on other languages supports the view that these distributional patterns are detectable beyond Yucatec. The AF alternation is sensitive to embedding depth in Jakalteko, for example, and Ixil, Sipakapense and Mam show an asymmetry between relative clause constructions and wh questions in the expected direction.

3.3 Revisiting syntactic ergativity

As I showed in chapter 2, the restriction of AF to subjects of transitive relative clauses has often been presented as evidence that AF is a reflex of underlying syntactic ergativity (Larsen and Norman, 1979; Larsen, 1981; Van Valin, 1981; England, 1983, 1988; Campana, 1992; Campbell, 2000). Invoking the existence of ergatively organized syntax to explain the restriction of AF to A-extractions is made problematic by the fact that the phenomenon of AF is not categorical. Many languages, including Yucatec, exhibit variation in the choice between the synthetic verb and the AF verb. I showed in the previous sections that the patterns of variation exhibited in the choice between the AF and the synthetic

verb in Yucatec instead support the alternative the view that the alternation is a subtype of RP/gap alternation.

Significantly, the restriction of the AF verb to extractions of transitive subjects, so often taken to be an indicator of syntactic ergativity, immediately follows on the RP/gap analysis. It only takes a quick inspection of dependent person marking paradigms in Mayan languages (Yucatec's is repeated below) to see why this is so. That intransitive dependent verbs should not participate in the alternation follows from a basic asymmetry in the Mayan pronominal paradigm: all Mayan languages have overt bound pronominals for all persons and core grammatical functions *except for third person singular set B*. Across the whole family, third person S and O are phonologically zero.

	Set A		Set B	
	Sg	Pl	Sg	Pl
1	in(w)	k(...-o'n)	-en	-o'n
2	a(w)	a(w)...-e'x	-ech	-e'x
3	u(y)/y	u(y)...-o'b	-ø	-o'b

Table 3.5: Yucatec's bound pronominal system

If the AF verb form functions to allow the dependency to be concluded by a gap (a zero-expressed argument), then the transitivity restriction on the AF alternation follows from the simple fact that finite intransitive verbs do not carry pronominal inflection. That is, dependencies involving an intransitive dependent verb are already terminated with a gap in such cases.

(3.52)　　a. *H-ø-p'áat　　le　　máak-o'*
　　　　　　　PRV-B3-leave　DEF　man-D2
　　　　　　　'The man left.'

　　　　　b. *le　　máak　h-ø-páat-o'*
　　　　　　　DEF　man　PRV-B3-leave-D2
　　　　　　　'the man that left'

We could phrase this in teleological terms and say that these languages do not 'need' a specialized gap strategy in these contexts because intransitive verbs do not inflect for subject pronouns. This is really just a way of anthropomorphizing an aspect of the synchronic

grammar which I will argue in chapter 5 is an outcome of diachronic grammaticalization processes. For now it suffices to observe that the AF alternation is conditioned by the asymmetry of its person marking paradigm. We do not find any verbal alternation precisely in those contexts where main verbs do not carry overt pronominal inflection. Thus, what may at first glance appear to be some deep rule of grammar (a special restriction on subject extractions) is actually driven by surface morphology. It is the *morphological* ergativity of the pronoun system which drives the AF alternation, not any underlying syntactic ergativity.

This point is important not only in order to dispel certain interpretations of Mayan syntax, but also because it will allow us to identify commonalities across other verb-alternating head marking languages which differ primarily in the morphological alignment patterns of their person marking systems. This is the topic of chapter 4. In the remainder of this chapter, I will briefly discuss Yucatec's RP/gap alternations in the context of the Accessibility Hierarchy.

3.4   The Accessibility Hierarchy and Yucatec resumption

The Accessibility Hierarchy of Keenan and Comrie (1977) and Comrie and Keenan (1979) states implicational universals governing what kinds of grammatical functions NP_{rel} can bear in the relative clause. The basic claim is that the grammatical functions of a language are arranged in a hierarchy such that if, in that language, NP_{rel} can bear a given grammatical function, it can also bear all functions that are higher on the hierarchy. The original formulation of the Accessibility Hierarchy is:

(3.53) Subject > Direct Object > Indirect Object > Obliques > Genitives > Objects of
         Comparison

Cross-linguistically, the distribution of gaps and pronouns has been observed to correspond to different positions on the hierarchy: gaps are more likely to be used with positions higher on the hierarchy, resumptive pronouns at the lower end.

On Keenan and Comrie's account, subject relativization is stated as an absolute universal, that is, all languages are assumed to be able to relativize subjects. Similarly, it is supposed that relativization with a gap is available in any language. Taken together, the claim is therefore that gapping is universally available for subject relativization. The

fact that cases of obligatory subject resumption can be identified in various languages (e.g. Urhobo (Keenan and Comrie, 1977), and also several Mayan ones, on the present analysis) undermines the universality of this claim, but it appears to be correct as a typological generalization. And the availability of the AF form in many Mayan languages reflects this typological tendency for subject gapping.

The particulars of the Mayan case raise some interesting questions. I have shown that subject resumption is possible in Yucatec: the synthetic (resumptive) verb form may occur in subject extracted dependencies. In terms of the implicational hierarchy of the Accessibility Hierarchy, this would predict that object resumption should also be available in Yucatec.

Let's take a look at some of the properties of pronominal objects in Yucatec to see what implications emerge regarding the Accessibility Hierarchy. As we have seen, singular object 'pronouns' are zero exponents in Mayan languages. In the absence of an NP object, transitive verbs are unambiguously understood as anaphoric to a third person entity (3.54).[14] The featural values of the zeros in these cases are entirely recoverable given the fact that the remainder of the pronominal paradigm is fully specified.

(3.54) *Le   x-chiiwol=o'   t-u-chi'-ah*
        DEF   FEM-tarantula=D2   PRV-A3-bite-CMP
        'The tarantula bit it/him/her'.

However, it appears that zero pronouns in Yucatec exhibit a distinct referential behavior from their overt counterparts. Unlike overt pronominal elements in Yucatec, unexpressed NP objects are not constrained to be anaphorically bound by a discourse participant in the linguistic context, nor do they even necessarily have to receive a definite interpretation. Hence the null object in (3.55) can refer either to the specific puma Juan saw, or to a different puma:

(3.55) *Juan=e'   t-uy=il-ah   hun-túul   koh   beyxan Pedro*
        Juan=TOP PRV-A3-see-CMP one-CL.AN puma also   Pedro
        t-uy=il-ah
        PRV-A3-see-CMP
        'Juan saw a puma and Pedro saw it too/ Juan saw a puma and Pedro saw one too.'

---

[14]This type of situation may be contrasted with true 'zero' or 'null' pronouns in other languages, which have no intrinsic specification for person, number or gender. In Chinese or Japanese for example, which lack verbal agreement morphology, zero pronouns can be used for various persons and numbers.

The reference does not even necessarily have to be singular:

(3.56) *Juan=e'   t-uy=il-ah   hun-túul   koh-o'ob beyxan Pedro*
     Juan=TOP PRV-A3-see-CMP one-CL.AN puma-PL also   Pedro
t-uy=il-ah
PRV-A3-see-CMP
'Juan saw pumas and Pedro saw them too/ Juan saw pumas and Pedro saw some
too.'

The difference between this null object and the overt pronominal markers is revealed when its behavior is contrasted with the *plural* object enclitic, which is an overt marker which functions as the plural marker both for NPs, and for verbs. In the absence of a nominal object, the plural set B marker on the verb functions as a bound pronoun:

(3.57) *Le   maestro=o' k-u=ka'ans-ik-o'ob*
     DET   teacher-D2   IMPF-A3=teach-INC-PL
'The teacher teaches them.'

Significantly, the equivalent of (3.55) with a plural marked verb induces a definite interpretation of the object:

(3.58) *Juan=e'   t-uy=il-ah   hun-túul   koh-o'ob beyxan Pedro*
     Juan=TOP PRV-A3-see-CMP one-CL.AN puma-PL also   Pedro
t-uy=il-ah-o'ob
PRV-A3-see-CMP
'Juan saw pumas and Pedro saw them too.'/ NOT: 'Juan saw pumas and Pedro saw
some too.'

In other words, the overt pronominal markers are referentially more restricted than their zero counterparts.The latter type, I submit, do not represent specified prononominal arguments, but rather simply represent the absence of an overt object, whose reference may be contextually determined.[15]

Given that overt, bound pronominal marking develops from the grammaticalization of independent pronouns, the lack of any overt third person pronominal inflection across all members of the Mayan family can probably be attributed to the lack of an appropriate

---

[15]The correlation between pronominal form type and referential specificity has been explored by Evans (2002) for Bininj Gunwok.

original lexical source for grammaticalization. That is, (pre) proto Mayan presumably did not possess a set of independent third person absolutive pronouns, and thus, pronominal inflection never developed in precisely these cases.[16] In the synchronic grammars in Mayan languages, similarly, we can suppose that pronominal elements in these cases are simply 'not there'. It therefore follows that in unbounded dependencies where NP$_{rel}$ is a singular object, this grammatical function is expressed by a gap.[17]

This would appear to represent an inversion of the Accessibility Hierarchy: a language where subject resumption is possible, but not object resumption. However, I believe that it is only appropriate to consider such an implicational hierarchy in the context of the lexical resources available to a given language. Yucatec does not possess singular object pronouns: the lack of object resumptives follows accordingly.

Moreover, when plural object pronouns are considered, it can be seen that Yucatec does in fact pattern in the typologically predicted way. In simple clauses, plural object pronouns are in complementary distribution with NP objects, as the comparison between the grammatical (3.59a) and (3.59b), and the unacceptable (3.59c) show:

(3.59)   a. *Le   maestro=o' k-u=ka'ans-ik-o'ob*
            DEF   teacher=D2   IMPF-A3-teach-INC-PL
            'The teacher teaches them.'

      b. *Le   maestro=o' k-u=ka'ans-ik   le   pàalal-o'ob=o'*
            DEF   teacher=D2   IMPF-A3-teach-INC   DEF   student-PL=D2
            'The teacher teaches the students.'

      c. **Le   maestro=o' k-u=ka'ans-ik-o'ob   le   pàalal-o'ob=o'*
            DEF   teacher=D2   IMPF-A3-teach-INC-PL   DEF   student-PL=D2
            'The teacher teaches the students.'

In object extracted relatives, the dependency may be concluded either with a resumptive plural object bound pronoun (3.60a) or with a gap (3.60b):

---

[16]The lack of third person independent pronouns is not particularly unusual from a typological perspective. Many languages lack third person personal pronouns. Instead, third persons may be referred to by full NPs, demonstratives (which may constitute a historical source for pronouns and verbal agreement, cf. e.g. Siewierska 2004), or by nothing at all. Given these facts, it is to be expected that in some instances, lack of third person pronominal inflection is due to the lack of an appropriate source for grammaticalization.

[17]As opposed, for example, to a 'null resumptive pronoun'.

(3.60) a. *In k'ahóol le   pàalal-o'ob k-u=ka'ans-ik-**o'ob***
       A1 know   DEF student-PL   IMPF-A3=teach-INC-PL
    'I know the students that he teaches (them).'

    b. *In k'ahóol le   pàalal-o'ob k-u=ka'ans-ik*
       A1 know   DEF student-PL   IMPF-A3-teach=INC
    'I know the students that he teaches.'

Thus, we find that overt resumption is possible in Yucatec both for subjects and for objects. In sum, the verbal alternations that we find in Yucatec long distance dependencies conform to the typological patterns of relativization codified by the Accessibility Hierarchy.

3.5   Summary

In this chapter I advocated the view that pronouns can be formally instantiated in a variety of ways, ranging from words, to morphological inflection on heads:

(3.61) strong pronoun > weak pronoun > clitic > affix

I demonstrated, furthermore, that, on the assumption that resumptive pronouns are simply ordinary pronouns, the type of a given language's person marking system will determine the form of its resumptive pronouns. A language like English, which has an independent pronominal system, has independent resumptive pronouns. A head-marking language like Yucatec, by contrast, which has a morphologically dependent pronominal system, has morphologically dependent resumptive pronominals.

    I suggested, finally, that languages possessing morphologically dependent resumptives, nevertheless may exhibit alternations with gaps in unbounded dependencies. The crucial difference is that for such languages, the alternation between gaps and resumptives relates to distinct paradigms of verb forms. I argued that the AF alternation in Mayan languages is an instance of precisely this. This is supported by the fact that the variation exhibited in this domain, both within Yucatec in terms of the choice between alternants, and across Mayan languages, in terms of patterns of categorical cut-off points, mirrors typologically attested asymmetric frequency distributions of resumptive pronouns and gaps.

Chapter 4

Beyond Mayan

4.1 Introduction

In this chapter I look beyond Mayan to a range of other head-marking languages which make use of special verb forms in unbounded dependencies. Like Mayan AF verbs, these are distinguished (minimally) from their main clause counterparts by the absence of person marking inflection. And, just like the Mayan AF alternation, the alternation between these special verbs, and their fully inflected counterparts pattern like RP/gap alternations. Thus, while the Mayan AF alternation has often been viewed in typological isolation, as a sui generis phenomenon, this chapter sets the Mayan AF alternation firmly in typological context, providing cross-linguistic support for the analysis of Mayan AF developed in the previous chapters, and consequently, revealing verbal alternations in head-marking languages to be a productive, yet under-explored corner of RP/gap typology.

4.2 Verbal alternations in unbounded dependencies

Verbal alternations which are sensitive to wh-extraction and other A'-movement processes, and which are characterized (minimally) by the absence of the regular person marking present in main clauses have been attested for a large number of languages, including Hausa (Tuller, 1986), Kinande (Schneider-Zioga, 1995) , Somali (Lecarme, 1989; Frascarelli, 1999, 2000; Mereu, 1999), Fiorentino and Trentino (Brandi and Cordin, 1989), Kikuyu (Clements, 1984), Halkomelem Salish (Wiltschko, 2006), Chamorro (Chung, 1982,

1998), Palauan (Georgopoulos, 1991), Tukang Besi (Donohue, 1999), Moore (Haïk et al., 1985) Konjo (Friberg, 1996), Chalcatongo Mixtec (Macaulay, 1996), Selayerese, (Finer 1997), Makaresse (Finer, 1997), Irish (McCloskey, 1990; Hendrick, 1988), Yimas (Foley, 1991), Apurina (Facundes, 2000) and Bare (Aikhenvald, 1995). The language specific facts relating to such verbal alternations have been reported separately for individual languages, and analyzed in a variety of different ways. In the generative literature the phenomenon has been reported for (subsets of) these languages under the term *wh-agreement* (e.g. Chung 1982), see also Georgopoulos (1991), or *anti-agreement* (since Ouhalla 1993). Here I will on occasion use the term *anti-pronominal* rather than anti-agreement to refer to the alternation, which more accurately captures the view of these alternations advocated here. Borrowing terms from the Indo-European tradition, I will also at times refer to the bare (pronominal-less) verb forms as *analytic*, and their pronominally inflected counterparts as *synthetic* (as in my use of the term for Mayan languages).

As a first example, consider Tarifit Berber. In Tarifit Berber, the local extraction of a subject in wh-questions (4.1a) , relatives (4.2a) and clefts (4.3a), requires a special verb form which is referred to by traditional grammarians as the 'participle'. This verb form does not bear subject person marking, unlike finite verbs of main clauses. The regular finite verb form bearing subject person marking, though obligatory in main clauses, is ungrammatical in these contexts, as shown by the starred (b) examples:

(4.1) Berber (Ouhalla, 1993)

         a. *Man   tamghart ay   yzrin   Mohand*
           which   woman   COMP see.PART Mohand
           'Which woman saw Mohand?'

         b. **Man tamghart ay   t-zra   Mohand*
           which   woman   COMP 3FS-saw Mohand

(4.2)   a. *tamghart nni   yzrin   Mohand*
           woman   COMP saw.PART Mohand
           the woman who saw Mohand

         b. **tamghart nni   t-zra   Mohand*
           woman   COMP 3FS-saw Mohand

(4.3)　a.　*Tamghart-a　ay　　yzrin　　　Mohand*
　　　　　woman-this　COMP　saw.PART　Mohand
　　　　　'It was this woman who saw Mohand.'

　　　　b.　**tamghart-a　ay　　t-zra　　　Mohand*
　　　　　woman-this　　COMP　3FS-saw　Mohand

The parallel with the Mayan case couldn't be clearer. In the same three construction types, clefts, wh-constructions and relative clause constructions, we find an alternative verb form occurring in dependent clauses in subject extractions which does not bear the expected person marking of main clause verbs.

　　　Chamorro provides a second example. In Chamorro, the questioning, clefting or relativizing of a subject requires a verb form infixed for *-um-* (4.4b), rather than the regular subject person marking of main clauses. The contrast is exemplified below between a main clause (4.4a), and a subject wh-question (4.4b):

(4.4)　Chamorro (Chung 1998, adapted from 52 and 53b, 236)

　　　　a.　*Ha-fa'gasi　si　　Juan　i　　kareta*
　　　　　3SG-wash　PN　　Juan　the　car
　　　　　'Juan washed the car'

　　　　b.　*Hayi　f-um-a'gasi　i　　kareta*
　　　　　who　　*um*-wash　　the　car
　　　　　'Who washed the car?'

Just as within the Mayan language family, verb alternating languages do not form a unitary class, in terms of the formal realization of the alternating verbs. While in some cases, as in the Berber and Chamorro examples above, the verbal alternation involves the presence of suppletive morphology in place of the regular person marking, such as Chamorro's -um-infix for subject extraction, in other languages the only distinguishing characteristic of the alternative verb form is the lack of person marking obligatory in main clauses.

　　　An example of this type is Chalcatongo Mixtec, a head-marking language in which subject person markers attach to the right edge of the finite verb (4.5a). In focus constructions, Wh questions and relative clause constructions, the subordinate verb appears without the person marker, as in the focus cleft in (4.5b):

(4.5) Chalcatongo Mixtec (Macaulay, 1996, 140)

    a. *ñãʔã   wã́ã xĩ́nũ=ñá*
       woman the   run=3F
       'The woman is running.'

    b. *ñãʔã   wã́ã xĩ́nũ*
       woman the   run
       'The woman is the one who is running.'

Papuan Malay provides a second example of this type. The third person marker *de-* is obligatory in main clauses:

    (4.6) Papuan Malay (Donohue, 1999, 257)

    a. *de=so=jalan*
       3s=PF=walk
       'He's already walked (away)'

    b. *de=so=lia   dia*
       3s=PF=see   3s
       'She's seen him'

In wh-questions, the verb does not bear the subject person marker *de-* of main clauses:

(4.7) *Siapa maka jue   itu   tadi*
     who   eat   cake that ealier
     'Who ate that cake?'

In sum, we find across a range of head-marking languages special verb forms occurring in unbounded dependencies which, are characterizable formally by (minimally) an absence of canonical person marking, and, distributionally, by their restricted occurrence in unbounded dependencies. As I will show below, despite the difference in the formal realization of some of these verb forms, the alternations that they participate in with synthetic verb forms in unbounded dependencies exhibit the expected behavior of RP/gap alternations.

4.3 Morphological RP/gap alternations: distributional evidence

As I showed in chapter 3, when both resumptive pronouns and gaps are available in a language, they generally have overlapping, but often non-identical distribution with respect to syntactic positions/grammatical functions. This distribution tends to conform to certain well-attested patterns. I discussed in chapter 3 how some of these patterns of distribution are detectable in the Yucatec AF alternation. In this section, I show how these same patterns of distribution are also apparent in anti-pronominal languages beyond Mayan. This is by no means an exhaustive survey; it is restricted to examples from languages where there are comprehensive enough descriptions of the relevant syntactic phenomena. But it will suffice to convey that, in a range of languages that are morphologically similar to Mayan languages in the relevant respect, the choice between verbal alternants is conditioned by the same types of factors, supporting the view that they are all exemplars of the same basic phenomenon.

4.3.1 Islands

If anti-pronominal alternations represent an alternation between a resumptive pronoun and a gap, we should expect to find that the synthetic (resumptive) verb form is available in island contexts, while the analytic (argument-gapped) construction is not, or at least, the gap should be highly disfavored. This turns out to be correct. Just as in Yucatec, anti-pronominal alternations pattern in the expected way with regard to island sensitivity, as the following examination of Irish and Palauan reveal.

Irish

Irish verbal paradigms consist of forms that are traditionally called synthetic and analytic (McCloskey and Hale, 1984, 489). The synthetic form carries subject pronominal inflection, while the analytic form does not. Each verbal paradigm consists of exactly one analytic form and a set, possibly null, of synthetic forms. In a single inflectional ending, the synthetic form encodes information about tense and mood, as well as the person and number of its subject. The analytic form encodes only information about tense and mood.

Unlike the synthetic form, the analytic form must combine with independent pronouns. The paradigm in (4.3.3) represents the conditional of the verb *chuir* 'put' in the Ulster dialect. In this particular paradigm, the synthetic forms are the 1st person singular and plural, and the second person singular (in boldface).[1]

	Singular	Plural
1	**chuirfinn**	**chuirfimis**
2	**chuirfeá**	chuirfeadh sibh
3	chuirfeadh sé (M)	chuirfeadh siad
	chuirfeadh si (F)	

Table 4.1: Alignment patterns for verb-alternating languages

Generally, it is ungrammatical to use the analytic member of a paradigm with an independent pronominal subject, if a suitable synthetic form exists. That is, because there exists a conditional first person synthetic form, *chuirfinn*, (4.8) is ungrammatical:

(4.8) *Chuirfeadh mé isteach ar an phost sin.*
 put.CONDIT 1 in on that job
 'I would apply for that job'.

Independent resumptive pronouns are illicit in the subject position immediately subjacent to the head of a relative clause or the interrogative phrase of a question (the so-called highest-subject restriction):

(4.9) **an fear aN raibh sé san otharlann*
 the man COMP was he in-the hospital
 the man that (he) was in hospital

The inflectional subject behaves just like an overt pronoun. It is also ungrammatical in this position:

(4.10) **na daoine aN mbldis san otharlann*
 the people COMP be.PAST.HABIT.3PL in-the hospital
 'the people that (they) used to be in hospital'

---

[1] The data and observations in this section on Irish are taken from McCloskey and Hale (1984).

In such instances, the analytic form must be used. Consider the cleft example in (4.11):

(4.11) *Chan   mise a   chuirfeadh   isteach ar an   phost sin.*
       COP+NEG me   COMP put.CONDIT in   on   that job
       'It's not me that would apply for that job'.

The antecedent for the NP*rel* is a synthetic first person pronoun. Even though there exists a synthetic form of *chuir*, the only possibility in this context is the analytic form *chuirfeadh*, which does not bear person marking.

    We can now turn to island environments. Independent resumptive pronouns in Irish are not subject to island constraints, whereas the binding between a relative clause head (or an interrogative phrase) and a gap is. Embedded questions, for instance, are absolute islands. (4.12a) is ungrammatical, with a gap, while (4.12b), with an independent resumptive pronoun, is acceptable.

(4.12)   a. **daoine nach   mbionn   fhios   agat ariamh an dtiocfaidh*
            people   COMP+NEG you-know HABIT   ever   Q come.FUT
            in am
            in time
            'people that you never know if __ will come on time'

         b. *daoine nach   mbionn   fhios   agat ariamh an dtiocfaidh*
            people COMP+NEG you-know HABIT   ever   Q come.FUT
            siad in am
            they in time
            'people that you never know if they will come on time'

As McCloskey observes, the inflectional subject behaves like an overt pronoun in this respect. The inflectional subject is acceptable inside islands which do not contain its antecedent. This is illustrated in (4.13):

(4.13) *daoine nach   raibh fhios againn   an dtiocfaidis   in am*
       people COMP +   NEG we-knew Q come.CONDIT.3PL in time
       'people that we did not know whether they would come on time'

In sum, the inflectional subject is indistinguishable in its syntactic behaviour from an overt pronoun in subject position. Accordingly, McCloskey and Hale (1984), working in Principles and Parameters, analyze the synthetic form as contributing a null pronominal argument

(i.e., pro). (Andrews, 1990), working in Lexical Functional Grammar, analyzes the synthetic form as contributing the PRED of its argument at functional structure. The analytic form, for its part, exhibits the distributional behavior of a gap.

Palauan

In Palauan, subject person marking, which, for the third person is *-ng*, is obligatory main clauses (4.14a). Verbs not bearing this subject person marker are ungrammatical in main clauses (4.14b):

(4.14)   a. **_ng-kiltmeklii_** *a ulaol a Peter*
          3S-clean-3S   floor   Peter
        'Peter cleaned the floor'

      b. **_*kiltmeklii_** *a ulaol a Peter*
         clean-3S   floor   Peter
        'Peter cleaned the floor'   (Georgopoulos, 1991)

In simple extractions, however, the dependent verb must occur without the obligatory subject person marking of main clauses:

(4.15) *a 'ad el*   **_mil'er-ar_**   *tia*   *el buk*
       man Comp PF.buy-3S.Obj Dem L book
   'the man who bought this book'   (Georgopoulos, 1991)

In island environments, the dependent verb occurs in the regular person marked form, that is, with the subject person marker attached. The two dislocated topic constructions below are examples of such island environments. In (4.16), the topic is bound to the subject within a sentential subject. (4.17) is an example of a wh-island:

(4.16) *a Mary a kltukl*   *e*   *kmo* **_ng-oltoir_** *er*   *a John*
       Mary   R-clear COMP   3S-love PREP   John
   'Mary, (it's) clear that she loves John'   (Georgopoulos, 1991, 80)

(4.17) *a del-ak*   *a diak ku-dengei*   *el*   *kmo* **_ng-ngera_** *a*
       mother-1S   NEG IRR-1S.IRR-know.Im COMP   Cl-what
       bo   *lo-ruul*   *el mo belsoil*
       IRR-FUT 3S.IRR-do L go   dinner
       'My mother, I don't know what she will cook for supper' (Georgopoulos, 1991, 81)

Georgopoulos actually offers examples such as (4.16) and (4.17) as evidence that gaps in Palauan are not sensitive to islands, of any type. This is because she analyzes the morphologically bound person marking as pure agreement markers. On the alternative view espoused here that the person marking information on the verb is functions pronominally, these dependencies are not truly concluded with gaps, but rather with resumptive pronominal inflection. It is notable that Georgopoulos does not provide any examples of islands in which the embedded verb does not bear subject pronominal marking.[2]

4.3.2 Embedding

Ouhalla (1993, 486) observes that there is apparently no language which displays an anti-pronominal alternation (in his terms, anti-agreement) in relation to long extraction (embedded environments) but not in relation to short extraction. This distributional pattern is predicted on the assumption that the alternation is a type of RP/gap alternation: as we saw in the last chapter, resumptives tend to occur in embedded environments, gaps in less complex ones.

Tarifit Berber

In Tarifit Berber, main clause verbs obligatory carry subject person marking. In simple extractions, this synthetic verb form with subject marking is ungrammatical (4.18a) and (4.19a). Instead, the participle verb form is required ((4.18b) and (4.19b):

(4.18) a. **man   tamghart   ay   t-zra   Mohand*
         'Which   woman   COMP   3FS-saw   Mohand'

    b. *man   tamghart   ay   yzrin   Mohand*
       which   woman   COMP   see.PART   Mohand
       'Which woman saw Mohand?'   (Ouhalla, 1993, 492–493)

(4.19) a. **tamghart   nni   t-zra   Mohand*
         woman   COMP   3FS-saw   Mohand

---

[2]See Bresnan (1998) for independent evidence from weak cross-over effects that subject person marking in Palauan functions pronominally.

    b. *tamghart   nni   yzrin   Mohand*
       woman   COMP saw.PART Mohand
       'the woman who saw Mohand'   (Ouhalla, 1993, 492–493)

In embedded contexts, however, the synthetic verb form with the inflectional subject occurs:

(4.20)   a. *man   tamghart ay   nna -n   qa   t-   zra Mohand*
            which woman   COMP said -3PL that 3FS saw Mohand
            'Which woman did they say saw Mohand?'

        b. *tamghart nni   nna -n   qa   t-   zra Mohand*
           woman   COMP said -3PL that 3FS- saw Mohand
           'The woman that they said saw Mohand'   (Ouhalla, 1993, 492–493)

The fact that these alternations are behaving like RP/gap alternations is made especially clear when we compare the situation with subject extractions to that of object extractions. Object pronouns in Berber take the form of clitics. The alternation that takes place in the syntax between the independent resumptive object pronominal and a gap parallels what occurs in the morphology, in the form of a verbal alternation, for subject extraction. Thus, object resumptives are ungrammatical in local extractions (4.21a), but are optional in embedded environments (4.21b):

(4.21)   a. *Tamhdart nni   (*ti) zri-gh*
            student   COMP her   saw-1s
            'The student I saw.'

        b. *Tamhdart nni   nna-n   qu   zri-gh   (ti)*
           student   COMP said.3PL that saw/1S her
           'The student that they said that I saw.'   (Ouhalla, 1993, 492–493)

In sum, the inflectional subject is indistinguishable in its syntactic behaviour from an overt clitic pronoun in object position, both are constrained against occurring in local extractions. Dependencies featuring the participle form, for their part, have the character of subject gap structures.

Breton

Breton exhibits a similar situation. In short extractions, an analytic form of the dependent verb is required, which does not bear subject inflection. The following two examples demonstrate this for a wh-question (4.22a) and a relative clause construction (4.22b). The analytic verb form *lenne* is obligatory, rather than the plural subject-inflected form *lennent*:

(4.22)   a. *Petore paotred a   lenne (*lennent) al   levrioù?*
            which boys   COMP read   read.3PL   the books
            'Which boys read the books?'

        b. *ar   vugale   a   lenne (*lennent) al   levrioù*
            the children COMP read   read.3PL   the books
            the children who read the books   (Hendrick, 1988)

Long extraction of the subject does not require the analytic form.

(4.23) *Setu ar   mere'hed hoc'heus lavaret **emaint** o   labourat e   Kemper*
        here the   women   have.2PL said   be.3PL PART work   in Kemper
        'Here are the women you said are working in Kemper.'   (Hendrick, 1988)

4.3.3 Construction type

It has been reported for a variety of languages that the unbounded dependencies which allow resumptives may be restricted: while they may occur in relative clause constructions, they are disallowed or dispreferred in wh-questions (Boeckx, 2003; Alexopoulou, 2009). The AF alternation in Yucatec and several other Mayan languages exhibited this effect. It is an effect which is also visible in other verb-alternating languages.

Papuan Malay

Main clause verbs in Papuan Malay obligatory carry a subject pronominal clitic (4.24).

(4.24) ***De**=so=jalan*
      3S=PF=walk
      'He's already walked (away).'   (Donohue, 1999)

(4.25a) and (4.25b) show that an independent pronoun may appear as the subject of a verb in the same sentence as the inflected verb, but only if the independent pronoun carries a contrastive intonation pattern. A free pronoun alone, in the absence of a clitic, does not create a grammatical clause (4.25c)

(4.25)   a. *Día(,) de=tra=tau*
              3S   3S=NEG=know
          '<u>He</u> doesn't know.'

        b. **Dia de=tra=tau*
             3S   3S=NEG=know
          'He doesn't know.'

        c. **Dia tra=tau*
             3S   NEG=know
          'He doesn't know.'   (Donohue, 1999)

In wh-questions, the verb does not bear the subject person marker *de-* of main clauses:

(4.26) *Siapa maka jue   itu   tadi*
       who   eat   cake that ealier
       'Who ate that cake?'   (Donohue, 1999, 257)

In relative clauses, however, it is optional:

(4.27)   a. *pace (yang) jalan itu*
             man   REL   walk that
          'the man who went'

        b. *pace (yang) de=jalan itu*
            man   REL   3S=walk that
          'the man who went'   (Donohue, 1999, 262)

Northern Italian dialects

In various Northern Italian dialects, such as Fiorentino, Trentino and Bolognese, subject markers are obligatorily attached to main clause verbs, both in isolation and in the presence of co-indexed NPs, as (4.28a) and (4.28b) show for Trentino:

(4.28) Trentino (Roberge, 1990, 86)

> *El   Mario *(el) magna.*
> the Mario he   eat-3SG
> 'Mario eats.

(4.29) *Le   putele   *(le)=ven.*
>   the girls   they=come.3PL
>   'The girls come.

In extraction contexts, the subject marker is ungrammatical in wh-questions in these dialects:

(4.30) Bolognese (Roberge, 1990)

> a. *Qui é   bel?*
>   who be.3SG beautiful
>   'Who is beautiful?

> b. **Qui l'=é   bel?*
>   who   he=be.3SG beautiful
>   'Who is beautiful?

(4.31) Trentino (Ouhalla, 1993, 481)

> a. *Quante   putele ha   parlá   con ti*
>   how.many girls   AUX.3SG spoken with you
>   'how many girls have spoken to you?'

> b. **Quante   putele le   ha   parlá   con ti*
>   how.many girls   they AUC.3SG spoken with you
>   'how many girls have spoken to you?'

However, it is possible (but not obligatory) to have a subject marker in the subject gap of a relative clause, as the following example from Bolognese demonstrates:

(4.32) Bolognese (Roberge, 1990)

> *la   ragazôla q'   la   zugheva*
> the litte.girl who she was.playing
> 'the little girl who was playing

Summary

The above survey reveals that, just like the Yucatec AF alternation, in a range of head-marking languages the choice between verbal alternants in unbounded dependencies is conditioned by the same types of factors that condition the choice between syntactic resumptives and gaps, supporting the view that that such alternations are all exemplars of a morphological subtype of RP/gap alternation.

	Syntactic RP/gaps	Morphological RP/gaps
Islands	√	√
Embedding	√	√
Construction type	√	√

Table 4.2: Conditioning factors for choice between alternants

4.4 Some generative perspectives

Subsets of anti-pronominal languages mentioned above have been discussed in the GB/Minimalist literature (see, i.a., Chung 1982; Chung and Georgopoulos 1988; Georgopoulos 1991; Baker 2008; Watanabe 1996; Schneider-Zioga 1996, 2000; Phillips 1995). Below, I review the two GB approaches, Chung's wh-agreement analysis, and Ouhalla's anti-agreement analysis, which probably represent the two most widely disseminated views of the phenomenon in this generative framework, and which have each generated a body of derivative literature on the subject. Both connect the special verb forms to gaps at some level of structure, and, in this respect, bear a similarity to the present account. However, they each (in distinct ways) adopt very different views of the role of the overt morphology in these verbal alternations from the one advocated in the present approach. It is therefore fruitful to evaluate their respective claims.

4.4.1 Wh-agreement

Chung (1982) argued originally for Chamorro, and later Palauan (Chung and Georgopoulos, 1988) that the special verb forms that occur in these languages in unbounded dependencies are cases of *wh-agreement*. Wh-agreement is a process by which the grammatical

relation (or abstract case), borne by a wh-trace is formally registered on a verb bearing some structural relation to that trace. In other words, the verb forms that show up in extraction contexts exhibit agreement with traces. This work was seminal in being among the first in the Principles and Parameters framework to call attention to the fact that inflectional morphology seems to be sensitive to something that forms a natural class in this framework: namely, A-bar-movement processes.

For Chung the infix *-um-* on the verb that I illustrated above in (4.4b), repeated below as (4.33b), is 'subject wh-agreement', which may be compared with the regular subject agreement of main clauses. In (4.34) below, *-in-* is 'object wh-agreement', which may be compared with the regular object agreement of main clauses.

(4.33) Chamorro (Chung 1998, adapted from 52 and 53b, 236)

        a. *Ha-fa'gasi si   Juan i   kareta*
           3SG-wash PN Juan the car
           'Juan washed the car'

        b. *Hayi f-um-a'gasi i   kareta*
           who   *um*-wash   the car
           'Who washed the car?'

(4.34)   a. *Ha-fahan si   Maria i   sanhilo'-ña gi   tenda*
          3SG-buy ART Maria ART blouse   OBL store
          'Maria bought her blouse at the store'

        b. *Hafa f-in-ahan-ña   si   Marai gi   tenda?*
           What buy-*in*-3SG.GEN ART Maria OBL store
           'What did Maria buy at the store?'

Although Chung is ostensibly aiming to explain the presence of morphologically distinct verbs in unbounded dependencies, on close inspection, her analysis in fact has only a very weak link to the actual morphology that she is trying to account for, which ultimately undermines its explanatory power. Dukes (2000) and Donohue and Maclachlan (1999) both offer thorough critiques of various aspects of Chung's wh-agreement analysis; I restrict the discussion below to a few general points which are particularly relevant for the case at hand.

First, unlike Chamorro, in a language like Palauan, as in a large number of head-marking languages, there is no exceptional overt morphology that could be analyzed as special overt wh-agreement. In such languages, as we have seen, the only defining feature of the verbal alternation is the *absence* of canonical subject person marking on the so-called 'wh-agreement' verb:

(4.35) Palauan (Georgopoulos, 1991)

a	*rengalek*	*a*	*rirell-ii*	*a*	*bresent*	*el*		*mo*	*er*	*a*	*sensei*
	children		made-3S		present	COMP		go	P		teacher

'It's the children who made a present for the teacher.'

On Chung and Georgopoulos' analysis the Palauan case is also treated as instance of wh-agreement, though there is nothing overt in the morphology to signal anything as such. This is presumably not problematic for Chung, because she makes ample use of 'covert' agreement morphemes; they occur frequently in her analysis of Chamorro as well. Chung points out, in many cases of extraction in Chamorro, the appearance of special overt morphology is optional or does not occur. The examples in (6) illustrate that the appearance of object wh-agreement is optional for some kinds of object extraction:

(4.36) Chamorro (Chung 1998, adapted from 72a and 72b, 242)

a. *Hafa   si   Maria   s-in-angane-nña   as   Joaquin?*
   what PN Maria in-say.to-AGR   OBL Joaquin
   'What did Maria tell Joaquin? '

b. *Hafa   si   Maria   ha-sangani   si   Joaquin?*
   what PN Maria AGR-say.to PN Joaquin
   'What did Maria tell Joaquin? '

In some cases it is prohibited. Subject 'wh-agreement' curiously does not occur when the dependent verb is irrealis (future). In (4.37), for example, the verb carries the third person *-u* of main irrealis clauses:

(4.37) *Hayi   para   u-sangani   yu'   ni   ansa?*
        who   FUT   3SG-say.to   me   OBL   answer
        'Who is going to tell me the answer?'

Because such dependencies, on Chung's analysis, nevertheless must be assumed to involve traces, Chung is forced to assume that such cases represent 'covert' wh-agreement, resulting from the 'morphological realization of a Chamorro-specific agreement rule' (1998, 239) and should be classed together with the cases that are overt. Dukes (2000) points out, however, that if overt and non-overt wh-agreement are parallel, then it is left unaccounted for why overt wh-agreement should block all other regular subject agreement, while non-overt wh-agreement never does. In other words, why does 'covert' wh-agreement get realized morphologically as canonical main clause agreement in Chamorro? Why does it manifest itself as an *absence* of canonical agreement in Palauan?

If the presence and nature of overt morphology is ultimately irrelevant for positing the existence of wh-agreement, then, taken to its logical conclusion, all languages, including English, can be argued to possess covert wh-agreement. This is not particularly explanatory. Rather than seeking an explanation for the Chamorro alternation by trying to formally unify the cases of overt morphology with the cases of 'covert' morphology, it seems that the explanandum should rather be to account for why the morphology differs in these different contexts.

Comparative Austronesianists have argued that the special morphology in Chamorro in fact has a straightforward historical explanation: it is the morphological residue of the old Philippine-type voice system, which is eroding in Chamorro main clauses (Dukes, 2000; Donohue and Maclachlan, 1999). I will return to this important point in some detail in chapter 5, when we explore the historical genesis of anti-pronomoninal alternations and their diachronic trajectories.

For now, it will suffice to observe that if overt morphology is irrelevant for the existence of wh-agreement, then we have no principled explanation for why it is in particular head-marking languages which exhibit verbal alternations in unbounded dependencies. On Chung's analysis, there is no logical connection between language's morphological type, and the possibility of exhibiting wh-agreement.

4.4.2 Anti-agreement

Drawing from data from Berber, two Northern Italian dialects, Turkish and Celtic, Ouhalla (1993) develops a binding theoretic account of the phenomenon, which he terms 'anti-agreement'. He argues that anti-agreement is a strategy used by 'Null Subject' languages

to avoid the licensing of a resumptive *pro* in the (closest) subject position. Anti-agreement, on this analysis, 'impoverishes' the rich agreement features of Null Subject languages sufficiently to prevent the licensing of a null *pro*. This reduces the problem to the more recognizable set of constraints operable over resumptive binding, in particular the Highest Subject Restriction (see McCloskey 1990 and Aoun and Li 1989). Schneider-Zioga (1995, 1996) also adopts a binding theoretic approach for anti-agreement in Kinande in particular.[3]

Though couched in GB terms, Ouhalla's original analysis of the anti-agreement phenomenon is similar in spirit to the basic analysis that I am advocating here. Both take the view that the phenomenon should be incorporated into resumption typology. Dependencies featuring 'anti-agreement' verb forms, the equivalent of AF verb forms in Mayan languages, are not concluded with resumptive pronouns. Where for Ouhalla it is the rich agreement of main clause verbs which licenses a 'null resumptive pro', on the present account, the person marking forms themselves contribute the pronominal information on main clause verbs, and which may, in long distance dependencies, function resumptively.

This is not a trivial difference, however. On Ouhalla's account, there is a necessary connection between a language being parametrized as a Null Subject language and the possibility of anti-agreement. In the GB framework, the concept of richness of a paradigmatic structure is crucial for the licensing of subject *pro* (Rizzi 1982, see Harbert 1995 for overview). Languages like Italian and Spanish are taken to have 'rich agreement', and therefore allow pro-drop (the licensing of *pro*). In the other direction, French and English are assumed to have impoverished agreement morphology, which cannot license *pro*.

The problem facing Ouhalla's analysis is part of a more general problem suffered by GB *pro* theory. As has been observed by many (see, i.a., Gilligan 1987; Cysouw 2003 and Siewierska 2004), it is difficult to find a precise definition in the literature of what exactly makes a personal agreement paradigm 'rich' enough to license a pro argument. This in turn, makes it difficult to establish precisely under what conditions Ouhalla's anti-agreement would be predicted to occur.

Let us consider some conceptions of *pro* in the GB tradition more closely. In *pro* theory, argument *pro* must be identified, that is, its ϕ feature content must be recoverable from its

---

[3]In recent years, binding theoretic approaches have given way to a variety of minimalist proposals for individual languages, including Baker (2008); Watanabe (1996); Schneider-Zioga (2000); Phillips (1995). Binding theoretic approaches have been abandoned mainly because, as Schneider-Zioga puts it, it isn't altogether clear what binding is under Minimalism).

syntactic surroundings. This is accomplished through co-indexation with a sufficiently rich AGR (Rizzi, 1986a; Borer, 1986). For example, in Rizzi's (1986a) theory of *pro*, identification is usually effected through binding by grammatical features on the licensing head. In order to be referential, an NP must be specified for person/number features. In the case of *pro*, therefore, it must be coindexed with features of person/number on its Case-governing head. Thus, the presence of rich agreement morphology is a necessary condition for this type of identification. As Jaeggli and Safir (1989, 32) describe it, in their discussion of identification: 'the most common notion of identification is by rich (or what may best be called 'strong') agreement, where inflectional affixes correspond to members of the conjugational paradigm'

As Rizzi himself notes, as have many subsequently, a language like Chinese is problematic for his proposal, since subject *pro* is allowed even though there is no agreement morphology to identify it. Some have tried to provide a unified analysis of these two language types by proposing that pro-drop occurs in languages either with verb rich agreement (e.g. Italian), or with no agreement (e.g. Chinese), but not in language with partial agreement paradigms. Jaeggli and Safir, for example, propose that it isn't richness of inflectional paradigms, but rather their morphological uniformity which plays a role in pro-theory.

However, the connection between morphological uniformity and the capacity to pro-drop is also problematic. As Cysouw (2003) observes, it is possible to conflate the reference between the three basic singular categories in agreement paradigms and still have pro-drop. While Spanish is often touted as a language with morphologically uniform paradigms, there does exist homophony in the Spanish inflection, in certain tenses. It is exemplified with the forms of the 'pretérito imperfecto' in (4.38), in which the third person and the first person are formally identical.

(4.38) *hablaba* I was speaking
        hablabas You were speaking
        hablaba He/she was speaking

Indeed, while standard Italian does not show any homophony in the singular, Spanish-type homophony is found in some regional Italic languages. An example is the 'imperfetto' inflection from Siciliano (Bigalke 1997, 60-61, reported in Cysouw 2003).

It is also possible to find languages as equally 'impoverished' as English, in which *pro* is licensed. One example is the Papuan language Wambon, which has an equivalent

degree of homophony in its singular paradigm as English (Cysouw, 2003). The lack of any intimate relation between agreement (and case) and pro-drop is also clearly exemplified in a detailed survey of the phenomenon in a range of South Asian languages by Butt (2001).

Thus, a uniform and one-to-one relationship between grammatical function and morphological inflection is not a necessary condition for the licensing of *pro*, forcing Cysouw, in his comprehensive survey of pronominal paradigms, to conclude that it is 'highly improbable that the structure of a pronominal paradigm correlates cross-linguistically with pro-drop characteristics' (2003, 53). Consequently, a more recent development in *pro* theory has been to disregard morphology and to treat agreement features as weak or strong according to the syntactic result required. As Corbett observes, ' 'weak' and 'strong' have for some become diacritics on nodes, specifying the syntactic outcome desired, rather than a reflection of morphological difference.' (2006, 99).

Given the tenuous relationship between morphology and the licensing of *pro* in GB, Ouhalla's null resumptive approach to the verbal alternations loses its footing, in a number of respects. First, if overt morphology is not reliably correlated with argument *pro* in GB, then it remains unclear why it is precisely the overt morphology that is affected in anti-agreement languages. Why, in that case, must overt inflection be 'impoverished', via anti-agreement, in order to prevent the licensing of a null resumptive, when there is no clear correlation between morphological impoverishment and (lack of) pro drop?[4]

Relatedly, if the licensing of argument *pro* is commensurate with the availability of pro-drop in a language, then why do languages such as Spanish or standard Italian, which allow subject pro-drop, not exhibit anti-agreement? Presumably, for Ouhalla, such languages also license a null subject *pro*, which must then illicitly function resumptively in unbound dependencies, without the intervention of anti-agreement.

---

[4]Indeed, even if we do, for argument's sake, assume that overt agreement features must be impoverished, in order to prevent resumptive *pro*, then one must still ask: what does it take for agreement features to become 'impoverished' on this analysis (that is, via anti-agreement). In some languages, as I have shown, anti-agreement is manifested as a complete lack of canonical person marking on the verb. In other languages it corresponds to suppletive morphological forms (Chamorro, Palauan).

4.5   The relevance of overt morphology

As the above overview revealed, in generative theories of anti-pronominal phenomena such as those developed in Chung (1998) and Ouhalla (1993), the connection between overt morphological inflection and the existence of anti-pronominal alternations is tenuous at best. As we have seen, Chung is forced to make extensive use of the notion of 'covert wh agreement', with the result that there is no logical connection between a language's morphological type, and the possibility of it exhibiting wh-agreement. On Ouhalla's theory, anti-agreement serves to 'impoverish' overt inflection in order to prevent the licensing of a 'null resumptive'. Yet because there is no cross-linguistic correlation between rich inflection and the licensing of *pro*, it ultimately remains unexplained why overt inflection should be affected in A-bar dependencies.

At its root, the problem for generative approaches based on abstract categories, covert morphology and parameter settings, is that the phenomenon is conditioned by none of these, but rather entirely by the presence of overt pronominal morphology on main clause verbs. By this I do not simply mean that languages which exhibit verbal alternations in A-bar dependencies necessarily are head-marking languages. While this is indeed the case, I intend the even stronger claim that special analytic verbs occur in unbounded dependencies precisely where the equivalent verb in a main clause context obligatorily carries an *overt* person marker. That is, the phenomenon is sensitive to the actual morphological alignment pattern of overt person markers. In this section I illustrate this important observation.

4.5.1   Argument role sensitivity

Schneider-Zioga observes in a footnote in a discussion of anti-agreement languages that while some anti-agreement languages, such as Palaun and Chamorro, have a distinct form of anti-agreement for subject extraction ad object extraction, "there are a number of anti-agreement languages that have anti-agreement for subject extraction only. At present we have no clear account of why there are languages of both types" Schneider-Zioga (2002, Fn. 13).

In fact, there are more types than the two Schneider-Zioga refers to. I showed in chapter 3 that the Mayan AF verb is not implicated in all subject extractions, but rather only when NP$_{rel}$ is the subject of a transitive clause. This asymmetry is in fact grounds for Stiebels

(2006), whose study is the only other one of which I am aware to consider the resemblence between Mayan AF and 'anti-agreement' languages, to dismiss the notion that there might be a fundamental commonality between Mayan AF and other anti-agreement languages, which (typically) exhibit the alternation for both A and S extraction.

When we inspect a wider range of languages, it emerges that the restriction to A-argument extraction is not, in fact, unique to Mayan languages. Coastal Salish languages, are just like Mayan, in only exhibiting an alternation in the case of the extraction of transitive subjects, as the following examples for Halkomelem and Squamish show. While the transitive verb in the relative clause alternates, the intransitive verb does not (these examples compare different languages, but the contrast is the same in the same in both):

(4.39) Halkomelem (Wiltschko, 2006, 9)

    a. *q'ó:y-t-es   te   Strang te   qwá:l*
       kill-trans-3ERG DER Strang DET mosquito
       'Strang killed the mosquito'

    b. *tl'ó   te   íle swíyeqe [q'óy-t   te   qwá:l]*
       3INDEP DET here man   kill-TRANS DET mosquito
       'This is the man who killed the mosquito'

(4.40) Squamish (Kroeber, 1999, 276)

    a. *na=ná-namʔ   ʔałi=s'álgm*
       AUX=RDP-go ART=monster
       'That monster is going about'

    b. *ʔałi=s'álgm   [na=ná-namʔ]*
       ART=monster AUX=RDP-go
       'That monster that is going about'

Indeed, one of the languages referred to in Stiebels' own discussion, and which is compared against the Mayan phenomenon, does not actually fit her own criterion for classification as a wh-agreement language: like Mayan languages, Chamorro, does not in fact alternate in the case of intransitive subject extraction:[5]

---

[5]This is easily confusable because Chung (1998) employs the gloss *Wh-Agr*, regardless of whether this corresponds to any overt morpheme.

(4.41) Chamorro (Chung, 1998, 103 and 238)

> a. *gaigi guini kada unu*
>   be   here   each   one
>   'Each one is here'

> b. *kuantu   na buteya gaigi gi   halum kahun ais*
>   how.many L bottle be   LOC inside box   ice
>   'How many bottles were in the refrigerator?'

As we saw above, Chamorro also exhibits a verbal alternation in the case of object extraction, thus, the alternations in Chamorro are sensitive to A and O extraction, but not S.

In Makassarese, we find yet another pattern. Alternate verb forms occur with the extraction of A, S and O. Extracted objects and intransitive subjects use a dependent verb form with no person suffix as in (4.42b), whereas the extraction of transitive subjects requires a dependent verb form with another general prefix (PREF) in place of the subject prefix, which does not carry person features; this is shown in (4.42c):

(4.42) Makassarese (Finer, 1997, 690)

> a. *na-pallu-i   berasak-a i   Ali*
>   3.ERG-cook-3.N rice-DEF   HUM Ali
>   'Ali cooks the rice'

> b. *berasak-a na-pallu-*(i)   i   Ali*
>   rice-DEF   3.ERG-cook-(3.N) HUM Ali
>   'Ali cooks the rice' ['It is the rice which Ali cooks']

> c. *i   Ali am-pallu-i   berasak-a*
>   HUM Ali PREF-cook-3.N rice-DEF
>   'Ali cooks the rice' ['It is Ali who cooks the rice']

Table (4.3) provides a summary of the different configurations of arguments which are targeted for verbal alternations in A-bar dependencies in 12 different head-marking languages.

As I showed in chapter 3, the restriction to transitive subjects in Mayan languages was correlated with the nature of the person marking paradigm on main clause verbs. In Mayan

A only	S & A	A & O	S, A & O	S & O	O only?
Mayan Salish	Berber Papuan Malay Kinande Irish Breton	Chamorro	Makaresse Yimas Tukang Besi	Selayerese	

Table 4.3: Argument roles triggering verbal alternations in 12 head-marking languages

languages, intransitive verbs, unlike transitive verbs, do not carry overt third person subject person markers. On the assumption that AF verbs are the gap counterparts of pronominally-inflected verb forms, we do not expect to find any alternating verb form precisely in those cases where there is no pronominal inflection on the canonical main clause verbs.

Extending this idea beyond Mayan, the more general typological prediction then, is that, for a given language, the anti-agreement alternation should be conditioned by the alignment of that language's morphologically dependent pronouns. That is, argument role sensitivity should reflect the particular alignment of overt pronominal inflection in a given language. In the next section I will demonstrate that this is indeed the case.

4.5.2   Morphological alignment patterns

In the following survey I will show how a language's morphological alignment pattern in main clauses is directly correlated with the environments in which verb alternations are triggered in unbounded dependencies.

Salish: A only

Pronominal paradigms in Coast Salish languages[6] exhibit a mix of phrasal clitic and suffix forms. The phrasal clitics (so-called indicative pronouns) function as subjects or objects and are enclitic to the clause predicate or to a preceding adverb or auxiliary elsewhere. Coast Salish languages use subject clitics in main clauses for both transitive and intransitive predicates. They are restricted to first and second person.

---

[6]These include Comox, Pentlatch, Sechelt, Squamish, Halkomelem, Nooksack, Straits, Twana and Lushootseed.

What is important to note is that in Coast Salish languages, the third person of intransitives is realized as zero. For third person of transitives, by contrast, in most coast Salish languages transitive verbs inflect for the 'ergative' subject suffix *-es*[7]. This is obligatory in main clauses. Thus, we find in Coastal Salish languages an alignment pattern in main clauses in which transitive subject pronouns are overt (suffixal), while S and O are phonologically null. The table below exemplifies the two sets of person markers:

	Indicative	Ergative
1	tsel	-
2	chexw	-
3	-	-es
1pl	tset	-
2pl	chap	-
3pl	-	-es

Table 4.4: Halkomelem Salish clitic and bound agreement paradigms

In extraction contexts, an alternation occurs just in case the NP*rel* is the subject of a transitive clause. In such cases the verb does not carry the inflectional subject of main clauses. In intransitive contexts, the subordinate verb does not change its form in any way.

(4.43) Comox, (Kroeber, 1999, 275)

    gat   kʷ=[taw-θi-uł]
    who   ART=tell-TR+2S.OBJ-PST
    'Who told you?'

(4.44) Comox, (Kroeber, 1999, 275)

    čəni   [ʔə=t'ut'θ-u-t-u-ł   tə=qigaθ]
    1s.INDEP CLEFT=shoot-LV-TR-PST ART=deer
    'It's me that shot the deer'

---

[7]The exception being Lushootseed, which marks third person subjects of all predicates, including transitive ones, by zero in main clauses (Kroeber, 1999).

It is interesting to note that many traditional descriptions of this phenomenon in Halkomelem are implicitly or explicitly accompanied by the functional explanation that the lack of ergative agreement in transitive-subject oriented relative clauses serves as a disambiguation device, distinguishing subject extracted relatives from object extracted relatives (see e.g. Suttles 2004, 76). We saw that the same functional explanation has been applied to the Mayan case, which exhibits exactly the same morphological alignment pattern for third person. In the context of the present discussion, this disambiguation effect can be understood as a side-effect, rather than the functional motivation for the emergence of such an alternation (otherwise, how do we explain the distinct alternation patterns in other anti-pronominal languages?) On the present morphologically-dependent RP/gap alternation account, the restriction of this alternation to transitive subjects follows, just as it does in Mayan, from the fact that third person intransitives, and third person objects are already phonologically zero.

Yimas: A and S/O

Yimas, a language of New Guinea, presents an interesting test-case, because it also exhibits an ergative alignment pattern of dependent person marking. Just as in Mayan and Coast Salish languages, in Yimas, third person S and O pattern alike, and distinctly from A. Yet in Yimas, not only A, but also S and O are realized with overt pronominal inflection in the third person. The following main clauses exhibit the contrast:

(4.45)   a. ***pu-n**-tay*
               3PL.ABS-3SG.ERG-see
             'He saw them'.

          b. ***pu**-wa-t*
             3PL.ABS-go-PERF
             'They went'.   (Foley, 1991)

We might expect then, that it may exhibit verbal alternations in all three extraction contexts (S, A and O). This is indeed the case. In the wh-extraction in (4.46a), where a subject is extracted from a transitive clause, the verb does not inflect for the ergative subject marker (its absence is represented below with ø). In the wh-extraction in (4.46b), in which a subject

is extracted from an intransitive clause, the verb also alternates, this time not inflecting for the intransitive subject pronominal of main clauses:[8]

(4.46) a. *Nawm [m-ø-kul-cpul-um]*
        who.pl NR-ø-2pl.ACC-hit-PL
        'Who hit you all?'

    b. *Nawn   [m-ø-na-ya-n]*
        who.SG NR-ø-DEF-come-Pres
        'Who is coming?'   (Foley, 1991)

Object extractions also trigger a verbal alternation; the dependent verb in object extractions does not carry object inflection:

(4.47) *krayn [m-ø-ka-tu-r-n*
     frog   NR-ø-1SG-ill-PERF-VI.SG
    'the frog which I killed'   (Foley, 1991)

Tukang Besi: A/S and O

The verb in a main clause in Tukang Besi is obligatorily prefixed to indicate the subject of the verb, and optionally (though usually), suffixed to index the object, if transitive:

(4.48) *No-'ita-'e   na   kene-no   te   ana*
     3R-see-3OBJ NOM friend-3POSS CORE child
    'The child saw its friend'   (Donohue, 1999, 51)

(4.49) *No-tinti na   ana*
     3R-run   NOM child
    'The child is running'   (Donohue, 1999, 51)

The following table provides the full paradigms for bound subject and object pronouns:

In extraction contexts, we see that conditions under which verbal alternations in unbounded dependencies in Tukang Besi occur reflect the language's pronominal alignment pattern. Extraction of S and A triggers a change in verb morphology. As shown in (5.39)

---

[8]NR in the Yimas glosses refers to a near distal deictic stem which functions like a complementizer, but is prefixed to the inflected verb of the relative clause.

	Realis subject	Object
1SG	ku-	-aku
2SG	'u- / nu-	-ko
3SG	no- / o-	-no
1PA	ko-	-kami
1PL	to-	-kita
2PL	i-	-komiu
3PL	no- /o-	-'e

Table 4.5: Tukang Besi subject and object bound pronominal forms (Donohue, 1999)

and (5.41), the verb in subject extracted relative clause bears no subject inflection, and takes the infix *-um-*.[9]

(4.50)   a. *No-balu=e'   na   lokate   wowine*
            3R-buy=3OBJ   NOM banana woman
            'The woman bought the bananas'

      b. *Te   emai na   b[um]alu te   loka*
            CORE   who   NOM buy.um   CORE banana
            'Who bought the bananas?'   (Donohue and Maclachlan, 1999, 129–130)

(4.51) *O-koruo   na   mia   [w[um]ila kua Lia] i   rearea   ai*
         3R-many NOM person go.UM   all Lia   OBL morning ANA
         'There were a lot of people who left for Lia this morning' (Donohue, 1999, 373)

In object extractions the verb does not bear object agreement, but instead inflects for the marker *ni-*:

(4.52) *Te   paira na   ni-basa nu   guro*
         CORE   what NOM NI-read GEN   teach
         'What did the teacher read?'   (Donohue and Maclachlan, 1999, 130)

---

[9]Tukang Besi *-um-* is cognate with the Chamorro *-um-* infix we saw above.

Berber: S and A

Berber possesses obligatory subject (S and A) marking on main verbs, which has been variously analyzed as pronominal clitics (Guerssel, 1995) or agreement affixes (Ouhalla, 1993, 2005).[10] There is no object pronominal inflection on the verb. Rather, object pronouns are realized as phrasal clitics.

The paradigm of subject person marking is illustrated in (4.5.2) with the perfective form of the verb *ls* 'to wear':

	Singular		Plural	
1	V-x	lsi-x	n-V	n-lsa
2M	t-V-t	t-lsi-t	t-V-m	t-lsa-m
2F	t-V-T	t-lsi-t	t-V-m	t-lsa-m-t
3M	i-V	i-lsa	V-n	lsa-n
3F	t-V	t-lsa	V-n-t	lsa-n-t

Table 4.6: Berber subject person marking paradigm (Ouhalla, 2005)

As I showed above, in subject extractions in Berber, the verb occurs in the participle form, which does not bear the obligatory subject marking of main clauses. This holds for both subjects of transitives (4.53a), and of intransitives (4.53b):

(4.53)   a. *tafruxt   ay   sqad-n   /*t-sqad*
            girl   COMP   send-PART   /3F-send
            'It was the girl who sent the letter'

      b. *u   ay   yxdln*
            who   COMP   arrived.PART
            'Who has arrived?'   (Ouhalla, 1993, 480)

As expected, object extraction does not trigger any change on the verb. The clitic pronoun cannot resume local dependencies, just as the subject inflected verb may not:

(4.54) *tamhdart   nni   (*ti)   zri-gh*
         student   COMP   her   saw-1S
         'The student I saw'

---

[10]Ouhalla's arguments against the pronominal clitic analysis are leveled at the morphological category of clitic, rather than the pronominal status of the forms.

4.6   Conclusion

In this chapter I have shown that the classification of the Mayan AF alternation as a morphological subtype of RP/gap alternation is directly supported by the fact that this subtype is cross-linguistically widespread: alternations equivalent to the AF alternation are exhibited in an impressive range of head-marking languages genetically unrelated to Mayan. I have shown that just like Mayan languages, such languages make use of special verb forms in dependent clauses in A-bar dependencies (wh-questions, relative clauses and clefts), which are distinguished (minimally) from their main clause counterparts by the absence of person marking inflection. Just as in Yucatec, the choice between the use of the synthetic verb form and the special 'analytic' verb form appears to be conditioned by the same factors that are implicated in RP/gap distributions cross-linguistically: sensitivity to islands, embedding and construction type, providing support for the notion that they are all exemplars of the same basic phenomenon.

   While the relationship between the Mayan AF phenomenon and other anti-pronominal languages has been dismissed in the literature on the basis of the fact that in Mayan the alternation is ostensibly unique in being restricted to A-extraction contexts (Stiebels, 2006), consideration of a larger cross-linguistic comparison set reveals the Mayan case to in fact fit neatly into the typology. We find that anti-agreement is conditioned entirely by surface morphology. Verbal alternations do not exist precisely in those contexts where main verbs do not carry overt pronominal inflection. Languages differ as to which grammatical functions (if any) are zero marked in the third person, yielding different configurations of apparent argument role sensitivities to extraction.

   This, of course, is not yet an explanation, it is merely a correlation. Why is the appearance of analytic verbs conditioned directly by the surface morphology of synthetic main clause verbs? In the next chapter I will show how the answer has its roots in diachronic change: synthetic verbs are simply the historical continuants of analytic verbs. It therefore couldn't be any other way.

Chapter 5

Frequency-driven grammaticalization

In descriptions of anti-pronominal alternations one often comes across statements that imply that the special dependent verb forms that occur in A-bar dependencies in head-marking languages are derived in some manner from their main clause counterparts. Topping (1973), for example, suggests that the special Chamorro forms are derived from the morphologically less marked verb type. Wiltschko's paper on Straights Salish is concerned with answering the question 'as to why ergative agreement is lost in subject centered relative clauses' (Wiltschko, 2006, 3). Elouazizi (2005, 120) describes for Berber that 'the canonical subject agreement inflectional morphology that is found in a declarative matrix clause undergoes a process of *radical agreement suppression*' in extraction contexts (original emphasis). Schneider-Zioga describes how in Palauan, 'subject extraction requires suppression of subject/verb agreement' (2002, 9).

Such statements may provide an adequate description of a synchronic system. They do, however, obfuscate the historical relationship between these verbal alternants, which I argue, is in fact defined by the inverse directionality. The thesis I develop in this chapter is that the special analytic dependent verb forms we find surfacing across head-marking languages in extraction contexts are conservative verb forms that have been retained in subordinate extraction contexts, after person marking grammaticalized to verbs in other contexts, above all in main clauses. From this perspective, it is the synthetic (= pronominally inflected) verbs that are the diachronic innovation in systems that originally only had analytic (= pronominal-less) verb forms. The reason for the exemption of dependent verbs from the grammaticalization process of person marking is a principled one: it relates to the

skewed frequency distributions of pronouns in different environments (above all in main vs. dependent clauses). Thus, the explanation I forward in this chapter is different from other analyses of antiagreement/wh agreement/Agent Focus in that it is explicitly proposed as a grammar-external explanation. My claim is that we need to look at patterns of language use, more precisely frequency distributions in language use, to be able to explain the formal properties and distribution of analytic verbs in A-bar dependencies.

Specifically this approach will provide a principled reason for why the alternations should be conditioned by the overt pronominal morphology of main clause verbs. As I showed in chapter 4, special analytic verbs only ever occur in unbounded dependencies precisely where the equivalent verb in a main clause context obligatorily bears overt person marking. This relation follows quite naturally on the hypothesis that pronominally inflected main clause verbs are the historical continuants of their analytic counterparts. In this chapter I will also show that the apparent heterogeneity of formal expression of these verb forms, and the 'exceptional morphology' that they often exhibit, derives from the historical processes which have shaped these alternations over time, through frequency sensitive grammaticalization processes.

5.1 The frequency-sensitivity of grammaticalization

The mechanisms by which linguistic structures emerge over time have been a focal point of study over the last several decades (e.g. Givón 1979; Lehmann 2002; Bybee et al. 1994; Heine and Kuteva 2002; Hopper and Traugott 1993). It has been observed for languages with a long documented history that lexical items within constructions can become grammatical items, and loosely organized elements in sentences can become more tightly fused. 'Grammaticalization' is the name given to this process. It is assumed to be the result of repetition across numerous speech events, during which strings of elements become automized as neuromotor routines (Bybee, 2002b; Haiman, 1994).

The beginning of the grammaticalization process is often characterized by semantic changes. Here we are concerned not with semantic change, but with the morphosyntactic side of grammaticalization, in which co-occurring elements fuse together and become phonologically attenuated. An example of such a process may be found in English present tense auxiliaries, which in speech tend to cliticize and attach to a host:

(5.1) I am > I'm

     I want to > I wanna

It is widely assumed that the more frequently two elements co-occur, the more likely it is that grammaticalization will take place (e.g. Traugott and Heine 1991; Krug 1998; Bybee and Scheibman 1999; Haspelmath 2004; Bybee 2002a). Bybee (2002a) refers to this as the "Linear Fusion Hypothesis". Haspelmath (2004) refers to it as the "Frequency Condition on Grammaticalization".

Haspelmath (2004) observes that frequency is relevant for grammaticalization in at least two different ways. First, the higher the frequency of use of a string of linguistic expressions, the more probably it is that a larger unit comprising these expressions will be formed by speakers. This can be understood as a special instance of the more general connection between memory storage and frequency of exposure. Second, higher frequency of use in turn results in greater predictability, which allows speakers to expend less articulatory effort. Thus, frequent expressions also reflect a greater tendency to be phonologically reduced by comparison with less frequent expressions (Zipf, 1935). In this way, the two main aspects of morphosyntactic grammaticalization, fused structures and formal attenuation, can be tied to frequency.

5.1.1   The grammaticalization of person marking

The change from an independent pronoun to a bound person marker that is phonologically and morphologically reduced is a signature example of morphosyntactic grammaticalization (Givón 1976; Lehmann 2002). As I described in chapter 3, the formal grammaticalization process is generally assumed to proceed as follows (cf. e.g. Lehmann 1988, 2002; Corbett 1995; Hopper and Traugott 1993; Siewierska 2004 and Fuß 2005). Over time, a formerly independent pronoun becomes a phonologically weak clitic and attaches to an adjacent element (typically the finite verb or auxiliary). The pronominal element then continues to erode, and is reanalyzed as an obligatory part of the verbal inflection.

(5.2) strong pronoun > weak pronoun > clitic > affix

As I illustrated in chapter 3, such a change has been well-documented for various Indo-European languages, and in some cases, the changes are still in process, or have taken

place during the recorded history of a language (e.g. non-standard French, Lambrecht 1981; Auger 1993, 1994; Gerlach 2002). The change can also be inferred for many further languages around the world, where dependent person markers exist which are obviously derived from independent pronouns still extant in other contexts in the language. Indeed, it is generally assumed that the great majority of bound pronouns in the world's languages originate from independent pronouns.

The effects of frequency are detectable in the grammaticalization of person marking in many ways. Frequency of usage of pronouns is known to vary hugely, depending on factors such as grammatical function and person category. For example, there is a documented tendency for languages to pronominalize subjects more than objects (Siewierska, 2004) and objects more than obliques. This in turn correlates with a greater tendency across languages for subject pronouns to develop into morphologically dependent markers over other grammatical functions:

(5.3) subject > object1 > object2 > oblique[1]

Siewierska (2004, 43) shows in a comprehensive survey that dependent person markers are much more common with arguments than with adjuncts, and among the arguments they are more common with subjects than with objects:

Dependent pro	Subject N=402	Object1 N=402	Object2 N=375	Oblique N=332
No. lgs	330	247	55	20
%	82	67	15	6

Table 5.1: Dependent pronominals and argument prominence

This asymmetry is observable not only with respect to cross-linguistic frequency, but also *within* languages: Siewierska shows that the availability of dependent person markers for a syntactic function lower on the argument prominence hierarchy entails the availability of dependent person markers for syntactic functions higher on the argument prominence hierarchy.

---

[1]Siewierska explains the labels as follows: 'Subject corresponds to A, object1 corresponds to P and to the argument of a ditransitive clause (either patient or recipient) that has the same person marking as that of the P in monotransitive clauses, object2 corresponds to the other ditransitive object and oblique corresponds to any argument associated with a specific semantic role which is not realized by the subject or object functions'.

It is also possible to detect asymmetries along the dimension of person. This presumably can also be traced to frequency: first and second persons are more frequently referred to than third person in discourse. Siewierska provides the following relative frequencies of subject, object and possessive singular for first and third person from a corpus of spoken English (only first person is given due to difficulty of distinguishing between different functions of *you*):

Person	Frequency
I	38 000
he	11 000
she	8 000
me	4 000
him	2 000
her	1 000
my	2 500
his	1 500
her	1 000

Table 5.2: Frequency of first- and third-person pronouns in conversational English, per million words, based on Biber et al's (1999) 40-million-word corpus (from Siewierska 2004, 267)

As the table shows, in conversation, first person subjects are nearly twice as frequent as third person subjects (*he* and *she*).

This tendency also appears to be mirrored across languages in terms of asymmetric patterns of grammaticalization of person markers, according to person type. It is common for languages to have dependent person marking with first and second person, but not with third person subjects (cf. e.g. Bybee 1985; Mithun 1991; Cysouw 2003; Siewierska 2004). Moreover, in languages which do possess both 1st and 2nd dependent person markers, as well as third person, it can often be shown that 1st and 2nd person markers grammaticalized prior to 3rd person forms (cf. Mithun 1991 for an overview of this in various native American languages).[2]

---

[2]Ariel (1998, 2000) explains the grammaticalization of first and second person pronouns to be driven by the tendency of speakers to use reduced markers to refer to highly accessible/salient discourse referents. This process is understood to apply particularly to first and second person subjects since these are always highly accessible as discourse participants.I suspect that 'accessibility' can probably be reduced to frequency, though I don't discuss this further here.

Obviously, third person pronouns do grammaticalize in many languages. Givón (1976) has proposed one scenario by which certain syntactic/discourse conditions can give rise to a heightened frequency of third person pronoun use. He argues that the dependent person marker can originate as an independent anaphoric person marker in topic shifted constructions (left- or right-dislocated) such as:

(5.4) Sally, she came early
       She came early, Sally

Givón proposes that through overuse, these topicalized constructions become reanalyzed as 'neutral syntax'. The dislocated topic becomes the subject, and the anaphoric person marker fuses to the verb. This scenario (so-called NP-detachment) has been applied to the non-standard French case, where examples like (5.5) preferentially receive a basic, non-dislocated interpretation, though transparently betraying their origin as topicalized left-dislocation constructions (they are surface identical to such constructions in standard French):

(5.5) *(Moi) je   porte   la   table*
       me   CLIT.1SG   carry   the   table
       I carry the table.
       Not: As for me, I carry the table. (Fuß, 2005)

It is also possible to detect frequency effects in the grammaticalization of *combinations* of pronouns. Haspelmath (2004) shows how frequency can account for the grammaticalization of certain combinations of pronouns in ditransitive constructions, leaving paradigmatic gaps. The so-called Ditransitive Person-Role Constraint can be formulated as follows: "Combinations of bound pronouns with the roles Recipient and Theme are disfavored if the Theme pronoun is first or second person and the Recipient pronoun is third person." This constraint is well-known from Spanish (5.6) and other Romance languages, and has, according to Haspelmath, been reported to hold either categorically, or probabilistically in over twenty other languages.

(5.6)   a. *Alicia **se lo** mostrará.*
             'Alicia will show it to her.' (3rd theme, 3rd recipient)

b. **Alicia **te le** mostrará*

'Alicia will show you to her.' (2nd theme, 3rd recipient).

Haspelmath provides corpus data to show that combinations like 'it/him to me' are cross-linguistically more frequent than combinations like 'me to it/him'. The reason for the frequency asymmetry has a straightforward semantic/pragmatic basis: the Recipient of a ditransitive construction is virtually always animate, and the Theme shows a strong tendency to be inanimate. Since first and second person pronouns are always animate but third person pronouns may be inanimate, it follows that the Theme NP will most often be third person, whereas the Recipient NP may be first, second or third person.

5.1.2   Frequency sensitivity and anti-pronominal alternations

Frequency driven grammaticalization can give rise to dependent person marking on verbs where there is a high frequency of pronoun-verb adjacency. We saw above that there are numerous ways in which asymmetric frequency distributions of pronoun-verb (and pronoun-pronoun) combinations can result in asymmetrical grammaticalization patterns of dependent person marking across languages (subjects but not objects, speech act participants but not third persons, etc). Though it is not often explicitly discussed in the literature, an important generalization that emerges when considering these proposed grammaticalization paths is that they involve processes that predict the rise of (various configurations of) dependent pronominal marking in *main clauses*. I submit that it is precisely in the case of dependent verbs in certain types of extraction contexts that there is unlikely to be a high frequency of co-occurrence of verbs and pronominal arguments and that this is the basis for the emergence of anti-pronominal alternations.

As we have seen, the typological record reveals that gaps are preferred/required over resumptive pronouns in certain positions. Gaps are more frequent, within and across languages, in direct argument positions (especially subject position, Keenan and Comrie 1977), and especially in locally extracted environments. Thus, while the appropriate conditions might arise for the grammaticalization of third person pronominal inflection on verbs in some environments (above all in main clauses, or in resumptive contexts), if a language frequently or obligatorily concludes certain types of A-bar dependencies with gaps, then dependent verbs in these contexts may be excluded from this head-marking process.

This is because the conditioning environment does not exist in these contexts: the lack of frequently adjacent pronoun-verb combinations.

As such, frequency-driven grammaticalization can give rise to a morphological distinction between main and dependent verbs. On this view, analytic verb forms found in various head-marking languages in various types of A-bar dependencies represent conservative verb forms which did not undergo the grammaticalization process of pronominal affixation. In this way, we can understand analytic verbs historically as the direct by-product of the grammaticalization of person marking affixes in other contexts (above all, main clauses). Or in other words, synthetic forms are the historical continuants of analytic forms. This series of developments is schematized in table (5.3).

	Main verb affected	Dependent verb unaffected
Stage 1 Independent pronoun	he ran	the man that __ ran
Stage 2 > Clitic	he=ran	the man that __ ran
Stage 3 > Affix	he-ran	the man that __ ran

Table 5.3: Representation of morphological differentiation between main and dependent verbs over time, via grammaticalization of person marking in main clauses

In the remainder of this chapter I will present several specific case studies which provide evidence for this basic diachronic trajectory. The first examines a cluster of Romance dialects in which subject person clitics are in the process of developing into affixes. This study allows us to see the very beginnings of the emergence of verbal alternations in nascent head-marking languages. The second study looks at the development of the Yucatec Mayan AF alternation in the imperfective, where the rise of imperfective aspect marking in main clauses has led to the increased morphological differentiation of main and AF verb forms in the modern language.

The subsequent studies all deal with cases where developments in main clauses have resulted in the appearance of apparently 'exceptional' morphology on dependent verbs in extraction contexts. The third study focuses on Western Malayo-Polynesian, where so-called 'wh-agreement' morphology has been shown to be the residue of the old Philippine type voice system which has eroded in main clauses in Chamorro and Tukang Besi through

the development of head marking in main clauses. The fourth study deals with a similar phenomenon in Central Mayan languages, where the special 'AF suffix' on dependent verbs in extraction contexts is identifiable as an old Proto Mayan suffix found on finite main verbs which has been lost in many modern languages in main clauses. I conclude with a lengthier study of the perfective AF alternation in Yucatec, whose evolution incorporates a number of the processes described in the previous studies.

5.2 Romance: grammaticalization in progress

Various French and Northern Italian dialects exhibit a change in progress whereby subject pronominal clitics are grammaticalizing into verbal affixes. Inspecting this change in progress provides a window into the first stages of morphological differentiation between main and dependent verbs with respect to the grammaticalization of person marking. We will see that while, in a number of respects, there is evidence in these languages of increasing fusion between the subject person marker and the verb in main clauses, significantly, dependent verbs in subject extraction contexts remain largely exempt from this process.

5.2.1 From clitics to affixes in main clauses

There has been much debate over the status of pronominal clitics in French (and the other Romance languages) for over a century (see i.a., Schuchardt 1885; Lambrecht 1981; Haiman 1991; Auger 1993, 1994; Friedemann 1997; Gerlach 2002; Fuß 2005; De Cat 2005; Roberge 1990; Miller 1992; Miller and Sag 1997; Monachesi 1999; Morin 1981; Cournane 2007). At the heart of this debate is that fact that these elements are undergoing a process of change, and so exhibit mixed properties characteristic of a change in progress. In the present discussion my aim is not to offer any conclusion about the categorical status of these markers. Rather, my point is to show that even those dialects whose subject person markers exhibit the most advanced stages of grammaticalization towards affix status in main clauses, nevertheless still frequently or obligatorily omit this affixal element on dependent verbs in extraction contexts.

I will use criteria for distinguishing affixes from postlexical clitics due to Zwicky and Pullum (1983), as well as on additional criteria based on Miller and Sag (1997) and Miller

(1992) to chart the progression of the morphologization of subject person markers in various French dialects (Standard French, Pied Noir French, Quebec French) and Northern Italian dialects (Trentino, Bolognese).

Syntactic visibility

One of Zwicky and Pullum's diagnostics for distinguishing between affix and clitic status concerns the syntactic visibility of the element:

(5.7) Syntactic rules can affect affixed words, but cannot affect clitic groups

In Standard French, most subject person markers are sensitive to syntactic context: they must precede the verb in declaratives, but must follow it in matrix questions with wh-ex-situ:

(5.8) Standard French (Fuß, 2005)

      a. *Tu   attends qui?*
         you wait   who
         'Who are you waiting for?'

      b. *Qui   attends-tu?*
         who   wait-you
         'Who are you waiting for?'

Evidence of a change in progress for at least one member of the paradigm is evidenced in standard French by the fact that for most verbs there is no acceptable form for the inverted first person singular pronominal form *je*, as opposed to other persons.

(5.9) Standard French (Miller and Sag, 1997)

      a. **Sors-je*
         going.out-I
         Intended: 'Am I going out?'

      a. **Chante-je*
         Sing-I
         Intended: 'Am I singing?'

In various of non-Standard French dialects this change has proceeded further, such that no subject person markers in the paradigm undergo subject-verb inversion in matrix questions. Instead, alternative yes-no strategies are used in which the person-marker and verb remain uninverted:

(5.10) Non-standard French (Lambrecht, 1981, 6)

> *Où   tu   vas?*
> where   you   go
> 'Where are you going?

Cournane (2007) states that in Quebec French inversion is similarly not possible with any subject person marker. Instead, yes-no question strategies, such as the use of question intonation with the declarative structure (5.11a), or the *est-ce-que* form (5.11b), leave the subject person marker and the verb in the same position:

(5.11) Quebec French (Cournane, 2007)

> a. *Je vas   chanter   un chanson?*
>   I   will-1SG sing-INF   a   song
>   'I'm going to sing a song?
>
> b. *Est-ce que je vas   chanter   un chanson?*
>   est-ce   que I   will-1SG sing-INF   a   song
>   'Am I going to sing a song?

Coordination

Miller (1992) argues that affixes cannot take wide scope over a coordination of hosts, and thus that coordination facts can help to determine the lexically-attached status of an element. In Standard French, subject person markers do not have to be repeated on each verb, as shown by the optionality of the subject person marker *elle* in (5.12):

(5.12) *Elle chante et   (elle) danse*
       she   sings   and   she   dances
       'She sings and dances.'

Wide scope over coordination is apparently abnormal in colloquial French: *% Je mange du pain et/pis bois du vin*. 'I eat bread and drink wine' (Miller, 1992), indicating once again a shift towards affixal status in Standard French.

In Quebec French this has become fixed: subject markers *must* be repeated on all coordinated verbs, suggesting affix status (also Pied Noir French, Cournane 2007):

(5.13) Quebec French (Cournane, 2007)

> *[A] chante pis   *(a) danse*
> she sings   and she   dances
> 'She sings and dances.'

Northern Italian dialects exhibit the same constraint, as shown here for Trentino:

(5.14) Trentino (Rizzi, 1986b)

> *La   canta   e   *(la) balla*
> she sings and she   dances
> 'She sings and dances.'

Obligatoriness of the element: clitic doubling in main clauses

In Standard French, subject person markers are not possible on the verb in main clauses in the presence of a DP: the two are in complementary distribution. (5.15a) shows that where there is a DP subject, a subject person marker does not occur. (5.15b) shows the presence of the subject person marker in the absence of the DP subject. (5.15c) shows the ungrammaticality of both the DP subject and the subject person marker co-occuring.

(5.15) Standard French (Cournane, 2007, 27)

> a. *Mon chat m'aime.*
>   my   cat   me.loves
>   'My cat loves me.'
>
> b. *Il   m'aime.*
>   he   me.loves
>   'He loves me.

    c. **Mon chat il   m'aime*
       my   cat   he me.loves
       'My cat he loves me (w/ normal sentence intonation. Ok as left-dislocated
       topic)

In non-standard French varieties, such as Quebec French (Cournane, 2007) and Pied Noir
French (Roberge, 1990) it is possible for the DP subject and the clitic to co-occur (so-called
'clitic doubling'). This is considered an optional operation (i.e. the clitic and/or double may
be omitted for the same semantic reading), as show for Quebec French in (5.16)

(5.16) Quebec French, (Cournane, 2007)

    a. *La   fille (a)   parle trop*
       The   girl 3SG.F talks   too.much
       'The girl talks too much.'

    b. *(La fille) a   parle trop*
       The   girl   3SG.F talks   too.much
       'The girl talks too much.'

In Trentino and other Northern Italian dialects subject markers are obligatory both in iso-
lation *and* in doubled constructions (Roberge, 1990, 86). In other words, person marking
has become obligatory on main verbs.

(5.17) Trentino (Roberge, 1990, 86)

    El   Mario *(el) magna
    the   Mario he   eat.3SG
    'Mario eats.'

(5.18) *Le   putele *(le) ven*
    the   girls   they   come.3PL
    'The girls come.'

Arbitrary gaps in paradigms

Arbitrary gaps in the set of combinations of pronominal affixes and verbs is typical of
inflection and not of cliticization (Zwicky and Pullum, 1983). Leveling of person, number

or gender distinctions is evident in many French and Italian dialects. Quebec French has reduced the Standard French paradigm for 3rd person plural, for which gender is no longer distinguished. Instead, the reflex of the masculine, 3rd plural marker *i* is used for both:

(5.19) Quebec French (Cournane, 2007)

      a. *Les hommes i   boivent*
          The men   3PL drink
          'The men drink.'

      b. *Les femmes i   boivent*
          The men   3PL drink
          'The men drink.'

In Bolognese, the same leveling of gender has occurred in the third person plural. In some other Northern Italian dialects, here shown by Genoese, this leveling of the subject marker forms has occurred even with singular forms, where an erstwhile masculine subject marker doubles a feminine DP:

(5.20) Zeneize (Genoese) (Haiman and Benincà, 1992, 193)

     U vene   a   Katajning
     he comes the Catherine
     'Catherine is coming.'

Summary

In sum, while subject markers in Standard French (with the possible exception of the first person singular, which exhibits the behavior of an affix) still appear to retain some of the properties and behavior of clitics, according to various diagnostics, the subject person markers of other French varieties exhibit a more advanced degree of grammaticalization towards affixal status. The subject person markers of Northern Italian dialects also behave more like affixes than clitics.

Property	SF	Quebec French	Pied Noir	Northern Italian
Syntactically invisible	(no)	yes	yes	yes
Occurs in VP coordinations	(no)	yes	yes	yes
Doubles clause internal DPs	no	optional	optional	yes
Paradigm leveling	no	yes	yes	yes

Table 5.4: Subject person marker properties in French and Northern Italian varieties

5.2.2   Impact on the subordinate domain

In Standard French, subject person markers like *elle* cannot generally appear in the subject gap of a relative clause, or a subject wh-question:

(5.21) *Ma fille qui   (*elle) parle espagnol.*
     my girl who she   speak Spanish
    'My daughter who speaks Spanish.'

When we examine person marking in A-bar dependencies in the dialect varieties discussed above, we find that even in those that were shown above to exhibit the most advance stages of formal grammaticalization towards affixal status in main clauses, nevertheless do not require, or, in some cases, do not permit, subject person marking on dependent verbs in subject extraction contexts, just as in Standard French.

Wh-extraction contexts are the least hospitable environments for subject person marking on dependent verbs. This is in keeping with the cross-linguistic tendency for gaps to be preferred over resumptive pronouns in wh-extractions (Boeckx, 2003). Thus we find that in Quebec French, dependent verbs occur without subject person marking in wh-extractions (5.22a), and are ungrammatical with such marking (5.22b):

(5.22) Quebec French (Roberge, 1990, 118)

    a. *Qui   sont tes   amis?*
       Who are   your friends
       'Who are your friends?'

    b. **Qui il   est allé?*
       who   he   is   gone
       'Who went?'

As in Quebec French, subject markers cannot appear in the subject gap of a wh-extraction in Pied Noir French, as shown in (5.23). Nor can they appear in Northern Italian dialects like Bolognese (5.24).

(5.23) Pied Noir French (Roberge, 1990, 120)

> **Qui il   vient?*
> who   he   come
> 'Who comes?'

(5.24) Bolognese

> **qui l'é   bel?*
> who   he.be beautiful
> 'Who is beautiful?'

One Northern Italian dialect, Moena, is reported to allow (but not require) a subject marker to appear on the verb in subject wh-extractions

(5.25) Moena (Haiman and Benincà, 1992, 193)

> *Chi e   lo po   i   ozitegn?*
> who   is he then the Occitans-MASC.PL
> 'Who are the Occitans?'

Subject relative clauses exhibit more variability, but here too, dependent verbs frequently occur without the subject person marking of main clauses. In one variety of Quebec French reported in Cournane (2007), subject person marking is illicit on dependent verbs in subject-extracted relative clauses:

(5.26) Quebec French (Cournane, 2007)

> **une fille qu-a   pleurait*
> a   girl who-she was.crying
> 'a girl who was crying'

The dialect of Quebec French reported in Auger (1994) and Bouchard (1982) by contrast, optionally allows a resumptive person marker:

(5.27) Quebec French (Bouchard, 1982, 104)

> *une fille qu(-a) pleurait*
> a   girl who-she was.crying
> 'a girl who was crying'

Similarly, in Pied Noir French it is optional to have a subject marker in the gap of a relative clause, as shown in (5.28):

(5.28) Pied Noir French (Roberge, 1990)

> *C'est   une femme qu'(-elle) était tres malheureuse avec son mari.*
> it.be-3S a   woman who-she   was very unhappy   with her husband
> 'She's a woman who was very unhappy with her husband.'

This is also true of Bolognese:

(5.29) *la   ragazôla q'-(la)   zugheva*
       the litte.girl who-she was.playing
       'the little girl who was playing'

5.2.3 Summary

Examining the grammaticalization of subject person markers in non-standard French and Northern Italian varieties reveals that even in cases where person markers have become obligatory or near-obligatory affixes on main clause verbs, this is not necessarily matched by obligatory affixation on dependent verbs in subject-extraction contexts. Such contrasts thus provide a window into the genesis of anti-pronominal alternations. They emerge as the result of the morphologization of person marking on main clause verbs.

	St. French	Q. French	Q. French 2	Pied Noir	N. Italian
Subject RCs	no	no	optional	optional	optional
Subject wh questions	no	no	no	no	optional

Table 5.5: Subject person marker properties in Romance varieties

5.3 From Colonial to Modern Yucatec

The colonial record in Yucatec provides evidence of a next possible stage in the morphological differentiation between verbal alternants in anti-pronominal languages. In this section I show how the grammaticalization of aspectual particles on main clause verbs has created a further morphological contrast between verb forms in Yucatec imperfective contexts.

5.3.1 Grammaticalization of tense/aspect marking in main clauses

It has been observed in the grammaticalization literature that innovative periphrastic tense/aspect constructions often grammaticalize into morphological forms first (or only) in main clauses, leading to formal (and often then semantic) differentiation between main and subordinate verbs. An example of this is the emergence of the Spanish imperfective subjunctive. Klein-Andreu (1991) discusses the change from the Latin Pluperfect Indicative into the Spanish Imperfect Subjunctive in *-ra*, which she attributes to the emergence of periphrastic perfects in Spanish (composed of the auxiliaries *haber* and *ser* plus the Passive Perfect Participle). The periphrastic construction gradually developed into the modern Present Perfect and Pluperfect Indicative, replacing the original Latin Pluperfect in *-ra* in past anterior functions in main clauses. Now the *-ra* forms remain only in certain types of subordinate clauses, where they have taken on past subjunctive meanings and functions.

Bybee refers to a similar process which took place in Arabic, where a new present progressive generalizes to become a present, while the old imperfect that is being replaced loses its main clause functions and is gradually restricted to subordinate clauses. A similar case of erstwhile presents getting trapped in subordinating constructions is also attested in changes from Classical Armenian to Modern Armenian (Bybee, 2002a).

In the Yucatec study below, we will see a similar process, whereby an innovative imperfective aspect marker has become obligatory in main clauses, but has not extended to dependent verbs in A-bar dependencies.

5.3.2   The Yucatec imperfective

In modern Yucatec, as I have shown, the AF verb in the imperfective aspect is differentiated from its main clause counterpart not just by the absence of subject pronominal marking, but also by the absence of the preverbal aspect marker *k-*.

(5.30)   a. *Le   máak **k-u=xok-ik**   le   libro=o'*
             DET   man   IMPF-A3=read-INC   DET   book=D2
             'The man reads the book

   b. *Máax   t-aw=al-ah   **xok-ik***
       who   PRV-A2=say-CMP   read-INC
       'Who did you say reads it?

Inspection of the colonial record reveals that the preverbal imperfective marker *k-* in main clauses is a relatively recent innovation. In early Colonial Yucuatec, the imperfective in main clauses was realized by the incompletive status suffix (*-ik*; 'ic' in the sources) alone. Thus the imperfective verbs *a ual-ic* 'you say' and *a uocl-ic* 'you steal' in (5.31) are inflected only for the incompletive suffix -ic, but do not bear the preverbal imperfective aspect marker k-, which is obligatory on all modern imperfective Yucatec verbs (cf. *k-aw-al-ik*, 'you say'):

(5.31) *tumen   **a ual-ic**   bal   tie   uacix   tumen   **a uocl-ic**   bal   tie*
          because A2 say-INC thing PREP or   because A2 steal-INC thing PREP
       u   bal   u   bae
          A3   thing   A3   self
          because you say something to him, or because you steal something from his
          belongings   (Anon. 1576/1620)[3]

At this stage in the language, imperfective verbs realized in subject extracted dependencies differed from imperfective main clause verbs only in the absence of the subject pronominal which was obligatory in main clauses, as it is in modern Yucatec.

---

[3]Transcribed and annotated by Whalen (2003).

(5.32) *Bay huntul-il Dios, huntul-il pixan* **cuxcinn-ic**   *u   cucut-il uinic-e*
     as   one-of   God, one-of   soul   gives.life-INC A3 body-of man-DEIC
     As there is only one God, there is only one soul that gives life to the body of man.
     (Anon. 1576/1620)[4]

Thus we find in Colonial Yucatec a situation which is not unlike the Romance dialects described above: Subject pronouns were expressed by clitic/affixal elements in main clauses. Highest subject relativization involved a gapped subject. The verbs themselves, in main and subordinate clauses in imperfective contexts, were not otherwise differentiated morphologically, save for the presence/absence of person marking.

Over the next two centuries we begin to find the optional appearance of the preverbal aspect particle *ki-* in main clauses. Beltran, writing in 1746, mentions the particle *ki*, stating that it is merely used for 'ornament or for greater signification' and denotes present time (Beltrán, 1746).[5] He notes that it combines with the subject pronoun to become *k-in, k-a, k-u,* etc.

By the early twentieth century, *k-* is not obligatorily, but appears more often than not. Tozzer (1921, 43)[6] writes that 'when *k-* or *ki-* is used the idea may in some cases be translated by the Spanish term *a veces* 'sometimes''. However, even in this period, Tozzer gives examples of imperfectives where the aspect particle *k-* does not appear:

(5.33) *U   puts-ik   Pedro winik*
     A3   hit-INC   Pedro man
    'The man is hitting Peter'

Tozzer (1921, 43) writes of the *ki* particle "These are prefixed to the forms of the pronoun and are sometimes so closely joined to the pronoun by phonetic changes that it is difficult to separate them from the form of the pronoun. In general, it can be said that the uncontracted forms are most common among the Mayas as well as among the Lacandones. The uncontracted forms seem to have been earlier than those where contraction has resulted. Among the Mayas near Valladolid the contracted forms were used almost exclusively"

---

[4]Transcribed and annotated by Whalen (2003).

[5]Andrade (1955, iv) assumes that the period of Modern Yucatec starts around the first half of the nineteenth century, and that Beltran's grammar, which was published in 1746, reflects the transitional stage between Classical and Modern Yucatec.

[6]Tozzer's data were collected between 1901-1905.

(5.34) 1sg   *ki-in > k-in*
     2sg   *ki-a > k-a*
     3sg   *ki-u > k-u*
     1pl   *ki-k > q*

Thus, we see that by the turn of the century, the Yucatec imperfective marker *ki* has become increasingly frequent in main clause. Significantly, it is is the person markers of main verbs which originally functioned as hosts for these elements. Over time, these markers have become inflectional elements attached to the subject person markers. The fact that this aspect marking did not develop on AF verbs in subject extraction contexts follows from the fact that AF verbs do not carry preverbal person markers.

5.3.3   Summary

The grammaticalization of obligatory preverbal imperfective aspect marking in main clauses is thus a very recent phenomenon. The consequence of this change is that it has produced a further morphological contrast between main clause verbs on the one hand, and dependent verbs in subject extracted relative clauses on the other: because these aspect markers grammaticalized to the subject pronominal clitics in main clauses, and because subject pronominal clitics did not typically occur resumptively in subject relative clauses, the modern language now exhibits two morphological paradigms of imperfective verbs, one, the conservative set, restricted to subordinated environments. This verb form carries no dependent pronoun or main clause aspect marking, and thus preserves the original gapping structure for subject extracted relatives.

5.4   'Exceptional Morphology' in Western Malayo-Polynesian

I showed in chapter 4 that a number of anti-pronominal languages exhibit special morphology which appears on dependent verbs in A-bar dependencies, apparently 'in place' of regular person marking. This has been documented perhaps most famously in Austronesian languages, where it has often been described as 'wh-agreement' (see chapter 4). In this section I report on some comparative Austronesian studies which show that the exceptional morphology is actually the retention of older main clause verbal morphology that

was lost in main clauses. Again, we shall see that the retention of the older morphology can be ascribed to the fact that the grammaticalization of head-marking did not take place on dependent verbs in A-bar dependencies.

Verbal alternations in Chamorro A-bar dependencies do not simply involve the presence or absence of main clause person marking. In addition to not bearing person marking, dependent verbs in transitive subject and in object extractions in Chamorro are also characterized by special infixal morphology. (5.35a) shows a main transitive clause, in which the verb obligatorily inflects for subject person marking. (5.35b) shows a subject wh-extraction. The verb now bears the infix *-um-*. In (5.36) below, which is a case of object extraction, the verb bears the infix *-in-*.

(5.35) Chamorro (Chung 1998, adapted from 52 and 53b, 236)

> a. *Ha-fa'gasi si Juan i kareta*
> 3sg-wash PN Juan the car
> 'Juan washed the car.'

> b. *Hayi f-um-a'gasi i kareta*
> who *um*-wash   the car
> 'Who washed the car?'

(5.36) a. *Ha-fahan si Maria i sanhilo'-ña gi tenda*
> 3SG-buy ART Maria ART blouse   OBL store
> 'Maria bought her blouse at the store'

> b. *Hafa f-in-ahan-ña si Marai gi tenda?*
> What buy-*in*-3sg.GEN ART Maria OBL store
> 'What did Maria buy at the store?'

Donohue and Maclachlan (1999) have shown conclusively that this morphology is actually a vestige of the old proto-Austronesian voice system, tthe so-called Philippine-type voice system (Himmelmann, 2002), which is still productive in both main and subordinate clauses in Philippine languages such as Tagalog, some languages of northern and central Borneo, northern Sulawesi and Madagascar, as well as most of the Formosan languages (Ross, 2002, 20).

In a Philippine type voice system, the semantic role of the syntactic pivot is marked by verbal affixes. In the Tagalog examples below, the infix *-um-* in (5.37a) is the 'actor voice'

and the infix *-in-* in (5.37b is the 'patient voice'. In each case, the syntactic pivot introduced by the specific phrase marker *ang* assumes the role indicated by the verbal affix:

(5.37) Tagalog (Himmelmann, 2002, 12)

    a. *H[um]anap na   ng   bahay ang   bata'*
       AV-search   now GEN house SPEC child
       'The child looked for houses/a house.'

    b. *H[in]anap na   ng   bata' ang   bahay*
       PV-search now GEN child SPEC house
       'The child looked for the house(s).'

In several Western Austronesian languages, including Tagalog, verbs in unbounded dependencies are required to occur in a particular voice form. If the extracted argument is the agent of the dependent verb, then the dependent verb must occur in the Agent Voice (5.38b). If the extracted argument is the patient of the dependent clause verb, then the dependent verb must occur in the Patient Voice (5.38c). In other words, only syntactic subjects can be extracted in Philippine type languages.

(5.38) Tagalog (Donohue and Maclachlan, 1999, 127)

    a. *B[um]asa ang   guro   ng   libro*
       Read.AV NOM teacher GEN book
       'The teacher read a book'

    b. *Sino ang   b[um]asa ng   libro?*
       who NOM read.AV   GEN book
       'Who read a book?'

    c. *Ano ang   b-in-abasa ni   Maria?*
       What NOM read.PV   ART Maria?
       'What is Maria reading?'

As Donohue and MacLachlan show, the exceptional morphology of A-bar dependencies in Chamorro is nothing other than Philippine Type voice morphology. If one compares the Tagalog subject extraction in (5.38b) with the Chamorro subject extraction in (5.35b), we see the same, cognate infix *-um-*. Similarly, the object extractions in both languages involve cognate infixes *-in-* (compare 5.38c and 5.35c). Table 5.6 provides the corresponding

morphemes for different extraction types, in Tagalog, a related Philippine-type language Kapamangan and Chamorro.[7]

	Tagalog	Kapamangan,	Chamorro
A extraction	mag-, -um-	maN-, -um-	-um-
S extraction	mag-, -um-	maN, -um-	-um-
O extraction	-in (-an, i-)	-an, i-	-in-

Table 5.6: Verbal Morphology in A-bar dependencies (Donohue and Maclachlan, 1999)

As Donohue and MacLachlan point out, the difference between Philippine type languages like Tagalog, and Chamorro, is that in Chamorro the voice morphology is restricted to subordinate environments. In Tagalog we do not see any equivalent alternation between main clause verbs and dependent clause verbs, because the voice system is still productive in main clauses, cf. (5.38a) and (5.38b). Thus, seen from a comparative Austronesian perspective, the noteworthy aspect of the Chamorro verb paradigms is not the appearance of *-um-* and *-in-* in extractions, but rather "the fact that *um* is missing from the Chamorro paradigm in transitive main clauses" (Donohue and Maclachlan, 1999).

Donohue and MacLachlan suggest that this gap in the Chamorro paradigm is consistent with a view of language change in which subordinate clauses are taken to be more conservative than main clauses (e.g. Givón, 1984; Campbell, 1986). That is, Chamorro has innovated a new transitive active voice in main clauses, while preserving the old voice morphology in subordinate clauses.[8]

However, it seems that we don't have to appeal to such a general principle of subordinate clause conservatism to explain the fact that the innovative verb form does not occur in Chamorro extraction contexts. While it is not clear yet from comparative evidence exactly

---

[7]Table 5.6 is incomplete. Chung (1998) also states that 'Wh agreement' with oblique extraction involves the optional use of *-in-* if the verb is unaccusative. This plausibly also has its roots in the old Philippine Type voice system. Four different Proto Malayo-Polynesian voices have been reconstructed. These include, in addition to the Actor and Patient voice, a Location and Circumstantial voice, both of which involved the use of the *-in-* prefix in the perfective (Ross, 2002) These latter two were obligatory in certain types of oblique extractions.

[8]Some of the old Austronesian voice morphology is in fact still visible in certain types of main clauses in Chamorro: *-in-* shows up in what is described in some contemporary grammars as a passive (though see Donohue and Maclachlan (1999) for arguments against a passive analysis of this construction); *-um-* occurs in what are described as 'anti-passive' clauses.

how the innovative pronominally inflected verb forms arose (the pronoun attaches to the verb stem in these constructions, not to any of the overt voice-marked forms), what is clear is that the Chamorro main clause innovation involves the development of head-marking on transitive verbs, that is, of person marking inflection out of originally independent pronouns in preverbal position (Zobel, 2002). By the same principle that we have seen already at work in varieties of Romance languages which have developed head-marking, dependent verbs may remain exempt from this process, precisely because the necessary precondition for morphologization is not there (frequent pronoun-verb adjacency): In Philippine type languages, subject extraction involves a gapped argument in the dependent clause (cf. (5.38b and 5.38c).

The preservation of Austronesian voice morphology on dependent verbs in A-bar dependencies is not restricted to Chamorro. It is a phenomenon that has been studied in a range of Indonesian languages by Zobel (1997). It is also evident in Tukang Besi, which reveals a very similar situation to Chamorro (Donohue, 1999; Donohue and Maclachlan, 1999). Like Chamorro, Tukang Besi exhibits an alternation between main and dependent verb forms, as we saw in chapter 4. Main transitive verbs inflect for subject and object person marking (5.39a), but this does not occur on dependent verbs. Instead, in subject extractions, the infix *-um-*appears (5.39b), and in object extractions, the infix *-ni-* appears (5.39c). These infixes are cognate with the Actor and Patient Voice infixes respectively of Philippine languages:

(5.39) Tukang Besi (Donohue and Maclachlan, 1999, 129-130)

    a. *No-balu=e'*   *na*   *lokate*   *wowine*
       3R-buy=3OBJ NOM banana woman
       'The woman bought the bananas.'

    b. *Te*   *emai* *na*   *b[um]alu* *te*   *loka*
       CORE who NOM buy.um   CORE banana
       'Who bought the bananas?'

    c. *Te*   *paira* *na*   *ni-basa* *nu*   *guro*
       CORE what NOM NI-read GEN teach
       'What did the teacher read?'

The difference between Chamorro and Tukang Besi is that while Chamorro has only grammaticalized person marking for transitive subjects, Tukang Besi has done so for both transitive and intransitive subjects. Thus, in the synchronic grammar, Tukang Besi exhibits

a verbal alternation between intransitive main clauses and intransitive subject extractions. Main intransitive verbs inflect for subject person marking (*no*, in 5.40), while dependent intransitive verbs in subject-extracted clauses exhibit the old proto-Austronesian AF voice infix *-um-* (5.41):

(5.40) Tukang Besi (Donohue, 2002)

> *No-wila   lego-lego*
> 3R-go   arms.swinging
> 'He was walking, arms swinging.'

(5.41) Tukang Besi (Donohue, 1999, 373)

> *O-koruo   na   mia   [w[um]ila kua Lia] i   rearea   ai*
> 3R-many NOM person go.UM   all Lia OBL morning ANA
> 'There were a lot of people who left for Lia this morning.'

Chamorro, by contrast, did not develop person marking on intransitive (realis) verbs. There is therefore no alternation between verb forms in main clauses and in subject extractions out of intransitive clauses, because intransitive main clause verbs also take the old AF voice with *-um-*:

(5.42) Chamorro, (Donohue and Maclachlan, 1999, 125)

> *H[um]anao siha*
> go.UM   3PL.NOM
> 'They went.'

Table 5.7 presents the differences between the conservative Tagalog case, and Chamorro and Tukang Besi. (PRON-$\sqrt{}$ represents the innovative construction with pronominal inflection attached to the verb stem). In Tagalog there is no difference between main clause (Actor-Voice) verbs and subject extractions. In Chamorro, which has developed person marking on main transitive clauses, the morphological distinction between verb forms exists only in the case of transitive subject extractions. In Tukang Besi, where person marking has developed for transitive and intransitive subjects alike in main clauses, there is a synchronic alternation between main clause verbs and dependent verbs for both intransitive and transitive subjects.

Environment	Tagalog	Chamorro	Tukang Besi
A	mag-, -um-	PRON-√	PRON-√
S	mag-, -um-	man-/-um-	PRON-√
A-extraction	mag-, -um-	-um-	-um-
S-extraction	mag-, -um-	-um-	-um-

Table 5.7: Voice morphology in Tagalog, Chamorro and Tukang Besi

5.4.1 The attrition of the proto-Mayan 'status' system

Comparative evidence and reconstruction allows us to identify a similar process in Mayan, whereby the attrition of the old Proto Mayan status system has resulted in the retention of exceptional morphology in extraction contexts. As we saw in Chapter 2, various Mayan languages exhibit a special suffix on AF verbs, which does not occur in main clause verbs. The form of this suffix varies, depending on the language, and on the verb class. For Q'eqchi', an Eastern Mayan language, (5.43a) shows the *-o* suffix of root transitives; (5.43b) shows the *-n* suffix of derived transitives:

(5.43) Q'eqchi' (Dayley, 1981, 20)

    a. *Laa'at x-at-sak'-o-k w-e*
       you   ASP-B2-hit-ATP-M A1-to
       'It was you who hit me.'

    b. *li winq li x-kamsi-n r-e...*
       the man who T-kill-ATP A3-to
       'The man who killed him...'

The origins of the *-n* suffix has a complicated and as of yet, unclear history, and I will not address it here. However, Kaufman (2002) does provide good reconstructed evidence for the origins of the *-o(w)* suffix found on root transitive AF verbs in Eastern Mayan languages such as Q'eqchi'. He shows that it is probably a reflex of the old Proto-Mayan 'plain status' suffix. I build on his insights here to yield a partial picture of the development of the root transitive AF verb in Eastern Mayan languages.

Proto Mayan has been reliably reconstructed as possessing four sets of verbal 'status' suffixes, 'imperative', 'perfect', 'plain' and 'dependent', each of which combined with a

subset of preverbal aspectual markers to yield finite verbs, and which had different allomorphs, depending on the verb class (Kaufman 1986, cited in England 1989).[9] Those used with transitive verbs have two or more variants each, depending (minimally) on whether the transitive stem is radical or derived. Root transitives are monosyllabic and of the shape CVC. Derived transitives are polysyllabic and end in a vowel or a glottal stop. Table (5.8) provides the reconstructed suffixes for monosyllabic root transitives and polysyllabic derived transitives according to Kaufman:

Status	Root transitive	Derived transitive
Plain	*-o-h__# , *-o-w /...	*-h or *-V
Dependent	*a-ʔ	*-Vŋ
Imperative	*a-h / __# , *-a-w	*-Vŋ
Perfect	*-o-ʔm	*-o-ʔm

Table 5.8: Proto Mayan transitive status suffixes

Kaufman (2002) proposes that the *-o* suffix found on AF verbs in Eastern Mayan languages such as Q'eqchi' (above), is a reflex of the old plain status, which originally occurred on main clause verbs. Accordingly, he suggests that proto Mayan possessed a special agent focus structure which involved deletion of the set A person markers from the plain status verb. In the terms of this study, this amounts to saying that subject extraction in proto Mayan involved a subject-gap structure; the verb forms themselves were not otherwise differentiated morphologically in main and subject-extracted contexts, save for the absence of person marking.

We can go a step further back and reasonably hypothesize that in pre-proto Mayan, subject pronouns were independent elements, and subject-extracted unbounded dependencies were terminated with gaps. By proto Mayan, independent subject person markers had become morphologically dependent. This grammaticalization process affected main clauses, but, crucially, did not affect subject extracted subordinate clauses, because, by hypothesis,

---

[9]Comparative studies of living Mayan languages, Colonial texts and grammars, and the hieroglyphic record have yielded a partial picture of the ancestral Mayan language. Much work has been undertaken on the reconstructed phonological system of proto Mayan (including Kaufman 1964; Campbell 1977; Kaufman 1978; Campbell 1985; Fox 1978) whose testimony has formed the basis of the current theories of language sub-groupings. Historical morpho-syntactic reconstruction is still very much more in its infancy, though significant advances have been made, especially with regard to verbal morphology and the make-up of the verbal complex.

resumptive pronouns were infrequent in these environments. Again this is very remininiscent of the Romance situation described above, where the alternation bewteen verb forms is defined solely by the absence or presence of person marking.

In Mayan languages, however, subsequent changes have taken place, which have caused further morphological distinctions between the two verb forms. Changes in main verb inflection on root transitives have produced a greater formal differentiation between AF and main clause verb types: The plain status has been lost in Q'eqchi' and Poqomam in main clauses, for example, so that the old plain status *-ow* is now only present on subject-extracted subordinate verbs.

Table (5.4.1) summarizes the hypothesized developments leading to the modern Qeqchi AF alternation for root transitives:

Stage	Development
Pre-proto Mayan	Subject pronouns are independent elements
Proto Mayan	Subject pronouns grammaticalize onto main verbs
	> Results in morphological difference between main and dep. verbs
Q'eqchi'	Plain status marking lost in main clauses
	> Results in 'exceptional' suffixation on dependent verb

Table 5.9: Developments from proto Mayan to Q'eqchi' for root transitives, main and dependent verbs

The relationship between AF verbs and the antipassive

I showed in chapter 2 that in Greater Quichean and Mamean languages there exists a formal similarity between antipassive and AF inflection. In Q'eqchi', for example, there are two different suffixes which derive antipassives, depending on the verb type. For root transitive verbs, which are underived, and monosyllabic, the antipassive suffix is *-o*. For derived transitive verbs (polysyllabic), the suffix is *-n*. (5.44a) gives an example of a (derived) transitive verb; (5.44b) its antipassive counterpart inflected with *-n*; and (5.44c) an antipassive verb derived from a root transitive with *-o* (repetitions of examples 2.20):

(5.44) Q'eqchi' (Dayley, 1981, 19-20)

    a. *x-at-ch'iila*
       T-B2-A1-scold
       'I scolded you.'

    b. *x-in-ch'iila-n*
       T-B1-scold-ATP
       'I was scolding.'

    c. *x-at-b'is-o-k*
       T-B2-measure-ATP-M
       'You were measuring.'

The suffixes on AF verbs are formally identical to those of the antipassive. (5.45a) shows the *-o* suffix of root transitives; (5.45b) shows the *-n* suffix of derived transitives:

(5.45) Q'eqchi' (Dayley, 1981, 20)

    a. *Laa'at x-at-sak'-o-k*   *w-e*
       you   ASP-B2-hit-AP-M A1-to
       'It was you who hit me.'

    b. *li   winq li   x-kamsi-n r-e...*
       the man who T-kill-AP- A3-to
       'the man who killed him...'

The few discussions of the historical development of these verb forms that exist in the literature have tended to assume a rather complicated series of mergers and reanalyses to account for the similarities between antipassive and AF suffixes, which are assumed to have entirely distinct etymologies. For example, Dayley (1981, 58) proposes that proto Mayan had an AF verb marked with the suffix *-(V)n* and an antipassive verb marked with **w*. The use of *-Vn* to mark the antipassive on derived transitive verbs in languages such as Q'eqchi' and K'iche' is understood to have come about through extending its use from the focus antipassive.

    An alternative explanation is that the formal similarities between AF and antipassive suffixes in Quichean and Mamean languages (especially the presence of the same root/ derived suffixal distinction) point to some distant common origin. If it is indeed the case, as I argue here, that AF verbs in Mayan languages, just like antipronominal verbs elsewhere, represent the retention of earlier verb forms that remained exempt from main clause

grammaticalization processes, then the similarity between antipassive suffixes and AF suffixes may suggest that antipassive verb morphology also derives from earlier main clause verb morphology before the grammaticalization of ergative person marking, and thus we find the common morphology appearing in both contexts. Something that points to the merit of such a possibility is the fact that in an unrelated language which exhibits an antipronominal alternation, Chamorro, the infix found on subject-extracted dependent verbs, *-um-* , also occurs in what are described in synchronic grammars as antipassive clauses (**?**).

Determining the validity of this claim will require a much better understanding of the grammaticalization paths of antipassives in general, and a better knowledge of comparative-historical Mayan morphosyntax than currently exists. I therefore leave this here as a speculative note.

5.5 The Yucatec perfective alternation

The final case study I present is of the Yucatec AF perfective. The history of the Yucatec AF perfective involves several of the processes described above (retention of older morphology on dependent verbs, development of main clause aspect marking (twice!)) and thus serves as a useful case study of morphological differentiation between main and dependent verbs in extraction contexts over time. The history is largely recoverable on the basis of reconstructed evidence and Colonial texts.

The AF perfective, just like the AF imperfective described above, does not carry preverbal aspect or person marking. It is differentiated additionally from the main clause perfective in the form of the status suffix. It inflects with *-eh* (only clause finally, 5.46a), rather than the completive status suffix *-ah* of main clauses (5.46b):

(5.46) a. *Le   máak=o' **t-u=xok-ah**   le   libro=o'*
           DET   man=D2   PRV-A3=read-CMP   DEF   book=D2
         'The man read the book'

   b. *Máax t-aw=a'l-ah   **xok-eh?***
      who   PRV-A2=say-CMP   read-eh
      'Who did you say read it?'

   c. *Máax t-aw=a'l-ah   **xok** le   libro=o'?*
      who   PRV-A2=say-CMP   read   DET   book=D2

'Who did you say read the book?'

In descriptions of Yucatec the perfective AF verb form is described as inflecting for the 'subjunctive' (otherwise known as the 'dependent') suffix. The Yucatec subjunctive suffix *-eh* is formally identical to the suffix found in perfective AF contexts in the modern language, and, like the AF suffix, only ever occurs utterance finally (that is, when the verb is not preceded by lexical material). The subjunctive suffix occurs in a variety of contexts, for example in irrealis contexts following the subordinator *káa* (5.47), or with certain higher aspectual or modal predicates, for example, the predictive aspect marker *bíin* (5.48):

(5.47) *In   k'áat káa u   bis-en   Cancun in   tàatah*
      A3   wish   that A3   take(SUBJ)-B3 Cancun A1   father
      'I want my father to take me to Cancun'   (Verhoeven, 2007, 132)

(5.48) *Bíin   u   ts'iíib+óol-t   x-ch'úup-tal   xib-o'b*
      PRED   A3   write+soul-APP(SUBJ) F-female-PROC-INC   male-PL
      'The men shall wish to become women' (Vapnarsky 1995, 89, cited in Bohnemeyer 2002)

There are some data that suggest, however, that while the modern subjunctive and the AF perfective are homophonous, they might in fact distinct origins. This is because in the Colonial period the subjunctive suffix and the perfective AF suffix do not appear to be formally identical. Colonial grammars reveal that Classical Yucatec possessed three different subjunctive allomorphs for transitive verbs, the choice depending on the verb class. Root transitives inflected for *-Vb'*, transitive verbs derived from active verbal nouns with *-t* inflected for *-e(h)*, and transitive derived from intransitive roots with the causative marker *-es* took zero (Coronel, 1620; de San Buenaventura, 1684).[10] This is summarized in (5.10).

This allomorphy can be verified by an inspection of early colonial texts. The following examples are all extracted from a text dating from around 1575, author unknown.[11] The first examples show root transitives inflected for the *-Vb* subjunctive:

---

[10]These allomorphs have modern reflexes in the Yucatecan languages Itzaj and Mopan.

[11]Annotated, transcribed and translated by Whalen 2003.

Transitive verb class	subjunctive suffix
Root transitives	Vb
s-derived transitives	∅ or e
t-derived transitives	e

Table 5.10: Colonial Yucatec subjunctive allomorphs

(5.49) *yoklal   ua bin   yn Dza-**b**   yt.   bin   yn pat-**ab**   u*
because if PRED A1 give-SUBJ and PRED A1 make.up-SUBJ A3
kaba-obe bin   tubuc tech
name-PL   forget.SUBJ   you
'because if I will give them, and I will make up their names, you will forget them.'

(5.50) *bala bin   a   uil-**ab**   hahil yn   than   lae*
thing PRED A2 see-SUBJ truth A1   word   now
'The thing is, you will see the truth of my word now'

(5.51) gives a further example of a root transitive with *-ub*, and also an *s*-derived transitive with *-e*:

(5.51) *tilic   a   xachet-ic uamac   bin   u   **luk-ub**   yetel u*
while A3   seek-INC whoever PRED A3 devour-SUBJ and   A3
tabes-e
ensnare-SUBJ
'while he seeks whomever he will devour and ensnare'

The following examples show *t*-derived transitives inflecting for *-e*:

(5.52) *Ua bin   u   **nana olt-e**   yt.   u   **paktumt-e**   tulacal*
if   PRED A3 contemplate-SUBJ and A3 consider-SUBJ all
'if he will contemplate and consider all of it'

(5.53) *bal   bin   ca   **belt-e***
what PRED A1.pl do-SUBJ
'What will we do?'

The AF forms, by contrast, do not appear to exhibit this allomorphy. Colonial grammars consistently give the AF perfective form with *-e* only, regardless of the verb class. Text examples appear to confirm the distinction between the perfective AF form and the subjunctive at this stage. Compare (5.49), the relevant part of which is repeated below as 5.54 with (5.55), which both involve the verb *Dza* 'give' (modern Yucatec *tsa*), and both of which are from the same text. *Dza* is a root transitive, and thus, in the subjunctive, inflects for *Vb*. In the AF context however, it inflects for *-e* (5.55):

(5.54) *yoklal   ua bin   yn Dza-**b***
      because if   PRED A1 give-SUBJ
      'because if I will give them'

(5.55) *Ca   u   kat-ah   emperador - paale macx yam Dza-**e***   *u*   *Dzabilah*
      then A3   ask-CMP   emperor   - child   who   first give-AF A3   offering
      Diose
      God
      'Then the emperor asked: "Child, who first gave his offering to God?"'

By 1746, Beltrán (1746, 112) writes that the subjunctive form in *-b* is still found but the more common form for the future is *-e* for root transitives. The Colonial record therefore appears to reveal a gradual levelling of the subjunctive allomorphs to *-e*. Tozzer (1921) though aware of the allomorphy from earlier grammars, was not able to find *-b* in use at all by 1905. By this stage the subjunctive is realized in Yucatec with *-e* only, across all verb classes. The consequence is the formal identitiy of the subjunctive *-e* and the AF perfective *-e*. Yet even by the turn of the century, with Tozzer's description, it seems that there is still a detectable difference between the AF perfective and the subjunctive. While Tozzer writes that the subjunctive *-e* only occurs utterance finally, there are examples of the AF perfective formed with *-e* even when lexical material follows: (Tozzer, 1921, 93):

(5.56)   a. *Maš   puts-e*
          who   hit-AF?
         'Who struck him?'

      b. *T-en   puts-e*
         I   strike-AF.
        'I am the one who struck him.'

(5.57)   a. *Maš mis-t-e   na*
             who sweep-APP-AF house?
             'Who swept the house?'

        b. *T-en mis-t-e   na*
             I   sweep-APP-AF house
             'I am the one who swept the house.'

In modern Yucatec, the formal merger with the subjunctive is complete. Both the perfective AF suffix *-eh*, and the subjunctive *-eh* only occur utterance finally. But if the formal identity of the modern subjunctive suffix and the AF perfective is a recent analogical phenomenon, what is the origin of the AF perfective?

A plausible source is the old proto Mayan plain status suffix that I already discussed above in the context of K'iche'. This suffix is assumed to have been lost in Yucatec, though I suggest tentatively that the modern Yucatec AF perfective is a modern reflex of the old plain status suffix. This would be consistent with the fact that eastern Mayan languages are also assumed to have reflexes of this suffix on root transitive AF verbs, as we saw above. It is also supported by the aspectual semantics of the Yucatec AF suffix. In the proto language, the plain status is reconstructed with a punctual/perfective meaning in main clauses, where it occured in the absence of any additional preverbal aspect markers (Kaufman 1986, cited in England 1989). The fact that the Yucatec AF verb shares this perfective semantics suggests that it might also be a reflex of the old proto Mayan plain status (it most certainly doesn't appear to have much in common temporal-semantically with the modern Yucatec subjunctive, which occurs in irrealis and future contexts).

We can thus sketch the following picture of the origin of the special AF perfective. It represents the retention of the archaic plain status verb, restricted to subordinate contexts involving agent extraction, in a parallel development to what we saw above for central Mayan languages. Subsequent changes in aspect marking in main clauses have since rendered this verb form morphologically distinct from its modern main clause perfective counterpart, just as they did in Central Mayan languages. While in Q'eqchi', this involved the loss of the plain status suffix, in Yucatec, a new completive suffix *-ah* was innovated in main clauses, grammaticalized from proto Mayan **-ah* 'ealier, before' (Kaufman 2001). Second, just as the aspect particle *ki* became obligatory in main clauses in the imperfective in Yucatec (see above), so too the aspect particle *ti* became obligatory in main clause verbs in the perfective.

In early Colonial Yucatec the perfective in main clauses was formed simply with the set A prononominal plus the verb inflected for the new completive status *-ah*:

(5.58) **u   key-ah-en**   *juan ca   ix*   **u   hadz-ah-en**
　　　A3   scold-CMP-B1 Juan then and A3   whip-CMP-B1
　　　'Juan scolded me and then whipped me'　　　　　　　　(Coronel, 1620, 127)

Thus, by this stage, there exists a morphological distinction between the older perfective in subject extraction contexts, inflected still with the old plain status *-e*, and the newer perfective of main clauses, now inflected with the innovative perfective *-ah*.

Over the following centuries a further change affects the morphology of main perfective verbs: we see the emergence of a new perfective particle *t(i)-*, which follows a parallel path of development to the imperfective particle *k-* discussed above. Beltrán (1746) gives the perfective without *t-*, though mentions the particle in passing as an optional marker:

(5.59) *In   cambez-ah*
　　　A1   teach-CMP
　　　'I taught.'　　　　　　　　　　　　　　　　　　　　　(Beltrán, 1746, 21)

Text examples by the late 18th century still predominantly lack the *t-* particle. The following example is taken from *The Book of Chilam Balam of Chumayel* (1782), translated and transcribed by Tozzer (1921):

(5.60) *Ken   Chilam Balam ka*   **in tsol-ah**   *u-t'an*
　　　I am   Chilam Balam and I   interpret-CMP A3-word
　　　'I am Chilam Balam and I interpreted his word.'

Early twentieth century grammarians (e.g. Lopez Otero 1914, 72), write of the *ti-* for the 'recent past tense', giving, respectively, *tin, ta, tu,* for the contracted forms with the 1st, 2nd and 3rd person singular subject person markers, respectively. At this stage the simpler form without the presence of the marker in main clauses is still available. By the time of Tozzer's description, *t-* regularly (but not obligatorily) occurred in perfective contexts: He states "With this form in *-ah* the nominal pronoun is usually compounded with *ti-* or *t-*". Tozzer gives the following regular contractions:

(5.61) *t-in* remains *t-in* or > *t'-in*   (first person singular)

   t-a remains *t-a* or > *t'-a*   (second person singular)

   t-u remains *t-u* or > *t'-u*   (third person singular)

   t-k > *t'*   (first person plural)

The grammaticalization of *t-* in perfective main clauses results in a further morphological divergence between main and subordinate perfective verbs. In the modern language, the AF perfective retains the old completive status marking; the main clause perfective is formed with the *t-* perfective marker and the new completive status suffix *-ah*. The chronology of the changes can be summarized as follows:

1. **Pre-proto Mayan**: finite perfective verbs are formed with plain status **-V(w)*. Pronouns are independent elements. Subject extraction involves gapping

2. **Proto-Mayan**: independent subject pronouns reduce to pronominal inflection
   > *results in morphological differentiation between main and dependent verbs*

3. **Pre-Yucatecan**: new completive status suffix is grammaticalized in main clauses from **-ah* ('already, earlier'), replacing the old plain status in main clauses.
   > *results in appearance of 'exceptional inflection' on dependent verbs: the old plain status is retained in subject extraction environments.*

4. **Colonial Yucatec**: emergence of recent past *ti-* in main clauses (optional), tendency to fuse to subject pronominal clitic

5. **Modern Yucatec**: *t-* becomes obligatorily fused to subject pronominal clitics in the perfective in main clauses
   > *results in further morphological differentiation between main and AF alternants*

6. AF *-eh* is lost in non clause final position (by analogy with the distribution of formally identical subjunctive suffix *-eh*)

5.6   The typology of anti-pronominal alternations

In the above sections I presented a series of case studies which illustrated some specific paths by which anti-pronominal alternations can arise in nascent head-marking languages. In this final section I briefly discuss one or two predictions that the theory of frequency-sensitive morphologization makes for the typological space of anti-pronominal alternations.

Stage	Main clause verb	subject-extracted subordinate verb
1	u(y) il-Vh	il-Vh
2	uy=il-Vh	il-Vh
3	uy=il-eh	il-eh
4	uy=il-ah	il-eh
5	(ti-)uy=il-ah	il-eh
6	t-uy=il-ah	il-(eh)

Table 5.11: Differentiation of Yucatec main and AF verb forms over time, for the verb *il* 'to see'

On the view that it is the relatively low frequency of adjacent pronoun-verb combinations in extraction contexts (where gaps are more frequent than resumptive pronouns) which gives rise to asymmetric patterns of pronoun grammaticalization, then the probability of an anti-pronominal alternation emerging is constrained by at least these two factors: the likelihood of person marking grammaticalizing in main clauses, and the likelihood of it not grammaticalizing in dependent clauses.

As I discussed above, the frequency-sensitive grammaticalization process of person marking can yield asymmetries with regard to the distribution of dependent person markers in main clauses, within and across languages. With regard to grammatical function, subjects tend to be pronominalized in discourse more than objects, and hence are more likely to grammaticalize onto main verbs than any other function (as discussed in Siewierska 2004). This results in a higher frequency of subject dependent person markers across languages than object dependent person markers. By contrast, resumptive pronouns appear to be particular disfavored across languages in subject position. These two distributional facts jointly lead to the prediction that anti-pronominal alternations should be particularly frequent across head-marking languages for subject-extractions, because this is where the frequency distributions of pronoun-verb combinations are particularly skewed.

From the very limited sample that I have discussed in this dissertation, this seems to be the case. Indeed, in the generative literature, anti-pronominal alternations are often treated as deriving from specific constraints on *subject* extraction (e.g. Ouhalla 1993). It has been a point of confusion in this research tradition that there exist languages that exhibit anti-pronominal alternations for the extraction of other grammatical functions (cf.

e.g. Schneider-Zioga 2002). This probably reflects the simple fact that anti-pronominal alternations for subject-extraction are most frequent (at least in the sample of languages currently available). As the table 5.12 indicates (repeated from table 4.3, chapter 4), there is a predominance of languages that exhibit anti-pronominal alternations for S- and A-extraction.

A only	S & A	A & O	S, A & O	S & O	O only?
Mayan	Berber	Chamorro	Makaresse	Selayerese	
Salish	Papuan Malay		Yimas		
	Kinande		Tukang Besi		
	Irish				
	Breton				

Table 5.12: Argument roles triggering verbal alternations in 12 head-marking languages

As table 5.12 also shows, other grammaticalization paths of person marking in main clauses are also possible, giving rise to the cross-linguistic differences in apparent morphological reflexes of 'argument role sensitivities' to extraction, as documented in chapter 4. Mayan languages, and some Salishan languages also, grammaticalized subject person marking onto transitive verbs only (for third person), hence the transitivity 'restriction' on the AF alternation (so often assumed to be a reflex of ergative syntax). Makaresse has grammaticalized person marking for S, A and O, thus giving rise to verbal alternations for all three grammatical functions in extraction contexts. We would expect that the relative frequencies of anti-pronominal alternations for these configurations of argument-roles would reflect the relative likelihood, cross-linguistically, for person markers to grammaticalize on main clause verbs according to these different alignment patterns. This is in keeping with the fact that there are no instances in my (albeit very limited) sample of languages which exhibit antipronominal alternations for object extractions only. This presumably can be accounted for by the fact that it is cross-linguistically very unusual for languages to develop dependent person marking on main clauses for objects only (Siewierska, 2004). More cross-linguistic data are required to see if these generalizations hold up with a wider sample of languages.

Of course, it is not only via grammatical function that asymmetries can arise with respect to the grammaticalization of person marking onto verbs. This process can also be

restricted in other ways, for instance with respect to the tense/aspect category of the verb, or the grammatical number of the pronoun. Presumably differences also manifest in some languages at the lexical level (this appears to be the case, for instance, in Irish, where different lexical verbs exhibited differet paradigms of pronominal inflection, McCloskey and Hale 1984). To date there exists very little information about the cross-linguistic patterns regarding these types of grammaticalization asymmetries. Accordingly, it remains an open question how such asymmetries are correlated with the typological space of anti-pronominal alternations.

5.7 Conclusions

In this chapter I developed the thesis that the special analytic dependent verb forms that occur in head marking languages in extraction contexts are conservative verb forms which remained exempt from the grammaticalization of head marking. On this view, it is the synthetic (= pronominally inflected) verbs that are the diachronic innovation in systems that originally only had analytic (= pronominal-less) verb forms. This proposal therefore inverts the claims of many synchronic grammarians that analytic verbs are derived from their synthetic main clause counterparts. The reason for the exemption of these verb forms has its roots in the well-documented fact that grammaticalization processes are frequency sensitive: the more frequently two elements co-occur, the more likely it is that grammaticalization will take place. I argued that the relatively low frequency of adjacent pronoun-verb combinations in extraction contexts (where gaps are more frequent than resumptive pronouns), by comparison with main clause verbs, can give rise to asymmetric patterns of pronoun grammaticalization, and thus lead to the emergence of these morphological alternations.

Chapter 6

Resumption variation and processing

Within and across grammars, gaps and resumptive pronouns are distributed according to certain well attested patterns. Over the course of this dissertation I have made use of these patterns in a number of ways. In chapter 3 they were used as a diagnostic for classifying Mayan AF as a morphological subtype of RP/gap alternation. In chapter 4 these same distributional parallels helped us to identify the commonality between Mayan AF and verbal alternations in head-marking languages more generally. And in chapter 5, the asymmetric frequency distributions of gaps and resumptive pronouns were shown to be directly implicated in the historical genesis of verbal alternations in emergent head-marking languages, via the frequency sensitivity of the grammaticalization process which created them. This still leaves a deeper question untouched: how do we explain the existence of these asymmetric frequency distributions? Why, within and across languages, are gaps particularly frequent in certain types of extraction environments? Why do resumptive pronouns occur more frequently in others?

   In this chapter I discuss and contribute to the idea that the distributional patterns we find in this domain are influenced by processing factors. The hypothesis that variation in production may be determined by external pressures such as processing constraints has been widely and successfully investigated for a variety of phenomena (inter alia Ferreira and Dell, 2000; Hawkins, 1994, 2004, 2007; Gries, 1999; Frank and Jaeger, 2008; Arnold et al., 2004; Jaeger and Wasow, 2007; Jaeger, 2006; Race and MacDonald, 2003; Temperley, 2003; Hofmeister, 2007; Hofmeister and Sag, 2009). On the view that variation in language use shapes language structure over time, processing preferences also emerge as

an important factor responsible for typological patterns across grammars. Hawkins (2004) refers to this as the 'Performance-grammar correspondence hypothesis', and to the research program within which the hypothesis is developed as 'Processing Typology' (Hawkins, 2007).

Resumption/gap variation may profitably be understood from the perspective of Processing Typology. I will show how the variation in this domain, both cross-linguistically and language internally proves problematic for categorical accounts of the distributional patterns. Instead, more of the critical data can be explained via processing. I will discuss various processing-based theories of RP/gap distributions, and present three experiments on Yucatec, which provide some of the first controlled production data in support of the performance-grammar correspondence hypothesis. These experiments also serve to offer interesting grist for the mill for teasing apart the predictions of different theories of dependency processing and resumption.

6.1 Categorical approaches to RP/gap distributions and the variation problem

Generative theories of grammar over the last forty years have understood the distributional patterns of resumptive pronouns and gaps in categorical terms, as deriving from specific and innate universal constraints on competence grammars. In movement based theories, for example, the sensitivity of gaps to islands is accounted for by positing constraints on movement. In the GB tradition such constraints have been formulated in terms of subjacency (Chomsky, 1981), preventing movements out of all environments that involve crossing more than one bounding node, and with the selection of bounding nodes being subject to parametric variation (Rizzi, 1982).[1] The immunity of resumptive pronouns to island constraints in turn follows from the fact that they are not derived via movement. To take another example, the oft-cited ban on subject resumptives (the 'Highest Subject Restriction') has been explained in binding theoretic terms as resulting from an antilocality constraint which

---

[1]In early transformational formulations unbounded dependency constructions were frequently derived by way of a rule which deleted bound pronouns in certain contexts when bound by the head of a relative clause construction, for instance (Morgan, 1972; Perlmutter, 1972; Bresnan and Grimshaw, 1978; McCloskey, 1979). Island constraints were discussed as a sensitivity to the difference between chopping rules (movement and deletion) and copying rules (those which leave a pronominal copy in the origin site).

affects pronouns, but does not apply to traces (Borer, 1984; McCloskey, 1990; Aoun and Choueiri, 1996):

(6.1) The A-bar-Disjointness Requirement

A pronoun must be A-bar-free in the least complete functional complex containing the pronoun and a subject distinct from the pronoun (McCloskey, 1990).

Such constraints are highly abstract and derive their meaning only within a theory of Universal Grammar. As Hawkins asks "Why should complex NPs be less hospitable environments for gaps than other structures? Why are there so-called wh-islands? Why are resumptive pronouns found in place of gaps in some syntactic positions and not others? We are simply told that this is the way things are" (1999, 245).

One can, of course, adopt the innatist position and view Universal Grammar as the ultimate locus of explanation for these constraints. Yet it turns out that it is rather difficult to identify the appropriate categorical universal constraints and parameters that would be required to capture the range of variation. The core of the problem is that many aspects of the distribution of resumptive pronouns and gaps are, quite notoriously, not categorical. There are two senses in which this is true. The first is cross linguistic: as far as I am aware, it is impossible to identify any universal constraints on resumptive/gap distributions, that is, to find constraints governing the distribution of resumptives/gaps which can be shown to hold across all languages. The second is language internal. Many putative categorical constraints argued to hold within a language in the domain of resumption can also shown to be gradient. Below, I discuss each of these in turn.

6.1.1 Cross-linguistic variation

Consider the Highest Subject Restriction. Although it is a robust typological tendency that gaps are more prevalent that resumptives in highest subject position, across languages, the highest subject 'restriction' is by no means a universal constraint. There are many attested cases of languages which do allow highest subject resumptives. Examples include Colloquial Spanish (Suñer, 1998), Yiddish (Ariel, 1999), Aoban (Keenan and Comrie, 1977), Urhobo (Keenan and Comrie, 1977), Yoruba (Cable, 2003) and Yucatec Mayan (this dissertation). In the other direction, there are languages that require neither the subject, nor

the object in the highest clause to be resumptive. These include Russian and Makassarese and also Literary Welsh, as shown in (6.2) (Sells, 1984):

(6.2) Welsh (Sells, 1984, 133)

*y llyfr y   darllenais i ef
the book COMP read   I it
'the book that I read'

Finally, there are also differences with embedded subjects. In Irish, subjects of embedded clauses may serve the resumptive function (McCloskey, 1990). In Welsh, they must. But in Lebanese Arabic (Aoun and Choueiri, 1996; Aoun, 2000; Aoun et al., 2001) even embedded subjects may not be realized resumptively.

It is also possible to detect construction specific differences in certain languages. In German, embedded dependencies involving gaps which cross complementizers are permitted in wh-questions (6.3a), but not in relative clauses (6.3b) (Kvam 1983, cited in Hawkins 1999):

(6.3)   a. *Wen glaubst du,   dass Maria gesehen hat*
           who think   you   that Maria seen   has
           'Who do you think that Mary saw?'

        b. **Die Person, die   du   glaubst, dass Maria gesehen hat*
           the   person who you think   that Maria seen   has
           'the person who you think that Maria saw'

The same distinction has been observed for Russian (Comrie, 1973). The mirror image of this tendency has been noted for resumptive pronouns: they are more likely to occur in relative clauses than in wh-questions (Boeckx, 2003). As Hawkins (1999) observes, most current generative theories assume a single parameter setting/constraint for a given language, and do not predict this kind of construction-specific difference.

A similar point can be made for island sensitivity. While, again, it is a robust tendency that gaps do not occur in island environments while resumptives may, it is possible to identify languages where resumptives are sensitive to islands. In Igbo, NPs and wh-clauses are islands for both gaps and resumptives (Goldsmith, 1981). Resumptives in Hebrew free

relative clauses also obey island constraints (Borer, 1984). Toba Batak allows neither gaps nor resumptives in complex NPs (Hawkins, 1999).

It is also not difficult to find cases where, conversely, gaps are not sensitive to islands. This has been documented most famously for Northern Germanic languages: Swedish (Allwood, 1982; Engdahl, 1982; Andersson, 1982), Danish (Erteschik-Shir, 1973) and Norwegian (Taraldsen, 1982), as well as in Akan (Saah and Goodluck, 1995).

The consequence of this cross-linguistic variation is that the specific constraints preventing or requiring resumptives in certain syntactic positions must be stipulated for each language on a case by case basis. This is, of course, satisfactory in so far as the objective is to generate a descriptively adequate grammar of resumption in an individual language, and in so far as the constraints can be demonstrated to be categorical. It does, however, result in an absence of motivation or explanation for the constraints, and it does not speak to the robust typological *tendencies* that nevertheless exist.

6.1.2 Within-language gradience

Close inspection of individual languages often reveals in any case that the constraints argued to hold within a language do not always hold absolutely, and so even the solution to the ostensibly simpler task of formulating the set of categorical constraints/parameters for a given language can be elusive.

The graded nature of acceptability of different types of filler-gap dependencies has long been observed in the theoretical literature (see Hofmeister and Sag 2009 for a comprehensive overview). Acceptability may be affected by non-structural factors, as Ross himself observed for syntactic islands in English (Ross, 1967). Various factors that do not affect the number of barriers or cyclic nodes crossed by the dependency may still have a noticeable effect on the acceptability of island constructions involving gaps in English. Chomsky (1973) observes an asymmetry in acceptability between examples like those in (6.4), in which the lexical realization of the complementizer is distinct:

(6.4) a. What crimes does the FBI know how to solve?

  b. *What crimes does the FBI know whether to solve?

The specificity or referentiality of the island-forming head noun can also ameliorate the

severity of island violations in English (Kluender, 1998) ('≥' stands for 'sounds more acceptable than'):

(6.5)   a. That's the article that we need to find someone who understands ≥

         b. That's the article that we need to find the reviewer who understands

Similar effects have also been found for Spanish in acceptability judgment tasks (Goodall, 2004). Gradience in acceptability judgements has also been reported in magnitude estimation experiments by Hofmeister et al. (2007), who show how the manipulation of the non-syntactic factors such as the informativity of the filler affects acceptability ratings of superiority violations of multiple wh-questions.

Similar gradience has been documented for resumption.[2] In Hebrew, for example, while highest subject resumptives are generally stated to be categorically impossible (6.6a), it has been noted in several instances that they become more acceptable where topicalized constituents intervene between the RC head and the resumptive (Givón, 1973; Doron, 1982; Ariel, 1999) (6.6b):

(6.6) Hebrew (Ariel, 1999, 223)

       a. *Ha makhela she ø/ *hi   hirshima   oti beyoter*
          the chorus   that [she] impressed me   most
          'the chorus that impressed me most'

       b. *Ha makhela she oti hi   hirshima   beyoter*
          the chorus   that me she impressed most
          'the chorus that impressed ME most'

Borer (1984) has offered a structural analysis of these facts, in which subject RPs are allowed in relative clauses which contain a topicalized NP by positing a topic node for Hebrew to the right of the COMP node. However, as (Ariel, 1999) points out, not all topicalizations result in equally acceptable subject RP structures. Givón (1973) and Ariel (1990) show that topicalized time adverbials have a weaker effect on the acceptability of subject RPs than topicalized referential NPs do (Givón notes furthermore that the longer the adverbial, the more acceptable the RP becomes):

---

[2]The literature on filler-resumptive dependencies and gradience is much less extensive, by comparison with that of filler-gap dependencies.

(6.7) *Ha   yalda ha-samuda she   tamid   ø/?hi meaxeret*
      The   sweet girl   who always ø/she is-late
      'The sweet girl who is always late'

The definiteness of the antecedent has also been reported to affect the acceptability of Highest Subject resumptives. Highest Subject resumptives become more acceptable if the antecedent is indefinite in Spanish (Tarallo, 1986; Silva-Corvalán, 1996; Suñer, 1998), Hebrew (Meiri and Nitzan 1995, reported in Ariel 1999), and Yiddish and English (Prince, 1990).

In some languages where resumptives are not tolerated in bare wh-questions, they become more acceptable in which-NP questions (Yucatec Mayan, chapter 3; Spanish, Suñer 1998).

6.2 RP/gap variation from the perspective of processing typology

An alternative and fruitful approach to accounting for asymmetric patterns of variation in language is to appeal to system external factors such as processing pressures (Hawkins, 1994, 2004, 2007; Gries, 1999; Ferreira and Dell, 2000; Race and MacDonald, 2003; Temperley, 2003; Jaeger, 2006; Jaeger and Wasow, 2007; Frank and Jaeger, 2008; Hofmeister, 2007; Hofmeister and Sag, 2009, inter alia). 'Processing typology' is the name coined by Hawkins in a recent progammatic paper, to describe the approach of using processing preferences not just to account for statistical patterns of variation within languages, but also, via their conventionalization, for the implicational patterns of categorical constraints which are detectable across grammars. As I described above, Hawkins refers to this as the 'Performance-grammar correspondence hypothesis':

(6.8) Performance-grammar correspondence hypothesis (PGCH): Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments (Hawkins, 2004).

The connection between usage and grammar follows from the observation that the statistical preferences exhibited by some languages may emerge as categorical grammatical

constraints in other languages. For instance, Givón (1979, 26ff) observes that many languages prohibit referential indefinite NPs in subject position, while others allow them but still show a clear preference for definite subject NPs. Bresnan et al. (2001) observe that some languages (such as Lummi) do not allow passives with first or second person agents, while other languages (such as English) show a significantly lower frequency of passives with first or second person agents. And Hawkins (1994, 2004) derives the typologically observed preference for SVO languages from the same principles of efficient processing that predict the short-before-long preference in head-initial VP languages like English and the long-before-short preference in head-final languages like Japanese and Korean. It may be noted, that on the classical generative view, according to which grammars are autonomous and are unaffected by performance factors, parallels such as these remain inexplicable.

Processing typology can be thought of as representing one particular dimension of the general usage-based approach to language advocated here, which understands language structures as emergent phenomena, influenced by the regularities of language use. On the view that variation in language use shapes language structure over time, processing preferences can be considered an important (though by no means the only) factor influencing linguistic variation. Diachronic change is understood to be the key mediating mechanism that allows performance factors to shape grammars (Bybee, 1988; Keller, 1994; Kirby, 1999; Nettle, 1999; Haspelmath, 1999, 2004; Croft, 2000a; Hawkins, 2004). Highly frequent patterns may become conventionalized, while statistically rare constructions may disappear, resulting in the loss of variability (what Haspelmath 2004 refers to as the 'diachronic filtering of grammaticalization').

Cross-linguistic patterns of RP/gap distributions have been something of a signature phenomenon for processing typologists. It has long been observed (e.g. Givón, 1976; Keenan and Comrie, 1977; Ariel, 1999; Hawkins, 1999, 2004, 2007), that the patterns of preference in performance that are detectable within a language that allows both gaps and RPs are the same patterns of variation that can be detected across languages, in terms of the fixed conventions of grammars.

Various attested typological patterns of RP/gap distribution have been shown to be mirrored within languages as statistical preferences. For example, the typological patterns of RP/gap distribution codified by the Accessibility Hierarchy (Keenan and Comrie, 1977) have been documented within languages in terms of statistical tendencies, as summarized

by Ariel (1999): Tarallo (1986) finds higher proportions of resumptives in lower syntactic positions in Brazilian Portuguese. Labelle (1990) finds that French children used RPs and resumptive NPs more the lower the NP rel is on the NP accessibility hierarchy. Pérez-Leroux (1995) finds similar facts for English and Spanish speaking children.

Definiteness effects are also evident. Shlonsky (1992) states that standard Arabic *requires* an RP when the head is indefinite. Similarly, in Greek resumptive pronouns are described as unacceptable in relative clauses with definite heads, but optional with indefinite ones (Tsimpli, 1999; Alexiadou and Anagnostopoulou, 2000). In Hebrew, Yiddish, Spanish and English, these indefiniteness effects show up as statistical tendencies. Tarallo (1986) for example, finds a preference for RPs when the head is indefinite (2.9 times more). Prince (1990) has similar findings for Yiddish and English - RPs are significantly more prevalent in restrictive relative clauses in which the head is indefinite. Silva-Corvalán (1996) has similar findings for Chilean Spanish, where most RPs occur in restrictive relative clauses when the head is indefinite.

As the above examples show, there is positive evidence in support of the performance-grammar correspondence hypothesis in the domain of resumptive/gap distributions. However, it is important to note that there is still a severe lack of controlled production data to provide strong quantitative and broad cross-linguistic support for the correspondence.

In table (6.1) I provide a list of the hard/soft constraint correspondences I have been able to glean from the typological literature on resumptive/gap alternations. The languages in the 'Hard constraints' column are those that have been reported to categorically ban resumptive pronouns in the condition to the left of the '<' sign, but allow them to the right (e.g, disallow them for highest subjects, but allow them for highest objects). The languages in the 'Soft constraints' column exhibit variation, but show a tendency to disprefer resumptives or exhibit a lower frequency of them in the condition to the left of the '<' sign relative to the condition to the right of the '<' sign. Yucatec is in parentheses in order to indicate that the evidence is based only on limited acceptability judgement data (but see §6.4).[3]

In §6.4 I present three experiments which provide some of the first controlled production studies to support the performance-grammar correspondence hypothesis in the domain of RP/gap distributions. Before we turn to the experiments however, it is first necessary to introduce the different processing-based theories of RP/gap variation that exist in the

---

[3]It also remains to be seen whether so-called categorical constraints reported for the various languages in the left column are actually more gradient than currently assumed.

	Hard constraints	Soft constraints
Embedding (Highest Subj < Embed Subj)	Irish Swahili Berber Welsh Hebrew Jakalteko	(Yucatec)
Construction type (Wh < RC)	Hebrew Levantine Arabic Mam Sipakapense Ixil	(Yucatec)
Grammatical function (Subj < obj)	Standard Arabic Gilbertese Slovenian	Brazilian Portuguese Colloquial Spanish
Antecedent definiteness (Indef < def)	Standard Arabic Greek	Hebrew Spanish Yiddish English

Table 6.1: Hard constraints reflecting soft constraints in RP/gap alternations

literature.

6.3   Processing theories of resumption

The notion that the distributional patterns of RPs and gaps have a connection to process-ing has an established history, rooted in two distinct research traditions. In the generative tradition, some scholars have sought to explain the phenomenon of so-called 'intrusive resumption' from a processing perspective. Unlike 'grammatical resumptives', intrusive resumptives (also known as 'processing resumptives') are typically regarded as an auxil-iary phenomenon, on many analyses not part of the grammar proper, but rather something that speakers resort to in certain performance conditions (particularly, increased processing load, Erteschik-Shir 1992; Dickey 1996; Asudeh 2004). English is language which has

been argued to only possess 'intrusive' resumptives.

Processing accounts of intrusive resumptives have been chiefly concerned with ascertaining whether such resumptives really are generated by the grammar, and, if not, determining how it is that they nevertheless occur in production. For some, intrusive pronouns are argued to arise due to online processing pressures, for example due to poor planning (Kroch, 1981), or due to the incremental nature of production which can produce locally licensed, but globally ungrammatical structures (Asudeh, 2004). Others, either on theoretical (Creswell, 2002), or experimental grounds (Ferreira and Swets, 2005) have argued that intrusive resumptives are generated by the grammar, but for some reason are degraded in acceptability.

To date, the processing approaches to intrusive resumptives have not been explicitly developed to predict the variance in the distribution of grammatical resumptives, presumably on the assumption that the two phenomena are categorically distinct. In fact, in the generative tradition, the problem is often carved up such that while 'intrusive' resumptives can be explained as emerging for processing related reasons, the behavior of grammatical resumptives, being generated by the grammar, is outside of the scope of processing considerations. Indeed, because, in this tradition, it is not coherent to say that a form applies more often or less often in a given environment, conditioned variation in the domain of so-called 'grammatical resumption' is not a possible object of study.

By contrast, typologists and functionalists who have focused mainly on so-called grammatical resumptives, tend to assume, either implictly or explicitly, that there is nothing categorically different between grammatical and intrusive resumption. On accounts such as Keenan and Comrie (1977); Hawkins (1999, 2004) and Ariel (1999), discussed below, in which the distribution of resumptives and gaps is thought to represent conventionalized processing preferences, with different languages located at different points on a continuum, intrusive resumption in English merely represents one extreme end of the spectrum of variance.

In this tradition, Keenan and Comrie (1977) were the first to posit a relationship between processing ease and the distribution of gaps and resumptives. They argued that the preference for gaps at the high end of the Accessibility Hierarchy (AH) was explainable by declining ease of processing down the AH. In support of this correspondence, they cited numerous child language studies that show that English children find subject relative clauses

easier to produce and understand than object relatives (e.g. Legum 1975). They also cited corpus studies (e.g. Keenan 1975) which show that textual frequencies for relativizations on all positions of the AH are consistent with this hypothesis.

The greater difficulty of object relatives by comparison with subject relatives has been shown for adults, using different measurement procedures including online lexical decision, reading times, and response accuracy to probe questions (e.g., Ford 1983; Holmes and O'Regan 1981; King and Just 1991; for a review, see Gibson 1998). Neurolinguistic support for the difference between subject relatives and object relatives using ERPs has been presented in King and Kutas (1992, 1993).[4]

Results like these show that resumptives tend to show up in hard-to-process environments. The parallels between typological resumptive distribution and experimentally demonstrated differences in processing difficulty in English unbounded dependencies is not restricted to the accessibility hierarchy. Increased linear or structural distance, for example, has been shown to increase the processing cost of English unbounded dependencies. In general, the longer the distance between the filler and the gap, the harder the dependency is to process (Gibson, 1998, 2000). Correspondingly, as we have seen, across languages, resumptives tend to occur more in embedded environments, where there is a greater distance between the head and the relativized site. The semantic weight of other referential entities along the dependency path has also been shown to increase processing difficulty in English filler-gap dependencies: complex NPs intervening between a filler and a gap impose more difficulty than semantically light, or highly salient NPs (Warren and Gibson, 2002; Gordon et al., 2004). Correspondingly, as we saw above (see example 6.6), the acceptability of subject resumptives in Hebrew is affected by the type of intervening topicalized constituent: intervening timeplace adverbials were preferable with gaps, while heavier referential NPs increased the accepability of resumptives.

But why do resumptives show up in hard-to-process environments? What role do resumptives play in facilitating processing difficulty in these contexts? In the remainder of thsi section I discuss three different theories of resumptive processing.

---

[4]More recent investigations have found that object relatives are not universally more difficult than subject relatives in English; collocational frequency, for example, or the pronominality of the embedded NP can render object RCs easier than subject RCs, see Reali and Christiansen (2007) for an overview. I will return to this point §6.8.

6.3.1 Three theories of resumption processing

The theories of resumption processing discussed below either implicitly or explicitly are framed in terms of what may be described as a resource-limitation/storage based paradigm.[5] According to storage-based theories, the processing of some syntactic structures is taken to require more resources than others do. This resource, generally understood to be some form of memory, is in short supply, which gives rise to greater processing difficulty for structures that are more resource sensitive (Baddeley and Hitch, 1974; Baddeley, 1992; Miyake and Shah, 1999; Andrade, 2001; Baddeley, 2007).

For resource-limitation theories, unbounded dependencies are difficult structures to process due to the particularly high working memory demands they place on the processor. Hawkins (2004, 173) sums up the task:

> Filler-gap dependencies are hard to process, and they are characterized by a heightened processing load and a constant effort to relate the filler to the appropriate gap site or subcategorizer... At the same time the filler must be held in working memory, all the other material on the path from filler to gap/ subcategorizer must be processed simultaneously, and the latter must be correctly identified and co-indexed.

Due to the rapid and incremental nature of language processing, retrieval from memory storage of the right information at the appropriate time must take place efficiently, in order to prevent cumulative processing difficulties. These theories predict that difficulty will be compounded whenever memory resources are put under further strain. As such they are able to account for various factors that have been shown to contribute to the difficulty of processing English filler-gap dependencies. Distance-based effects are explained, because the more time between the initial encoding of some stimulus and the retrieval of that stimulus from memory, the more difficult the retrieval will be (Gibson, 1991, 1998,

---

[5]This is the dominant, but not the only paradigm. A second type of theory has focused particularly on the role of expectations and predictability in syntactic processing (Hale, 2001; Levy, 2005). This idea is closely related to constraint-satisfaction processing models such as those of MacDonald (1993); MacDonald et al. (1994); Tanenhaus et al. (1995) and McRae et al. (1998), and to production models which consider predictability from an information theoretic perspective, according to which speakers structure utterances so as to optimize information density (Aylett and Turk, 2004; Genzel and Charniak, 2002; Jaeger, 2006; Levy and Jaeger, 2007).

2000; Grodner and Gibson, 2005). The semantic weight effect is also predicted on storage based accounts, because semantically heavy referential entities increase the working memory load during the processing of the dependency.

Below I outline the four most spelled out theories of RP/gap alternations that exist in the literature.

Role marking

In early functional approaches, resumptive pronouns were argued to facilitate processing by providing a more explicit surface structure for parsing. Givón (1973, 1975), for example, surmised that resumptives render the RC a 'semantically complete or grammatically simpler sentence'. From this perspective, the role of resumptive pronouns is understood to facilitate processing by explicitly identifying the role of the relativized argument in the relative clause. In support of this view, Givón (1973) shows that in Hebrew, resumptive pronouns become obligatory as soon as the neutral word order within the relative clause is disrupted.

Similarly, Keenan (1972, 1975) and Keenan and Comrie (1977) view resumptive RC strategies as presenting more of the logical structure of the relative clause. Keenan and Comrie (1977) state that RP strategies 'successfully express the basic meaning of the RC in contexts where the meaning is otherwise difficult to perceive'. In other words, a resumptive pronoun gives formal expression to the extraction site in an RC, which is particularly beneficial when the comprehension costs are higher. Tarallo (1986) and Silva-Corvalán (1996) have also theorized that resumptive pronouns facilitate relative clause processing where the relative clause would otherwise be ambiguous or difficult to process for reasons of recoverability.

The notion that RPs function to explicitly identify the grammatical/semantic role of the NP_{rel} is picked up by Hawkins (1994, 1999) in his more elaborated theory of dependency processing. He reiterates the observation that filler-gap constructions are difficult structures to process, placing high demands on working memory, and simultaneous processing decisions. In light of this difficulty, he proposes that the human processor prefers filler-gap domains to be as small as possible (his principle of Minimize Filler Gap Domains). Processing is facilitated the more syntactically/semantically reduced a filler gap domain

(FGD) is. In this way there is less additional processing simultaneously with gap identi-fication. The domain minimization principle accounts for the direction of preferences in languages where there are different available choices for relativization, and, through their conventionalization, for implicational hierarchies across grammars.

For example, he proposes that Keenan and Comrie's Accessibility Hierarchy can be explained by the domain minimization principle. The Accessibility Hierarchy involves in-creasingly complex domains for relativization, calculated in terms of the number of nodes and structural relations that need to be computed in order to match the head of the RC with the position relativized on (the gap). Because, on his theory, an FGD is defined to include a gap's dependent arguments (those on which a gap depends for syntactic or semantic prop-erty assignments), an asymmetry between the complexity of subject and object relatives is predicted: a direct object requires the cooccurrence of a subject, and is asymmetrically dependent on it syntactically and semantically. A subject, by contrast, can occur with one-place predicates, and it is not dependent on a direct object in two-place predicates. The FGD for a direct object gap will therefore always contain an overt subject, and is thus more complex, than the FGD for a subject gap, which need not contain the object. Similarly, an indirect object is assumed to require the cooccurrence of both subject and object, thus indirect object gaps are, in turn, more complex than direct object gaps.

In addition to minimizing filler-gap domains, grammars may avoid filler-gap structures entirely, by conventionalizing structural alternatives. Resumptive pronouns, on Hawkins' view, are argued to facilitate processing, because an empty category does not need to be inferred from its environment, but rather, is expressed formally in the surface structure. On this view, like those of earlier functionalists such as Givón (1973, 1975), and Keenan and Comrie (1977), the resumptive pronoun functions to explicitly identify the head's role in the relative clause, making it 'as clear as it can possibly be'.

While this might seem to predict that resumptives should be preferred across the board, in all positions, Hawkins suggests that a competing principle of economy motivates the preference for gaps in simple environments, where recoverability of the relevant informa-tion is straightforward. This results in a competition between reduced form processing and explicit dependency marking, with increased structural complexity favoring the explicit resumptive strategy.

Domain Minimization

Building on Hawkins (1994) and Hawkins (1999), Hawkins (2004) proposes a general theory of processing complexity in which complexity is measured by the size of the processing domain of a structure and its interaction with general efficiency principles for domain processing (e.g. minimize domains, maximize online processing). Efficiency is increased in accordance with three principles:

   i Minimizing domains (MD) within which certain properties are assigned.

   ii Minimizing the linguistic forms (MF) and maximizing the role of contextual information

   iii Maximizing online processing (MOP) (selecting and arranging linguistic forms to provide the earliest possible access to as much of the syntactic and semantic representation as possible.)

As in previous work, the basic idea for filler-gap/RP dependencies is that the more relations of combination or dependency are involved between the filler and the subcategorizer, the higher the complexity of the dependency and the greater the processing load.

    A crucial departure from his earlier work concerns his analysis of resumptive pronoun and gapping structures. He assumes that gap and resumptive structures involve two distinct dependency types. While resumptive dependencies involve only a relation of coindexation between the locally realised argument (the resumptive) and the filler, gap structures, in addition to the coindexation relation between the gap and the filler, involve a second relation: that of lexical cooccurrence of the filler and the gap within the lexical domain of the predicate. Dependencies involving resumptives are therefore less complex than those involving gaps, because the former only involve one relation of co-indexation. This becomes particularly relevant where the dependency is made more complex, for example, where it is lengthened by multiple levels of embedding. Because information relating to lexical cooccurrence involves every single node intervening between the filler and the gap, the size of the processing domains of gaps (but not pronouns) keeps increasing with embedding. Thus, in more embedded positions, pronouns are correctly predicted to be preferred over gaps.

The gap structure involves more dependency assignments, but it does have a compensating advantage: there is less linguistic form to process (Minimize Forms). Explicit marking (resumption) involves fewer dependency assignments but more processing of linguistic form.

Locality theory

Alexopoulou and Keller's (2007) study is framed in terms of Gibson's (1998) theory of linguistic complexity, the Syntactic Prediction Locality Theory, a storage based model of the comprehension mechanism. It is specifically focussed on the comprehension and acceptability of so-called 'intrusive resumptives', but I mention it here because they suggest that their account may also be extendable to grammatical resumptives, in so far as the two might be viewed as more of a continuous phenomenon. Like Hawkins (2004) they adopt the view that resumptive dependencies and gap dependencies constitute two structurally different dependency types, each associated with different processing costs. They share Hawkins' two central hypotheses (though with different vocabulary) that a) the relation between the filler and the pronominal is anaphoric and therefore is not subject to the locality restrictions of syntactic dependencies involving gaps, and b) the relation between the gap and the filler is a syntactic one, registered through the dependency chain. In a cross-linguistic magnitude estimation study (Greek, German and English) testing the acceptability of object extractions in wh-questions, they found that the presence of resumptives improves the acceptability of wh-extractions out of islands (though they are never restored to full acceptability). They also found that across all three languages the acceptability of pronominals improves with embedding (for *that-* and *whether-*clauses).[6]

On the basis of their results they speculate that "the presence of a resumptive makes the parser abandon the syntactic/cyclic resolution of the dependency and revert to an anaphoric dependency. That is, the pronominal searches for its antecedent, not through the syntactic cyclic route, but in the previous discourse, as in the case of intrasentential anaphora". Because anaphoric dependencies are not subject to locality, this is a legitimate resolution of the dependency. Resumptive dependencies are never restored to full acceptability, however, because the parser has already attempted a cyclic resolution of the dependency which

---

[6]Gaps in English remained significantly better than pronominals in all conditions, while in German and Greek, the resumptive pronouns could be as acceptable as gaps.

is more costly.

Referential form processing and accessibility theory

A different perspective on resumptive pronoun distributions treats the choice between RPs and gaps as driven by the same principles that underlie referential form choice and processing in discourse (Ariel, 1999). One of the basic insights that has come out of research on discourse-level referential form choice is that the specificity of a referential expression is inversely correlated with the salience of the referent in the context of the discourse. (Givón, 1976; Ariel, 1990; Gundel et al., 1993). For example, pronouns, which convey little information about their referent, are primarily used for referents that are salient in the discourse. In contrast, proper names and definite descriptions, which typically unambiguously identify their referent, are often used to introduce new referents to the discourse or to make reference to discourse referents that are not salient (Ariel, 1990; Gundel et al., 1993; Garrod and Sanford, 1994; Gordon and Hendrick, 1998). On many views, 'salience' or 'accessibility' is understood as ease of retrievability in working memory. The multiplicity of referential forms is treated as solution to difficulties imposed by memory constraints: referential form choice on this view is a function of retrievability and pragmatic principles (Almor and Nair, 2007). Wherever an antecedent is difficult to retrieve from memory, more information is needed to aid that retrieval process. Where retrieval costs are minimal, excessive information is perceived as marked, because information is assumed to have a pragmatic purpose.

Ariel's Accessibility Theory views referring expressions as marking varying degrees of mental accessibility of the antecedent. On this approach, each referential form is tied to a unique point along an 'accessibility scale' such that higher positions on the scale correspond to greater accessibility of the antecedent. Full names and definite descriptions are low points on this scale, demonstrative expressions (this, that) are on the middle, and pronouns (he, she) are on the high end. Zeros, or gaps, are the highest point. Referent accessibility is argued to be affected by (i) distance from the last mention of the referent, (ii) competition with other possible referents, (iii) contextual salience of the referent, and (iv) unity (whether the previous mention of the referent is in the same or previous sentence, or in the same or previous paragraph). The way referential form is used to convey the accessibility of the referent is tied to the use of Gricean pragmatic principles (Grice, 1975).

Speakers choose the referential form so as to encode the accessibility of the referent and thus aid listeners to successfully identify that referent.

Ariel suggests that accessibility effects detectable in discourse-level referential form choice and processing are also operative intra-sententially, governing the distribution of resumptives and gaps. She observes that in the case of unbounded dependencies, the relativised NP position exhibits only a subset of the range of referential expressions available at the inter-sentential level, being restricted for the most part to pronouns and zeros (gaps). This follows from the fact that relative clause heads are highly salient, since they are marked as necessitating a relativised position coreferent with them. On her account, just as in discourse, where less accessible antecedents are correlated with more informative subsequent referents, so too in unbounded dependencies, less accessible antecedents trigger less attenuated anaphoric forms (resumptives). The same factors are relevant, and, according to her proposal, can account for the distribution of gaps and resumptives: i) distance, ii) competition with other referents, iii) salience of the referent, iv) unity (in this case, whether the relative clause is tightly bound to matrix clause). For example, on her view, indefinite antecedents are less accessible because they introduce a new referent into the discourse, and as such have not been activated by previous mention. This can explain the preference for resumptives with indefinite relative clause heads within and across languages. Embedding is predicted to correlate with a greater preference for resumption, due to the increased distance between the antecedent and the anaphor, which decreases the accessibility of the antecedent.

Summary

Role marking theories attribute the function of the resumptive to easing the processing load by explicitly marking the relativization site. In other words, the resumptive facilitates identification of the relativization site. Domain minimization and Locality theories ascribe a facilitating role to resumptives because they are assumed to be associated with a different (simpler) dependency relation which has a reduced processing cost. Accessibility theory assumes the resumptive provides a more informative cue for accessing the antecedent. That is, the function of a resumptive has primarily to do with facilitation of retrieval of the antecedent.

6.4 An experimental investigation of Yucatec RP/gap alternations

While a good deal of experimental research has been undertaken on filler-gap dependencies, filler-resumptive structures have received decidedly less attention. This is an obvious consequence of the fact that English, the target language of most processing research (see Jaeger and Norcliffe 2009), does not have productive resumption.[7] To date all of the experimental studies on resumption have focussed exclusively on the phenomenon of intrusive resumption.[8] For productive RP/gap alternations, the empirical evidence on which processing theories are based has come mainly from small corpus analyses, observations of written and spoken discourse, and impressionistic acceptability judgments. There are, as far as I am aware, no experimental production studies of productive RP/gap alternations.

The Yucatec experiments that I present below are a modest first step at procuring more controlled, quantitative data on a productive RP alternating language. In its own right, Yucatec presents an interesting test-case for processing theories of resumption. In Yucatec, along with a subset of morphological RP languages, the presence of the resumptive element can create *ambiguous strings*. As we will see below, this divergent property of Yucatec can help us determine between some of the competing theories of resumption processing.

6.4.1 The ambiguity problem

As we have seen, in a language like Yucatec, while the gap-alternant only has one available interpretation (6.9a), the resumptive-alternant has two interpretations (6.9b). Either the subject pronominal can be interpreted as bound to the antecedent, or (in appropriate contexts) it can be interpreted as disjoint from the antecedent, with the antecedent instead co-indexed with the object argument of the relative clause verb:

(6.9)   a. *le   winik ts'uts'ik le   x-ch'uupal-o'*
             the man   kiss-INC   DET   F-girl-D2
             Only: 'the man that kisses the girl'

---

[7]But see Prince (1990) for a corpus study which shows that they are more prevalent in relative clauses than is often thought.

[8]There are only three in total, as far as I am aware: (Dickey 1996; Ferreira and Swets 2005 and Alexopoulou and Keller 2007).

b. *le   winik k-u=ts'uts'ik   le   x-ch'uupal-o'*
   the   man   IMP-A3=kiss-INC   DET   F-girl-D2
   the man that kisses the girl/the man that the girl kisses

For role marking-based approaches to RP/gap alternations, where resumptives are assumed to facilitate processing by explicitly marking the grammatical function of the relativized argument, the facts of an RP alternating language like Yucatec are entirely out of step. The competition Hawkins (1999) proposes, for example, between economy and explicitness isn't a competition in the AF case: the AF verb form is both more economical (it is shorter, at least marginally), and it's more informative (it's unambiguous). Why then do we nevertheless find variation between resumptive and gap structures in Yucatec?

Disambiguation considerations are also built into Hawkins' (2004) domain minimization approach, though perhaps less obviously at first glance. As we saw above, Hawkins proposes three general efficiency principles: (i) Minimize Domains, (ii) Minimize Linguistic Form, (iii) Maximize Online Processing. This last says that linguistic forms should be selected so as to provide the earliest possible access to as much of the syntactic and semantic representation as possible. The relative efficiency of two alternants is determined by how they score according to each of these three principles. For the Yucatec case, while, under certain conditions, the resumptive form might win over the AF form in terms of Minimize Domains (see Appendix A, for a full calculation), the AF alternant will always score higher in terms of the other two efficiency principles, Minimize Form and Maximize Online Processing. Again, this is because the resumptive form is both longer, and can produce globally ambiguous strings. Hawkins does not offer any metric for assessing the relative strength of these three principles when they produce mutually conflicting results; at the very least we might assume that a much weaker preference for resumptives in ambiguous contexts would be predicted, given the stipulation of the MOP.

The Yucatec case thus leads to rather unclear predictions for Hawkins (2004), which recognizes a contributing but not exhaustive role for an efficiency principle which prefers unambiguous structures. For role-marking theories, according to which the sole function of a resumptive is to facilitate the identification of an extraction site, the prediction is very clear: Yucatec's gapping structure will be preferred where resumptives are preferred in other languages, because it is the gapping structure that unambiguously identifies the relativized argument.

This immediately recalls the recoverability oriented accounts of Mayan AF, which, as I showed in chapter 2, are the only theories of Mayan AF in the Mayanist tradition developed to account for the variation between the two verb forms. It is worth reiterating that disambiguation oriented theories in the Mayanist tradition do not form a unitary class, and so themselves make divergent predictions about the variation. For some, the primary function of the AF verb is to disambiguate where context will not (Mondloch, 1978; Gutiérrez-Bravo and Monforte, 2009). On harmonic alignment accounts, by contrast, AF is argued to be sensitive to proto-typical relations between grammatical role and participant properties such as topicality/definiteness/animacy (e.g. Aissen, 2003). The AF verb is the more marked verb form which is used just in case the intended interpretation of the string is the non-canonical one, that is, where the subject is ranked lower than the object on the animacy/topicality/definiteness scale.

6.5 Experiment 1: Embedding and Definiteness

Most broadly, the purpose of experiment 1 is to test whether patterns of RP/gap distribution that have been attested across grammars are observable in production choices in Yucatec. The focus is on two factors that have been reported in the typological literature to affect the likelihood of resumption: definiteness of the antecedent and dependency length/complexity.

Various languages have been reported in the literature to have categorical constraints pertaining to both these factors. Perhaps most famous (from the perspective of generative theory) is the set of languages which exhibit the so-called 'Highest Subject Restriction', according to which subject resumptives are banned from the most local clause (in the position subjacent to the antecedent) but are possible in more embedded environments (see, e.g. McCloskey 1990). Languages that have been reported as possessing this categorical constraint include Irish, Swahili, Berber, Welsh and Hebrew (see chapter 3, and also §6.2 above).

In chapter 3 I provided some preliminary data from acceptability judgments which showed that in Yucatec the resumptive structure is preferred particularly in embedded environments. But do we find this to be the case in spoken production? If we find an effect of this factor in experiment 1, this will provide the first controlled and statistically reliable

evidence from an on-line production study that production preferences do in fact mirror attested typological patterns of RP/gap distributions. It will therefore offer firm support to a grammar-external explanation for a widely cited 'UG' constraint.

Various languages are reported to exhibit definiteness effects on resumption choice, including Spanish, Yiddish, Hebrew and even English (see §6.2) though the data have all come from small corpus studies and impressionistic judgments. In at least two languages, a definiteness effect has been reported as categorical: in Standard Arabic a resumptive pronoun is apparently obligatory when the antecedent is indefinite. In Greek, gaps are obligatory with definite heads, while resumptives become optional with indefinite heads. Again, if we find an effect of definiteness in the Yucatec study, this will provide the first evidence from spoken production for a correspondence between performance and grammar in this domain.

More specifically, experiment 1 tests the predictions of several of the RP processing theories outlined above.

ROLE MARKING theories predict that resumptive pronouns will be preferred under conditions of increased complexity, because they function to explicitly identify the extraction site. Because it is the AF (subject gap) form which is the most informative in the Yucatec case, we might expect the inverse relation for Yucatec: the AF form, being more explicit, may be preferred under more complex conditions, i.e., where the dependency involves embedded clauses.

DOMAIN MINIMIZATION predicts a stronger preference for resumptives in more complex dependencies. Because information relating to lexical co-occurrence involves every single node intervening between the filler and the gap, the size of the processing domains of gaps (but not pronouns) keeps increasing with embedding. Thus, in more embedded positions, pronouns are predicted to be preferred over gaps. However, because Yucatec's AF structure is unambiguous, while the RP form is ambiguous, this effect may be reduced or non-existent in Yucatec, because Hawkins' competing principle of Maximize Online Processing states that structures are preferred which resolve semantic relationships more immediately. If we do find an effect of embedding in Yucatec, then, in the context of Hawkins' theory, this would show that MOP is less relevant for efficient processing than Minimize Domains. Because

processing cost on this theory is directly tied to domain complexity, Hawkins' approach does not directly predict that referential properties of the head NP will effect production choices.

ACCESSIBILITY THEORY predicts that both embedding and indefiniteness of the antecedent will be correlated with higher rates of resumption, because both these factors influence the accessibility of the antecedent.

HARMONIC ALIGNMENT makes no prediction for structural complexity, because the only factors predicted to influence the choice between the AF and the resumptive form are the referential properties of the relative clause participants. The AF form should be preferred (or required, if a categorical analysis is assumed) only when the RC head is outranked by the RC object, in other words, only when the antecedent is indefinite (and the RC object is definite). This therefore makes the opposite prediction to accessibility theory with respect to definiteness.

6.5.1 Location and participants

All experiments reported here were undertaken with student participants from the Universidad de Oriente (UNO), a state university in the Yucatán. This university is located just outside of the town of Valladolid, and is attended by students living throughout the state of Yucatan, with the bulk of the population concentrated in Valladolid and surrounding villages. The students who participated in the studies were all bilingual (Spanish/Yucatec) speakers, who speak Yucatec in the home. They were all computer literate.

28 undergraduate students were paid MX$40 to complete experiment 1, which lasted between half an hour and forty five minutes.

6.5.2 Methodology

The three experiments discussed in this chapter employed a computer-based translation task, programmed and run with Exbuilder (designed by E. Longhurst, for M. Tanenhaus

lab, University of Rochester).[9] In this task, subjects, wearing a head set with a head-mounted microphone, listen to recorded sentences spoken in Spanish (experiments 1 and 3) or Yucatec (experiment 2) and have to translate them out loud into the other language. Each trial of the task involves the following steps. After pressing the space bar to initiate the experiment they first hear a sentence. They have the option to play the sentence again as many times as they wish by clicking on a 'repeat' icon. When they are ready to speak, they advance with another space-bar press. A green icon of a mouth appears which signals that they can begin speaking. Once they have produced their Spanish translation, they proceed to the next sentence by means of another space bar press. Responses are automatically recorded onto the computer in WAV format, and automatically labeled with subject, experiment, item and condition numbers.

The task begins with an initial training session with the experimenter in the room in order for the participant to get accustomed to the computer task. Instructions are presented orally to the subject by the experimenter in Spanish. The training session consists of 6 sentences (unrelated to the experimental items).[10] Once the experimenter is satisfied that the subject has understood the task, and is comfortable with the computer commands, he/she is left to complete the actual experiment.

Prior to the initiation of the experiment, some basic background information is collected about each subject, including age, gender, approximate number of hours of Yucatec spoken per day, and the village in which they grew up (in order to test for dialectal variation). None of these factors proved to be significant predictors of the variation in the experiments reported here.

It is worth noting that production studies of English typically involve recall-based tasks.[11] Recall tasks were successfully piloted with Yucatec subjects. However, it was decided ultimately that translation tasks were a more appropriate means of eliciting production data, for several reasons.

First, translation tasks are ecologically sound in a bilingual community where translation is a familiar practice. Second, by removing the speaker from a conversational setting

---

[9]I owe Florian Jaeger and Andrew Watts a debt of gratitude for doing the programming for this and the other two experiments, and to Florian Jaeger and Philip Hofmeister for their assistance with the design.

[10]These consisted of simple intransitive sentences with prepositional phrases.

[11]See e.g., Bock and Brewer (1974); Bock and Irwin (1980); Ferreira and Dell (2000); Fox Tree and Meijer (1999).

and using a cued-recall situation to plant a message in the speakers mind, recall tasks compromise naturalness in order to obtain experimental control. Like recall tasks, the translation tasks used in the Yucatec experiments also removed speakers from a conversational setting (speakers are interacting with computers), and as such it might be argued that such a task might also not create an appropriate communicative environment which would, for example, be realistic enough to elicit ambiguity avoidance effects (relevant in experiment 3). However, translation tasks do place an intrinsic emphasis on interpretation in a way that recall tasks do not. One might suppose that subjects will be more responsive to potential ambiguities when producing a translated utterance, than they would when repeating a previously heard utterance. More generally, translation tasks provide a very useful check that subjects fully understand what they have heard.

Using translation tasks does create a different set of issues, however. While the experimental designs for the input were counter balanced 2 X 2 Latin-square design, subjects did not always consistently maintain the independent variables in their translations. That is, they did not, for example, consistently produce a definite determiner where there was a definite determiner in the input. In effect, what the translation task does is allow for the controlled elicitation of a spoken corpus of Yucatec relative clauses. Luckily, skewed data do not pose a problem in the analyses reported here, because the data are modeled using logistic regression. Because logistic regression directly models probability without assuming a particular distribution of data, it is robust for skewed and otherwise non-normally distributed data.

6.5.3 Materials

The stimuli in experiment 1 consisted of 24 relative clause constructions, which varied with respect to embedding (simple vs. embedded) and the definiteness of the relative clause head (indefinite vs. definite). For each item, the relative clause was embedded as the object of a transitive matrix clause, whose subject was always a proper name. In the embedded conditions, the embedded clause subject was also always a proper name. The RC object NP was always definite and inanimate.

Four stimuli sets were constructed in such a way that participants heard all items, but only one version of any given item. All participants heard the same number of each condition. The experiment consisted of 24 items, plus 32 fillers, which were arranged such that

DEF-SIMPLE	Rodrigo se burló **del comensal** que derramó la bebida
	(Rodrigo made fun of **the diner** that spilled the drink)

INDEF-SIMPLE	Rodrigo se burló **de un comensal** que derramó la bebida
	(Rodrigo made fun of **a diner** that spilled the drink)

DEF-EMBED	Rodrigo se burló **del comensal que María dijo** derramó la bebida
	(Rodrigo made fun of **the diner that Maria said** spilled the drink)

INDEF-EMBED	Rodrigo se burló **de un comensal que María dijo** derramó la bebida
	(Rodrigo made fun of **a diner that Maria said** spilled the drink)

at least one filler always intervened between any two of the experimental items.

6.5.4 Coding

Responses were coded by the experimenter, who has competent comprehension of Yucatec. The translated productions were coded for the two predictors: definiteness of the antecedent, presence/absence of embedding and the dependent variable (verb type). A subset of responses were checked for consistency against full transcriptions of the response sentences undertaken by native Yucatec linguistic students from the UNO who are trained in Yucatec orthography and transcription and who did not participate in the experiment.

Translated productions were included for analysis provided that the speaker produced a transitive relative clause construction. All trials in which the speaker did not produce a relative clause construction, for example, due to paraphrasing or incomplete sentences were eliminated. In addition, subjects were excluded who were determined to be non-variable speakers.[12]

Initial examination of the data revealed an important observation. What was expected to be a binary response variable (AF vs. full verb form), turned out to be ternary: in addition to the availability of the AF form (6.10a) and the full verb form (6.10b), a third relevant response type was observed, in which the relative clause subject was expressed by

---

[12]This subject type could be subdivided into two types, those who consistently produced the synthetic verb form, and those who consistently produced the AF verb form.

an independent resumptive pronoun *le'eti*, in the preverbal focussed position (6.10c):

(6.10)   a. *Rodrigo-e'   t-u=leets'-ah   le   komensal t-uy=a'al-ah*
              Rodrigo-TOP PRV-A3-make.fun-CMP the diner   PRV-A3-say
              *Maria **wek** le   ukul-o'*
              Maria spill DET drink-D2
              'Rodrigo made fun of the diner that Maria said spilled the drink'

           b. *Rodrigo-e'   t-u=leets'-ah   le   komensal t-uy=a'alah Maria*
              Rodrigo-TOP PRV-A3-make.fun-CMP the diner   PRV-A3-say Maria
              ***t-u=wekah**   le   ukul-o'*
              PRV-A3=spill DET drink-D2
              'Rodrigo made fun of the diner that Maria said spilled the drink'

           c. *Rodrigo-e'   t-u=leets'ah   le   komensal t-uy=a'alah Maria*
              Rodrigo-TOP PRV-A3-make.fun-CMP the diner   PRV-A3-say Maria
              ***le'eti**   wek le   ukulo-'*
              PRON.3sg spill DET drink-D2
              'Rodrigo made fun of the diner that Maria said spilled the drink'

Because this third strategy involves the use of an independent pronominal element, it was decided that these responses should be included in the analysis. How does this independent resumptive strategy pattern compared to the use of the dependent resumptive form, and the AF (subject-gap) form? All the existent literature on resumptive/gap alternations presupposes the existence of only a binary alternation. What happens when there are three potential means of terminating the dependency? In order to compare the three alternants, a series of binary mixed model logistic regressions were fit to subsets of the data, with subjects and items as random factors. In the results section, I report the results of each binary analysis separately, and then proceed to a more general discussion.

6.5.5 Results

The results of all experiments reported here are analyzed using mixed-effect logit models (Bates and Sarkar, 2007), which are appropriate for analyzing categorical data (Baayen, 2007; Jaeger, 2008). They can be understood as predicting the probability of a specific response (e.g. a verb type) in the different conditions (see Agresti 2002; Jaeger 2008). Mixed logit models are an extension of logistic regression that allows for the modeling

of random subject and item effects. This is necessary in order to generalize beyond the subjects and items in the current study (Clark, 1973).

In mixed logit models coefficients are given in log-odds (the space in which mixed logit models are fitted to the data). If a significant coefficient is positive, this shows that the tested verb type is more likely in the tested level of the variable than in the other. For example, in the first experiment, Embedding has two levels, SIMPLE and EMBEDDED. If the tested level is EMBEDDED, then if the coefficient of Embedding for EMBEDDED relatives is positive, then the AF verb is more likely for embedded relative clauses than for simple relative clauses. Negative coefficients show the opposite. If the coefficient of Embedding for EMBEDDED relatives is negative then the AF verb is less likely to be produced in embedded relatives than in simple relatives.

For each result, I report the coefficient (B) for each independent variable, and its level of significance. I also give the difference in odds (e^B) between conditions (odds are the exp (log-odds)).

(a) AF verb vs. resumptive verb

The dependent variable for analysis (a) was the verb type: AF (gap structure) vs. the resumptive verb (i.e., independent resumptive pronoun responses were excluded). A mixed logit model with Embedding type and Definiteness as fixed effects, and item and subject as random effects, showed two main effects and no interactions. In the model both embedding and the definiteness of the head NP are highly significant predictors of the verb form choice: the AF (gap) structure was less likely to be produced in the embedded condition than in the simple condition (B = -1.12 (SE = .34), $p{<}0.001$; e^B =.32) and was less likely to be produced with an indefinite head than with a definite head (B = -0.83 (SE = 0.33, $p{<}0.01$; e^B=.43). In other words, embedding and indefiniteness of the antecedent both independently favor the dependent resumptive structure. The overall quality of the model is good. The concordance statistic C = 0.88 and Somers Dxy = 0.76. These figures indicate the ability of the model to discriminate between all pairs of the observations that differ in verb type.

(b) AF verb vs. independent resumptive pronoun

In analysis (b) the response variable was the AF verb (gap structure) vs. the independent

pronoun (i.e., dependent resumptive responses were excluded). A mixed logit model with Embedding type and Definiteness as fixed effects, and item and subject as random effects again showed two main effects and no interactions. In the model both embedding and the definiteness of the head NP are highly significant predictors of the verb form choice: the AF (gap) structure was less likely to be produced in the embedded condition than in the simple condition (B = -4.44 (SE = .47), $p<$0.000; e^B =.01) and was less likely to be produced with an indefinite head than with a definite head (B = -1.13 (SE = 0.44, $p<$0.01; e^B=.32). In other words, embedding and indefiniteness of the antecedent both independently favor the independent resumptive structure. The model quality is excellent: Somer's C = 0.9508285; Dxy = 0.9016569.

(c) Independent resumptive pronoun vs. resumptive verb

The final analysis modeled the likelihood of the independent resumptive pronoun vs. the dependent resumptive form. In the model, embedding is a highly significant predictor of choice of resumptive type. The independent resumptive pronoun structure is more likely to occur in the embedded condition than the simple condition (B = 3.3 (SE = .39, $p<$0.000; e^B =27.1). Definiteness is not a significant factor (p>0.82). The model is an excellent overall fit; C = 9171411; Dxy = 0.8342822.

6.5.6 Discussion

The three binary regression analyses revealed that both embedding and definiteness are significant predictors of the use of the AF verb in production. The AF verb (the gap alternant) is significantly less likely in the embedded condition than in the simple condition, and significantly less likely in the indefinite-antecedent condition than in the definite antecedent condition.

These results are interesting and relevant in many respects. First, and most broadly, they demonstrate that Yucatec variation in production mirrors the patterns of RP/gap distributions attested across grammars. The 'Highest Subject Restriction', which is apparently a categorical constraint in many languages (including Jakaltek, which is genetically related to Yucatec), is reflected probabilistically in Yucatec in terms of production preferences. The definiteness restriction (resumptives are required with indefinite antecedents/gaps are required with definite antecedents), which has been reported categorically in at least two

languages, and probabilistically in corpus studies for several others, is also visible probabilistically in Yucatec production. As such, these results provide some of the first robust support from production data for the Performance-Grammar Correspondence hypothesis in the domain of RP/gap variation.

When we compare the results across the three analyses, a very interesting picture emerges. In analysis a), which modeled the likelihood of gaps vs. dependent resumptives, the dependent resumptive was more likely in the embedded condition than in the simple condition. In analysis c), which modeled the likelihood of dependent resumptives vs. independent resumptives, the independent resumptive turned out to be more likely in the same condition. That is, the asymmetrical frequency distribution of gaps vs. dependent resumptives matches that of dependent resumptives vs. independent resumptives. Thus, the level of formal attenuation of the referential expression concluding the dependency is appears to be inversely correlated with the complexity of the dependency, as measured by embedding depth (and with the salience of the antecedent, as measured by definiteness, though with the caveat that definiteness wasn't a significant predictor of the choice between the bound resumptive and the independent resumptive structure).

$$\text{gap} < \text{dependent resumptive} < \text{independent resumptive}$$
$$\text{less complex} < \text{more complex}$$

Figure 6.1: Referential form attenuation and dependency complexity in Yucatec relative clauses

These results are consistent with results reported for referential form choice in discourse, where specificity of referential form is correlated with factors such as distance and the referential salience of the antecedent (shorter distances and greater salience of the antecedent = more attenuated forms, Givón 1976; Ariel 1990; Gundel et al. 1993). I will come back to this point below when I discuss these results in the context of Accessibility Theory.

Let us consider now more specifically how the different theories of RP/gap processing discussed above fare in handling these results. For role-marking theories, the Yucatec results are the opposite of what is predicted. Despite the fact that the AF form provides a more informative, unambiguous structure than the dependent resumptive structure, we do

not find a preference for the AF form in conditions of heightened complexity. These results therefore do not support the view that resumptive pronouns facilitate processing primarily by providing a more explicit surface structure for parsing.

From the perspective of Harmonic Alignment Theory (Aissen, 2003), the correlation with structural complexity shows that participant properties are not the sole determining factor driving variation between the AF and the synthetic verb in Yucatec. Thus, while the choice between AF and synthetic verb forms in Tzotzil has been argued to be driven by the relative ranking of *participant* properties, in Yucatec at least, we find a strong effect of *structural* complexity affecting verb form choice. It remains to be seen whether such effects are also visible in Tzotzil. In the context of the thesis that I have developed in this dissertation, that AF alternations in Mayan languages are morphologized resumptive-gap alternations, borne out of earlier syntactic resumptive-gap alternations in proto Mayan, then we might predict Tzotzil to also be sensitive to embedding depth. This would be supported by the fact that the embedding effect has been reported categorically for at least one other language in the family, Jakaltek (where the AF verb is required in Highest Subject contexts, but is optional in embedded environments, Craig 1977).

Moreover, the fact that resumptives were more frequent with indefinite heads than definite ones directly contradicts the predictions of Harmonic Alignment. Because the RC objects were always kept definite across all conditions, it is predicted, on a Harmonic Alignment account, that the AF form will be more likely when the head NP is indefinite, because in such cases the object outranks the extracted subject along the dimension of definiteness. Of course, it must be noted that objects were also inanimate, across all conditions in this experiment. Because objects did not outrank extracted subjects along the dimension of animacy, it might be argued that this has a counteractive effect. However, we would expect then, that this would be relevant across all conditions, not just in those cases where the head NP was indefinite. It remains to be seen whether the Harmonic Alignment effects reported for Tzotzil stand up under more statistically rigorous conditions, so I don't comment on the Tzotzil case further here.

The Yucatec results present an interesting challenge for approaches such as Hawkins' Domain Minimization and Alexopoulou and Keller's (2007) Locality Theory which both view resumptives and gaps as marking two structurally distinct dependency types: gaps mark 'syntactic' dependencies, resumptives mark 'anaphoric' dependencies. In both these

theories, the relative processing difficulties are attributed to the costs associated with processing different dependency structures. Because this reduces the phenomenon to a binary contrast between structural types, these 'dual dependency' theories do not easily allow for the possibility that there may be a cline of resumptive types. Or rather, it could accommodate the presence of different resumptive types, but it would predict the same effect for all resumptive types, because they are, by definition, associated with the same sort of structural dependency.

The Yucatec facts thus cast doubt on the validity of dual dependency approaches, because we have seen that the two resumptive structures, involving bound and independent resumptives, show the same distributional pattern that we find for gaps and dependent resumptives. Moreover, because the notion of processing effort is very narrowly defined according to these theories, restricted to syntactic variables such as 'domain complexity' or 'locality', it doesn't directly predict the antecedent-definiteness effect shown in the Yucatec results. Finally, there is, I believe, a more general question that arises when considering dual dependency approaches. Why should gaps be associated with one abstract structure, and resumptives with another? The assumption of an abstract *structural* distinction between filler-resumptive and filler-gap dependencies is a direct legacy of the generative-transformational tradition (movement vs. base-generation). As we have seen, this theoretical distinction receives its support from the purportedly categorical differences in the syntactic behaviour of resumptives and gaps. However, we reviewed in this chapter the number of cross linguistic and language internal exceptions to such postulated categorical distinctions between gaps and resumptives undermines the view that they mark different structural dependencies. In turn, it is doubtful whether processing models aiming to account for the difference between the production and parsing of gaps and resumptives should be based on hypothesized abstract-structural differences of this sort.

Ariel's Accessibility Theory differs from dual dependency approaches in this respect, because it doesn't postulate that resumptives and gaps are associated with different dependency structures, and measure processing cost based on those structural differences. Rather, what is relevant is the amount of form used at the foot of the dependency to refer back to the antecedent. Wherever an antecedent is difficult to retrieve from memory, more information is needed to aid that retrieval process. Thus, greater formal attenuation is associated with more accessible antecedents. In this respect, it is able to accommodate the

Yucatec results, because it doesn't presuppose the existence of a binary syntactic contrast between dependency structures (gap/resumptive), but rather allows for a cline of referential forms (just as in discourse).

The Yucatec studies most certainly indicate the need to bring discourse-level processing principles into the account of what have often been taken to be autonomously sentence-internal phenomena. However, Ariel's Accessibility Theory does raise rather more questions than it solves. As a whole it provides no independent and cognitive based notion of processing cost. As Almor and Nair (2007) note, this leads to a somewhat circular line of argumentation, where cost is defined on the basis of the accessibility scale, which is what is being explained in the first place. For example, for the case at hand, indefinites are located at the low end of the accessibility scale and thus are assumed to induce a greater processing cost for subsequent retrieval. But why are indefinites less accessible? It becomes rather easy to conveniently label anything as low-accessible when it happens to correlate with high frequency of resumption.

Indeed, in some respects the notion of indefiniteness being intrinsically tied to low accessibility is somewhat counter-intuitive. Indefinites have a use condition requiring very low or zero salience of the referent *prior* to the reference event. But once the indefinite has been introduced, the referent is, presumably, very salient. At the discourse level, at least, indefinites probably often introduce a high expectation that their referent is highly relevant to the ongoing discourse and hence a likely referent for some subsequent pronoun.[13] How appropriate then, is it to talk about an indefinite NP as being a less accessible antecedent than a definite NP in the case of filler-gap/RP dependencies? How might we best explain the cross-linguistic definiteness effects we find with respect to resumption distribution? I will return to these questions in the general discussion.

6.6 Experiment 2: Definiteness and Event Bias (Interpretation)

Experiment 1 showed that despite the fact that the AF form provides a more informative, unambiguous structure, this does not result in a preference for this form in conditions of

---

[13]Thanks to Ivan Sag, p.c., for pointing this out to me.

heightened complexity. However, some recoverability-oriented theories of Mayan AF sug-
gest that the AF form is used mainly where context would otherwise leave the construction
globally unambiguous (e.g. Mondloch (1978). In experiment 1, because the relative clause
embedded NP was always inanimate, there was no chance of global ambiguity, because
the embedded NP could never be interpreted as the subject of the relative clause (drinks
can't spill diners, for example). It is possible then, that under conditions of heightened
ambiguity, we might find a disambiguation effect, in keeping with the prediction of a large
body of functionalist literature on Mayan, and with ambiguity oriented accounts of RP/gap
alternations (Givón, 1973, 1975; Keenan and Comrie, 1977; Tarallo, 1986; Silva-Corvalán,
1996).

Experiment 2 and 3 thus jointly test whether disambiguation plays a role in the choice
between the subject gap (AF) structure and the subject-resumptive structure under such
conditions. A first step to testing disambiguation theories of AF production requires es-
tablishing under what conditions the string is particularly biased towards an object relative
reading in comprehension: from a disambiguation perspective, we would expect to see the
strongest preference for choosing the AF form where object interpretations are strongly
favored.

Experiment 2 was therefore an interpretation task, designed to create contexts where
object relative readings should be particularly pragmatically favored, by manipulating the
plausibility of the event described by the relative clause. Given two human participants
in particular roles (e.g. *patient* and *doctor*) and a verb denoting an event in which only
one of these participants is prototypically the agent (e.g. *cure*), are subjects more likely to
interpret strings as object relatives where this produced the prototypical match between the
agent role and the occupation of the participant?

That is, is a string such as:

(6.11) *Waane'* *t-uy=ohelt-ah*   *le*   *k'oha'an* *t-u=ts'aak-ah*   *huntúul*
     Juan   PRV-A3-know-CMP DEF patient   PRV-A3=cure-CMP a
     doktor *yéetel* *u*   *k'ab*
     doctor   with   A3   hand
     'Juan knew the patient that cured a doctor with his hands/Juan knew the patient that
     a doctor cured with his hands'

more likely to be interpreted as a object relative ('Juan knew the patient that a doctor
cured'), because this aligns "doctor" with the agent role in the curing event?

Simultaneously, experiment 2 allowed the opportunity to test once again the predictions of Harmonic Alignment theory. For Harmonic Alignment Theory, the plausibility of the event should not affect the interpretation of the string, because it is only the relative ranking of RC participants which is relevant, not the interaction between participant properties and the event type denoted by the verb. Thus, this theory would predict that (6.11) would be interpreted as a subject relative, because the extracted argument is more definite than the non-extracted argument, and hence outranks it.

Interpretations were elicited by having subjects translate the ambiguous Yucatec strings into Spanish. The experiment took advantage of the fact that while the Yucatec strings are ambiguous, in Spanish, subject relatives are clearly distinguished structurally from object relatives.

6.6.1 Participants

21 student volunteers were paid $MX40 to participate in experiment 2, which lasted between half an hour and 45 minutes.

6.6.2 Materials

The stimuli were created first in Spanish, and were then translated into Yucatec by a bilingual (Yucatec/Spanish) consultant. They were then recorded spoken aloud onto a digital recorder by the same consultant in San Francisco. The speaker was instructed to read each sentence clearly at a natural pace, and to take care not to vary the intonation or rate of speech between conditions or items. Colloquial Yucatec features a lot of elision of consonants, and there is a tendency, especially in rapid speech, for definite determiners (which occur at the left edge of the noun phrase) to phonologically reduce and attach to the right edge of the linearly preceding word. After consultation with the native speaker, it was decided that the stimuli should be read without employing any elision of definite determiners, in order to ensure that the determiners would be properly heard by subjects (their presence being crucial for one of the experimental conditions), even though this resulted in a somewhat more 'careful' rendition of the Yucatec sentences.

All items were transitive sentences in which the main clause object was modified by a transitive relative clause which could be interpreted as the antecedent to either the subject

or the object of the relative clause. The subject of the main clause of each item was a proper name; the relative clause head, and the embedded object were both NPs denoting types of people, typically professions (nurse, doctor, priest, etc), which were culturally familiar. Each item was concluded with a prepositional phrase. Fillers, of similar length to the experimental items, using intransitive verbs with various types of adjuncts were also recorded.

The experiment used a 2 x 2 design, crossing the relative definiteness of the relative clause NPs with the plausibility of the event given a subject relative clause reading. All conditions involved resumptive forms (as opposed to AF forms) - thus, they were all potentially ambiguous between a subject and an object interpretation. In the definiteness condition, the definiteness of the head NP and the embedded NP alternated such that when the head was definite, the embedded NP was indefinite and vice versa. In the plausibility condition, the event denoted by the relative clause was either most plausible, given a subject relative reading, or most plausible, given an object relative reading. Plausibility was manipulated by altering the relative ordering of the two RC NPs. Thus, in the sample item below, 'fireman' is most plausibly the agent of the rescuing event and 'girl' as the patient of the rescuing event. Thus 1 and 2 are plausible on a subject relative reading (*Maria hugged a/the fireman who rescued the/a girl*), where the head NP 'fireman' is interpreted as the agent of the relative clause. 3 and 4, with the ordering of the two NPs switched, are most plausible on an object relative reading (*Maria hugged a/the girl who the/a fireman rescued*), where the head NP 'girl' is interpreted as the patient of the relative clause verb.

1. PLAUS SUBJ-DEF HEAD
   Maria hugged **the fireman** [$_{RC}$ rescued **a girl** from the building]

2. PLAUS SUBJ-INDEF HEAD
   Maria hugged **a fireman** [$_{RC}$ rescued **the girl** from the building]

3. PLAUS OBJ-DEF HEAD
   Maria hugged **the girl** [$_{RC}$ rescued **a fireman** from the building]

4. PLAUS OBJ-INDEF HEAD
   Maria hugged **a girl** [$_{RC}$ rescued **the fireman** from the building]

1. X-Mariae' tu meek'ah le h-bombero tu hóosah huntul chan x-ch'uupal te edifisio-o'

2. X-Mariae' tu meek'ah huntul h-bombero tu hóosah le chan x-ch'uupal te edifisio-o'

3. X-Mariae' tu meek'ah le chan x-ch'uupal tu hóosah huntul h-bombero te edifisio-o'

4. X-Mariae' tu meek'ah huntul chan x-ch'uupal tu hóosah le h-bombero te edifisio-o'

Four stimuli sets were constructed in such a way that participants heard all items, but only one version of any given item. All participants heard the same number of each condition. The experiment consisted of 24 items, plus 32 fillers, which were arranged such that at least one filler always intervened between any two of the experimental items.

6.6.3 Coding

Responses were coded by the experimenter, who is fluent in Spanish. The translated productions were coded for the two predictors, definiteness of the antecedent, and plausibility type of the input (plausible-as-object RC vs. plausible-as-subject RC). Translated productions were analyzed provided that the speaker produced a transitive relative clause construction. Because what was of interest was the interpretation of the Yucatec strings, both passive relative clause constructions and object relatives were coded as patient relatives, because in both of these, the patient role is aligned with the RC head.

All trials in which the speaker did not produce a relative clause construction, for example, due to paraphrasing or incomplete sentences were eliminated. This resulted in 377 observations in total, across 24 items and 21 subjects.

6.6.4 Results

Significance was assessed by fitting a mixed model logistic regression to the data, with subjects and items as random factors. The binary dependent variable for experiment 2 was the interpretation of the Yucatec relative clause (subject relative clause or object relative clause).

In the model there was a highly significant effect of event bias on interpretation in the intuitive direction: speakers are less likely to interpret the Yucatec string as an object relative clause where this interpretation is biased against. (B = -7.55 (SE = 1.37), p<0.000; e^B = 0.0005)

Definiteness, by contrast, was not a strong predictor of interpretation. There was no significant interaction between the definiteness of the head NP and the definiteness of the embedded NP (p<0.91). The definiteness of the head NP was not independently significant (p<0.7). The definiteness of the embedded NP was a marginally significant predictor (p<.098): strings where the embedded NP was indefinite were less likely to be interpreted as object relative clauses vs. when the embedded NP was definite (B = -1.12 (SE = 0.68), p<0.098; e^B = 0.32). In other words, indefinite embedded NPs were themselves more likely to be interpreted as objects of the relative clause verb, not as subjects of the relative clause verb. The concordance statistic C = 0.98 and Somers Dxy = 0.96, showing an excellent overall quality of the model.

6.6.5 Discussion

Perhaps unsurprisingly, Yucatec speakers are strongly influenced by event bias in their interpretation of ambiguous relative clauses. Strings strongly biased towards an object relative reading are more likely to receive an object relative reading.

Experiment 2 provided no support for Harmonic Alignment. If Harmonic Alignment influenced interpretation in Yucatec, then we should have found a significant effect in the interaction between definite/indefinite head- and embedded NPs. But the interaction between the definiteness of the head NP and that of the embedded NP was not a significant predictor of the variation in interpretation.

The fact that there was (a marginally significant) independent effect of definiteness of the embedded NP on the interpretation of the relative clauses is an interesting result. From the perspective of the present theory, according to which the set A person marker on the verb is pronominal, the preference for interpreting indefinite embedded NPs as RC objects (thereby yielding a subject-relative reading for the clause), is consistent with the notion that cataphoric relations between the set A pronominal and a postverbal indefinite NP should be particularly disfavored (see chapter 3). This trend is also generally predicted in terms of the cross-linguistic preference for indefinites to be rhematic (i.e. to be a "comment" rather than a "sentence topic") (Mithun, 1991).

6.7 Experiment 3: Definiteness and Event Bias (Production)

Experiment 2 showed that event bias strongly predicts how an ambiguous transitive relative clause in Yucatec is interpreted. Experiment 3 was designed to test whether, given these same event biases, Yucatec speakers are more likely to use the unambiguous AF verb form particularly where the intended interpretation is biased against. Disambiguation oriented accounts would predict that the AF verb form would be used more in cases where the intended meaning was the pragmatically marked one.

It simultaneously tested for an effect of definiteness of the head NP and the embedded NP on verb form choice. We already found a significant effect of definiteness on the choice of verb form in experiment 1. We would therefore expect this result to be replicated in experiment 3. Harmonic Alignment would predict that the AF verb form would be produced more (or, always, on categorical accounts), where the RC head did not outrank the RC object, and that event plausibility should have no effect. We found no evidence for Harmonic Alignment in experiment 1 or 2; we should therefore not expect to see any in experiment 3.

6.7.1 Participants

25 UNO undergraduates were paid MX$40 per hour to participate in the study.

6.7.2 Materials

Experiment 3 was essentially a replica of experiment 2, but this time the direction of translation was reversed: subjects heard Spanish sentences, and were required to translate them into Yucatec. Whereas experiment 1 was an interpretation task (the Yucatec items were all ambiguous); experiment 3 did not involve any ambiguity in the input: all Spanish items were unambiguously subject relative clauses. Because subjects participated in both experiments, experiment 3 was run last and experiment 2 first, so that there was an interval of at least three weeks between them.

Spanish stimuli were adapted from the Yucatec versions of experiment 2, and checked with native Spanish speakers. The stimuli were recorded by a bilingual (Yucatec/Spanish) speaker in Valladolid, Mexico. As in the previous experiment, the speaker was instructed

to read each sentence clearly at a natural pace, and to take care not to vary the intonation or rate of speech between conditions or items.

Experiment 3 used a 2 x 2 design. The relative definiteness and the plausibility of the event were manipulated, resulting in four conditions for each item:

1. PLAUS-DEF HEAD

   María abrazó **al bombero** que rescató a **una niña** del edificio

   (Maria hugged the fireman that rescued a girl from the building)

2. PLAUS-INDEF HEAD

   María abrazó **a un bombero** que rescató a **la niña** del edificio

   (Maria hugged a fireman that rescued the girl from the building)

3. IMPLAUS-DEF HEAD

   María abrazó **a la niña** que rescató **a un bombero** del edifico

   (Maria hugged the girl that rescued a firman from the building)

4. IMPLAUS-INDEF HEAD

   María abrazó **a una niña** que rescató **al bombero** del edificio

   (Maria hugged a girl that rescued the fireman from the building)

Four counter-balanced stimuli sets were constructed in such a way that participants heard all items, but only one version of any given item. The experiment consisted of 24 items, plus 32 fillers.

6.7.3 Coding

Responses were coded by the experimenter. A subset of responses was checked for consistency against full transcriptions of the response sentences undertaken by native Yucatec linguistic students who did not participate in the experiment.

Translations were coded for definiteness of the two relative clause NPs, the event-plausibility of the Spanish RC (plausible/implausible), and the dependent variable (AF/ resumptive verb form). Translated productions were analyzed provided that the speaker produced a transitive relative clause construction. All trials in which the speaker did not

produce a relative clause construction, for example, due to paraphrasing or incomplete sentences were eliminated. Responses in which NP_{rel} was realized as a focussed independent pronoun were also excluded from the analysis (unlike experiment 1), due to the small number of tokens. In addition, subjects were excluded who were determined to be non-variable speakers, that is, who did not produce any alternation between verb forms, across all items and conditions.

6.7.4 Results

A mixed model logistic regression was fitted to the data, with subjects and items as random factors. The binary dependent variable for experiment 2 was the choice of verb form (resumptive vs. AF).

In the model, event bias is not a significant predictor of verb choice ($p<0.36$) In other words, event descriptions which are pragmatically unlikely, do not increase the likelihood of using the explicitly unambiguous verb form. Definiteness of the head and of the RC object NP were both significant predictors. The AF form is less likely to occur with an indefinite head than with a definite head (B = -2.38 (SE = 0.56), $p<0.000$; e^B = 0.09). Independently, the AF verb is also less likely to be produced when the RC object is indefinite than when it is definite (B = -2.52 (SE = 0.64), $p<0.000$; e^B = 0.08). Interactions were not significant. The overall quality of the model is very good. C = 0.977, Somers Dxy = 0.953.

6.7.5 Discussion

Experiment 3 shows that there is no correlation between event bias and verb choice. Yucatec speakers do not exhibit any preference for producing the disambiguating AF verb more in contexts where the intended interpretation is pragmatically biased against. Despite the fact (as per experiment 2) that in these contexts, hearers are more likely to interpret strings with the resumptive verb form as object relatives rather than subject relatives, this has no influence on the choice between verb forms in production.

This result is consistent with a substantial body of research on language production that has shown that production decisions are not necessarily directly influenced by concern for ambiguity avoidance – even in cases where it might have a large effect on the success of

communication. It has been shown in production studies of English that speakers don't consistently avoid syntactic ambiguity with word order (Arnold et al., 2004), or with optional 'that' complementizers (Ferreira and Dell, 2000).

Experiment 2 provides no support for recoverability-oriented theories of any type. For Harmonic alignment we would expect that indefinite objects should decrease the likelihood of the AF verb form only when the head NP is indefinite. However, the interaction between definiteness of the head NP and definiteness of the embedded NP was not significant in the model. In this respect, the production results match the interpretation results of experiment 1: Harmonic Alignment (at least in the domain of the relative definiteness of participants) influences neither production nor comprehension of Yucatec relative clauses.

We do find, however, that definiteness of the head NP and definiteness of the object NP are both independently significant predictors of the choice between verb forms. Indefinite head NPs and indefinite object NPs are both more likely to occur with the resumptive verb form.

The fact that indefinite heads are correlated with a higher frequency of the resumptive form is is consistent with the typological record, and with experiment 1. What about the definiteness of embedded objects? I am aware of no corpus or experimental data, nor any theoretical literature that bears on the issue of what influence embedded NPs may have on the likelihood of resumption. In a memory-based model of processing, one might suppose that anything which increases the complexity of the dependency, and thus places higher demands on working memory decreases the ease of the retrievability of the antecedent, and thus creates a preference for resumptive structures. On the assumption that embedded objects are included in the dependency domain in Yucatec,[14] then indefiniteness of the embedded object NP can be argued to increase the complexity of the dependency, and thus exert a greater processing cost. This would be consistent with recent work has explored the effect of the embedded noun phrase type on sentence complexity (Gordon et al., 2001, 2004; Mak et al., 2002; Warren and Gibson, 2002) in English. For example, Warren and Gibson (2002) examined whether referential properties of the second noun phrase affect the complexity of center-embedded sentences. They found that the processing difficulty

---

[14]Hawkins (2004) argues that generally, the filler-gap domain for subject relatives extends only to the subcategorizer in the RC, but not to the object, except possibly in verb initial languages where the object is also required to disambiguate. This technically seems to hold for Yucatec, given the significant effect of the definiteness of the object NP, though it is seems unlikely that the reason really has to do with disambiguation.

in nested sentences depends on the degree to which the embedded subject was old or new in the discourse according to the Giveness Hierarchy (Gundel et al., 1993). Gordon et al. (2001) found that English object relatives were made easier when the embedded (subject) noun phrase was the indexical pronoun *you* and were more difficult when it was a proper name.[15] This may also explain the interesting preferences observed impressionistically in chapter 2, but not tested experimentally here: that the AF form is preferred over the resumptive form where the RC object is a pronominal.

6.8 Summary and general discussion

The three Yucatec experiments presented above collectively yielded the following findings:

- Variation in Yucatec relative clause production mirrors patterns of RP/gap distribution attested across grammars for two conditions. Resumptives are more likely in embedded conditions and where the antecedent is indefinite.

    - This provides some of the first robust support from production data for the Performance-Grammar Correspondence hypothesis in the domain of RP/gap variation. In so doing it points to a grammar-external explanation for a widely cited 'UG' constraint (The Highest Subject Restriction)

- Yucatec exhibits variation between *three* subject relativization strategies (Exp 1): gap, dependent resumptive pronoun and independent resumptive pronoun. The asymmetric frequency distributions of gaps relative to dependent resumptives parallels that of dependent resumptives relative to independent resumptives.

    - This finding is not easily accommodated by theories in which the relative processing costs of gaps and resumptives derive from a binary syntactic contrast between syntactic and anaphoric dependency types (e.g. Hawkins 2004 and Alexopoulou and Keller 2007).

---

[15]The authors interpreted the results from a similarity-based interference perspective: memory interference during encoding and retrieval is prevented because the matrix and the embedded noun phrases produce non-interfering representations.

- Despite the fact that the AF form provides a more informative, unambiguous structure than the dependent resumptive structure, there is no preference for the AF form in conditions of heightened complexity (Exp 1), nor in contexts where the intended interpretation is pragmatically biased against (Exp 3). Thus, despite the fact that in these contexts, it was shown (Exp 2) that hearers are more likely *interpret* strings with the full verb form as object relatives rather than subject relatives, this has no influence on the choice between verb forms in production.

  – These findings are not predicted by role-marking approaches to resumptive pronoun processing, nor by ambiguity avoidance based accounts of Mayan AF.

- The AF verb was independently more likely in simple structures than in embedded structures (Exp1), more likely with definite antecedents than indefinite antecedents (Exp 1, 2).

  – These results are not predicted by Harmonic Alignment, according to which the AF verb should occur when the antecedent is indefinite and outranked by a definite RC object.

- The AF verb was also more likely with definite RC objects than with indefinite RC objects (Exp 3).

  – This is a novel finding. No theory of resumption processing has so far considered the possibility of elements *following* the extraction site affecting the choice between resumptive and gap. I am aware of no corpus or experimental data for other languages which speaks to this issue.

These experiments raise more questions than they solve. As we saw, the choice between alternants in relative clauses (gap, dependent resumptive, independent resumptive) is determined by the level of complexity/length of the dependency and antecedent definiteness in Yucatec. This is reminiscent of results reported for referential form choice in discourse, where the specificity of referential form is correlated with factors such as distance and the referential salience of the antecedent (Givón, 1976; Ariel, 1990; Gundel et al., 1993), and suggests that some of the processing principles identified to account for referential form choice in discourse might be profitably extended to the inter-sentential level also.

Ariel (1999) attempts to do just this, by extending Accessibility Theory, a theory of referential form choice at the discourse level, to filler-gap/resumptive dependencies (Ariel, 1999). Less attenuated forms (resumptives) are used when the antecedent is less accessible. However, as I argued in the discussion for Experiment 1, Accessibility Theory incorporates a rather circular line of argumentation, because cost is defined on the basis of the accessibility scale, which is what is is trying to explain to begin with. Indefinites are located at the low end of the accessibility scale and thus are assumed to induce a greater processing cost for subsequent retrieval. But why are indefinites less accessible?

It lies well beyond the scope of this dissertation to develop a theory of RP/gap dependency processing. Here I will simply point to some possible avenues for future research on this topic. With regard to the antecedent-definiteness effect, one possibility is that it is not some intrinsic accessibility-property of indefinites that is of relevance here, but rather what the presence of an indefinite NP says about the likelihood of it subsequently being referred to. Recent work by Wasow et al. (2005) may be of relevance in this regard. They show that relativizers are far more frequent in English object relative clauses when the head NP is introduced by an indefinite 'a' or 'an', rather than a definite 'the'. They attribute this to the relative predictability of an upcoming relative clause, given an indefinite vs. a definite NP head. Relative clauses are more predictable with definite heads than with indefinite heads. In a context in which a relative clause is highly predictable, the listener gets less useful information from having the beginning of the relative clause explicitly marked. Hence, relativizers do less to facilitate comprehension where non subject relative clauses are predictable. Or, from a production perspective, where a relative clause is more predictable, speakers can afford to reduce articulatory effort, and therefore omit the relativizer.

Why are relative clauses more predictable with definite heads? There is a straightforward semantic/pragmatic reason for this. Restrictive relative clauses characteristically serve to limit the possible referents of the NPs in which they occur. Wasow et al. (2005) point out that certain determiners, nouns, and adjectives have semantic properties that make further restriction of the denotation preferable. Universal assertions expressed with the universal quantifiers *all* and *every*, for example, are generally true only of restricted sets. Thus, (6.12a) is true for many more VPs than (6.12b). The use of a relative clause therefore allows speakers to avoid making overly general claims.

(6.12)   a. Every linguist we know VP

    b.   Every linguist VP

For existentials that opposite is true. (6.13a) is true for many more VPs than (6.13b), since (6.13a) is true if VP holds of any linguist, whereas (6.13b) is true only if it holds of a linguist we know.

(6.13)   a. A linguist VP

           b. A linguist we know VP

Finally, definite determiners generally indicate that the referent of the NP it is introducing is contextually unique – that is, the intended referent can be uniquely identified given the linguistic and non-linguistic context. But identifying a unique referent often requires providing more information about it than is expressed by the noun alone. A relative clause is a way of doing this. Thus, (6.14b) can more successfully refer to a unique individual than (6.14a):

(6.14)   a. the linguist

           b. the linguist I told you about

Thus, the different semantic/pragmatic restrictions of different modifiers has an effect on the relative probability of a following relative clauses. How is this relevant for RP/gap distributions with indefinites? If relative clause modification is more probable with a definite NP, then, accordingly, this increases the probability that that NP will be subsequently referred to intrasententially (i.e., in the relative clause). Conversely, if relative clauses are less likely with indefinite NP heads, then it is concomitantly less likely that they will be referred to again intrasententially. If referential form attenuation is preferred wherever the referent is most predictable, then this accounts for the skewed frequency distributions of resumptive pronouns and gaps with respect to indefinite vs. definite heads.

    On a predictability based approach, we could view the difference between a resumptive and a gap as a specific instance of the more general phenomenon, amply documented, of the tendency of speakers to minimize production effort (Zipf, 1935) as long as this does not interfere with other constraints on communication.[16]

---

[16]In this respect, recent production models which consider predictability from an information theoretic perspective, might be profitably applied to RP/gap distributions (Aylett and Turk, 2004; Genzel and Charniak, 2002; Jaeger, 2006; Levy and Jaeger, 2007). On these accounts, speakers structure utterances so as to optimize information density. Information optimization accounts predict that speakers are more likely to omit optional elements, the more predictable (and hence redundant) the information conveyed by those elements is.

A predictability-based theory might be profitably extended to the asymmetric frequency distributions that have been reported in the typological literature with respect to wh-questions and relative clauses and resumption. Recall that resumptives are more likely in relative clauses than in wh questions. Given that languages often possess dedicated wh-words which occur in extraction contexts, it seems reasonable to assume that wh-words convey a higher predictability of an upcoming intersential dependency than nominals which may or may not be modified by relative clauses. This would be easily verifiable by corpus studies.

A predictability-based theory would also put us in the position to pose a further set of empirical questions. It has been noted for English at least that subject relatives are not uniformly preferred. Though they are frequent as modifiers of object NPs, they turn out to be rare as modifiers of subject NPs (Fox and Thompson, 1990; Geisler, 1998). This is presumably because objects are more likely than subjects to be lexically expressed and thus modifiable (Michaelis and Francis, 2007). This in turn is due to the fact that subjects are generally topical and thus generally discourse-old and often pronominalized (Mithun, 1991). If this is the case cross-linguistically, then would we predict a typological preference for resumptive pronouns in subject relative clauses modifying matrix subject NPs, because those relative clause types are less frequent and thus less predictable? It depends what level of frequency is relevant. If it is about the global frequency of different relative clause types, then yes. But it is possible that it is the relative likelihood of a relative clause given the prior mention of a nominal which is relevant. That is, while relative clauses might on the whole be rare as modifiers of subject NPs because subject NPs are most frequently pronominalized, it may be that when a subject *is* lexically expressed, then there is a high probability that it will be modified by a relative clause. Lexically expressed subjects are presumably particularly common in cases of topic shift. These are precisely the conditions under which relative clause modification would be particularly likely (in order to restrict the possible reference set of the shifted topic).

This possibility is supported by the interesting relationship between wh-questions and relative clause frequencies on the one hand and gap/RP distributions on the other. While subject relative clauses are very prevalent cross-linguistically and in conversational speech, it turns out that subject wh-extractions are apparently very rare overall in English conversation (Homer, 2000). If this is a cross-linguistic tendency, then this shows an interesting inverse relationship between the overall frequencies of different dependency types, and the

relative frequency of gaps across those different dependency types (gaps appear to be more prevalent in subject wh-questions than in subject relative clauses, despite the overall low frequency of subject wh-questions, at least in English). This suggests that resumption/gap distributions are not correlated with overall frequency of relative clauses vs. questions, but rather with the contextual probability of intersentential reference, given a particular antecedent. That is, while subject wh-extractions may be rare overall, when a wh-word does occur in a sentence, it is highly predictive of an upcoming intersentential dependency.

These are all important questions that require substantial corpus and experimental research to tease apart. As of yet, there exist very few cross-linguistic or language specific generalizations about resumptive distributions beyond the very coarse-grained factors already mentioned in this chapter: grammatical function, construction type, embedding. Much more nuanced information is necessary to develop an adequate account of the interaction of factors affecting RP/gap distributions in production, and how, over time, this impacts on the shape of grammars.

Chapter 7

Conclusions

"Unfortunately, or luckily, no language is tyrannically consistent. All grammars leak...",
goes another famous dictum of Sapir's (1921, 38). I hope to have shown in this dissertation
that studying the 'leaks' can be very revealing. In addressing the phenomenon of Mayan
AF, and anti-pronominal alternations in general, I have argued that too much weight has
been placed on grammar as an explanatory locus. The mysteries left behind by synchronic
generative analyses can be unlocked by using usage and grammaticalization to explain
more of the critical data. Invoking Universal Grammar as the ultimate locus of explanation
is, as Hall writes, 'a convenient though unjustified termination point for enquiry' (1992, 7).

7.1   Summary of findings

Taxonomy

The AF alternation in Yucatec is a morphological subtype of RP/gap alternation. This clas-
sification not only solves the riddles surrounding its formal properties (presence/absence
of person marking) and its distribution (its restriction to A-extraction contexts); it also
provides the necessary basis for typological comparison. The variation exhibited in this
domain, both within Yucatec in terms of the choice between alternants, and across Mayan
languages, in terms of patterns of categorical cut-off points, mirrors typologically attested
asymmetric frequency distributions of resumptive pronouns and gaps.

Typology

Synthetic/analytic verbal alternations in A-bar dependencies are typologically wide-spread among head-marking languages. Just as in Mayan, the choice between verbal alternants is conditioned by the same factors that are implicated in RP/gap distributions cross linguistically, providing support for the notion that they are all exemplars of the same basic phenomenon.

Diachrony

Frequency distributions in language use explain the formal properties and distribution of analytic verbs in A-bar dependencies in head-marking languages. Analytic verbs are conservative verb forms which remained exempt from the grammaticalization of head marking. The reason for the exemption of these verb forms has its roots in the fact that grammaticalization processes are frequency sensitive: the more frequently two elements co-occur, the more likely it is that grammaticalization will take place. The relatively low frequency of adjacent pronoun-verb combinations in extraction contexts (where gaps are more frequent than resumptive pronouns), by comparison with main clause verbs, can give rise to asymmetric patterns of pronoun grammaticalization, and thus lead to the emergence of these morphological alternations.

Processing

The experimental findings presented in this dissertation support the view that the asymmetric frequency distributions of gaps and RPs (within and across languages) in turn can be explained by processing preferences. Three experiments on Yucatec provided support for the Performance-Grammar Correspondence Hypothesis (Hawkins, 2004) in the domain of RP/gap distributions. The results of the production experiments showed that RPs are more likely in embedded environments, and where the antecedent is indefinite. Thus, the Highest Subject Restriction, which is apparently a categorical constraint in many languages (including Jakalteko, which is genetically related to Yucatec), is reflected probabilistically in Yucatec in terms of production preferences. The definiteness restriction (resumptives are required with indefinite antecedents/gaps are required with definite antecedents), which has been reported categorically in at least two languages, and probabilistically in corpus

studies for several others, is also visible probabilistically in Yucatec production.

More specifically the results of these experiments showed that choice of verb form in Yucatec is not determined by ambiguity avoidance. Neither is choice of verb form in Yucatec determined by Harmonic Alignment. It also showed that Yucatec speakers make use of three different strategies for terminating dependencies: gap < dependent RP < independent RP. This cline is reminiscent of referential form clines at the discourse level. In both cases, the specificity of referential form is correlated with factors such as distance and the referential salience of the antecedent, suggesting that the same cognitive principles lie behind both types of variation.

7.2   Implications and open questions

7.2.1   Variation and change in the Mayan language family

In this dissertation I have shown that variation in Yucatec production can be profitably analyzed from a processing perspective. Choice in production in the domain of AF in Yucatec is influenced by dependency complexity and by the referential properties of the relative clause participants (and, presumably, a slew of other factors not tested here).

As of yet variation in production in other Mayan languages remains largely unchartered territory. Aissen (1999, 2003) has made some extremely thought provoking inroads in her explorations of Tzotzil AF variation, though the studies involve small sample sizes and are thus not statistically robust. For most other languages, current descriptions of the data do not tell us to what degree variation is possible in the domain of AF. Do the languages that reportedly require AF in all A-extraction contexts really have such a categorical rule? Yucatec was also proposed to obligatorily require the use of the AF verb in extraction contexts, but closer inspection revealed this not to be the case. In languages that do have variable-AF, do we find the same general patterns of variation that exist in Yucatec? Can we detect language specific structural differences which are correlated with differences in AF distribution?

Comparative-historical morpho-syntax is still very much in its infancy in Mayan linguistics and in this dissertation I have only scratched at the surface of the complicated historical processes which have given rise to the specific morphological instantiation of

anti-pronominal alternations in the modern languages. I have argued that these alternations have their roots in original syntactic RP/gap alternations prior to the grammaticalization of main clause transitive subject person marking. Interacting with this process is a complicated series of evolutions of the tense/aspect morphology of main clause verbs, which has yielded further morphological differentiations between main and dependent verbs over time, in different ways in different branches of the family. A great deal of work remains to be done in order to uncover the evolutionary paths by which individual Mayan languages arrived at their modern state.

7.2.2   Formal and referential properties of pronouns and their relation to RP distributions

It is striking to observe that productive resumptive languages tend to be dependent pronoun languages (i.e., languages which possess pronominal clitics/affixes), rather than independent pronoun languages. I am unaware of any cross-linguistic surveys which confirm this generalization, but it certainly seems to be the case from the extant cross-linguistic literature on productive resumption. This leads us to the question: what is the relationship between the formal properties of a language's RPs and their frequencies in intrasentential dependencies? The Yucatec results described in this dissertation put an additional spin on this issue, as it revealed the possibility of a three way alternation between gaps, dependent RPs and independent RPs within a single language.

How, moreover, does the *information* content of a pronoun affect its distribution? It is well known that the grammaticalization of person markers is often characterized by semantic changes. As formal erosion progresses, there may be a reduction or loss of information about the referential identity of person markers, for example, number or gender features. The end point of such a process is the complete loss of referentiality on the part of the person marker, and the obligatory presence of the nominal with which it agrees, as has happened in English (Siewierska, 1999, 225). It remains an open question what the interaction is between this process of semantic reduction and of formal reduction over time. It therefore also remains an open question how both types of reduction in turn might affect RP distributions over time. The connection between the formal and referential properties of RPs and their distributions in intrasentential dependencies is a very rich area for future typological and experimental investigation.

7.2.3 Performance-grammar correspondences among genetically related languages

In chapter 3 I pointed to some rather tantalizing evidence that Yucatec's probabilitistic variation appears to be mirrored in categorical constraints in other Mayan languages. This was shown for differences between construction types. In Yucatec the AF verb is more frequent in wh-questions. In Mam, the AF verb is obligatory with questions, but optional with focus clefts and relative clauses (England, 1983). In Ixil the AF verb is obligatory with both wh-questions and focus clefts, but optional with relative clauses (Ayres, 1983). In Sipakapense, AF is obligatory with wh-questions and focus clefts, but the synthetic form is obligatory with relative clauses (Dayley, 1981). I also showed that the probabilistic embedding effect in Yucatec (the AF verb is more frequent in simple dependencies than in embedded dependencies) is reflected in Jakalteko, where the AF verb is apparently obligatory in simple dependencies, but optional in embedded ones (Craig, 1977). These correlations support the view that the same functional factors which determine the variation in usage in Yucatec, also have shaped the evolution of grammatical constraints in other languages in the family.

These parallels are particularly significant in that they offer some interesting preliminary data for the relation between performance and grammar *within a genetically related set of languages*. To date, I am unaware of other studies which have charted such parallels between probabilitistic tendencies and categorical constraints within a single language family. If processing preferences in performance shape grammars over time, then this should be first and foremost visible within a set of genetically related languages.

7.3 Final thoughts

A recent informal survey finds that psycholinguistically controlled production research has been conducted on fewer than thirty of the world's 5,000 to 10,000 languages. A sizable literature on sentence production (more than five papers) seems to exist for only seven languages: English, Dutch, German, French, Spanish, Italian, and Japanese, of which six fall into two language families (Germanic and Romance), both of which have developed from Indo-European and have been spoken in close geo-cultural proximity for many centuries (Jaeger and Norcliffe, 2009). Research on cross-linguistic variation in sentence production

assumes a particular importance due to the rapidly dwindling sample of data. It is estimated that 90% of the world's languages will be extinct or moribund by the end of the twenty-first century (Romaine, 2007). While many language communities realistically are already too small to conduct controlled variation studies, the majority still have enough speakers (70% of all languages are still spoken by more than 1000 speakers, about 40% by more than 10,000 speakers, Romaine 2007, 121).

I hope that this dissertation will help to push forward field-based language variation research as a paradigm. This is indispensable for future work that aims to close the vast typological gaps inevitable in variationist and psycholinguistic research based primarily on English. Moreover, it is critically important for bringing a diversified sample of data to bear on the deeper questions surrounding the connection between usage and grammar.

Appendix A

Hawkins' Efficiency Principles

Hawkins (2004) proposes three efficiency principles by which the relative processing costs of two alternants can be measured: maximize online processing, minimize form and minimize domains. Below I present the calculation of the relative complexity of the domains for the resumptive and gap versions of (A.1a; no embedding) and (A.1b; embedding), in order to determine what his theory predicts for the Yucatec case.
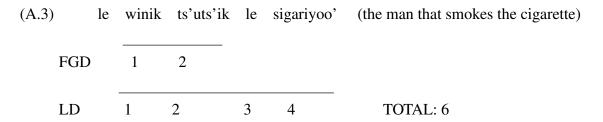
(A.1)   a. *le   winik   (k-u)=ts'uts'ik   le   sigariyo-o'*
          the   man   (IMP-A3)=smoke-INC   DET   cigarette-D2
          the man that smokes the cigarette

      b. *le   winik   t-uy=a'al-ah   Juan   (k-u)=ts'uts'ik   le*
          the   man   PRV-A3=say-CMP   Juan   (IMP-A3)=smoke-INC   DET
          sigariyo-o'
          cigarette-D2
          the man that Juan said smokes the cigarette

In order to determine domain complexity in filler-gap/filler resumptive structures, two dependency domains are calculated, and the subtotals summed. Hawkins writes that while in general, a Filler-Gap Domain for subject relative extends from the antecedent to the subcategorizer, in the case of verb initial languages, it must be extended to include the gap's nondependent arguments, because in such cases the filler and the subcategorizor alone are ambiguous between a subject gap and an object gap interpretation. If there is no case marking (and where verb agreement does not distinguish the arguments), the sequence *cats that saw rats* in a VSO language could be assigned a subject relative interpretation [cats$_i$ [that

saw O$_i$ rats]], or an object interpretation [cats$_i$ [that saw rats O$-i$]]. In such cases, according to Hawkins, the FGDs for both interpretations will include the disambiguating overt NP within the relative clause, since this NP is required for gap identification and processing.

We must apply Hawkins' reasoning to Filler-Resumptive domains in Yucatec, which must also be extended to include the object, because the string is not otherwise disambiguated. If we treat the set A pronoun as a part of the verbal node, then the resumptive form will have a filler-resumptive domain of 4, for the simple relative clause example in (A.1). If we treat the pronominal element as a separate word, this yields a filler-resumptive domain of 5. This structure will have a lexical domain of 3 (4 if the clitic is a separate word), which spans the subcategorizer and its arguments (it does not include the RC head, because, on Hawkins' analysis, the resumptive pronoun allows for local processing of the lexical domain). The following is a graphic representation of the domain nodes, calculated for the pronominal element as a separate node (in the remainder of the structures, the alternative totals, with the pronominal element counted as a separate word, will be given in parentheses):

(A.2)   le   winik   ku=ts'uts'ik   le   sigariyoo'   (the man that smokes the cigarette)

FGD   1   2   3   4

LD   1   2   3   TOTAL: 7 (9)

The alternative gap structure will have a Filler-Gap Domain of 2: it doesnt need to extend to include the object, because, unlike the resumptive form, it unambiguously receives a subject relative interpretation. It will have a larger lexical domain than its resumptive counterpart however (4 vs. 3), because it must access the RC head:

(A.3)   le   winik   ts'uts'ik   le   sigariyoo'   (the man that smokes the cigarette)

FGD   1   2

LD   1   2   3   4   TOTAL: 6

The AF form wins, just, with a minimal domain of 6, vs. 7 for the resumptive structure (or 9 if the subject pronoun is counted as a separate node). Moreover, it also involves less

form, which renders it less costly in terms of Hawkins' second principle, Minimize Form, and is more efficient in terms of the third principle, that of Maximizing Online Processing. This states that a structure will be dispreferred in proportion to the number of properties that are unassignable to it on-line, compared with the number (un)assigned in an alternative structure or sequence.

In the Yucatec case, processing the full form requires leaving the RC head unassigned for longer[1]

	Resumptive form	Gapped form
MD	-	+
MF	-	+
MaF	-	+

Table A.1: Resumptive vs. gap structure, no embedding

Hawkins' theory predicts that resumptives will be preferred in embedded structures, because in gap structures, the lexical domain increases with every level of embedding. In the Yucatec case, the resumptive structure with one level of embedding (**??**) has a Filler-Resumptive domain of 6 (7 if the pronoun is a separate word), and a Lexical Domain of 3:

(A.4)   le   winik   tuy=a'alah   Juan   ku=ts'uts'ik   le   sigariyoo'

FGD   1   2   3   4   5   6

LD   1   2   3   TOTAL: 9 (10)

The gap counterpart has a smaller FGD (4), but a larger LD of 6, because it must include all of the embedded nodes, and the filler.

(A.5)   le   winik   tuy=a'alah   Juan   ts'uts'ik   le   sigariyoo'

FGD   1   2   3   4

---

[1]Depending on the animacy of the postverbal NP, and the lexical semantics of the verb, the structure may remain globally ambiguous even after processing the post-verbal NP.

LD	1	2	3	4	5	6	TOTAL: 10

For the embedded structure the resumptive form wins, but only just, in terms of maximizing domains (or draws, if the pronoun is counted as a separate word). But the AF structure still wins in terms of Minimize Form, and it still wins in terms of Maximize Online Processing.

	Resumptive form	Gapped form
MD	(+)	(-)
MF	-	+
MaF	-	+

Table A.2: Resumptive vs. gap structure, embedding

Thus, when coupled with other principles based on disambiguating grammatical relations (the MaOP), Hawkins' Efficiency principles yield no strong predictions for the relative processing cost of gapping vs. resumptive structures in Yucatec. The prediction of a preference for resumptives in embedded contexts in Yucatec would rest on the assumption that domain minimization exerts a greater processing pressure than the MaP and the MF combined. Hawkins' does not offer any metric for assessing the relative weight of his principles when they are in competition.

Appendix B

Experimental Stimuli

B.1 Experiment 1 stimuli

1   Juan ignoró al locutor que anunció las noticias
Juan ignoró a un locutor que anunció las noticias
Juan ignoró al locutor que Linda dijo anunció las noticias
Juan ignoró a un locutor que Linda dijo anunció las noticias

       [Juan ignored the/an announcer that (Linda said) announced the news]

2   María condenó al ladrón que rompió la tienda
María condenó a un ladrón que rompió la tienda
María condenó al ladrón que Leonardo pensó rompió la tienda
María condenó a un ladrón que Leonardo pensó rompió la tienda

       [Maria convicted the/a thief that (Leonardo thought) broke the store]

3   Pedro abrazó al tejedor que terminó el tapiz
Pedro abrazó a un tejedor que terminó el tapiz
Pedro abrazó al tejedor que Rosario dijo terminó el tapiz
Pedro abrazó a un tejedor que Rosario dijo terminó el tapiz

       [Pedro hugged the/a weaver that (Rosario said) finished the tapestry]

4   Laura persiguió al criminal que robó la caja
Laura persiguió a un criminal que robó la caja

Laura persiguió al criminal que Renaldo dijo robó la caja
Laura persiguió a un criminal que Renaldo dijo robó la caja

[Laura chased the/a criminal that (Renaldo said) stole the box]

5   Julia sacudió al administrador que negó el rumor
Julia sacudió a un administrador que negó el rumor
Julia sacudió al administrador que Perla pensó negó el rumor
Julia sacudió a un administrador que Perla pensó negó el rumor

[Julia shook the/an adminstrator that (Perla thought) denied the rumor]

6   Roberto agarró al preso que destruyó la cosecha
Roberto agarró a un preso que destruyó la cosecha
Roberto agarró al preso que Oscar dijo destruyó la cosecha
Roberto agarró a un preso que Oscar dijo destruyó la cosecha

[Roberto caught the/a prisoner that (Oscar said) destroyed the crop]

7   Jaime felicitó al matemático que resolvió el problema
Jaime felicitó a un matemático que resolvió el problema
Jaime felicitó al matemático que Nina pensó resolvió el problema
Jaime felicitó a un matemático que Nina pensó resolvió el problema

[Jaime congratulated the/a mathematician that (Nina said) solved the problem]

8   Carla buscó a la sirvienta que limpió la casa
Carla buscó a una sirvienta que limpió la casa
Carla buscó a la sirvienta que Luisa dijo limpió la casa
Carla buscó a una sirvienta que Luisa dijo limpió la casa

[Carla looked for the/a servant that (Luisa said) cleaned the house]

9   Gabriela protegió al aldeano que transportó la carga
Gabriela protegió a un aldeano que transportó la carga
Gabriela protegió al aldeano que Xavier pensó transportó la carga
Gabriela protegió a un aldeano que Xavier pensó transportó la carga

[Gabriela protected the/a village that (Xavier thought) transported the cargo]

10   Miguel saludó al investigador que escribió el libro

Miguel saludó a un investigador que escribió el libro
Miguel saludó al investigador que Marta pensó escribió el libro
Miguel saludó a un investigador que Marta pensó escribió el libro

[Miguel greeted a/the researcher that (Marta thought) wrote the book]

11   Adriano regañó al panadero que quemó el pastel
Adriano regañó a un panadero que quemó el pastel
Adriano regañó al panadero que Pancho pensó quemó el pastel
Adriano regañó a un panadero que Pancho pensó quemó el pastel

[Adriano scolded the/a baker that (Pancho thought) burnt the cake]

12   Rafael esperó a la vendedora que tostó el grano
Rafael esperó a una vendedora que tostó el grano
Rafael esperó a la vendedora que Rafael dijo tostó el grano
Rafael esperó a una vendedora que Rafael dijo tostó el grano

[Rafael waited for the/a saleswoman that (Rafael said) toasted the grain]

13   Valeria rescató al acróbata que subió el árbol
Valeria rescató a un acróbata que subió el árbol
Valeria rescató al acróbata que Nicolás pensó subió el árbol
Valeria rescató a un acróbata que Nicolás pensó subió el árbol

[Valeria rescued the/a acrobat that (Nicolas thought) climbed the tree]

14   Victor espió al refugiado que reveló el secreto
Victor espió a un refugiado que reveló el secreto
Victor espió al refugiado que Pepita dijo reveló el secreto
Victor espió a un refugiado que Pepita dijo reveló el secreto

[Victor spied on the/a refugee that (Pepita said) revealed the secret]

15   Raquel habló con el campesino que compró la semilla
Raquel habló con un campesino que compró la semilla
Raquel habló con el campesino que Felipe pensó compró la semilla
Raquel habló con un campesino que Felipe pensó compró la semilla

[Raquel spoke with the/a peasant that Felipe thought bought the seed]

16   Manuel habló con el trabajador que reparó la television
     Manuel habló con un trabajador que reparó la television
     Manuel habló con el trabajador que Ramón dijo reparó la television
     Manuel habló con un trabajador que Ramón dijo reparó la television

     [Manuel spoke with the/a worker that (Ramon said) repaired the television]

17   Lucio olió a la actriz que usó el perfume
     Lucio olió a una actriz que usó el perfume
     Lucio olió a la actriz que Ximena pensó usó el perfume
     Lucio olió a una actriz que Ximena pensó usó el perfume

     [Lucio smelled the/an actress that (Ximena thought) wore the perfume]

18   Rodrigo se burló del comensal que derramó la bebida
     Rodrigo se burló de un comensal que derramó la bebida
     Rodrigo se burló del comensal que Latoya dijo derramó la bebida
     Rodrigo se burló de un comensal que Latoya dijo derramó la bebida

     [Rodrigo made fun of the/a diner that (Latoya said) spilled the drink]

19   Alejandro contestó al profesor que hizo la pregunta
     Alejandro contestó a un profesor que hizo la pregunta
     Alejandro contestó al profesor que Paulina pensó hizo la pregunta
     Alejandro contestó a un profesor que Paulina pensó hizo la pregunta

     [Alejandro answered the/a professor that (Paulina thought) asked the question]

20   Alvaro gritó a la lavandera que manchó el vestido
     Alvaro gritó a una lavandera que manchó el vestido
     Alvaro gritó a la lavandera que Trinidad dijo manchó el vestido
     Alvaro gritó a una lavandera que Trinidad dijo manchó el vestido
     [Alvaro yelled at the/a washerwoman that (Trinidad said) stained the dress]

21   Celia silenció al espectador que causó el ruido
     Celia silenció a un espectador que causó el ruido
     Celia silenció al espectador que Tomás pensó causó el ruido
     Celia silenció a un espectador que Tomás pensó causó el ruido

     [Celia silenced the/a spectator that (Tomas thought) caused the noise]

22   Jose silbó al corredor que ganó la carrera
      Jose silbó a un corredor que ganó la carrera
      Jose silbó al corredor que Marco dijo ganó la carrera
      Jose silbó a un corredor que Marco dijo ganó la carrera

      [Jose whistled at the/a runner that (Marco said) won the race]

23   Rosa peleó con el guerrillero que escondió la bomba
      Rosa peleó con un guerrillero que escondió la bomba
      Rosa peleó con el guerrillero que Pablo pensó escondió la bomba
      Rosa peleó con un guerrillero que Pablo pensó escondió la bomba

      [Rosa fought with the/a guerilla that (Pablo thought) hid the bomb]

24   Fernando golpeó al mensajero que perdió el documento
      Fernando golpeó un mensajero que perdió el documento
      Fernando golpeó al mensajero que Rosita dijo perdió el documento
      Fernando golpeó un mensajero que Rosita dijo perdió el documento

      [Fernando hit the/a messenger that (Rosita said) lost the document]

B.2   Experiment 2 stimuli

1   Waane' tu yoheltah le doktor tu ts'aakah hun túul k'oja'an yéetel u k'ab
    Waane' tu yoheltah hun túul doktor tu ts'aakah le k'oha'an yéetel u k'ab
    Waane' tu yoheltah le k'oha'an tu ts'aakah hun túul doktor yéetel u k'ab
    Waane' tu yoheltah hun túul k'oha'an tu ts'aakah le doktor yéetel u k'ab

    [Juan found out about the/a doctor [cured a/the patient with his hands]]
    [Juan found out about the/a patient [cured a/the doctor with his hands]]

2   X Mariae' tu meek'ah le h bombero tu hosah hun túul chan x ch'uupal te edificioo'
    X Mariae' tu meek'ah hun túul h bombero tu hosah le chan x ch'uupal te edificioo'
    X Mariae' tu meek'ah le chan x ch'uupal tu hosah hun túul bombero te edificioo'
    X Mariae' tu meek'ah hun túul chan x ch'uupal tu hosah le bombero te edificioo'

    [Maria hugged the/a fireman [rescued a/the girl from the building]]
    [Maria hugged the/a girl [rescued a/the fireman from the building]]

3   Pedroe' tu t'anah le h periodista tu entrevistartah hun túul h musiko te cha'ano'

Pedroe' tu t'anah hun túul h periodista tu entrevistartah le h musiko te cha'ano'
Pedroe' tu t'anah le h musiko tu entrevistartah hun túul h periodista te cha'ano'
Pedroe' tu t'anah hun túul h musiko tu entrevistartah le h periodista te cha'ano'

[Pedro called the/a journalist [interviewed a/the musician at the concert]]
[Pedro called the/a musician [interviewed a/the journalist at the concert]]

4   X Laurae' t'àanah yéetel le doktor tu eksaminartah hun túul h k'oha'an táanil u
     yúuchul le operasiono'
     X Laurae' t'àanah yéetel hun túul doktor tu eksaminartah le h k'oha'an táanil u
     yúuchul le operasiono'
     X Laurae' t'àanah yéetel le h k'oha'an tu eksaminartah hun túul doktor táanil u
     yúuchul le operasiono'
     X Laurae' t'àanah yéetel hun túul h k'oha'an tu eksaminartah le doktor táanil u
     yúuchul le operasiono'

     [Laura spoke with the/a doctor [examined a/the patient before the operation]]
     [Laura spoke with the/a patient [examined a/the doctor before the operation]]

5   Daniele' tu yilah le profesor tu aprobartah hun túul h xòok te universidado'
     Daniele' tu yilah hun túul profesor tu aprobartah le h xòok te universidado'
     Daniele' tu yilah le h xòok tu aprobartah hun túul profesor te universidado'
     Daniele' tu yilah hun túul h xòok tu aprobartah le profesor te universidado'

     [Daniel saw the/a professor [graded a/the student at the university]]
     [Daniel saw the/a student [graded a/the professor at the university]]

6   Robertoe' tu ts'ibtah yóok'ol le h kansah tu hats'ah hun túul x xòok te eskuelao'
     Robertoe' tu ts'ibtah yóok'ol hun túul h kansah tu hats'ah le x xòok te eskuelao'
     Robertoe' tu ts'ibtah yóok'ol le x xòok tu hats'ah hun túul h kansah te eskuelao'
     Robertoe' tu ts'ibtah yóok'ol hun túul x xòok tu hats'ah le h kansah te eskuelao'

     [Roberto wrote about the/a teacher [whipped a/the student at the school]]
     [Roberto wrote about the/a student [whipped a/the teacher at the school]]

7   X Lusiae' tu saludartah le heneral tu kastigartah hun túul h ch'úuk le túun yúuch le
     guerrao'
     X Lusiae' tu saludartah hun túul heneral tu kastigartah le h ch'úuk le túun yúuch le
     guerrao'
     X Lusiae' tu saludartah le h ch'úuk tu kastigartah hun túul heneral le túun yúuch le
     guerrao'

X Lusiae' tu saludartah hun túul h ch'úuk tu kastigartah le heneral le túun yúuch le guerrao'

[Lucia greeted the/a general [punished a/the spy during the war]] [Lucia greeted the/a spy [punished a/the general during the war]]

8   X Rakele' tu kehartuba te maago tu asustartah hun túul pàal te fiestao'
X Rakele' tu kehartuba ti hun túul maago tu asustartah le pàal te fiestao'
X Rakele' tu kehartuba te pàal tu asustartah hun túul maago te fiestao'
X Rakele' tu kehartuba ti hun túul pàal tu asustartah le maago te fiestao'

[Raquel complained about the magician [frightened a boy at the party]]
[Raquel complained about the magician [frightened a boy at the party]]

9   Miguele' tu xíimb'altah le sikologo tu yáantah hun túul x byuda káa ts'ó'ok le beloryoo'
Miguele' tu xíimb'altah hun túul sikologo tu yáantah le x byuda káa ts'ó'ok le beloryoo'
Miguele' tu xíimb'altah le x byuda tu yáantah hun túul sikologo káa ts'ó'ok le beloryoo'
Miguele' tu xíimb'altah hun túul x byuda tu yáantah le sikologo káa ts'ó'ok le beloryoo'

[Miguel visited the/a psycholoist [helped a/the widow after the funeral]]
[Miguel visited the/a widow [helped a/the psychologist after the funeral]]

10   Josee' tu xokah yóok'ol le gueriyero tu yoklah hun túul x kó'olel te estasionamyentoo'
Josee' tu xokah yóok'ol hun túul gueriyero tu yoklah le x kó'olel te estasionamyentoo'
Josee' tu xokah yóok'ol le x kó'olel tu yoklah hun túul gueriyero te estasionamyentoo'
Josee' tu xokah yóok'ol hun túul x kó'olel tu yoklah le gueriyero te estasionamyentoo'

[Jose read about the guerilla [kidnapped a woman in the parking lot]]
[Jose read about the woman [kidnapped a guerilla in the parking lot]]

11   Manuele' tu saludartah le h taksista tu manehartah hun túul pasahero te ospitalo'
Manuele' tu saludartah hun túul h taksista tu manehartah le pasahero te ospitalo'
Manuele' tu saludartah le pasahero tu manehartah hun túul h taksista te ospitalo'

Manuele' tu saludartah hun túul pasahero tu manehartah le h taksista te ospitalo'

[Manuel greeted the/a taxi driver [drove a/the passenger to the hospital]]
[Manuel greeted the/a passenger [drove a/the taxi driver to the hospital]]

12   X Anae' b'áatenah yéetel le h ahente tu kastigartah hun túul preso te kampamentoo'
       X Anae' b'áatenah yéetel hun túul h ahente tu kastigartah le preso te kampamentoo'
       X Anae' b'áatenah yéetel le preso tu kastigartah hun túul h ahente te kampamentoo'
       X Anae' b'áatenah yéetel hun túul preso tu kastigartah le h ahente te kampamentoo'

[Ana fought with the/an agent [punished a/the prisoner in the camp]]
[Ana fought with the/a prisoner [punished an/the agent in the camp]]

13   X mariae' u k'ah óol le h wàach tu ts'onah hun túul h sibil te kayeo'
       X mariae' u k'ah óol hun túul h wàach tu ts'onah le h sibil te kalleo'
       X mariae' u k'ah óol le h sibil tu ts'onah hun túul h wàachte kayeo'
       X mariae' u k'ah óol hun túul h sibil tu ts'onah le h wàach te kayeo'

[Maria knew the/a soldier [shot a/the civilian in the street]]
[Maria knew the/a civilian [shot a/the soldier in the street]]

14   Edgare' tu ché'ehtah le polisia tu yalkab'tah hun túul h òokol te ich merkadoo'
       Edgare' tu che'ehtah hun túul polisia tu yalkab'tah le h òokol te ich merkadoo'
       Edgare' tu ché'ehtah le h òokol tu yalkab'tah hun túul polisia te ich merkadoo'
       Edgare' tu che'ehtah hun túul h òokol tu yalkab'tah le polisia te ich merkadoo'

[Edgar laughed at the/a policeman [chased a/the thief through the market]]
[Edgar laughed at the/a thief [chased a/the policeman through the market]]

15   Albertoe' kahlah ti u tséel le h guardia tu kalantah hun túul x artista te baalao'
       Albertoe' kahlah ti u tséel hun túul h guardia tu kalantah le x artista te baalao'
       Albertoe' kahlah ti u tséel le x artista tu kalantah hun túul h guardia te baalao'
       Albertoe' kahlah ti u tséel hun túul x artista tu kalantah le h guardia te baalao'

[Alberto lived next to the/a guard [protected a/the actress from the bullet]]
[Alberto lived next to the/a actress [protected a/the guard from the bullet]]

16   X Paulae' tu ts'íibtah te lisensiado tu tsol t'aantah hun túul kliente yóok'ol le hú'uno'
       X Paulae' tu ts'íibtah ti hun túul lisensiado tu tsol t'aantah le kliente yóok'ol le hú'uno'

X Paulae' tu ts'íibtah te kliente tu tsol t'aantah hun túul lisensiado yóok'ol le hú'uno'

X Paulae' tu ts'íibtah ti hun túul kliente tu tsol t'aantah le lisensiado yóok'ol le hú'uno'

[Paula wrote to the/a lawyer [advised a/the client about the document]]
[Paula wrote to the/a client [advised a/the lawyer about the document]]

17 Franciscoe' tu kontratartah le h entrenador tu entrenartah hun túul h báaxal táanil le báxalo'

Franciscoe' tu kontratartah hun túul h entrenador tu entrenartah le h báaxal táanil le báxalo'

Franciscoe' tu kontratartah le h báaxal tu entrenartah hun túul h entrenador táanil le báxalo'

Franciscoe' tu kontratartah hun túul h báaxal tu entrenartah le h entrenador táanil le báxalo'

[Francisco employed the/a trainer [trained a/the player before the game]]
[Francisco employed the/a player [trained a/the trainer before the game]]

18 X karlae' tu takubah te x pul yá'ah tu pul yá'ahtah hun túul nohoch máak te k'íiwiko'

X karlae' tu takubah ti hun túul x pul yá'ah tu pul yá'ahtah le nohoch máak te k'íiwiko'

X karlae' tu takubah te nohoch máak tu pul yá'ahtah hun túul x pul yá'ah te k'íiwiko'

X karlae' tu takubah ti hun túul nohoch máak tu pul yá'ahtah le x pul yá'ah te k'íiwiko'

[Carla hid from the/a witch [bewitched a/the old man in the plaza]]
[Carla hid from the/a old man [bewitched a/the witch in the plaza]]

19 Federikoe' tu mèentubah amigoil te huez tu sentensiartah hun túul x kó'olel le semana máano'

Federikoe' tu mèentubah amigoil ti hun túul huez tu sentensiartah le x kó'olel le semana máano'

Federikoe' tu mèentubah amigoil te x kó'olel tu sentensiartah hun túul huez le semana máano'

Federikoe' tu mèentubah amigoil ti hun túul x kó'olel tu sentensiartah le huez le semana máano'

[Federico made friends with the/a judge [sentenced a/the woman last week]]
[Federico made friends with the/a woman [sentenced a/the judge last week]]

20   X Isabele' tu yáantah le x enfermera tu bakunartah hun túul x ch'úupal káa ts'ó'ok le aksidenteo'
X Isabele' tu yáantah hun túul x enfermera tu bakunartah le x ch'úupal káa ts'ó'ok le aksidenteo'
X Isabele' tu yáantah le x ch'úupal tu bakunartah hun túul x enfermera káa ts'ó'ok le aksidenteo'
X Isabele' tu yáantah hun túul x ch'úupal tu bakunartah le x enfermera káa ts'ó'ok le aksidenteo'

[Isabel helped the/a nurse [vaccinated a/the girl after the accident]]
[Isabel helped the/a girl [vaccinated a/the nurse after the accident]]

21   X Gabriela tu ignorartah le kalá'an tu poch'ah hun túul nojoch kó'olel te baaro'
X Gabriela tu ignorartah hun túul kalá'an tu poch'ah le nojoch kó'olel te baaro'
X Gabriela tu ignorartah le nojoch kó'olel tu poch'ah hun túul kalá'an te baaro'
X Gabriela tu ignorartah hun túul nojoch kó'olel tu poch'ah le kalá'an te baaro'

[Gabriela ignored the/a drunk [insulted a/the old woman in the bar]]
[Gabriela ignored the/a old woman [insulted a/the drunk in the bar]]

22   Karlose' kulah tu tséel le h artista tu dibujartah hun túul modelo u tyá'al le exposisiono'
Karlose' kulah tu tséel hun túul h artista tu dibujartah le modelo u tyá'al le exposisiono'
Karlose' kulah tu tséel le modelo tu dibujartah hun túul h artista u tyá'al le exposisiono'
Karlose' kulah tu tséel hun túul modelo tu dibujartah le h artista u tyá'al le exposisiono'

[Carlos sat next to the/an artist [sketched a/the model for the exhibition]]
[Carlos sat next to the/a model [sketched an/the artist for the exhibition]]

23   X Beatrise' meyahnah yéetel le kamarografo tu filmartah hun túul aktor te tumbèen estudioo'
X Beatrise' meyahnah yéetel hun túul kamarografo tu filmartah le aktor te tumbèen estudioo'
X Beatrise' meyahnah yéeteleh le aktor tu filmartah hun túul kamarografo te tumbèen estudioo'

X Beatrise' meyahnah yéeteleh hun túul aktor tu filmartah le kamarografo te tumbèen estudioo'

Beatriz worked with the/a cameraman [filmed an/the actor in the new studio]]
Beatriz worked with the/an actor [filmed a/the cameraman in the new studio]]

24 Davide' tu k'á'a óoltah le guiya tu bisah hun túul x turista te buuso'
Davide' tu k'á'a óoltah hun túul guiya tu bisah le x turista te buuso'
Davide' tu k'á'a óoltah le x turista tu bisah le guiya te buuso'
Davide' tu k'á'a óoltah hun túul x turista tu bisah hun túul guiya te buuso'

[David recognized the/a guide that took a/the tourist to the bus]]
[David recognized the/a tourist that took a/the guide to the bus]]

B.3 Experiment 3 stimuli

1 Juan se enteró del médico que curó a un lisiado con sus manos
Juan se enteró a un médico que curó al lisiado con sus manos
Juan se enteró del lisiado que curó a un médico con sus manos
Juan se enteró de un lisiado que curó al médico con sus manos

[Juan found out about the/a doctor that cured a/the patient with his hands]
[Juan found out about the/a patient that cured a/the doctor with his hands]

2 Maria abrazó al bombero que rescató a una niña del edificio
Maria abrazó a un bombero que rescató a la niña del edificio
Maria abrazó a la niña que rescató a un bombero del edificio
Maria abrazó a una niña que rescató al bombero del edificio

[Maria hugged the/a fireman that rescued a/the girl from the building]
[Maria hugged the/a girl that rescued a/the fireman from the building]

3 Pedro llamó al periodista que entrevistó a un músico en el concierto
Pedro llamó a un periodisat que entrevistó al músico en el concierto
Pedro llamó al músico que entrevistó a un periodista en el concierto
Pedro llamó a un músico que entrevistó al periodista en el concierto

[Pedro called the/a journalist that interviewed a/the musician at the concert]
[Pedro called the/a musician that interviewed a/the journalist at the concert]

4   Laura habló con el cirujano que examinó a un paciente antes de la operación
        Laura habló con un cirujano que examinó al paciente antes de la operación
        Laura habló con el paciente que examinó a un cirujano antes de la operación
        Laura habló con un paciente que examinó al cirujano antes de la operación
        [Laura spoke with the/a surgeon that examined a/the patient before the operation]
        [Laura spoke with the/a patient that examined a/the surgeon before the operation]

5   Daniel contactó al profesor que calificó a un estudiante en la universidad
        Daniel contactó a un profesor que calificó al estudiante en la universidad
        Daniel contactó al estudiante que calificó a un profesor en la universidad
        Daniel contactó a un estudiante que calificó al profesor en la universidad

        [Daniel contacted the/a professor that graded a/the student at the university]
        [Daniel contacted the/a student that graded a/the professor at the unviersity]

6   Roberto escribió sobre el maestro que azotó a una estudiante en la clase
        Roberto escribió sobre un maestro que azotó a la estudiante en la clase
        Roberto escribió sobre la estudiante que azotó a un maestro en la clase
        Roberto escribió sobre una estudiante que azotó al maestro en la clase

        [Roberto wrote about the/a teacher that whipped a/the student in the class]
        [Roberto wrote about the/a student that whipped a/the teacher in the class]

7   Lucía respetó al general que interrogó a un espía durante la guerra
        Lucía respetó a un general que interrogó al espía durante la guerra
        Lucía respetó al espía que interrogó a un general durante la guerra
        Lucía respetó a un espía que interrogó al general durante la guerra

        [Lucia respected the/a general that interrogated a/the spy during the war]
        [Lucia respected the/a spy that interrogated a/the general during the war]

8   Raquel se quejó del terrorista que mató a un peatón en la calle
        Raquel se quejó de un terrorista que mató al peatón en la calle
        Raquel se quejó del peatón que mató a un terrorista en la calle
        Raquel se quejó de un peatón que mató al terrorista en la calle

        [Raquel complained about the/a terrorist that killed a/the pedestrian in the street]
        [Raquel complained about the/a pedestrian that killed a/the terrorist in the street]

9   Miguel visitó al psicólogo que ayudó a una viuda después del funeral
        Miguel visitó a un psicólogo que ayudó a la viuda después del funeral

Miguel visitó a la viuda que ayudó a un psicólogo después del funeral
Miguel visitó a una viuda que ayudó al psicólogo después del funeral

[Miguel visited the/a psychologist that helped a/the widow after the funeral]
[Miguel visited the/a widow that helped a/the psychologist after the funeral]

10 Jose leyó sobre el guerillero que secuestró a una mujer en el estacionamiento
Jose leyó sobre un guerillero que secuestró a la mujer en el estacionamiento
Jose leyó sobre la mujer que secuestró a un guerillero en el estacionamiento
Jose leyó sobre una mujer que secuestró al guerillero en el estacionamiento

[Jose read about the/a guerilla that kidnapped a/the woman in the parking lot]
[Jose read about the/a woman that kidnapped a/the guerilla in the parking lot]

11 Manuel saludó al taxista que llevó a un pasajero al supermercado
Manuel saludó a un taxista que llevó al pasajero al supermercado
Manuel saludó al pasajero que llevó a un taxista al supermercado
Manuel saludó a un pasajero que llevó al taxista al supermercado

[Manuel greeted the/a taxi driver that took a/the passenger to the supermarket]
[Manuel greeted the/a passenger that took a/the taxi driver to the supermarket]

12 Ana peleó con el agente que torturó a un preso en el campamento
Ana peleó con un agente que torturó al preso en el campamento
Ana peleó con el preso que torturó a un agente en el campamento
Ana peleó con un preso que torturó al agente en el campamento

[Ana fought with the/a agent that tortured a/the prisoner in the camp]
[Ana fought with the/a prisoner that tortured a/the agent in the camp]

13 Mara conoció al soldado que disparó a un civil en el parque
Mara conoció a un soldado que disparó al civil en el parque
Mara conoció al civil que disparó a un soldado en el pargue
Mara conoció a un civil que disparó al soldado en el parque

[Maria knew the/a soldier that shot a/the civilian in the park]
[Maria knew the/a civilian that shot a/the soldier in the park]

14 Edgar se rió del policía que persiguió a un ladrón a través del mercado
Edgar se rió de un policía que persiguió al ladrón a través del mercado
Edgar se rió del ladrón que persiguió a un policía a través del mercado

Edgar se rió de un ladrón que persiguió al policía a través del mercado

[Edgar laughed at the/a policeman that chased a/the thief through the market]
[Edgar laughed at the/a thief that chased a/the policeman through the market]

15 Alberto residió al lado del guardia que protegió a una actriz de la bala
Alberto residió al lado de un guardia que protegió a la actriz de la bala
Alberto residió al lado de la actriz que protegió a un guardia de la bala
Alberto residió al lado de una actriz que protegió al guardia de la bala

[Alberto lived next to the/a guard that protected an/the actress from the bullet]
[Alberto lived next to the/an actress that protected a/the guard from the bullet]

16 Paula escribió al abogado que aconsejó a un cliente sobre el documento
Paula escribió a un abogado que aconsejó al cliente sobre el documento
Paula escribió al cliente que aconsejó a un abogado sobre el documento
Paula escribió a un cliente que aconsejó al abogado sobre el documento

[Paula wrote to the/a lawyer that advised a/the client about the document]
[Paula wrote to the/a client that advised a/the lawyer about the document]

17 Francisco empleó al entrenador que entrenó a un jugador antes del partido
Francisco empleó a un entrenador que entrenó al jugador antes del partido
Francisco empleó al jugador que entrenó a un entrenador antes del partido
Francisco empleó a un jugador que entrenó al entrenador antes del partido

[Francisco employed the/a trainer who trained a/the player before the game]
[Francisco employed the/a player who trained a/the trainer before the game]

18 Carla se escondió de la bruja que embrujó a un viejo en la plaza
Carla se escondió de una bruja que embrujó al viejo en la plaza
Carla se escondió del viejo que embrujó a una bruja en la plaza
Carla se escondió de un viejo que embrujó a la bruja en la plaza

[Carla hid from the/a witch that bewitched an/the old man in the plaza]
[Carla hid from the/an old man that bewitched a/the witch in the plaza]

19 Frederico se hizo amigo del juez que condenó a una mujer la semana pasada
Frederico se hizo amigo de un juez que condenó a la mujer la semana pasada
Frederico se hizo amigo de la mujer que condenó a un juez la semana pasada
Frederico se hizo amigo de una mujer que condenó al juez la semana pasada

[Frederico made friends with the/a judge that condemned a/the woman last week]
[Frederico made friends with the/a woman that condemned a/the judge last week]

20   Isabel ayudó a la enfermera que vacunó a una muchacha después del accidente
Isabel ayudó a una enfermera que vacunó a la muchacha después del accidente
Isabel ayudó a la muchacha que vacunó a una enfermera después del accidente
Isabel ayudó a una muchacha que vacunó a la enfermera después del accidente

[Isabel helped the/a nurse that vaccinated a/the girl after the accident]
[Isabel helped the/a girl that vaccinated a/the nurse after the accident]

21   Gabriela escuchó al borracho que insultó a una vieja en el bar
Gabriela escuchó a un borracho que insultó a la vieja en el bar
Gabriela escuchó a la vieja que insultó a un borracho en el bar
Gabriela escuchó a una vieja que insultó al borracho en el bar

[Gabriela listened to the/a drunk that insulted an/the old woman in the bar]
[Gabriela listened to the/an old woman that insulted a/the drunk in the bar]

22   Carlos se sentó al lado del artista que dibujó a una modelo para la exposición
Carlos se sentó al lado de un artista que dibujó al modelo para la exposición
Carlos se sentó al lado de la modelo que dibujó a un artista para la exposición
Carlos se sentó al lado de una modelo que dibujó al artista para la exposición

[Carlos sat next to the/an artist that sketched a/the model for the exhibition]
[Carlos sat next to the/a model that sketched a/the artist for the exhibition]

(B.1) Beatriz trabajó con el camarógrafo que filmó a un actor en el nuevo estudio
Beatriz trabajó con un camarógrafo que filmó al actor en el nuevo estudio
Beatriz trabajó con el actor que filmó a un camarógrafo en el nuevo estudio
Beatriz trabajó con un actor que filmó al camarógrafo en el nuevo estudio

[Beatriz worked with the/a photographer that filmed an/the actor in the new studio]
[Beatriz worked with the/an actor that filmed a/the photographer in the new studio]

(B.2) David reconoció al guía que llevó a un ciego al autobús
David reconoció a un guía que llevó al ciego al autobús
David reconoció al ciego que llevó a un guía al autobús
David reconoció a un ciego que llevó al guía al autobús

[David recognized the/a guide that carried a/the blind person to the bus]
[David recognized the/a blind person that carried a/the guide to the bus]

Bibliography

Abbott, M. (1991). Macushi. In D. C. Derbyshire and G. K. Pullum (Eds.), *Handbook of Amazonian Languages*, Volume III, pp. 23–160. Berlin: Mouton de Gruyter.

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley.

Aikhenvald, A. Y. (1995). *Bare*. München: Lincom.

Aissen, J. (1987). *Tzotzil Clause Structure*. Dordrecht: Reidel.

Aissen, J. (1992). Topic and focus in Mayan. *Language 68*(1), 43–80.

Aissen, J. (1999). Agent focus and inverse in Tzotzil. *Language 75*, 451–485.

Aissen, J. (2003). Differential coding, partial blocking and bi-directional OT. In P. Nowak and C. Yoquelet (Eds.), *BLS*, Volume 29, pp. 1–16.

Alexiadou, A. and E. Anagnostopoulou (2000). Asymmetries in the distribution of clitics: The case of Greek restrictive relatives. In F. Beukema and M. den Dikken (Eds.), *Clitic phenomena in European languages*, pp. 47–70. John Benjamins.

Alexopoulou, T. (2009). Truly intrusive: resumptive pronominals in questions and relative clauses. To appear in *Exploring the Left Periphery*, *Lingua*.

Alexopoulou, T. and F. Keller (2007). Locality, Cyclicity and Resumption: At the Interface between the Grammar and the Human Sentence Processor. *Language 83*(1), 110–160.

Allwood, J. (1982). The complex NP constraint in Swedish. In E. Engdahl and E. Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*, pp. 15–32. Stockholm: Almqvist and Wiksell.

Almor, A. and V. A. Nair (2007). The Form of Referential Expressions in Discourse. *Language and Linguistics Compass 1*, 84–99.

Andersson, L. (1982). What is Swedish an exception to? Extractions and island constraints. In E. Engdahl and E. Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*, pp. 33–45. Stockholm: Almqvist and Wiksell.

Andrade, J. (2001). The contribution of working memory to conscious experience. In J. Andrade (Ed.), *Working Memory in Perspective*. Psychology Press/Taylor and Francis.

Andrade, M. J. (1955). A Grammar of Modern Yucatec. Microfilm Collection of Manuscripts on Middle American Cultural Anthropology, no. 41. Regenstein Library, University of Chicago.

Andrews, A. D. (1990). Unification and morphological blocking. *Natural Language and Linguistic Theory 8*, 507–57.

Aoun, J. (2000). Resumption and last resort. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada 16*.

Aoun, J. and L. Choueiri (1996). Resumption and last resort. Available online at http://www.usc.edu/dept/LAS/linguistics/semitic/.

Aoun, J., L. Choueiri, and N. Hornstein (2001). Resumption, movement and derivational economy. *Linguistic Inquiry 32*, 371–403.

Aoun, J. and Y. H. A. Li (1989). Scope and constituency. *Linguistic Inquiry 20*, 219–251.

Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. London: Routledge.

Ariel, M. (1998). Three grammaticalization paths for the development of person verbal agreement in Hebrew. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap*, pp. 93–111. Stanford: CSLI.

Ariel, M. (1999). Cognitive universals and linguistic conventions: the case of resumptive pronouns. *Studies in Language 23*(2), 217–269.

Ariel, M. (2000). The development of person agreement markers: from pronoun to accessibility markers. In S. Kemmer and M. Barlow (Eds.), *Usage-based models of language*, pp. 197–261. Stanford: CSLI.

Arnold, J. (1998). *Reference Form and Discourse Patterns*. Ph. D. thesis, Department of Linguistics, Stanford University.

Arnold, J., T. Wasow, A. Asudeh, and P. Alrenga (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language 51*(1), 55–70.

Asudeh, A. (2004). *Resumption as Resource Management*. Ph. D. thesis, Department of Linguistics, Stanford University.

Auger, J. (1993, 16-20 August). On the History of Relative Clauses in French and some of its Dialects. In *Historical Linguistics 1993: Selected papers from the 11th International Conference on Historical Linguistics*, Los Angeles.

Auger, J. (1994). *Pronominal Clitics in Quebec Colloquial French: A Morphological Analysis*. Ph. D. thesis, University of Pennsylvania.

Aylett, M. and A. Turk (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence and Duration in Spontaneous Speech. *Language and Speech 47*(1), 31–56.

Ayres, G. (1983). The Antipassive "Voice" in Ixil. *International Journal of American Linguistics 49*(1), 20–45.

Baayen, R. H. (2007). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Baddeley, A. (1992). Working Memory: The Interface between Memory and Cognition. *Journal of Cognitive Neuroscience 4*(3), 281–288.

Baddeley, A. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Baddeley, A. and G. J. Hitch (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Volume 8, pp. 47–89. New York: Academic Press.

Baker, M. C. (1991). On some subject-object non-asymmetries in Mohawk. *Natural Language and Linguistic Theory 9*, 537–576.

Baker, M. C. (2008). On the Nature of the Antiagreement Effect: Evidence from Wh-in-Situ in Ibibio. *Linguistic Inquiry 39*(4), 615–632.

Bates, D. M. and D. Sarkar (2007). Lme4: Linear mixed-effects models using s4 classes. R-package version 0.99875-6.

Beckner, C., R. Blythe, J. Bybee, M. H. Christiansen, W. Croft, N. C. Ellis, J. Holland, J. Ke, D. Larsen-Freeman, and T. Schoenemann (2008). Language is a Complex Adaptive System. Ms.

Beltrán, P. (1859/1746). *Arte del Idioma Maya*. Mérida, México: Imprenta de J.D. Espinosa.

Berinstein, A. (1985). *Evidence for Multiattachment in K'ekchi Mayan*. New York: Garland.

Bigalke, R. (1997). *Siciliano*, Volume 129 of *Languages of the World/Materials*. München: Lincom.

Blair, R. (1964). *Yucatec Maya Noun and Verb Morphosyntax*. Ph. D. thesis, Indiana University, Bloomington.

Blair, R. and R. Vermont Salas (1965/1967). *Spoken (Yucatec) Maya*. Chicago: University of Chicago, Department of Anthropology.

Bloomfield, L. (1933). *Language*. New York: Henry Holt and Company.

Boas, F. (1911). *Handbook of American Indian Languages*. Washington.

Bock, J. K. and W. F. Brewer (1974). Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology 103*, 837–843.

Bock, J. K. and D. E. Irwin (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior 19*, 467–484.

Boeckx, C. (2003). *Islands and Chains: Resumption as Derivational Residue*. Linguistik Aktuell. John Benjamins.

Bohnemeyer, J. (2002). *The grammar of time reference in Yukatek Maya*. Munich: LIN-COM.

Bohnemeyer, J. (2004). Split intransitivity, linking, and lexical representation: the case of Yukatek Maya. *Linguistics 42*(1), 67–107.

Bohnemeyer, J. (2009). Linking without grammatical relations in Yucatec Maya: Alignment, extraction and control. In J. Helmbrecht, Y. Nishina, Y.-M. Shin, S. Skopeteas, and E. Verhoeven (Eds.), *Form and Function in Language Research: Papers in Honour of Christian Lehmann*. Mouton de Gruyter.

Bopp, F. (1816). *Über das Conjugationssytem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persisichen und germanischen Sprache*. Frankfurt am Main: Andreäische Buchhandlung.

Borer, H. (1984). Restrictive relatives in modern Hebrew. *Natural Language and Linguistic Theory 2*, 219–260.

Borer, H. (1986). I-Subjects. *Linguistic Inquiry 17*, 375–416.

Bouchard, D. (1982). Les constructions relatives en français vernaculaire et en français standard: étude d'un paramètre. In C. Lefebvre (Ed.), *La syntaxe comparée du français standard et populaire: aproches formelle et fonctionnelle*, Volume 1. Éditeur officiel du Québec.

Brandi, L. and P. Cordin (1989). Two Italian dialects and the Null Subject Parameter. In O. Jaeggli and K. Safir (Eds.), *The Null Subject Parameter*, pp. 111–142. Dordrecht: Kluwer.

Bresnan, J. (1998). Morphology competes with Syntax: Explaining Typological Variation in Weak Crossover Effects. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (Eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, pp. 59–92. Cambridge, MA: The MIT Press and MIT Working Papers in Linguistics.

Bresnan, J. (2001). The emergence of the unmarked pronoun. In G. Legendre, J. Grimshaw, and S. Vikner (Eds.), *Optimality-theoretic Syntax*, pp. 113–142. Cambridge: The MIT Press.

Bresnan, J., S. Dingare, and C. D. Manning (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG 01 Conference*. CSLI Publications.

Bresnan, J. and J. Grimshaw (1978). The syntax of free relatives in English. *Linguistic Inquiry 9*, 331–391.

Bresnan, J. and S. Mchombo (1987). Topic, pronoun and agreement in Chichewa. *Language 63*(4), 741–782.

Bricker, V. R. (1978). Antipassive constructions in Yucatec Maya. In N. C. England, C. G. Craig, and L. Furbee-Losee (Eds.), *Papers in Mayan Linguistics*. Columbia: Museum of Anthropology.

Bricker, V. R. (1979). Wh-questions, relativization, and clefting in Yucatec Maya. In L. Martin (Ed.), *Papers in Mayan Linguistics*, pp. 107–136. Columbia, Miss: Lucas Brothers.

Bricker, V. R. (1981). The source of the ergative split in Yucatec Maya. *Journal of Mayan Linguistics 2*, 83–127.

Bricker, V. R., E. P. Yah, and O. D. de Po'ot (1998). *A Dictionary of the Maya Language as Spoken in Hocobá, Yucatán*. Utah: University of Utah Press.

Butt, M. (2001, August). Case, Agreement, Pronoun Iincorporation and Pro-Drop in South Asian Languages. Talk handout for the Workshop on The Role of Agreement in Argument Structure, Utrecht.

Bybee, J. (1985). *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, J. (1988). The diachronic dimension in explanation. In J. A. Hawkins (Ed.), *Explaining language universals*, pp. 350–379. Oxford: Blackwell.

Bybee, J. (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication 2*, 249–278.

Bybee, J. (2002a). Main clauses are innovative, subordinate clauses are conservative: Consequences for the nature of constructions. In J. Bybee, M. Noonan, and S. A. Thompson (Eds.), *Complex sentences in grammar and discourse: essays in honor of Sandra A. Thompson*, pp. 1–19. John Benjamins.

Bybee, J. (2002b). Sequentiality as the basis of constituent structure. In T. Givón and B. Malle (Eds.), *The Evolution of Language out of Pre-Language*, pp. 107–132. John Benjamins.

Bybee, J., R. Perkins, and W. Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bybee, J. L. and J. Scheibman (1999). The effect of usage on degrees of constituency: the reduction of *don't* in English. *Linguistics 37*(4), 575–596.

Cable, S. (2003). Resumption and weak crossover in Yoruba. Abstract posted at http://web.mit.edu/scable/www/work/papers/Yoruba-resumptive-wco.html.

Campana, M. (1992). *A movement theory of ergativity*. Ph. D. thesis, McGill.

Campbell, L. (1977). *Quichean Linguistic Prehistory*. Los Angeles: University of California Press.

Campbell, L. (1985). *The Linguistics of Southeastern Chiapas*. Provo, Utah: Brigham Young University Press.

Campbell, L. (1986). Claims about syntactic change and Finnish historical syntax. *Journal of the Atlantic Provinces Association 8*, 72–93.

Campbell, L. (2000). Valency changing derivations in K'iche'. In R. M. W. Dixon and A. Y. Aikhenvald (Eds.), *Changing Valency: Case Studies in Transitivity*. Cambridge: Cambridge University Press.

Campbell, L. and T. Kaufman (1990). Lingüística Mayence: ¿Dónde nos encontramos ahora? In *Lecturas sobre la lingüística Maya*, pp. 51–58. La Antigua: Centro de Investigaciones Regionales de Mesoamérica.

Carlin, E. (1993). *The So Language*. Cologne: Institute of African Studies, University of Cologne.

Chambers, J. K. (1995). *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Cambridge, MA: Blackwell.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1973). Conditions on Transformations. In S. R. Anderson and P. Kiparsky (Eds.), *A Festschrift for Morris Halle*, pp. 232–286. New York: Holt, Reinhart and Winston.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, The Netherlands: Foris.

Chung, S. (1982). Unbounded Dependencies in Chamorro Grammar. *Linguistic Inquiry 13*, 39–78.

Chung, S. (1998). *The design of agreement*. Chicago: Chicago University Press.

Chung, S. and C. Georgopoulos (1988). Agreement with Gaps in Chamorro and Palauan. In M. Barlow and C. A. Ferguson (Eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions*, pp. 251–267. Stanford: CSLI.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior 12*, 335–59.

Clark, H. H. and C. Sengul (1979). In search of references for nouns and pronouns. *Journal of Memory and Cognition 7*, 35–41.

Clements, G. N. (1984). Binding domains in Kikuyu. *Studies in the Linguistic Sciences 14*, 37–56.

Comrie, B. (1973). Clause structure and movement constraints in Russian. In C. Corum, T. C. Smith-Stark, and A. Weiser (Eds.), *You take the high node and I'll take the low node*, pp. 291–304. Chicago: Chicago Linguistic Society.

Comrie, B. (1981). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.

Comrie, B. and E. L. Keenan (1979). Noun phrase accessibility revisited. *Language 55*(3), 649–664.

Cooreman, A. (1982). Topicality, ergativity and transitivity in narrative discourse: evidence from Chamorro. *Studies in Language 6*, 343–374.

Cooreman, A. (1987). *Transitivity and Discourse Continuity in Chamorro Narratives*. Berlin: Mouton de Gruyter.

Corbett, G. (1995). Agreement. In J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (Eds.), *Syntax: ein internationales Handbuch zeitgenössischer Forschung*, Volume 2, pp. 1235–1244. Berlin: Walter de Gruyter.

Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.

Coronel, J. (1998/1620). *Arte en Lengua de Maya*. Universidad Nacional Autónoma de México.

Cournane, A. (2007). The Grammaticalization of Subject Clitics in Quebec French: An Analysis of Reanalysis. MA Thesis, Concordia University.

Craig, C. (1977). *The Structure of Jacaltec*. Austin: University of Texas Press.

Creswell, C. (2002). Resumptive pronouns, wh-island violations, and sentence production. In *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+ 6)*, pp. 101–109. Universitá di Venezia.

Croft, W. (2000a). *Explaining language change: an evolutionary approach*. London: Pearson.

Croft, W. (2000b). Parts of speech as language universals and as language-particular categories. In P. Vogel and B. Comrie (Eds.), *Approaches to the typology of word classes*, pp. 65–102. Berlin: Mouton de Gruyter.

Cysouw, M. (2003). *The Paradigmatic Structure of Person Marking*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Day, C. (1973). *The Jacaltec Language*. Bloomington: Indiana University.

Dayley, J. P. (1981). Voice and ergativity in Mayan languages. *Journal of Mayan Linguistics 2*, 3–82.

De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua 115*, 1195–1219.

de San Buenaventura, G. (1996/1684). *Arte de la Lengua Maya*. Universidad Nacional Autónoma de México.

de Vries, M. (2001). Patterns of relative clauses. In *Linguistics in the Netherlands*, pp. 231–243. John Benjamins.

Dickey, M. W. (1996). Constraints on the sentence processor and the distribution of resumptive pronouns. In M. W. Dickey and S. Tunstall (Eds.), *Linguistics in the Laboratory*, Volume 19 of *Occasional Papers in Linguistics*, pp. 157–192. University of Massachusetts.

Dixon, R. M. W. (1979). Ergativity. *Language 55*, 59–138.

Donohue, M. (1999). *A Grammar of Tukang Besi*. Berlin: Mouton de Gruyter.

Donohue, M. (2002). Voice in Tukang Besi and the Austronesian focus system. In F. Wouk and M. Ross (Eds.), *The History and Typology of western Austronesian voice and voice-marking*, pp. 81–100. Canberra: Pacific Linguistics.

Donohue, M. and A. Maclachlan (1999). What agreement in Chamorro? In C. Smallwood and C. Kitto (Eds.), *Proceedings of the Austronesian Formal Linguistics Association VI*, Volume 16.2 of *Toronto Working Papers in Linguistics*, pp. 121–132.

Doron, E. (1982). On the Syntax and Semantics of Resumptive Pronouns. *Texas Linguistics Forum 19*, 1–48.

Dryer, M. (1997). Are grammatical relations universal? In J. Bybee, J. Haiman, and S. A. Thompson (Eds.), *Essays on language function and language type*, pp. 115–143. Amsterdam: Benjamins.

Dukes, M. (2000). Agreement in Chamorro. *Journal of Linguistics 36*, 575–588.

Edmonson, M. S. (1986). *Heaven born Merida and its destiny: The book of Chilam Balam of Chumayel*. Austin: University of Texas Press.

Elouazizi, N. (2005). Anti-agreement Effects as (Anti)-connectivity. In J. Alderete, C. Han, and A. Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics*.

Engdahl, E. (1982). Restrictions on unbounded dependencies in Scandinavian languages. In E. Engdahl and E. Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*, pp. 151–174. Stockholm: Almqvist and Wiksell.

England, N. (1983). *A grammar of Mam, a Mayan language*. Austin, Texas: University of Texas Press.

England, N. (1988). Mam voice. In M. Shibatani (Ed.), *Passive and Voice*. Amsterdam: John Benjamins.

England, N. C. (1989). Mam (Mayan) Clause Structures: Subordinate versus Main Clauses. *International Journal of American Linguistics 55*(3), 283–308.

Erteschik-Shir, N. (1973). *On the nature of island constraints*. Ph. D. thesis, MIT, Cambridge, MA.

Erteschik-Shir, N. (1992). Resumptive Pronouns in Islands. In H. Goodluck and M. Rochemont (Eds.), *Island Constraints*, pp. 89–108. Dordrecht: Kluwer.

Evans, N. (2002). The true status of grammatical object affixes: evidence from Bininj Gunwok. In N. Evans and H.-J. Sasse (Eds.), *Problems of polysynthesis*, pp. 15–50. Berlin: Akademie Verlag.

Facundes, S. (2000). *The language of the Apurinã people of Brazil (Maipure/Arawak)*. Ph. D. thesis, University of New York at Buffalo.

Ferreira, F. and B. Swets (2005). The production and comprehension of resumptive pronouns in relative clause "island" contexts. In A. Cutler (Ed.), *Twenty-first Century Psycholinguistics: Four Cornerstones*, pp. 263–278. Mahway, NJ: Lawrence Erlbaum Associates.

Ferreira, V. and G. S. Dell (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology 40*(4), 296–340.

Finer, D. L. (1997). Contrasting Ā-dependencies in Selayarese. *Natural Language and Linguistic Theory 15*, 677–728.

Foley, W. (1991). *The Yimas Language of New Guinea*. Stanford: Stanford University Press.

Ford, M. (1983). A method of obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior 22*, 203–218.

Fox, B. A. and S. A. Thompson (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language 66*, 51–64.

Fox, J. A. (1978). *Proto-Mayan accent, morpheme structure conditions, and velar innovations*. Ph. D. thesis, University of Chicago.

Fox Tree, J. E. and P. J. A. Meijer (1999). Building syntactic structure in speaking. *Journal of Psycholinguistic Research 28*, 71–92.

Frank, A. and T. F. Jaeger (2008). Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci08)*, pp. 939–944.

Frascarelli, M. (1999). Subject, nominative case, agreement and focus. In L. Mereu (Ed.), *The Boundaries of Morphology and Syntax*, pp. 195–215. Amsterdam/Philadelphia: John Benjamins.

Frascarelli, M. (2000). *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Dordrecht: Kluwer Academic.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, and A. M. Zwicky (Eds.), *Natural Language Parsing*, pp. 129–189. Cambridge: Cambridge University Press.

Friberg, B. (1996). Konjo's peripatetic person markers. *Papers in Austronesian Linguistics 3*, 137–71.

Friedemann, M.-A. (1997). *Sujets Syntaxiques, Positions, Inversions et pro*. Berlin: Peter Lang.

Fuß, E. (2005). *The Rise of Agreement: A Formal Approach to the Syntax and Grammaticalization of Verbal Inflection*. Amsterdam: John Benjamins.

Garrod, S. and A. Sanford (1994). Resolving sentences in discourse context: how discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, pp. 675–698. New York: Academic Press.

Garvey, C. and A. Caramazza (1974). Implicit causality in verbs. *Linguistic Inquiry 5*, 459–64.

Geisler, C. (1998). Infinitival Relative Clauses in Spoken Discourse. *Language Variation and Change 10*, 23–41.

Genzel, D. and E. Charniak (2002). Entropy Rate Constancy in Text. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 199–206.

Georgopoulos, C. (1991). *Syntactic variables: Resumptive Pronouns and A' Binding in Palauan*. Dordrecht: Kluwer.

Gerlach, B. (2002). *Clitics Between Syntax and Lexicon*. Amsterdam: John Benjamins.

Gernsbacher, M. A. (1990). *Language Comprehension and Structure Building*. Hillsdale, New Jersey: Erlbaum.

Gibson, E. (1991). *A computations theory of human linguistic processing: Memory limitations and processing breakdown*. Ph. D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition 68*, 1–76.

Gibson, E. (2000). The dependency locality theory: A distance based theory of linguistic complexity. In Y. Miyashita, A. Marantz, and W. O'Neil (Eds.), *Image, language, brain*, pp. 95–126. Cambridge, MA: MIT Press.

Gilligan, G. M. (1987). *A cross-linguistic approach to the pro-drop parameter*. Ph. D. thesis, University of Southern California.

Givón, T. (1973). Complex NPs, word order and resumptive pronouns in Hebrew. In C. Corum, T. C. Smith-Stark, and A. Weiser (Eds.), *You take the high node and I'll take the low node: Papers from the comparative syntax festival: The difference between main and subordinate clauses*, pp. 135–146. Chicago Linguistic Society.

Givón, T. (1975). On the role of perceptual clause in Hebrew relativization. *Afroasiatic Linguistics 2*(8), 1–17.

Givón, T. (1976). Topic, pronoun and agreement. In C. N. Li (Ed.), *Subject and Topic*, pp. 149–88. New York: Academic Press.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.

Givón, T. (1984). *Syntax: A Functional-Typological Introduction*, Volume 1. Amsterdam: John Benjamins.

Givón, T. (1990). *Syntax: A Functional-Typological Introduction*, Volume 2. John Benjamins.

Givón, T. (1994). The pragmatics of de-transitive voice: Functional and typological aspects of inversion. In T. Givón (Ed.), *Voice and Inversion*, pp. 3–46. Amsterdam: John Benjamins.

Goldsmith, J. (1981). The Structure of *wh*-Questions in Igbo. *Linguistic Analysis 7*(4), 367–393.

Goodall, G. (2004). On the syntax and processing of Wh-questions in Spanish. In V. Chand, A. Kelleher, A. J. Rogríguez, and B. Schmeiser (Eds.), *WCCFL 23: Proceedings of the 23rd Wester Coast Conference on Formal Linguistics*, Somerville, MA, pp. 237–250. Cascadilla Press.

Gordon, P. C. and R. Hendrick (1998). The representation and processing of coreference in discourse. *Cognitive Science 22*, 389–424.

Gordon, P. C., R. Hendrick, and M. Johnson (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition 27*, 1411–1423.

Gordon, P. C., R. Hendrick, and M. Johnson (2004). Effects of Noun Phrase Type on Sentence Complexity. *Journal of Memory and Language 51*(1), 97–114.

Greenberg, J. H. (1978). How does a language acquire gender markers? In J. H. Greenberg, C. A. Ferguson, and E. Moravcsik (Eds.), *Universals of Human Language*. Stanford: Stanford University Press.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics III: Speech acts*, pp. 41–58. New York: Academic Press.

Gries, S. T. (1999). Particle movement: A cognitive and functional approach. *Cognitive Linguistics 10*(2), 105–145.

Grimshaw, J. and V. Samek-Lodovici (1998). Optimal subjects and subject universals. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (Eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, pp. 193–219. Cambridge: MIT Press and MIT Working Papers in Linguistics.

Grodner, D. and E. Gibson (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science 29*, 261–291.

Guerssel, M. (1995). Berber Clitic Doubling and Syntactic Structure. *Revue qu/'eb/'ebecoise de linguistique 24*, 111–133.

Gundel, J. K., N. Hedberg, and R. Zacharski (1993). Cognitive status and the form of referring expressions in discourse. *Language 69*(2), 274–307.

Gutiérrez-Bravo, R. and J. Monforte (2008). On the nature of word order in Yucatec Maya.

Gutiérrez-Bravo, R. and J. Monforte (2009). Focus, agent focus and relative clauses in Yucatec Maya. In H. Avelino, J. Coon, and E. Norcliffe (Eds.), *New perspectives in Mayan Linguistics*, Volume 59. MIT Working Papers in Linguistics.

Haïk, I., H. Koopman, and D. Sportiche (1985). INFL en mooré et le liage dans le système a'. In *Rapport de recherches du groupe de linguistique africaniste*, Montréal.

Haiman, J. (1991). From v/2 to subject clitics: Evidence from Northern Italian. In E. Traugott and B. Heine (Eds.), *Approaches to Grammaticalization*, Volume 2, pp. 135–157. Amsterdam: John Benjamins.

Haiman, J. (1994). Ritualization and the development of language. In W. Pagliuca (Ed.), *Perspectives of Grammaticalization*, pp. 3–28. Amsterdam: John Benjamins.

Haiman, J. and P. Benincà (1992). *The Rhaeto-Romance Languages*. London and New York: Routledge.

Hale, J. T. (2001). A Probabalistic Early Parser as a Psycholinguistic Model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Hall, C. J. (1992). *Morphology and mind: a unified approach to explanation in linguistics*. London and New York: Routledge.

Hankamer, J. (1973). Unacceptable ambiguity. *Linguistic Inquiry 4*, 17–68.

Hanks, W. F. (1990). *Referential Practice, Language and Lived Space among the Maya*. Chicago: Chicago University Press.

Harbert, W. (1995). Binding theory, control and pro. In G. Webelhuth (Ed.), *Government binding theory and the minimalist program*. Blackwell.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft 18*(2), 180–205.

Haspelmath, M. (2002). Functional and formal explanations in linguistics. Handout from M. Haspelmath's class at the Düsseldorf Summer School.

Haspelmath, M. (2004). Explaining the Ditransitive Person-Role Constraint: a usage-based account. *Constructions* (2).

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language 75*(2).

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology 25*, 87–107.

Heine, B. and T. Kuteva (2002). On the evolution of grammatical forms. In *The Transition to Language*, pp. 376–397. Oxford: Oxford University Press.

Hendrick, R. (1988). *Anaphora in Celtic and Universal Grammar*. Dordrecht: Kluwer Academic.

Himmelmann, N. P. (2002). Voice in western Austronesian: an update. In F. Wouk and M. Ross (Eds.), *The History and Typology of western Austronesian voice and voice-marking*, pp. 7–16. Canberra: Pacific Linguistics.

Hofmeister, P. (2007). *Representational Complexity and Memory Retrieval*. Ph. D. thesis, Stanford University.

Hofmeister, P., T. F. Jaeger, I. A. Sag, I. Arnon, and N. Snider (2007). Locality and accessibility in wh-questions. In *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter.

Hofmeister, P. and I. A. Sag (2009). Cognitive constraints on syntactic islands. Submitted, *Language*.

Holmes, V. M. and J. K. O'Regan (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior 20*, 417–430.

Homer, K. (2000). *A Discourse Constraint on Subject Information Questions*. Ph. D. thesis, University of Colorado at Boulder.

Hopper, P. J. and E. C. Traugott (1993). *Grammaticalization*. Cambridge, England: Cambridge University Press.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reducation in Spontaneous Speech*. Ph. D. thesis, Stanford University.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language 59*, 434–46.

Jaeger, T. F. and E. J. Norcliffe (2009). The Cross-linguistic Study of Sentence Production. *Language and Linguistics Compass 3*(4), 866–887.

Jaeger, T. F. and T. Wasow (2007). Processing as a Source of Accessibility Effects on Variation. In *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society*.

Jaeggli, O. and K. J. Safir (1989). The null subject parameter and parametric theory. In O. Jaeggli and K. J. Safir (Eds.), *The Null Subject Parameter*. Dordrecht: Kluwer Academic Publishers.

Jelinek, E. (1984). Empty categories and non-configurational languages. *Natural Language and Linguistic Theory 2*, 39–76.

Kaufman, T. (1986). Outline of comparative Mayan grammar, I: morphology and particles.

Kaufman, T. (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 N. C. England and S. R. Elliott (Eds.), *Lecturas sobre la lingüística Maya*, pp. 59–114. Guatemala: CIRMA.

Kaufman, T. (2002). Comparative mayan studies.

Kaufman, T. S. (1964). Materiales lingüísticos para el estudio de las relaciones internas y externas de la familia de idiomas mayaos. In E. Vogt (Ed.), *Desarrollo cultural de los mayas*, pp. 81–136. Universidad Nacional Autónoma de México.

Kaufman, T. S. (1978). The current state of Mayan comparative phonology. Presented at the 3rd Mayan workshop, Cobán, Guatemala.

Keenan, E. (1972). Relative Clause Formation in Malagasy. In P. M. Peranteau, J. N. Levi, and G. C. Phares (Eds.), *The Chicago Which Hunt: Papers from the Relative Clause Festival*. Chicago: Chicago Linguistic Society.

Keenan, E. L. (1975). Logical Expressive Power and Syntactic Variaton in Natural Language. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press.

Keenan, E. L. (1985). Relative clauses. In T. Shopen (Ed.), *Language Typology and Syntactic Description Volume 2: Complex Constructions*. Cambridge: Cambridge University Press.

Keenan, E. L. and B. Comrie (1977). Noun phrase acceptability and universal grammar. *Linguistic Inquiry 8*(1), 63–91.

Keller, R. (1994). *Language change: The invisible hand in language*. London: Routledge.

King, J. and M. A. Just (1991). Invididual differences in syntactic processing: The role of working memory. *Journal of Memory and Language 30*, 580–602.

King, J. and M. Kutas (1992). ERP responses to sentences that vary in syntactic complexity: Differences between good and poor comprehenders. Poster presented at the Annual Conference of the Society for Psychophysiological Research, San Diego, CA.

King, J. and M. Kutas (1993). Bridging gaps with longer spaces: Enhancing ERP studies of parsing. Poster presented at the Sixth Annual CUNY Sentence Processing Conference, University of Amherst.

Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Klaiman, M. H. (1988). Affectedness and control: A typology of voice systems. In M. Shibatani (Ed.), *Passive and Voice*, pp. 25–84. Amsterdam: John Benjamins.

Klein-Andreu, F. (1991). Losing Ground: a discourse-pragmatic solution to the history of *-ra* in spanish. In S. Fleischman and L. R. Waugh (Eds.), *Discourse-pragmatics and the verb: Evidence from Romance*, pp. 164–178. London: Routledge.

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. *Syntax and Semantics 29: The Limits of Syntax*, 241–279.

Krämer, M. and D. Wunderlich (1999). Transitivity alternations in Yucatec, and the correlation between aspect and argument roles. *Linguistics 37*, 431–480.

Kroch, A. (1981). On the role of resumptive pronouns in amnestying island constraint violations. In *Papers from the 17th regional meeting of the Chicago Linguistics Society*, Chicago. Chicago Linguistics Society.

Kroeber, P. D. (1999). *The Salish language family: reconstructing syntax*. Studies in the anthropology of North American Indians. University of Nebraska Press in cooperation with the American Indian Studies Research Institute, Indiana University.

Krug, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics 26*(4), 286–320.

Kuryłowicz, J. (1964). *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.

Kvam, S. (1983). *Linkverschachtelung im Deutschen und Norwegischen*. Tübingen: Max Niemeyer Verlag.

Labelle, M. (1990). Predication, *wh*-movement and the development of relative clauses. *Language acquisition 1*(1), 95–119.

Labov, W. (1963). The social motivation of a sound change. *Word 19*, 273–309.

Lambrecht, K. (1981). *Topic, Antitopic and Verb-agreement in Non-Standard French*. Pragmatics and Beyond, Vol. II:8. John Benjamins.

Larsen, T. W. (1981). Functional correlates of ergativity in Aguacatec. In *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society*, pp. 136–153.

Larsen, T. W. and W. M. Norman (1979). *Ergativity: towards a theory of grammatical relations*, Chapter Correlates of ergativity in Mayan grammar, pp. 347–370. London: Academic Press.

Lecarme, J. (1989). Focus en somali: syntax et interpretation. Paper presented at the 4th International Congress of Somali Studies, Mogadiscio.

Legum, S. (1975). Strategies in the Acquisition of Relative Clauses. *Technical Note*, 2-75-10, Southwest Regional Laboratory.

Lehmann, C. (1988). On the function of agreement. In M. Barlow and C. A. Ferguson (Eds.), *Agreement in natural language*, pp. 55–65. Stanford: CSLI.

Lehmann, C. (1993a). The genesis of auxiliaries in Yucatec Maya. In A. Crochetiere, J. Boulanger, and C. Ouellon (Eds.), *Proceedings of the International Congress of Linguistics 15 (1992)*, Volume 2, pp. 313–316.

Lehmann, C. (1993b). Predicate classes in Yucatec Maya. *Función 13-14*(195-272).

Lehmann, C. (1998). *Possession in Yucatec Maya*. Munich: LINCOM Europa.

Lehmann, C. (2002). *Thoughts on Grammaticalization* (2 ed.). Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt.

Levy, R. (2005). *Probabilistic models of word order and syntactic discontinuity*. Ph. D. thesis, Stanford University.

Levy, R. and T. F. Jaeger (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, and T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS) 19*, pp. 849–856. Cambridge, MA: MIT Press.

Linder, K. P. (1987). *Grammatische Untersuchungen zur Charakteristik des Rätoromanischen in Graubünden*. Tübingen: Gunter Narr.

Lois, X. and V. Vapnarsky (2003). *Polyvalence of root classes in Yucatecan Mayan Languages*. Studies in Native American Languages. Munich: LINCOM Europa.

Lopez Otero, D. (1914). *Gramatica Maya: Metodo teorico practico*. Founded, in part, on notes by Audomaro Molina. Merida: Private Press.

Lucy, J. (1994). The role of semantic value in lexical comparison: Motion and position in Yucatec Maya. *Linguistics 32*(4), 623–656.

Macaulay, M. (1996). *A Grammar of Chalcatongo Mixtec*. University of California Publications in Linguistics 127. Berkeley: University of California Press.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language 32*, 692–715.

MacDonald, M. C., N. J. Pearlmutter, and M. S. Seidenberg (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review 101*(4), 676–703.

Mak, W. M., W. Vonk, and H. Schriefers (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language 47*, 50–68.

McCloskey, J. (1979). *Transformational syntax and model theoretic semantics: A case-study in Modern Irish*. Dordrecht: Reidel.

McCloskey, J. (1990). Resumptive pronouns, A-binding and levels of representation in Irish. In R. Hendrick (Ed.), *Syntax of the Modern Celtic Languages*, Volume 23 of *Syntax and Semantics*, pp. 199–248. New York and San Diego: Academic Press.

McCloskey, J. (2007). Resumption. http://ohlone.ucsc.edu/ jim/PDFFiles/syncom.pdf.

McCloskey, J. and K. Hale (1984). On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory 1*, 487–533.

McDonald, M. C. and B. MacWhinney (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language 29*, 469–492.

McQuown, N. A. (1967). Classical Yucatec (Maya). In R. Wauchope (Ed.), *Handbook of Middle American Indians*, Number 201-247. University of Texas Press.

McRae, K., M. J. Spivey-Knowlton, and M. K. Tanenhaus (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language 38*, 283–312.

Meillet, A. (1912). L'evolution des formes grammaticales. *Scienta (Rivista di scienza) 12*(26.6).

Meiri, N. and A. Nitzan (1995). Resumptive pronouns and accessibility theory. Seminar paper, Linguistics Dept. Tel Aviv University.

Mereu, L. (1999). Agreement, pronominalization and word order in pragmatically-oriented languages. In *The Boundaries of Morphology and Syntax*. Amsterdam/Philadelphia: John Benjamins.

Michaelis, L. and H. S. Francis (2007). Lexical Subjects and the Conflation Strategy. In *Topics in the Grammar-Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*. Amsterdam: Benjamins.

Miller, P. H. (1992). Postlexical Cliticization vs. Affixiation: Coordination Criteria. In *Proceedings of the 28th Meeting of the Chicago Linguistics Society*, Chicago, pp. 382–396. CLS.

Miller, P. H. and I. A. Sag (1997). French clitic movement without clitics or movement. *Natural Language and Linguistic Theory 15*, 573–639.

Mithun, M. (1991). The Development of Bound Pronominal Paradigms. In W. P. Lehmann and H.-J. J. Hewitt (Eds.), *Language Typology 1988: Typological Models in Reconstruction*, Amsterdam Studies in the Theory and History of Linguistic Science, 81, Amsterdam, pp. 85–104. John Benjamins.

Mithun, M. (2003). The information status of pronominal affixes. In *Transactions of the Philological Society*, Volume 10, pp. 235–278. London: Blackwell.

Miyake, A. and P. Shah (1999). Models of Working Memory: An Introduction. In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pp. 1–27. Cambridge: Cambridge University Press.

Monachesi, P. (1999). *A lexical approach to Italian cliticization*. Lecture Notes series NO. 84. Stanford: CSLI.

Mondloch, J. (1978). Disambiguating subjects and objects in Quiche Mayan. *Journal of Mayan Linguistics 1*, 3–19.

Morgan, J. L. (1972). Some aspects of relative clauses in English and Albanian. In P. M. Peranteau, J. N. Levi, and G. C. Phares (Eds.), *The Chicago Which Hunt: Papers from the Relative Clause Festival*. Chicago: Chicago Linguistic Society.

Morin, Y.-C. (1981). Some Myths about Pronominal Clitics in French. *Linguistic Analysis 8*, 95–109.

Müller, F. M. (1875). *Vorlesungen über die Wissenschaft der Sprache*. Leipzig: Julius Klinkhardt.

Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language 62*, 56–119.

Norcliffe, E. (2006). The Focus Construction in Jakalteko: A Biclausal Account. Paper presented at the 80th Annual Meeting of the Linguistics Society of America, Albuquerque.

Norcliffe, E. (2009). Revisiting Agent Focus in Yucatec. In H. Avelino, J. Coon, and E. Norcliffe (Eds.), *New perspectives in Mayan Linguistics*, Volume 59. MIT Working Papers in Linguistics.

Nordlinger, R. (1998). *A Grammar of Wambaya, Northern Territory (Australia)*. Pacific Linguistics C-140. Canberra: ANU.

Norman, W. (1977). Topic and focus in Mayan. Paper presented at the Mayan Workshop II, San Cristóbal de las Casas, Chiapas, México.

Olson, C. (1992). Gumawana (Amphlett Islands, Papua New Guinea): grammar sketch and texts. *Papers in Austronesian Linguistics 2*, 251–430.

Ouhalla, J. (1993). Subject extraction, negation and the anti-agreement effect. *Natural Language and Linguistic Theory 11*, 477–518.

Ouhalla, J. (2005). Agreement features, agreement and antiagreement. *Natural Language and Linguistic Theory 23*, 655–686.

Owen, M. G. (1968). *The semantic structure of Yucatec verbs*. Ph. D. thesis, Yale University.

Pérez-Leroux, A. T. (1995). Resumptives in the acquisition of relative clauses. *Language Acquisition 4*, 105–138.

Perlmutter, D. (1972). Evidence for shadow pronouns in French relativization. In P. M. Peranteau, J. N. Levi, and G. C. Phares (Eds.), *The Chicago Which Hunt: Papers from the Relative Clause Festival*. Chicago: Chicago Linguistic Society.

Pfeiler, B. (1995). Variación fonológica en el maya yucateco. In *Vitalidad e influencia de las lenguas indígenas en Latinoamérica*, pp. 488–497. México, D.F.: Universidad Nacional Autónoma de México, Instituto de Investigaciones Antropológicas.

Phillips, C. (1995). Ergative Subjects. In C. Burgess, K. Dziwirek, and D. Gerdts (Eds.), *Grammatical Relations: Theoretical Approaches to Empirical Issues*, pp. 341–358. Stanford: CSLI.

Poletto, C. (1995). The diachronic development of subject clitics in North-Eastern Italian dialects. In A. Battye and I. Roberts (Eds.), *Clause Structure and Language Change*, pp. 295–324. Oxford: Oxford University Press.

Poletto, C. (1997). Pronominal syntax. In M. Maiden and M. Parry (Eds.), *The Dialects of Italy*, pp. 137–145. Cambrige: Cambridge University Press.

Poletto, C. (1999). The internal structure of Agrs and subject clitics in the Northern Italian dialects. In H. van Riemsdijk (Ed.), *Clitics in the Languages of Europe*, pp. 581–620. Berlin: Mouton de Gruyter.

Prince, E. (1990). Syntax and discourse: A look at resumptive pronouns. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, pp. 482–497.

Race, D. S. and M. C. MacDonald (2003). The use of "that" in the Production and Comprehension of Object Relative Clauses. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*.

Reali, F. and M. H. Christiansen (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language 57*, 1–23.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, L. (1986a). Null Objects in Italian and the Theory of pro. *Linguistic Inquiry 17*, 501–557.

Rizzi, L. (1986b). On the status of subject clitics in Romance. In O. Jaeggli and C. Silva-Corvalan (Eds.), *Studies in Romance Linguistics*, pp. 391–419. Dordrecht: Foris.

Roberge, Y. (1990). *The Syntactic Recoverability of Null Arguments*. Montreal: McGill and Queens University Press.

Roberts, I. and A. Roussou (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Romaine, S. (2007). Preserving endangered languages. *Language and Linguistics Compass 1*, 115–132.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. Ph. D. thesis, MIT.

Ross, M. (2002). The history and transitivity of western Austronesian voice and voice-marking. In F. Wouk and M. Ross (Eds.), *The History and Typology of western Austronesian voice and voice-marking*, pp. 17–62. Canberra: Pacific Linguistics.

Rude, N. (1985). *Studies in Nez Perce grammar and discourse*. Ph. D. thesis, University of Oregon, Eugene.

Saah, K. K. and H. Goodluck (1995). Island effects in parsing and grammar: Evidence from Akan. *Linguistic Review 12*, 381–409.

Sapir, E. (1921). *Language: an introduction to the study of speech*. New York: Harcourt Brace.

Schneider-Zioga, P. (1995). Specifier/head agreement in Kinande. *Cahiers linguistiques d'Ottawa 23*, 67–93.

Schneider-Zioga, P. (1996). An argument for Agr phrase. In V. Montapayne and A. D. Green (Eds.), *Proceedings from the Eastern States Conference on Linguistics 1996*, Cornell University.

Schneider-Zioga, P. (2000). Anti-agreement and the Fine Structure of the Left Periphery. In *University of California Irvine Working Papers in Linguistics*, Volume 6.

Schneider-Zioga, P. (2002). The Case of Anti-Agreement. In *Proceedings of AFLA 8*, Cambridge, MA. MIT Working Papers in Linguistics.

Schuchardt, H. (1885). *Gegen die Junggrammatiker*. Berlin: Robert Oppenheimer.

Sells, P. (1984). *Syntax and Semantics of Resumptive Pronouns*. Ph. D. thesis, University of Massachusetts, Amherst.

Sharvit, Y. (1999). Resumptive pronouns in relative clauses. *Natural Language and Linguistic Theory 17*, 587–612.

Shlonsky, U. (1992). Resumptives pronouns as a last resort. *Linguistic Inquiry 23*(2), 443–468.

Siewierska, A. (1999). From anaphoric pronoun to grammatical marker: why objects don't make it. *Folia Linguistica 33*(2), 225–251.

Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press.

Silva-Corvalán, C. (1996). Resumptive pronouns: A discourse explanation. In C. Parodi, C. Quicoli, M. Saltarelli, and M. L. Zubizarreta (Eds.), *Aspects of Romance Linguistics*, pp. 383–395. Washington D. C.: Georgetown University Press.

Simpson, J. (1991). *Warlpiri Morpho-Syntax*. Dordrecht: Kluwer Academic.

Skopeteas, S. and E. Verhoeven (2005). Postverbal argument order in Yucatec Maya. *Sprachtypologie und Universalienforschung 53*(1), 71–79.

Skopeteas, S. and E. Verhoeven (2009). Distinctness effects on VOS order: Evidence from yucatec maya. In H. Avelino, J. Coon, and E. Norcliffe (Eds.), *New perspectives in Mayan Linguistics*, Volume 59. MIT Working Papers in Linguistics.

Smailus, O. (1989). *Gramámatica del Maya Yucateco Colonial*. Hamburg: Wayasbah.

Smith-Stark, T. (1978). The Mayan Antipassive: Some Facts and Fictions. In N. C. England (Ed.), *Papers in Mayan Linguistics*, pp. 169–187. Columbia: University of Missouri.

Stiebels, B. (2006). Agent focus in Mayan languages. *Natural Language and Linguistic Theory 24*(2), 501–570.

Suñer, M. (1998). Resumptive restrictive relative clauses: A crosslinguistic perspective. *Language 74*, 335–364.

Suttles, W. (2004). *Musqueam Reference Grammar*. Vancouver: University of British Columbia Press.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science 268*, 1632–1634.

Taraldsen, K. (1982). Extraction from relative clauses in Norwegian. In E. Engdahl and E. Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*. Stockholm: Almqvist and Wiksell.

Tarallo, F. (1986). Functional and structural properties in a variable syntax. In D. Sankoff (Ed.), *Diversity and diachrony*, pp. 249–260. John Benjamins.

Temperley, D. (2003). Ambiguity Avoidance in English Relative Clauses. *Language 79*(3), 464–484.

Thompson, C. (1989). *Voice and obviation in Athabaskan and other languages*. Ph. D. thesis, University of Oregon, Eugene.

Tonhauser, J. (2003). F-constructions in Yucatec Maya. In *Proceedings of Semantics of Under-Represented Languages in the Americas (SULA)*, Amherst, MA, pp. 203–223. Graduate Linguistics Students Association Publications.

Topping, D. M. (1973). *Chamorro Reference Grammar*. PALI language texts: Micronesia. Honolulu: University of Hawaii Press.

Tortora, C. (2003). *The Syntax of Italian Dialects*. Oxford: Oxford University Press.

Tozzer, A. M. (1921). *A Maya Grammar*, Volume 9 of *Papers of the Peabody Museum of American Archaeology and Ethnology*. Cambrige, MA: Harvard University.

Traugott, E. C. and B. Heine (1991). Introduction. In E. C. Traugott and B. Heine (Eds.), *Approaches to Grammaticalization*, Volume I, pp. 1–14. Amsterdam: Benjamins.

Tsimpli, I. M. (1999). Null operators, clitics and identification: A comparison between Greek and English. In A. Alexiadou, G. Horrocks, and M. Stavrou (Eds.), *Studies in Greek syntax*, pp. 241–262. Dordrecht: Kluwer.

Tuller, L. A. (1986). *Bijective relations in Universal Grammar and the syntax of Hausa*. Ph. D. thesis, University of California, Los Angeles.

Van Valin, R. D. (1977). *Aspects of Lakhota Syntax*. Ph. D. thesis, University of California, Berkeley.

Van Valin, R. D. (1981). Grammatical relations in ergative languages. *Studies in Language 5*, 361–394.

Verhoeven, E. (2007). *Experiental Constructions in Yucatec Maya*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Warren, T. and E. Gibson (2002). The influence of referential processing on sentence complexity. *Cognition 85*, 79–112.

Wasow, T., T. F. Jaeger, and D. Orr (2005). Lexical variation in relativizer frequency. To appear in *Proceedings of the Workshop on Expecting the Unexpected: Exceptions in Grammar, 27th Anual Meeting of the German Linguistics Association, University of Cologne, Germany*.

Watanabe, A. (1996). *Case absorption and wh-agreement*. Dordrecht: Kluwer Academic Publishers.

Watkins, C. (1969). *Indogermanische Grammatik III.1: Geschichte der indogermanischen Verbalflexion*. Heidelberg: Carl Winter.

Weiß, H. (1998). *Die Syntax des Bairischen. Studien zur Grammatik einer natürlichen Sprache*. Tübingen: Niemeyer.

Whalen, G. (2003). *An Annotated Translation of a Colonial Yucatec Manuscript: On Religious and Cosmological Topics by a Native Author*. FAMSI.

Wiltschko, M. (2006). On Ergative Agreement and Anti-Agreement in Halkomelem Salish. In S. Bischoff, L. Butler, P. Norquest, and D. Siddiqi (Eds.), *Studies in Salishan*, Volume 7, pp. 241–273. MIT Working Papers on Endangered and Less Familiar Languages.

Zavala Maldonado, R. (1997). Functional analysis of Akatek voice constructions. *International Journal of American Linguistics 63*, 439–474.

Zipf, G. K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. Houghton Mifflin.

Zobel, E. (1997). Voice systems in Nusa Tenggara Barat. Paper presented at The Austronesian Focus workshop, Eighth International Conference on Austronesian Linguistics, Taipei, Taiwan.

Zobel, E. (2002). The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax. In F. Wouk and M. Ross (Eds.), *The History and Typology of western Austronesian voice and voice-marking*, pp. 405–434. Canberra: Pacific Linguistics.

Zwicky, A. and G. Pullum (1983). Cliticization vs. inflection: English N'T. *Language 59*, 502–513.