DEPARTMENT OF LINGUISTICS
UNIVERSITY OF DHAKA

Analyzing the Syntax-Information Structure Interface of Focus Constructions in Bengali

Registration Number: 02/2006-3007
Re-registration Number: 56/2012-2013

Ph D Thesis
Submitted by

Syed Shahrier Rahman

November 2013

ABSTRACT

The present research has explored the focus marking in Bengali syntax. It provides a systematic overview, description and investigation of aspects related to information structure (IS) in the grammar of Bengali by applying methodology that allows cross-linguistic comparison of the data.

The present study is a combination of spontaneous data collection and experiments. In course of experiments it has been highlighted on focus word order patterns. In addition to the intrinsic scientific knowledge that can be gained, this study needs to find the correlation between information packaging and sentential focus in Bengali. The major motivation of this research is to know how the information structure of Bengali focus constructions affects the position of subjects, objects, adjuncts and verbs. The data collected indicate a marked preference for SOV (subject-object-verb) in a wide variety of discourse contexts, a preference that differs from those claimed to apply in similar contexts in Bengali. Nevertheless, intwo separate object narrow-focus conditions; Bengali subjects displayed a preference for SOV word order as compared to OVS (object-verb-subject) or OSV (object-subject-verb).

The presence of clitics implies the projection of focus and the extension of the preverbal field into the left periphery. The cliticization data gathered for Bengali in main clauses and subordinate clauses context

suggest a number of preverbal positions in which preverbal subjects, affective phrases, and topic elements may appear. In course of this research, it has been suggested that Bengali allows a Clause-internal Focus position between TP and vP. A Focused phrase must move overtly to SpecFoc. The subject in Spec TP is required to be 'presupposed', the subject in Spec vP is 'not presupposed'. The EPP feature of 'T' in Bengali is the feature [+N, +presupposed]. The reason is, only presupposed NPs can be licensed in Spec TP. The non-overt 'Expl(etive) pro' that I assume is merged in Spec TP when there is no subject in Spec TP, will have a feature [+N, +presupposed].

In this dissertation, it has been shown systematically how focus is the 'driving force' of Bengali Syntax. The clause structure of Bengali is more like a 'discourse structure', which is surprisingly, part of the computational system of Bengali. The word order and scope effects are thus the corollaries of the semantic partition of the clause structure that has been proposed.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Jeenat Imtiaz Ali for the continuous support of my Ph.D study and research, for his patience, motivation and enthusiasm. His guidance helped me in all the time of research and writing of this thesis.

In particular, I am ever grateful to my teacher Prof. Dr. Bhishmadeb Choudhury for enlightening me the first glance of research and for supporting me in all crucial issues of my life.

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here. Above all, I would like to thank my family for their support at all times.

I would like to acknowledge the academic and technical support of the University of Dhaka and its staff. The library facilities and computer facilities of the University have been indispensable. I remember the generous and spontaneous cooperation of the respondents who have participated in my field work. I also thank the Department of Linguistics for their support and assistance since the start of my research work in 2006.

I would also like to thank my colleagues and friends as they encourage me a lot to accomplish this huge work. For any errors or inadequacies that may remain in this work, of course, the responsibility is entirely my own.

Syed Shahrier Rahman

LIST OF ABBREVIATIONS & TRANSLITERATION SYMBOLS

Abbreviations

Acc	Accusative Case
CCG	Combinatory Categorial Grammar
CD	Communicative Dynamism
CF	Contrastive Focus
CL	Clitic
CP	Complementizer Phrase
Dat	Dative Case
DET	Determiner
DO	Direct Object
DP	Determiner Phrase
EMPH	Emphatic
EPP	Exponential Projection Principle
FF	Focus Fronting
FinP	Finite Phrase
FOC	Focus
FP	Focus Phrase
FUT	Future Tense
GEN	Genitive
INF	Infinitive
INFL	Inflection
IP	Inflectional Phrase
IS	Information Structure

LF	Logical Form
MP	Minimalist Programme
NEG	Negative
nonF	Non Finite
NP	Noun Phrase
NSR	Nuclear Stress Rule
OBJ	Object
OSV	Object-Subject-Verb
OVS	Object-Verb-Subject
PAS	Passive
PF	Phonological Form
Pl	Plural
PP	Prepositional Phrase
PPA	Possible Pragmatic Assertions
PRF	Perfect
PROG	Progressive
PRS	Present Tense
PRT	Particle
PST	Past Tense
QR	Quantifier Raising
QUAL	Qualitative
QUAN	Quantitative
SDRT	Segmented Discourse Representation Theory
SOV	Subject-Object-Verb

Spec	Specifier
S-structure	Sentence Structure
STopic	Sentence Topic
SUB	Subject
SVO	Subject-Verb-Object
TOP	Topic
TP	Tense Phrase
VOS	Verb-Object-Subject
vP	Little verb Phrase
VP	Verb Phrase
VSO	Verb-Subject-Object
X^{max}	Maximal projection of X
XP	X Phrase
YP	Y Phrase

Transliteration Symbols

Vowel letters		**Consonant letters**					
অ	o’	ক	k	ড	D	ম	m
আ	a	খ	kh	ঢ	Dh	র/ঋ	r
ই/ঈ	i	গ	g	ণ/ন	n	ড়	R
উ/ঊ	u	ঘ	gh	ত	t	ঢ়	Rh
এ/	e	ঙ/ং	ng	থ	th	ল	l
ঐ	oi	চ	ch	দ	d	শ/ষ	sh
ও	o	ছ	chh	ধ	dh	স	s
ঔ	ou	জ/য	j	প	p	হ	h
অ্যা	ae	ঝ	jh	ফ	ph	য়	y
		ট	T	ব	b		
		ঠ	Th	ভ	bh		

TABLE OF CONTENTS

1 CHAPTER ONE: INTRODUCTION 10

1.1 OUTLINE 11
1.2 CORE FACTS 14
1.3 BENGALI LANGUAGE: DEFAULT WORD ORDER 18
1.4 BENGALI FOCUS CONSTRUCTIONS 22
1.5 CLITICIZATION IN BENGALI: A FOCUS TOOL 26
1.6 MAIN QUESTION 35
1.7 HYPOTHESIS 36
1.8 CONCLUSION 37

2 CHAPTER TWO: THE RELATIONSHIP BETWEEN SYNTAX AND INFORMATION STRUCTURE 38

2.1 INFORMATION STRUCTURE 40
2.2 TOPIC/THEME 42
2.3 FOCUS/RHEME 45
2.4 SYNTACTIC ACCOUNTS OF THE SYNTAX-INFORMATION STRUCTURE INTERFACE 57
2.4.1 Information Structure Dichotomies 60
2.4.2 The Interface and Phases 61
2.5 SYNTAX-INFORMATION STRUCTURE INTERFACE MODEL 64
2.6 IMPLICATIONS: DISCOURSE FACTORS IN THE COLLECTION OF SPEAKER JUDGMENTS 71
2.7 CONCLUSION 73

3 CHAPTER THREE: METHODOLOGY 76

3.1 MIXED METHOD: AN OVERVIEW 77
3.2 RESEARCH DESIGN 81
3.3 PARTICIPANTS: TASKS 1 AND 2 82
3.3.1 Participant variables 83
3.3.2 Procedures: Tasks 1 and 2 85
3.4 TASK 1: SCALED PRAGMATIC APPROPRIATENESS TASK 86
3.4.1 Condition A 87
3.4.2 Condition B 88
3.4.3 Condition C 90
3.4.4 Condition D 91
3.4.5 Condition E 92
3.4.6 Condition F 93
3.4.7 Condition G 95
3.5 TASK 2: WORD ORDER PREFERENCE TASK 96
3.5.1 Condition A 97
3.5.2 Condition B 98
3.5.3 Condition C 98
3.5.4 Condition D 99
3.5.5 Condition E 100
3.6 TASK 3: RECORDED FIELD INTERVIEW 101
3.7 CONCLUSION 103

4 CHAPTER FOUR: ANALYSIS OF QUANTITATIVE AND QUALITATIVE MEASURES 105

4.1 INTRODUCTION 106
4.2 TASK 1 107

4.2.1 Condition A 107
4.2.2 Condition B 108
4.2.3 Condition C 110
4.2.4 Condition D 111
4.2.5 Condition E 112
4.2.6 Condition F 113
4.2.7 Condition G 118
4.2.8 Summary of Task 1 discourse conditions 119
4.2.9 Task 1 by sociolinguistic factors 120
4.2.10 Task 1 by gender 128
4.3 TASK 2 131
4.3.1 Condition A 132
4.3.2 Condition B 133
4.3.3 Condition C 134
4.3.4 Condition D 135
4.3.5 Condition E 136
4.3.6 Task 2 by sociolinguistic factors 138
4.3.7 Task 2 statistical results by gender 142
4.3.8 Task 2 Summary 143
4.4 FOLLOW-UP TASK FOR TASK 2 (TASK X) 144
4.5 TASK 3 146
4.6 SUMMARY AND DISCUSSION: QUANTITATIVE MEASURES 151
4.6.1 A note on clitic response ratings 152
4.7 METHODOLOGICAL CONSIDERATIONS 154
4.8 CONCLUSION 157

5 CHAPTER FIVE: FOCUS AND INFORMATION-STRUCTURE IN BENGALI 159

5.1 THE SYNTACTIC BEHAVIOR OF BENGALI FOCUS 160
5.2 POSITION OF THE FOCP IN BENGALI 166
5.2.1 Semantic Division of the Clause Structure 166
5.2.2 Proof that Bengali has a Clause-int(ernal) Focposition between TP and vP 167
5.2.3 Specificity Effect 168
5.2.3.1 Specificity explained 173
5.2.3.2 Explaining Focus 178
5.3 ENCODING DISCOURSE FEATURES IN THE LEXICON 178
5.4 TOPIC FOCUS AND OBJECT SHIFT 181
5.4.1 Topic Field 181
5.4.2 Object shift 185
5.4.2.1 Position of an Indefinite 186
5.5 CONCLUSION 189

6 CHAPTER SIX: CONCLUSION 190

7 APPENDIX 193

8 REFERENCES 197

Chapter One

Introduction

Bengali (also known as Bangla), the official language of Bangladesh, is a South Asian Indo-Arian language. Typologically, it is an agglutinative (Majumdar, 1920, p. 206) language mainly spoken in the Indian sub-continent. Besides Bangladesh, it is used as official (and regional official) language of the states of West Bengal, Tripura and Assam of the Republic of India. At present, the accepted standard language in West Bengal and (all parts of) Bangladesh are identical. The present research has explored the focus marking in Bengali syntax. It provides a systematic overview, description and investigation of aspects related to information structure in the grammar of Bengali by applying methodology that allows cross-linguistic comparison of the data. In order to achieve the aims mentioned above, the following objectives have been taken into account.

- Determine the attributes of the focus word order used by the native speakers of Bengali.
- Derive the underlying information structure of the above mentioned focus constructions.
- Find a relevant focus construction model to explain the nature of Bengali syntax-information structure (IS) interface.

By accomplishing these objectives, I shall be able to explain the nature of Bengali focus construction on the basis of the information structure.

1.1 Outline

Bengali speakers enjoy the liberty of using a variety of sentential word orders, which is also available in many other South-Asian languages. It

seems that Bengali syntax offers additional choices for word ordering within certain basic constituents, and native speakers of this language can choose any of them without any kind of syntactic and semantic restriction. However, the provision of such optionality is completely against the spirit of the economy principle as found in Minimalist Programme (MP). The minimalist framework assumes that the sole motivation for movement is the checking of formal features. In contrast, the concept of optionality in word orders has to allow two or more structures which are equivalent in interpretation. Nevertheless, it is not possible without modifying economy by claiming that those structures stem from different numerations. In fact, this is not also plausible in minimalist framework. As a result, the economy principle totally defies the optionality and tries to explain the presence of word order variations in a different way.

Miyagawa (1997) offers a proposal that argues against true optionality with respect to cases of scrambling in Japanese. He shows that each of the possible word-orders, usually referred to as being 'free', carries a slightly different interpretation, and is motivated by different syntactic and semantic considerations. Adger (2003) also demonstrates that subject scrambling in German, a movement operation that appears to be optional, in fact carries a difference in meaning with respect to presuppositions. Inspired by such arguments, it is possible to pose an important question about the correlation between focus word order

preferences by Bengali speakers and their supplied information structures.

The present study is a combination of spontaneous data collection and experiments. In course of experiments I have highlighted on focus word order patterns because in addition to the intrinsic scientific knowledge that can be gained, I need to find the correlation between information packaging and sentential focus in Bengali. In the course of this outline, my motivation is to know how the information structure of Bengali focus constructions affects the position of subjects, objects, adjuncts and verbs. This will also help me to sketch out the least marked order in this language, the order which is used to disambiguate, and the most frequent order. In this study, I shall assume that the role of information packaging is relevant to answering this question. The best way to find such facts is to collect relevant data from an authentic database like oral corpora. However, such a database is yet to be available in Bengali. Hence, I have collected primary data from native adults.

This dissertation consists of five chapters. The first chapter will work as an introduction to the main research area. The core facts related to Bengali focus and other syntactic issues will be demonstrated here briefly. The second chapter will explain the theoretical base for the information structure and focus constructions. In this chapter, I shall figure out the nature and properties of the focus and information structure as well. At the end of this chapter, I have been able to draw a

data collection design with the help of those theoretical issues. Since I need a well-designed framework for collecting spontaneous responses from Bengali native adults, the third chapter will provide the description of this experimental design and the methodology for collecting data. In the fourth chapter, I have started out to analyze the data. I also need to check the descriptive statistics of different components of the data set and illustrate findings with the help of graphs and tables. I have discussed findings and results in the fifth chapter. In the discussion area, a relevant focus construction model has been figured out to explain the nature of Bengali syntax-information structure (IS) interface. Finally I draw the conclusion and offer links for further research.

1.2 Core Facts

Bengali focus constructions and its information structure– are two central concerns of this research. Throughout, I use the term focus in the following sense: the focus of a sentence is the portion that contributes the most salient or relevant information (Aboh, 2007, p. 1). Focus generally provides new or contrasting information (Selkirk E. , 1984, p. 200). More specifically, focus indicates that a number of (contextually specified) alternatives are under consideration. Focus may be used to supply a range of pragmatic functions, such as identifying which alternative(s) has a pertinent property (new-information focus), designating a contrast between alternatives (contrastive focus), etc. It is important to mention that such a description does not make indication to the explicit means used by a language to point out focus. While

languages like English regularly apply intonation to mark focus, many languages make use of morphological and syntactic strategies (Erteschik-Shir, 2007). For example, movement to a clause-initial (or left-peripheral) position is a commonly used focus strategy cross-linguistically (Aboh, 2007). Hungarian is often maintained to have a preverbal position reserved for (exhaustive) foci (Kiss, 1998). However, many languages apply strategies where a functional morpheme designates that an element (usually) adjacent to it supplies new or contrasting information. In Bengali, positional variations of constituents in the syntactic domain, morphosyntactic cliticiztions and of course in-situ or ex-situ intonation strategies are used to manifest focus constructions and their information packaging as well. Since I delimit the present research area within the application of Bengali syntax, I shall try to concentrate only on the movement and variations of focus constituents and clitics. In the following sections of this chapter, I plan to introduce these elements and show their uses in Bengali language.

A particular focus construction may use more than one of the above strategies. Such as, in English it-clefts, the focused constituent occupies an assigned position (syntactic strategy), and is accented (intonational strategy). Many Asian languages combine a syntactic strategy with a morphological strategy to indicate DP focus. Bengali is a language of this type. I take focus to be a syntactic feature which can be manifested prosodically, morphosyntactically, or not at all. Absence of an overt realization of focus does not imply its non-existence. For instance, a

syntactic component that is pragmatically aimed to contrast with alternatives is still considered to be focused, regardless of whether it is overtly distinct or not (Sarkar, 2006). The most visible paradigm for identifying constituents as focused is question-answer dialog (Selkirk E. , 1996). For example, in English, merely new information can be prosodically prominent in an ordinary question-answer dialog. As a result (1b), but not (1c), is a suitable answer to (1a). Usually, focus on a constituent in an answer obligatorily corresponds to the wh-expression in the question.

(1) a. Who does Sumon saw?

b. Sumon saw Shobuj.

c. # SUMON saw Shobuj.

In (2), *read* contrasts with *write*, and both elements are (intonationally) focused.

(2) I didn't READ, I WRITE.

Whereas some languages may diverge in how contrastive and new-information focus is manifested. Here I present the concepts of contrastive and new-information focus, since Bengali identifies them morphosyntactically in nearly the similar ways.

I have already mentioned that a focus particle is a lexical item whose meaning interacts with the focus/background partition of sentences in which it occurs (Sudhoff, 2010, p. 6). Sentences containing focus

particles such as English *only* and *also* can have diverse truth-conditions depending on which constituent in their scope is focused. Therefore, several particles are labeled as focus-sensitive in a sentence. In the following example, (3a) but not (3b) is referred as false if Sumon also bought the book. (3b) but not (3a) is referred false if Sumon also borrowed a pencil box.

(3) a. Sumon only [BORROWED the book]

b. Sumon only [borrowed THE BOOK]

Along with focus construction, information structure also has a close connection with the discourse functions of a language. Two informational structural descriptions of propositions, namely 'topic' and 'focus', are very prominent in this formal organization of linguistic expression. I have already presented a brief description of focus. As a part of information structure, 'topic' is used to convey the aboutness of a linguistic expression and it helps the new information to be conveyed. On the other hand as I have mentioned above, 'focus' provides either new or contrastive information by emphasizing a syntactic domain of a sentence. In the alternative case, it is possible to state that the sentential discourse is organized by information packaging devices, such as topic, focus etcetera. The topic of a sentence which bears the aboutness phenomenon can roughly be defined as what the sentence is about. Usually the subject of a sentence functions as the topic of that sentence (Partee, 1992). For this reason, in the most common cases, it makes the left peripheral zone of a sentence the preferred position for the

topichood. However, it is also possible for a topic to occupy other positions in a sentence. Focus, on the other hand, is treated as the source of the most informative part of the sentence. It relates to what the speaker in the particular situation regards as unknown to the hearer. Phonologically, focus is the source of the main prosodic prominence of the sentence (Chomsky, 1971). One of the prominent roles of focus is to introduce new entities into the discourse. In other way, a sentential phrase denoting an entity already mentioned in the previous discourse can appear as the focus of the sentence. Hence there is a major split into Information and Identificational Focus (Kiss, 1998). The disparity between the two is closely connected to both the syntactic and semantic ideas (Kiss, 1998). Prosodically, identificational focus is typified by emphatic stress, while information focus is differentiated by a falling tone or sentential stress (Reinhart, 1995). Structurally, it has been noted cross-linguistically that identificational focus has a propensity to involve movement where as information focus is usually assigned *in situ* (Reinhart, 1995; Kiss, 1998).

1.3 Bengali Language: Default Word Order

Bengali is a predominately SOV language, with post-nominal determiners and adpositions. Alignment is mainly nominative/accusative. Subject/object person agreements are comprehended by cross-reference affixes on the (inflected) verb. Finite verbs agree in person with their subject. Tense and mood are marked by verbal suffixes. Aspectual information is determined by a verb-stem

alternation. It is possible to assume that the position of topic or focus in a sentence is applied and activated on the basis of a hypothetically neutral word order. However, I can draw an outline of this highly conceptual phenomenon and make a strong assumption about the neutral word order for Bengali. Indeed, it is impossible to determine such a word order that is absolutely neutral. Since any kind of utterance is always at least to some extent contextually motivated, I can say that the word order is more general and has the ability to be applied in many different conversations. Else, the word order does not have any specific context, and the speaker uses this word order with all new information when he or she is motivated to utter sentences out of the blue. In general, SOV prevails as a common possible preference in Bengali native speakers' intuition. To check this intuition, I can figure out a test to know what people say neutrally in a given situation. In this test, I introduce three different situations where the first one has a topic; the second one demands some new information and the third one requires all new information. I want to see what the native speaker in Bengali thinks about the word order in these specific situations:

(4a) Situation 1# The speaker is requested to say something about Sumon

Speaker: Sumon sharadin boi po'Re. S Adjunct O V

Sumon all day long book.pl read.PRS.3

'Sumon reads books all day long.'

Situation 2# The speaker is asked: what did Sumon buy yesterday?

Speaker: Sumon go'tokal ekti boi kinechilo S Adjunct O V
Sumon yesterday a book buy.PST.3
'Sumon bought a book yesterday.'

Situation 3# The speaker is asked: what is going on?

Speaker: Sumon o'nekkhon dhore boi poRche S Adjunct O V
Sumon a long time for book read.PRS.Prog.3
'Sumon has been reading a book for a long time.'

As I have found SOV word orders in all three types of situations. In contrast, I also consider other possible word orders in Bengali like SVO and OSV to determine that SOV order is felicitous in various contexts where as other orders are not felicitous in all of them. First, I try to apply OSV order in the context of example (4a) where someone is asked to tell something about Sumon:

(4b) Speaker:	?boi sharadin Sumon po'Re	O	Adjunct S V
Or,	?sharadin boi Sumon po'Re	Adjunct	O S V
Or,	?boi Sumon po'Re sharadin	O S	V Adjunct

It is really interesting that OSV order cannot go with any of the above mention sentences and make them weird. Even it does not work with different positions of adjunct. Therefore, it is possible to understand that

OSV cannot be the default word order in Bengali. Now I try SVO order for the same context:

(4c) Speaker:	?Sumon sharadin poRe boi	S Adjunct V OFOC
Or,	?sharadin Sumon pore boi	Adjunct S V OFOC
Or,	?Sumon poRe boi sharadin	S V OFOC Adjunct

Unlike OSV, the SVO order makes all three sentences object-focused. Therefore, I have got focused objects with this word order which is not required in this context. Moreover, the speaker is asked to tell about *Sumon* not about the *boi* 'book'. So, SVO cannot even produce expected information structure for this context. Instead of using SVO, another word order sequence like OVS can also be examined. In such a case, OVS order marks all three sentences subject-focused which are not also expected. According to these examples, I can say that other word orders except SOV cannot go with all contexts.

As a result, it is possible to assume that SOV is the default or the most general word order for Bengali. The native speakers of this language can use SOV order in different situations. It is generally assumed that a speaker structures or packages the information in such a way that there is an optimal exchange of information. Since my previously mentioned examples show that Bengali word order scrambling is used as a tool to avail the optimal exchange of information; it creates a possibility to think that scrambled word orders are motivated by explicit and specific constraints on information packaging.

1.4 Bengali Focus Constructions

'Focus' is commonly taken as an information highlighted element in a proposition. It is well known that languages may manifest one or more from the following range of morphosyntactic or prosodic options as reflexes of focus:

a.	Focus in-situ	Bengali, English, Hungarian...
b.	Focus ex-situ	
	•Clause- initial	Bengali, English...
	• Pre-verbal	Bengali, Hungarian...
	• Post-verbal	Bengali...
c.	Focus markers	Bengali...
d.	Focal stress	Bengali, English...

From this brief typology, it can be seen that languages frequently select more than one option from this set. For example, both English and Hungarian display in-situ and ex-situ focus. A language that displays focus fronting or clefting is also likely to mark the displaced constituent with main sentential stress. Some languages, as I shall see with Bengali, display both displacement and the presence of focus-marking morphemes (by using clitics such as: /o/, /i/ which I shall discuss in the next section) within the same construction.

A theoretical issue that arises again here: given the minimalist view of

language as a perfectly economical system, I do not expect more than one means of achieving one interpretive goal. If a language has more than one focusing strategy, can each of these be established as corresponding to a distinct interpretive goal, or are interpretive 'decisions' forced by pragmatic factors? It is clearly considerations of this nature that motivate approaches such as that of Kiss (1998), and which also, in part, motivate the present study. If I am able to set up that, in a given language, each divergent focusing tactic results in a distinct reading, and then I can work towards supporting the view of language as an entirely economical system in which redundancy and optionality do not subsist. If, in contrast, the empirical facts turn out to be inconsistent with this hypothesis, I am then forced to accept that optionality exists in the syntax, as well as at the interpretive level.

The following data which are posed after this short argument have been given in question-answer pairs; this discourse context allows me to explain the difference between semantic focus types. From these data, it comes into view that focus fronting is suitable as either new information or exhaustive listing focus, and that the type of focus is resolved by discourse context. The presence/absence of focus marker does not change the type of focus, only its 'impact'; native speakers express it as 'adding emphases'. Note that, even though the examples given here are limited to object focus, Bengali generously permits the focus fronting of any constituent, including VP.

The following examples also show that there is evidence for focus in-situ in Bengali. As the term 'in-situ' suggests, in these constructions, the focused constituent appears in its base position, INFL appears in the neutral form, and the focus is indicated by main sentential stress. The contexts given indicate that, as with ex-situ focus in Bengali, in-situ focus is appropriate as either new information or exhaustive listing focus.

Identificational focus works as a quantifier and involves an operator that expresses exhaustive identification (Stoyanova, 2008). As it behaves like an operator, it can move into a scope position in the specifier of a functional projection and is able to bind a variable (Kiss, 1998). This type of focus represents the subset of the set of contextually given elements and can bear a [±contrastive] feature. Identificational focus interprets as [+contrastive] while it operates on a closed set of entities. According to Rooth (1985), evoking alternatives is the primary function of focus, and the 'contrast set' evoked by the focus provides the locus for focus sensitive operators such as only, even, and also. Other researchers (e.g. (Horn, 1981), (Vallduví, 1992)) take information status to be primary, and treat contrast as secondary and derivative. Some type of linguistic prominence codes information focus and contrastive focus across languages, a fact that no doubt has contributed to a blurring of the distinction between these two categories.

Moreover, as a precondition of the contrastive version of identificational focus; members of its entities should be acquainted with the participants of the discourse. Besides this semantic guideline, contrastive focus maintains close connection with syntactic elements, especially wh-questions, and requires licensing positions in different languages. The analysis proposed in Rizzi (1997) suggests a fixed component, involving the heads specifying force and finiteness, and an accessory component involving the heads of topic and focus, which are activated when needed (i.e. when there is a topic or focus constituent to be accommodated in the left periphery of the clause). Therefore, I would like to say that in Bengali, focus can be activated for movement according to the necessity. However, it is not yet decided whether Bengali has a designated place for its focused elements. In the following set of examples, I get a simple wh-question (5a) at the beginning. A straightforward answer to (5a) is the normal word order sentence (5b). Since I get new information from this answer, I identify this as information focus. Now, I suppose that the answer which has been given is wrong and someone is asked to correct it. In this situation, I shall be able to get identificational or contrastive focus. In this way, (6a) contrasts with (5b) and identifies the correct proposition with different ordering. (6b) also does the same thing, however in that case the sentential order is different.

(5) a. Sumon ki rakhlo? SOV

Sumon what put.PST.3

'What did Sumon put?'

b. Sumon boi-ti rakhlo SOV (new information)

Sumon book the put.PST.3

'Sumon put the book'

(6) a. na, ko'lomti Sumon rakhlo OSV (exhaustiveness with contrast)

no, pen the Sumon put.PST.3

'No, Sumon put the pen'

b. na, Sumon rakhlo ko'lomti SVO (exhaustiveness and contras)

no, Sumon put.PST.3 pen the

It has been seen that contrastive focus also shows word order variation in Bengali. According to (Kiss, 1998) English contrastive focus is preposed into Spec-FP, whereas syntactic study on contrastive focus interpretation of Bengali is not very clear in the established literature. A final question which arises in relation to focus in-situ is the following: since Bengali permits both ex-situ and in-situ focus, is it possible to find both co-occurring in a single construction? Only relevant empirical evidences will be able to find any solution to this query.

1.5 Cliticization in Bengali: A Focus Tool

/o/ and /i/ are commonly used emphatic clitics in Bengali which also mark focus in this language. Native speakers of Bengali use them as *too* and [+ emphatic] respectively. Initially, /o/ and /i/ tend to adjoin as enclitics to an element of type X^{o} which is then the focus of the clitic

(Bayer, 2011, p. 3). The following examples will show the common patterns of these clitics:

(7a) Sumon- o poR- be
Sumon- too read- FUT3
'Also SUMON will read'

b) Sumon poR- be- o
Sumon read- FUT3- too
'Sumon will also read'

In (7a) /o/ adjoins with a nominal constituent and in (7b) shows that an inflected verb can also bear such a clitic. The following example shows that it can also attach overtly to an inflected noun.

(8) Sumon Shobuj- ke- o chineche
Sumon Shobuj- [Acc] -too know-PRF3
'Sumon has known also Shobuj'

So far, it has been illustrated that /o/ follows the inflected verb, and inflected nominal constituent. However, it will not be able to be placed between the stem and the ending. The example given bellow is showing this contrast and the unacceptability:

(9a) chin- i- o
know- [1]- too
'(I/I) also KNOW'

b) *chin-o-i

(10a) chin- chh- i- o

know- [prog] [1]- too

'(I/We) am/are also KNOWING'

b) *chin-o-chh-i

Now consider instances where the clitic can be added before the inflectional ending is attached.

(11a) Sumon Shobuj-ke chine- che-o

Sumon Shobuj[obj] know- PRF.3-too

'Shumon has also known Shobuj'

b) Sumon Shobujke chine-o-che

c) chin- e- ch- i- l- am- o

know- [PRT]- [PROG]-[link]- [PST]- 1]- too

'(I) have also KNOWN'

d) chin-e-o-ch-i-l-am

Even though the clitic can be appended between the stem and the inflection (see example: 11b, 11d), it cannot be inserted between affixes. In Bengali, the clitic insertion process selects binary options– the clitic either comes right after the stem, or it must come after all the affixes are added (Bayer, 2011, p. 5). This also rules out the following example as well.

(12a)*chin-e-ch-i-o-lam

b) *chin-e-ch-i-l-o-am

In contrast, nominal constituents in Bengali do not allow clitics to be inserted between the stem and the inflection. Therefore, example-13 is ungrammatical. /o/ appears here between a noun stem and the case-marker as well as the inflectional ending *-ke*.

(13) *Sumon Shobuj-o-ke chinche

'Sumon has known also Shobuj'

This rule is also applicable for Bengali verbal nouns a compound like element commonly available in this language. A verbal noun is derived by affixing the suffix -a to a V-stem. Such as a Bengali verb stem: gao 'sing' adjoins 'a' and derive gaoa 'singing'. A verbal noun can also add one or more N^{o}-object with it (e.g., gan gao-a 'song singing'). As (14a) below illustrates, /o/ can attach to the verbal noun and select its focus inside. On the other hand, (14b) reveals that clitics fail to adjoin to the focused N incorporated. For these examples, imagine a context in which someone states that (s)he liked somebody's listening of music very much.

(14a) tader [gan gao- a]- o bhalo laglo

their song sing-ing- too pleased-PRF.1

'(I) was pleased by also their singing of SONG'

b) *tader [gan-o gao-a] bhalo laglo

It is possible to assume that Bengali clitics enforce quantificational properties on their morphological/syntactic domain, comparable to *only* and *even* in English. As Rooth (1985) argues, a phrase narrowly focused

by *only*, *even*, etc., must be inferred relating to a quantificational domain. Let me assume that (15a) bellow is an S-structure; (15b) is the logical form derived from internal language, and (15c) is a rough semantic representation which transduces the logical form into a proposition with a universal quantifier having scope over it.

(15a) I saw only Sumon.

b) [only Sumon]$_i$ [I saw x$_i$]

c) *For all x* [I saw x→x = Sumon]

'Only Sumon', like 'Sumon-i' is a quantifier which must be assigned scope over the clause (proposition) at the level of logical form. Scope assignment, however, is constrained in language-specific ways. For instance, *even* in English cannot appear in an unconstrained fashion, although there is no prima-facie semantic reason which could prevent this.

(16a) *They have killed [$_{NP}$* even [$_{NP}$ my pet]]

b) **They have killed [$_{NP}$* my [$_{NP}$ even [$_{N}$pet]]

In English, *only*, *even*, etc. and their respective correspondents exhibit different island effects. According to Koster (1986), with the exception of VP, all maximal projections XP of lexical categories are virtual bounding nodes. A bounding domain can, however, be extended when XP is governed by an element which conforms to the basic orientation of government in the language. The direction that counts as basic in a VO-language such as English is rightward while it is leftward in OV-

languages such as Bengali. The following example in English will help me to realize the rightward system.

(17a) Sumon would [even [talk to Shobuj]] (adjoined to VP)

b) Sumon would talk [to [even [SOBUJ]]]

→ →

Adopting the rule of quantifier raising (QR) as suggested in May (1977), English allows for even *Sobuj* to undergo QR because P governs in the same direction as V: the PP ceases to be a bounding node, and (17b) is well-formed. The derivation of well-formed logical forms seems to be constrained by the following principles (Bayer, 2011, p. 10).

(18a) Focus-sensitive quantifiers (*only, even* etc) must have access to a domain of quantification.

b) Raising to S (or at least to predicate-level) provides a domain for quantification.

c) Governed quantifiers must be (canonically) governed in a dynasty (g_i,....,g_a) of uniformly oriented governors, up to a tree height where a quantification domain is found.

Bengali does not have a system as English does. It has exclusively post-positions. Since it is an OV-language, PPs should not lead to island effects.

(19a) Sumon Shobuj- er- o Songe alap korbe

Sumon Shobuj- [GEN]-too with discuss do [FUT]

'Sumon will discuss with also Shobuj'

b) Sumon- er- i biShoye alocona hochilo
Sumon- [GEN]-[EMPH]about dicussion PST.PROG-take-place
'Discussion took place about SUMON'

There are other postpositions, however, which do not allow simillar construction. Besides this, Bengali adjectives take NP-comlements to the left. As expected, island effects are absent here, too.

(20) Sumon- er- o bhokto
Sumon- [GEN]- too fan
'Fan of Sumon, too'

Quantified NPs which are adjoined to X^{max} can freely undergo quantifier rising because they are ungoverned. The X^{max} to which they belong does not count as a bounding node. The Bengali possessor-NP appears to be adjoined i.c., not in [Spec, NP]-position as in the case in English. The following example describes this fact:

(21a) $[_{NP1}[_{NP2}$ amar chacha- r] $[_{NP1}$ oi kalom]]
my uncle [GEN] that pen
'That pen of my uncle'

b) $[_{NP1}[_{NP2}$ my uncle's] $[_{N1}$ pen]]

c) *my uncle's that pen

Both the grammar of syntactic movement and the grammar of scope assignment behave accordingly, as described in the next.

(22) NP-split in the syntax

a) Tumi [kon chele ta- r ko'lom] dekhecho?

You which boy- [DET]-[GEN] pen seen-PERF.2

'which boy's pen did you see?'

b) [*kon- chele- ta- r*] tumi [*kalom*] dekhecho?

This is not possible in English. It would violate what has become known as the 'left branch condition'.

(23a) [*whose pen*] did you see?

b) * [*whose*] did you see [*pen*]?

Exactly the same constraint seems to be at work in logical form, as indicated by the grammaticality difference between (24) and (25).

(24) NP-split in logical form

[amar CHACHA r- o] jomi bikri hoe gache

[my uncle (GEN)-too] land sold become gone.PRF

'Also my UNCLE's land was sold'

(25a) *[my UNCLE even's] land was sold

b) *[my uncle TOO's] *land was sold*

It is possible to assume that N does not lexically govern a structural position such as [Spec, NP]. Despite the canonical direction of government that holds in a language, this guarantees that NP is an island for the possessor-NP in (21a).

I can now make the following generalization. Only those quantified elements X can undergo quantifier raising which are either adjoined to some YP or linked in a dynasty. A dynasty is built by a chain of successive governors when they govern in the same canonical direction.[$_x$ X particle] or [$_x$ particle X] is c(anonically)-governed, if the minimal maximal category dominating X contains a governor G which precedes X in a right-barnching language and follows X in a left-branching language. In the following examples, these requirements are not met:

(26a) **ami* [$_{NP}$ nil- o bari] dekhechilam

I blue- too house see-PST

'I saw a house which is also BULE'

b) *They have killed [$_{NP}$ my [$_N$ pet]]]

c) *Sumon [$_{NP}$ Shobuj- o- ke] chinche]

Sumon Shobuj- too- [Acc] know-PRF

'Sumon has known also SHOBUJ'

d) *tader [$_{Np}$[$_N$ [$_{vP}$[$_N$gan- o] [v goa]]- a]]] bhalo laglo

their song- too sing- ing pleased-PST

(26a-c) are ruled out because the NP is the minimal maximal category dominating the quantified X, and X fails to be c-governed in NP. The category dominating the quantified *nil-o* in (26a) is the NP *nil-o bari*, analogously, the relevant category in (26b) and (26c) is the NP my even *pet* and *Shobuj-o-ke* respectively. Notice that /o/ in (26c) c-commands

only the uniflected noun. After the attachment of the case-suffix -ke, which I do not consider to be a governor, the quantified element can only be the bare noun. The dominating minimal maximal category is the NP across which the quantificational domain cannot be extended. Under the assumption that *gan-o gao-a* in (26d) (=14b) is a VP which undergoes nominalization due to the affixation of -a, po'r c-governs gan-o, but the dominating and intervening NP blocks the formation of a dynasty with the verb (bhalo) laglo. Obviously, the nominalized VP itself does not count as an appropriate quantification domain.

1.6 Main Question

During the discussion, I have raised a couple of questions regarding Bengali focus constructions which need empirical support to be answered. So far, it has been observed that Bengali language permits different grammatical word orders. Syntactic interpretation of Bengali focus constructions and its correlation with information packaging still needs empirical support to come to a conclusion. Before doing this, I can devise the main question for this research with the help of our observation and precise properties of Bengali syntax. Therefore, it is important to figure out the main research question of this study:

How do the information structures coincide with focus constructions in the sentence patterns used by Bengali native speakers?

However, there is several specific research questions are allied with this main query. These allied questions are:

1. What are the basic characteristics of Bengali native speakers' focus preference?
2. How do the Bengali native speakers convey the information within the scope of focus structures?

I shall search answers to these questions by doing experiment with Bengali native adults. By now, it has been studied that there are intricacies entailed in settling on the exact status of Bengali focus constructions and their relevant information structure. I presume that the syntax–information structure interface is a vital step in the constructing of an acceptable focus structure. Precisely explaining this interface necessitates organized and empirical data on what kinds of sentence structures are suitable in what types of discourse contexts to serve as a guide. Therefore, I have designed quantitative and qualitative tasks to collect such data for Bengali, which will then supply the part of the empirical base for my analysis of Bengali focus constructions.

1.7 Hypothesis

Since my research question has already been devised, it motivates me to formulate the null hypothesis for our present research experiment. The null hypothesis is designated as H_0A:

H_0A: Bengali focus constructions are purely stylistic and different focus word orders carry same information packaging.

If I successfully reject this hypothesis with the help of the result of my experiments, then I need to test an alternative hypothesis which is designated as:

H_1B: Bengali focus construction is sensitive to information structure and the information structure motivates the syntactic position of focus particles.

1.8 Conclusion

The answers to these questions will be figured out by doing experiment with Bengali native adults. In the next chapter, I am going to discuss the theoretical issues and review relevant literatures regarding focus constructions and information structure.

Chapter Two

The Relationship between Syntax and Information-structure

I have already assumed in the previous chapter that the word order scrambling in Bengali is typically context-dependent and an explicit communicative context motivates to determine the word order of a Bengali sentence. Moreover this analysis based assumption will also increase the chance to nullify H_0A, the hypothesis which has been taken into account in the previous chapter. Therefore, the aim of this chapter is to discuss about focus word orders and their accompanying contexts. In section 2.1, I introduce some of the basic concepts and dichotomies used to describe information structure and focus. In section 2.2, I discuss some of the definitions of topic/ theme that have been proposed in the literature, and present the definition that I adopt in this dissertation. In section 2.3, I examine the accompanying definitions of focus/rheme in the literature, and discuss the assumptions that I make for this concept in this dissertation. In section 2.4, I briefly discuss the way in which the syntax-discourse interface has been described from a syntactic perspective, and some of the theoretical problems involved in such analyses. In this section I also examine syntactic accounts of the syntax-information structure interface. In section 2.5, I examine relevant literatures for the syntax-information structure interface. In section 2.6, I conclude the chapter by discussing the implications of discourse factors in speaker judgments, as well as the research questions guiding the methodology that I present in Chapter 3.

2.1 Information Structure

It is possible to divide sentences into informational units. An informational split is different from that of subject-predicate. This split as starts with the base which is used by the speaker and the hearer to communicate together. Vallduví & Engdahl (1996, p. 460) explain information packaging as the "structuring of sentences by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a particular context or discourse." A well established fact in the literature on information structure is that sentences bear a less informative part (Topic or Theme), and a more informative part (Focus or Rheme) as well. Different dichotomies have been proposed to comprise this split in the information structure: Theme-Rheme, Topic-Comment, Topic-Focus, Focus-Presupposition, Focus-Open Proposition, and Focus-Background (Hockett, 1958); (Erteschik-Shir, 2007); (Prince, 1986); (Rochemont, 1986); (Ward, 1988); (Vallduví E. , 1990), among others). Vallduvi (1990) studies that, even though all these proposals accept the basic idea of a split, they differ in where the split occurs. In literature, a large number of proposals that take apart the Topic/Theme from the rest of the sentence diverge on how to identify the topic. The same frequently results from definitions of focus or rheme. However, despite such disparity in explanation, the definitions of these terms are commonly taken for granted.

In the previous chapter, I have already experienced how focus is one of the most delicate issues in the study of grammar. I may assume that the

location of the focused constituent within a sentence is predictable on the basis of the syntax of a given language. Moreover, the definition of all constituents that might be interpreted as focused might be understood through language-specific skills. The major (though not only) division across languages as to the distribution of focus is related to the degree of freedom that languages allow for its location. Focus is represented in different ways in different languages. However, focus itself has different classifications according to its properties and appearance (Azad, 1994). Two types of focus that are generally defined as new information focus and identificational focus. This second type of focus which is also known as contrastive focus expresses exhaustive identification and occupies the specifier of a functional projection (Kiss, 1998). On the contrary, as I saw before, the new information focus conveys new information and involves no syntactic recording. Consider the following examples:

(1) Identificational Focus: It was to Sumon that they lent the book. (Exhaustiveness)

Information Focus: They lent the book to Sumon

From this example, I can again see that information focus is based on contextual new information for a certain linguistic expression. The following table (2.1) will help me to get the idea about the interactions among the contrastive and non-contrastive topic and focus:

	Topic	Focus
Non-contrastive	aboutness topic [topic]	new information focus [focus]
Contrastive	contrastive topic [topic, contrast]	contrastive focus [focus, contrast]

Table1: The Interactive Relation Between Topic and Focus

The table expresses that topic and focus are basic notions in information structure that can be enriched to yield a contrastive interpretation. In other words, a contrastive topic and a contrastive focus are an aboutness topic and a new information focus, interpreted contrastively. The following sections will discuss these issues in detail.

2.2 Topic/Theme

Mathesius (1975 [1961], p. 30) describes the theme as "the element about which something is stated". In the literature theme, or topic, generally share schemes of aboutness, discourse-oldness, shared knowledge, or discourse salience. Within the framework of communicative dynamism (CD), Firbas (1964, p. 272) suggests that the theme holds the lowest level of CD. According to him, the theme can convey not only known information but also it can transmit even new, unknown information. On the contrary, Contreras (1976, p. 16) believes

that the elements which develop the notion of theme are realized by the speaker to be present in the addressee's consciousness. By introducing the dichotomy Topic-Comment, Hockett (1958, p. 201) illustrates the topic as "what the speaker is going to talk about". Halliday (1967, p. 212) extends this idea by describing the theme as "what I am talking about now". According to Gundel's (1988 [1974]) proposal, topic exemplifies the subject-matter of the sentence but indicates that it is correlated with given (non-focal) information in the sentence, and never has primary stress. Sgall et al (1986, p. 80) portray the topic as "the items the speaker supposes to be activated in the hearer's memory at a given point of time". Erteschik-Shir (2007), among others, depicts topics as old or presupposed information; however, Reinhart (1995) does not support the classification of topics as old information. She exemplifies discourse as a shared procedure of constructing a "context set" consisting of subsets of propositions. Sentence topics are one subset of this context set which are able to classify referential entries. On the other hand, Dahl (1974) proposes a tripartite structure which involves the topic-comment and focus-background. By improving this notion, Vallduvi (1990) finds an overlap in Dahl's (1974, p. ex. 3) two dichotomies as in (2).

(2)

	topic	comment	
What does Sumon eat?	-Sumon	eats	cake.
background		focus	

In (2), *Sumon* is both topic and focus, and *eats* forms part of both the comment and background. Vallduvi suggests a different sort of trichotomy as a solution of this overlapping issue. He identifies it as a redundancy. According to Vallduvi, topics are stated as *links*. Topics only appear in the initial position of a sentence and they are used to turn on the hearer's 'knowledge store'. This proposal illustrates that the sentence is divide primarily into Focus-Ground. This splitting process continues and the Ground is divided again into Link-Tail. In domain of a sentence, Vallduvi finds focus as 'the only non-elidable part'. He also identifies the Tail is non-topic and non-focus. By this analysis then, *Sumon eats* would form the link, and *cake* would be the focus.

Since pronominal elements may be topical in a sentence, Lambrecht (1994) argues against the proposal of Vallduvi by restricting topics to discourse referents. According to him, such referents are what a proposition is about. Others have described topics in different terms such as presupposition, background, or open proposition. Jackendoff (1972) illustrates presupposition as information assumed by the speaker to be shared with the interlocutor. Dahl (1974) as well Vallduvi's (1990, p. 58) identifies the background as the non-focused part of the sentence. Therefore, Vallduvi elaborate this finding by categorizing it as the complement of the focus. Besides all these explanations, it is also noticeable that different languages mark topics differently; such as: phonologically, morphologically, lexically, and syntactically. Since a

number of ways are followed to mark topic cross-linguistically, Casielles (2004) suggests that different definitions of topic are needed for illustrating and identifying the characteristics of topic elements in human languages. According to her, the nature of topical elements of a language should be realized by considering the specific characteristics of that particular language.

The basic definition of *topic* that I assume is based on the notions of "aboutness" or "discourse old" discussed above. Topical, discourse-old elements tend to appear in preverbal positions. As most accounts of topic occur in a dichotomy with terms like *comment* or *focus*, they inevitably involve an accompanying definition of the term *focus* as well. In the following section, I present some of these characterizations of focus in the literature.

2.3 Focus/Rheme

According to Casielles (2004, p. 127), focus or rheme elements are "more informative and less topical", they are also identified as prosodically prominent. In general, they are involved to represent new information within the discourse, and also tend to occur toward the rightmost edge of a sentence. However, such a simple definition fails to accommodate the exceptions and complications of them. Rochemont (1986) differs with the idea that focus associates with new information based on the existence of focused pronouns (3), which he treats as old information.

(3) Who did they invite?

Sobuj said they invited HER.

In this example, the pronoun *her* must refer to known, old information. Therefore, since *her* may be focused, Rochemont claims that the correlation between focus and new information does not established. Casielles (2004) suggests that albeit a pronoun may refer to a prominent, discourse-old entity, it does not exclude it from being new information and the focus of a sentence. In (3), the informative component is the direct object. In the discourse, it is active enough to warrant being expressed by the pronoun *her*. Rochemont's preliminary definition of Focus (new information) is based on the idea of c-construability. A c-construable element has a semantic antecedent in the discourse δ. It is to be mentioned here that focus elements are not c-construable. There is a difficulty lies in this analysis regarding the focused pronoun, as in (4).

(4) Shumon helped Maya, then SHE helped HIM.

As (4) recommends that a discourse element be both focused and c-construable, Rochemont suggests discarding a definition of focus based on new information, and instead recommends two types of focus: Presentational focus and Contrastive focus. Presentational focus is illustrated as non-rightmost, non-contrastive stress, and is used with verbs like *appear*, which in the unmarked variety, have an accented subject (5a).

(5) a. The case was judged. Then a LAWYER appeared.

b. The case was judged. Then a lawyer APPEALED.

By his analysis, verbs like *appear* differ from others like *appeal* in that they seem to shift their status as the focus of new information to their subject. However, Casielles (2004) points that verbs like *appear* coincide with the set of unaccusative verbs, which need not necessarily influence focus projection. Chomsky (1971) advised that focus may project in sentences like 6(a) in which any bracketed elements may be part of the focus in English.

6 (a) He was (warned (to look out for (an ex-convict (with (a red SHIRT))))).

However, Casielles studies that this is only possible with rightmost focus in English. Focus that does not appear in a rightmost position in a sentence 6 (b) cannot project, and must remain narrow.

6 (b) Sumon followed Shobuj into the bedroom.

For Rochemont, non-rightmost accented elements are not contrastive focus elements. The notions of marked and unmarked accent are crucial for focus projection (see e.g. Cinque (1990), Reinhart (1995), Nash (1995), Zubizarreta (1998)). Unmarked accent is generated by the grammar, falls on the rightmost constituent, and identifies the unmarked focus of the sentence. Only this type of focus may project. For Rochemont, Contrastive focus is defined by a rather complex calculus (7).

(7) An expression P is a Contrastive Focus in a discourse δ, δ = {φ_1, ... φ_n}, if, and only if,

(i) P is an expression in φ_i,

and

(ii) if P/φ_i is the result of extracting P from φ_i, then P/φ_i is c-construable, and φ_i is

not c-construable.

Casielles takes issue with Rochemont's definition of c-construability because it would treat certain reflexive pronouns as contrastive (8).

(8) Who did Sumon hit?

He hit HIMSELF.

While (8) does not show to be contrastive – at least requireing further information on the discourse perspective – the fact that *himself* has a semantic antecedent would make it c-construable, and therefore contrastive. For Casielles, an added difficulty to this investigation is that certain expressions may be both presentational and contrastive focus. Following the definition in (8), both (9) and (10) should be contrastive since, in Casielles's view, the non-focused part is c-construable in both examples.

(9) A: Sumon's physical condition is a source of constant concern to Shobuj.

B: Sumon's physical condition is a source of constant concern to SHOBUJ.

(10) Shajib scolded Shobuj, and then he SMACKED him.

Nevertheless, the only center of attention that is c-construable by Rochemont's study is the one in (10) because it has an antecedent. The verb *SMACKED* in (10), however, qualifies as Contrastive Focus in that the non-focused segment of the expression is c-construable and the entire sentence in (10) is not. It becomes licensed as presentational focus in that the focus is not c-construable. While Rochemont locates this overlap in focus types a enviable result of his calculus, Casielles discovers this challenging, not only because only-contrastive, only-presentational, and contrastive and presentational are not undoubtedly defined, but also because the supplementary idea of direct and indirect c-construability are introduced as significant, but only for non-focused components (Rochemont (1986, p. 103). Finally, Casielles eliminates the concept of a recommended division of focus which allows for most presentational foci to be contrastive foci at the same time.

Gundel (1988 [1974]) talks about three means of focus by describing their types: psychological, semantic, and contrastive focus. Psychological focus refers to the center or focus of attention (AI focus in Hajičová (1987)), which would be topical by many of the analyses of topic. Casielles (2004) refers to psychological focus as the current center of attention in a discourse, which is more akin to topical elements. Semantic focus refers to new information being stated, or "the part of

the sentence that answers the relevant wh- question (implicit or explicit) in the particular context in which the sentence is used" (Gundel (1994, p. 461). This semantic focus can be marked by pitch accent, word order (including special focus positions), focus-marking particles, or any combination of these. Semantic focus includes context-active, discourse-old elements such as the pronoun *SHE* in (11).

(11) Sumona said it was SHE (=Sumona) who called.

By this study then, semantic focus may fall on a formerly mentioned component without changing its rank as focus. Therefore, as said by this explanation, not all semantically focused material need be entirely new to a discourse. This is the idea of focus that Casielles (2004) approves. Gundel's contrastive focus (CF) differs from that of Rochemont in that her CF refers to a strategy (phonological or syntactic) for making an element prominent in order to focus an interlocutor's attention on said element. Due to the fact that Gundel's CF falls primarily on topics, and due to the potential confusion that can result from such a definition, Casielles prefers to call Gundel's CF "emphatic stress".

As argued earlier, Vallduví (1990) splits the sentence into Focus and Ground. Ground stands for the unfocused segment of the sentence, and is further segregated into link and tail. For him, focus is the informative part of the sentence, and is the only part of the sentence that may not be elided. I will not go into the particulars of Vallduví's information packaging calculus, but in his system, Focus comes in two varieties:

Retrieve-add focus and Retrieve-substitute focus. When a sentence lacks a tail in its information structure, the relevant information is retrieved by adding focus (thus retrieve-add), and when it has a tail, information is retrieved by substituting focus in the relevant position within the structure. Structures with tails (retrieve-substitute) correspond with narrow focus, and structures lacking them (retrieve-add) correspond to wide focus. Casielles (2004) has a problem with Vallduví's treatment of structures with tails, particularly when focus is either retrieve-substitute focus, or narrow focus. She claims all are Focus-Background structures (in her terms), or Focus-Ground, viewing the distinction between link and tail as irrelevant in these cases. She argues that when one has an instance of narrow focus, Vallduví's distinction does not take into account that the rest of a sentence is necessarily part of the background (Vallduví's ground).

Kiss (1998) bases her analysis on Hungarian, suggesting two types of focus: identificational focus and information focus. According to her analysis, identificational focus bears syntactic and semantic properties lacking in information focus sentence. The following are the basics of her proposal, as listed in Casielles (2004).

(12) Information Focus

a. merely marks the non-presupposed nature of the information

b. allows for any type of phrase

c. does not take any scope

d. does not involve any movement

e. can be either smaller or larger (i.e. it can project)

(13) Identificational Focus

a. expresses exhaustive information

b. does not allow for all kinds of phrases (excluding universal quantifiers, also-phrases, and even-phrases)

c. takes scope

d. moves to the specifier of a functional projection

e. is always coextensive with an XP available for operator movement (does not project), although it can be iterated

Casielles (2004) points a resemblance in this distribution with Rochemont's Presentational vs. Contrastive Focus, and Vallduví's Retrieve-add vs. Retrieve-substitute system, both of which (ignoring some crucial differences) effectively draw a line between narrow and wide focus. By applying this in Bengali I can find that it allows universal quantifiers (14) and *even*-phrases (15) in identificational focus contexts.

(14) SHOB KHELNA nite cheyechilo meye-ti

All toys take.INF.3 want.PST.3 girl the

The girl wanted to take ALL THE TOYS.

(15) AEMONKI AEKTA KHELNA-O she nite cheyechilo

Even a toy -too she take.INF.3 want.PST.3

She wanted EVEN A TOY.

Kiss's categorizations of focus are based on Hungarian, which bears significant distinctions from other languages. Even though (14) and (15) bear similarities to identificational focus in Bengali, they do not express the same propositions: for example, *aemonki aekta khelnao* in (15) is hardly exhaustive.

For Lambrecht (1994), topic and focus do not form a dichotomy; rather, they are separate in relations. For Lambrecht, topic has to do with the aboutness of a proposition, while focus has to do with the conveying of new information (his pragmatic assertion). All declarative sentences transmit information: therefore, all declaratives have a focus, but not all have topics. Focus is information that is added to, not superimposed upon, a pragmatic presupposition. "The focus is, therefore, the element of information whereby the presupposition and the assertion differ from each other... It is the unpredictable element in the utterance" (op. cit.:158-159).

The types of focus functions in his analysis are Predicate-Focus (16), Argument- Focus (17), and Sentence-Focus (18). These focus types correspond to the sentence types Topic-comment, Identificational, and Event-reporting, respectively.

(16) (What did the children do next?) Predicate-focus/Topic-comment

The children went to SCHOOL.

(17) (Who went to school?) Argument-focus/Identificational

The CHILDREN went to school.

(18) (What happened?) Sentence focus/Event-reporting

My CAR broke down.

In predicate-focus sentences, the predicate forms the focus. In Argument-focus sentences, the focus is the missing argument. In Sentence-focus sentences, the focus includes the subject and the predicate. Casielles (2004) differs with Lambrecht's depiction of Sentence-focus, preferring to integrate this sentence type with his Predicate-focus type, as they are both wide-focus in nature. She further opposes Lambrecht's proposal by saying that Sentence-Focus lacks a topic, in line with Erteschik-Shir's (1997) "here and now" stage-topic (discussed further below).

Reinhart (2006) recommends that focus is coded in the phonological form (PF) of a sentence. She puts forward that the identification of a focus unit may be resolved for each derivation via a set of *possible pragmatic assertions (PPA)*. In the case of foci, at the point where syntax and stress are visible, at the interface between syntax and pragmatics a reference set of possible foci are generated. The discourse then selects the member appropriate to the given context. Reinhart formalizes this proposal (19) for a stressed object in English (bold face indicates a stressed constituent).

(19) *Focus Set*

The focus set of a derivation D includes all and only the constituents that contain the main stress of D.

(20) a. [IP Subject [VP V **Object**]]

b. [IP Subject [VP **Object** V]]

c. *Focus set:* {IP, VP, Object}

While in theory any of the members of the set in (20c) may be chosen as the focus of the utterance, at the interface only one may be chosen. At that point, the discourse conditions will choose which set(s) are appropriate.

Given the difficulties in arriving at a consensus on the meaning of focus, I follow López (2009) in defining regular focus as in Jackendoff (1972): focus determines a variable left open in the previous discourse. López provides the following example (p. 28, ex. 31).

(21) - What did John bring? [x | John brought x]

- John brought the wine. [x=the wine, 'the wine' is focus]

The preliminary discourse in (21) leaves open the variable x. As a result, the part of the question that resolves this variable (=*the wine*) is the focus/rheme of the sentence reply. In the following chapters, I frequently make use of the term *narrow focus*, which I use when either the subject or object is the unresolved variable.

In many Romance languages a focused element can be moved to the front of a clause or sentence, as in the Catalan example ((24), López's example (31)).

(22) [Context: You gave him the spoons.]

- ELS GANIVETS li vaig donar.
the knives CL.Dat.3 PST.1 give

THE KNIVES I gave him.

In (22), the context does not leave a variable open to be resolved. Focus fronting (FF) then creates this variable (λx you gave him/her x), which in turn opens up the set {x | x=things I may give him/her}. At the same time, FF provides this value for x (=the knives), thus creating a contrast with the preceding context. The interpretive import of focus fronting (FF) then is contrastive, and may not answer a wh- question, explicit or implicit. Crucially, this definition of contrast departs from other definitions of contrast discussed earlier on this chapter.

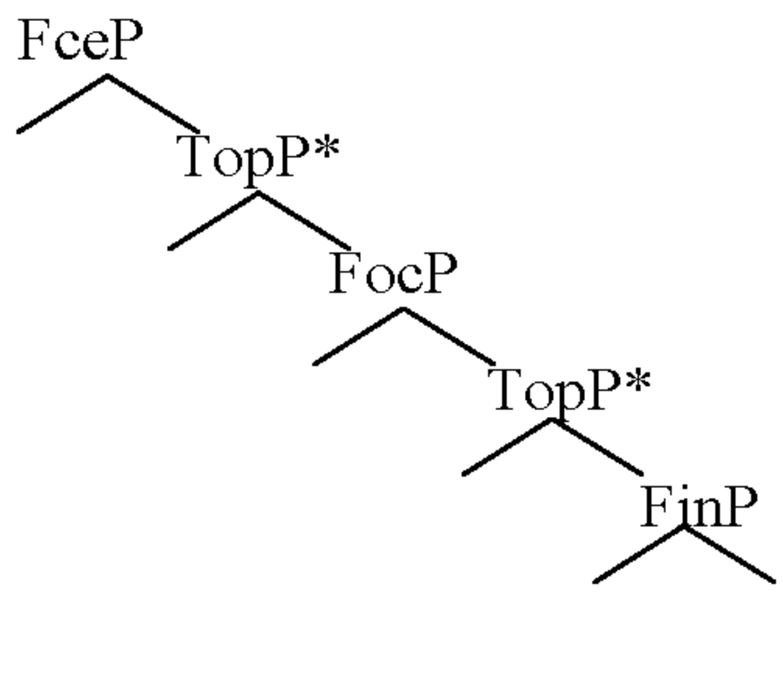
In this section, I have provided some examples of the various notions of focus that have been presented in the literature. Although this review has been far from exhaustive, I have discussed some of the important issues that must be taken into consideration. The notion of focus that I adopt in this dissertation follows the very basic notion assumed by López: regular focus provides resolution of a variable left open in the preceding discourse. The constituent providing such resolution is also referred to as the *rheme* (especially in López (2009)). When I use the

terms *subject narrow-focus* or *object narrow- focus*, I am referring to a particular variable that is resolved in the discourse. Focus fronting should be distinguished from regular focus in that it simultaneously opens a variable and closes it.

2.4 Syntactic accounts of the syntax-information structure interface

A number of researchers have described the syntax-information structure relation in purely syntactic terms. Rizzi (1997) proposed the expanded CP field as an interface layer between the propositional content (IP/TP) and the superordinate structure. For embedded clauses, the superordinate structure is the higher clause; for matrix (root) clauses, this is the discourse. This proposal incorporates Topic and Focus both as features and as labels of heads and projections in the narrow syntax. Note that in Rizzi (2004), focus and topic are referred to as criterial features. For Rizzi, the left-periphery has the following structure (23):

(23)

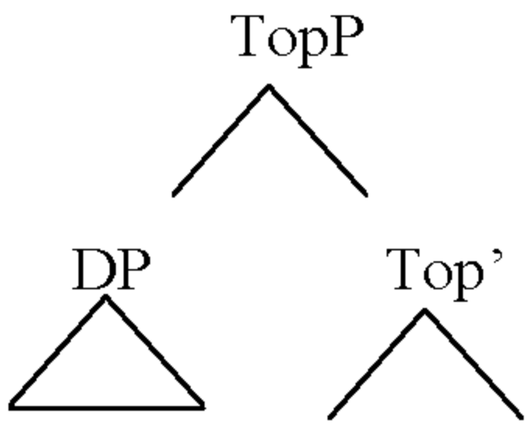


By this analysis, TopP is a recursive element which may appear prior to or following a focused element. As many languages only allow for one focus element per sentence (Hungarian is a notable exception in this respect), FocP is not afforded an asterisk for recursivity. Topic and

Focus projections are only activated when their corresponding features are present in a given numeration as a phonologically null lexical item. When this happens, the feature occupies the head of the projection, and the topicalized or focused element occupies the specifier, as in (24b).

(24) a. The book I gave to John

b.



The book [+Top] Top [+Top] ...

By this analysis then, the phonologically null Topic feature is posited to be part of the lexicon, and is internally merged in the head of Topic projection that receives the label of the functional feature. The lexical item *book*, which is marked with a [+Top] feature is attracted to Spec, TopP by the corresponding [+Top] feature on the Top head, and thus checks this feature. Therefore, for any left-peripheral Topic or Focus element, two crucial assumptions must be made: 1) that the features [+Top] and [+Foc] must exist in the lexicon as phonologically null lexical items, and 2) that some subset of phonologically realized lexical items may be [+Top]- or [+Foc]-marked in the Numeration prior to entering the syntactic derivation. An alternative to option 2 is that some number of phonologically realized lexical items also have a corresponding [+Top]- or [+Foc]- marked entry within the lexicon. This is an extremely implausible scenario since it would triple the lexical learning burden on the part of a child acquiring the language.

In the earliest generative analyses, F-marking was proposed by Jackendoff (1972) as an "artificial construct" to account for focused elements. Pollock (1989) proposed the functional projection FocP and a corresponding [+F] feature. However, the existence of this syntactic feature [+F] (Focus) has been challenged in the literature. Despite making use of a [+Focus] feature in her analysis, Zubizarreta (1998) observes that [±F] as a lexical feature is conceptually problematic since it would violate the Inclusiveness Condition (see also Szendrői (2001) (2004)). Inclusiveness involves the manners by which a node may acquire a feature – in this case, the discourse feature Focus. Following Chomsky (1995, p. 228), a non-terminal node inherits features from its daughter; while a terminal node may be assigned a feature from the lexicon. Therefore, the assignment of [+F] features to a constituent would have to happen in the lexicon. However, Zubizarreta proposes [F] not as a lexical feature, but as a derived phrase marker, which remains undefined until after Σ-structure.

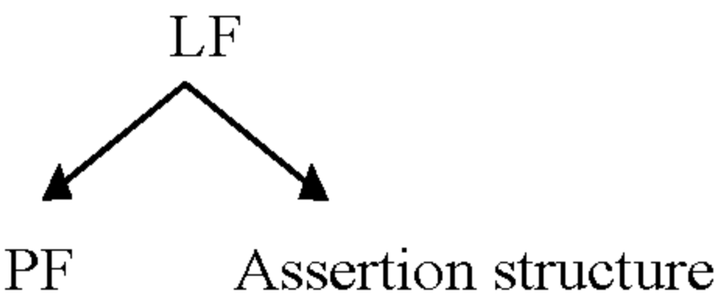
(25)

↓

(sets of phrase markers, feature checking)

Σ-structure ↓ (unique phrase marker)

(F-marking, NSR, FPR, p-movement)



In this model, phrase features remain essentially inert at the stage in

which features are checked. It is after Σ-structure (and prior to LF and PF) that F-marking, the NSR, the FPR, and p-movement take place in her analysis.

Despite the debate surrounding topic and focus features, numerous analyses have made use of them for lack of a more attractive alternative. Casielles (2004), which I discuss in the next section, is one such analysis.

2.4.1 Information Structure Dichotomies

Casielles (2004) examines the information structure dichotomies (e.g. New–Old information, Topic–Focus, Topic–Comment, Theme–Rheme, etc.) discussed above, and after thorough analysis, she arrives at two basic dichotomies which become the backbone of her proposal: Sentence Topic (STopic)–Focus and Focus–Background. Casielles draws a division between topic and background based on the following phonological, syntactic, and discourse features.

(26)

Sentence Topic (STopic)	Background
+ single	± single
+ sentence-initial	± sentence-initial
+ referential	± referential
± discourse-old	+ discourse-old
± unaccented	+ unaccented

Both STopics and Background elements are topical in her analysis.

Casielles suggests, however, that STopics could (and perhaps should) be referred to as preverbal subject topic, and Background as wide topic (p. 99, fn. 41). Unlike Background, STopics may be [- discourse-old], suggesting that STopics are present in thetic, "out of the blue" sentences, which bears similarity to Cardinaletti's (2004) subject of predication. Casielles admits trouble classifying thetic sentences (Sasse (1987)) by the two dichotomies she proposes. She notes that Lambrecht (1994) has a third sentence type for thetic sentences called Event-reporting, by which the whole sentence is focused when answering the question "What happened?". Another possibility that Casielles considers is that they are STopic- Focus sentences with a null STopic, but she notes that Erteschik-Shir (2007) disagrees with such a notion, and instead posits a "here and now" stage-topic to describe such sentences (which is supposed to correspond with Kratzer's (1989) spatio-temporal argument). The limit to this possibility, however, is that only stage-level predicates can be stage-topics in such a system (and would thus exclude individual-level predicates – perhaps not a problem since they are not eventive). Casielles also hypothesizes that thetic sentences may be instances of STopic-Comment structures, a structure that does not appear within her classification of sentence types.

2.4.2 The Interface and Phases

Parafita Couto (2005) examines the interface of information structure and syntax as it pertains to focus. Due to the existence of sentences such as (27a, cf. 27b), she, too, suggests that the [+Focus] feature must exist

in the grammar.

(27) a. tumi ashbe bole ami ghor shajiechi
you come.FUT.2 since I room (the) decorate.PERF.1

b. ashbe bole tumi ami ghor shajiechi
come.FUT.2 since you I room (the) decorate.PRF.1
Since YOU will come, I have decorated the room.

For Parafita Couto, movement of the type in (27b) is rightward p(rosodic)-movement to a phase edge. In this proposal, each phase edge is the locus for focus encoding. PF and semantics have access to the syntactic module at these phase-edges. Such access is necessary to ensure that the emerging structure meets the demands of the unfolding discourse. By her analysis then, phase edges are landing sites for p-moved XPs. This proposal is attractive in that it obeys Chomsky's (2005) notion of phases, which allows for multiple Spell-out over the course of a given derivation, thus granting PF cyclic access to non-phase-edge material at the end of each syntactic phase. Note that similar, but unrelated proposals possess interesting similarities in this respect. In Steedman's (2000) Combinatory Categorial Grammar (CCG) approach to the interface, for example, intonational boundaries coincide with major syntactic boundaries (see also Selkirk (1996) for a similar approach). Within this particular framework, surface structure, information structure, and intonation coincide within a given clause. However, just because focused elements may be moved by what appears

to be phase-related p-movement phenomena does not justify the existence of a [±focus] feature in the lexicon.

Szendrői ((2001) (2004)) argues that the inclusion of such pragmatic features in the lexicon violates the Inclusiveness Condition in Chomsky (1995) since, for a [+Focus] feature to be assigned to a constituent in a given Numeration, it would have to be a feature on that lexical item. She notes that there is no way in which this could be so, thus suggesting that [+F] is no more than a diacritic inserted to account for characteristics unrelated to a lexical property of a lexical item(see also Brunetti (2004), Emonds (2004) and Reinhart (2006) for critiques along similar lines). She proposes that Focus denotes and encodes an information status relation of constituents relative to the rest of an utterance. The same holds for Topic. However, the encoding of this relation via diacritics or features may not occur in the syntactic computation without violating inclusiveness.

Spyropoulos & Revithiadou (2007) focuses on reconciling phase theory and Multiple Spell-Out (Uriagereka (1999); see also Kratzer & Selkirk (2007) for a similar prosody-syntax interface treatment). They examine left-peripheral clitic-doubled objects and preverbal subjects in Greek. They show that prosodic islands match syntactic islands in the case of clitic-doubled objects, thus suggesting a syntax-prosody interface point. Crucially, however, this island correspondence does not hold for subjects. Therefore, they propose that clitic-doubled objects are (in Uriagereka's terminology) separate derivational cascades, assembled

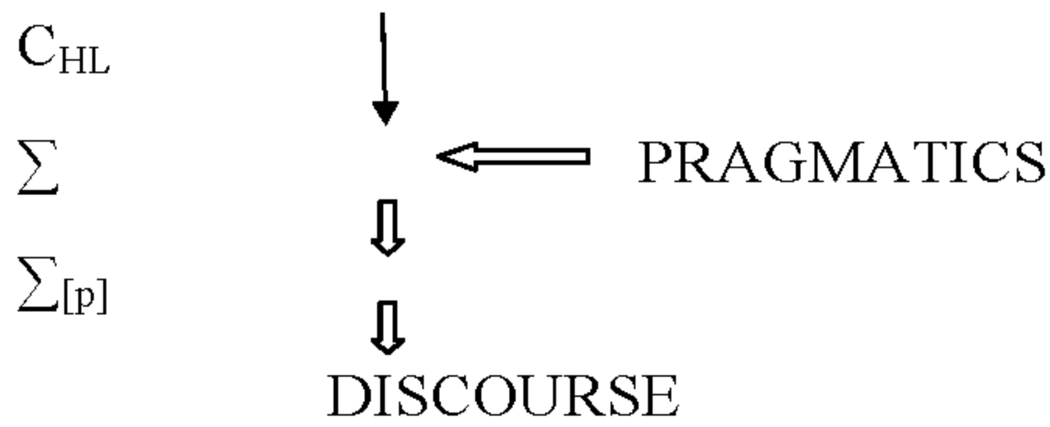
and spelled-out before they reach the main derivational cascade. Preverbal subjects in Greek, however, may be extracted from and are susceptible to prosodic restructuring. Based on such evidence, they propose that preverbal subjects in Greek form part of the main syntactic and prosodic derivation.

The preceding interface analyses agree in proposing that some sort of syntactic interface coincides with phase edges. This concept is central to López's (2009) analysis of the syntax-pragmatics interface in Spanish and Catalan. A crucial difference to his proposal lies in the model of the grammar that he proposes, which makes more concrete predictions regarding grammaticality and acceptability.

2.5 Syntax-Information structure Interface model

López's (2009) argues that discourse relations are settled by their syntactic configuration. According to the following schema (28), pragmatics module assign information structure functions and it looks over the structure of the relevant syntactic domain.

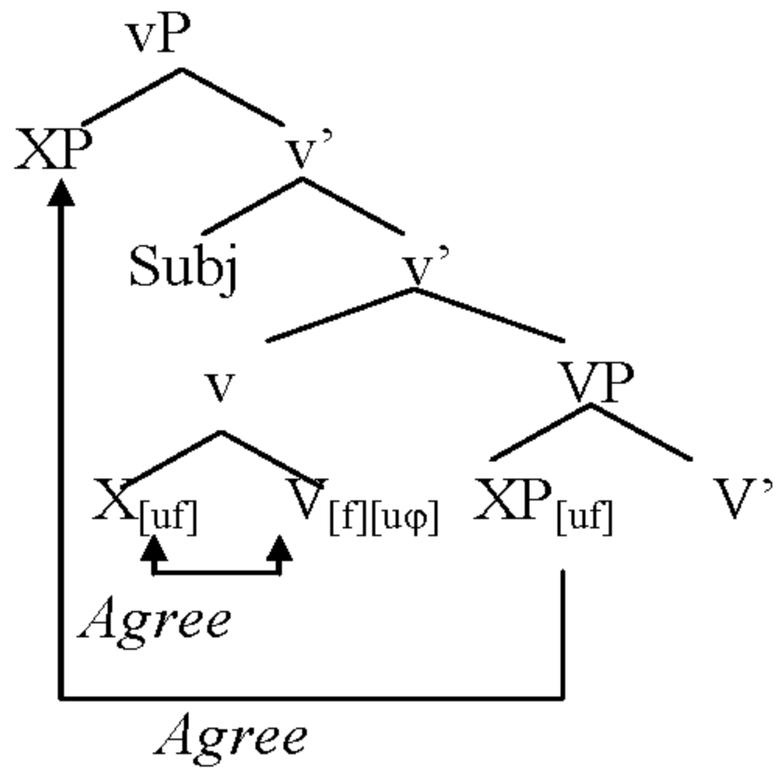
(28)



According to this proposal, pragmatic values may only be changed

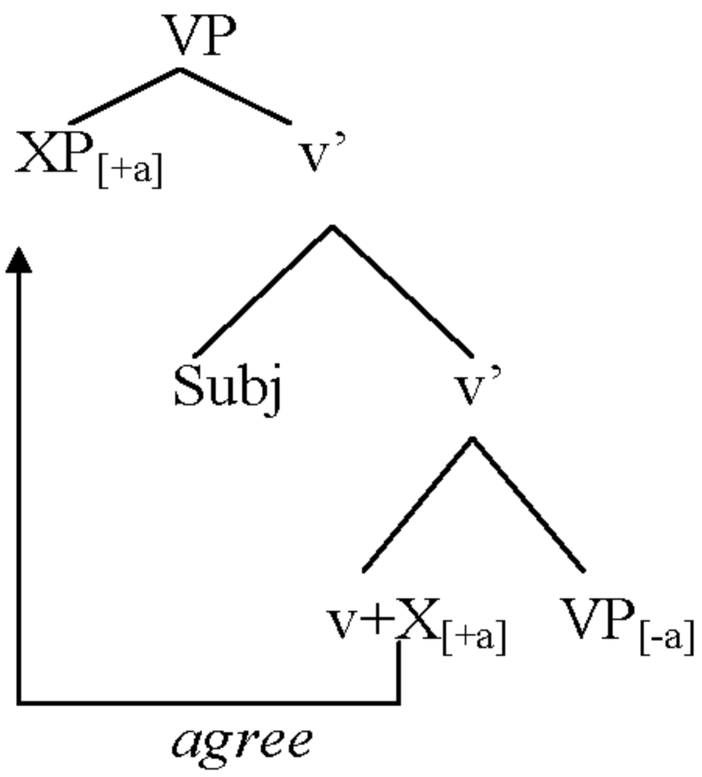
within the frameworks of the syntactic phase. It is worth to mention here that a phase is a syntactic domain. In general, a simple sentence is decomposed into two phases: CP and vP. Therefore, pragmatic values can correlate with these phases. If it is not the case, they are unaltered by further syntactic movement – the pragmatic value perseveres with a constituent assigned a value as it continues to move within the computational system. In López's analysis, the pragmatics module allocates interpretive values related to discourse anaphoricity and contrastiveness. These pragmatic features [±a] (anaphoric) and [±c] (contrastive) are not assigned to lexical elements in the numeration as they enter the derivation; rather, they are assigned derivationally as the pragmatics module "reads" the output from the syntactic module. Therefore, constituents appearing in certain structural positions at phase end get assigned interpretive pragmatic features [±pf]. The potential permutations of these (quasi) post-syntactic pragmatic features explain the discourse function of a constituent. For López, the term *rheme* refers to *regular focus*, a term that covers narrow-focus. It is mentionable that regular focus differs from *contrastive focus* mainly in syntactic terms. Contrastive focus is fronted, while regular focus occurs *in situ*. Assignment of the [+a] feature is assigned to a clitic X, which López assumes to already be in a feature dependency Agree relation with the verb prior to phase end. The following feature matrix describes the Agree relations at play prior to assignment of pragmatic features (29).

(29)



The feature [f] on v is recommended to be analogous to Case, and is valued by the clitic X. The object XP then does not have its [uf] satisfied yet. Following merge of the external argument, the remaining unvalued feature on the object XP triggers movement of the object XP to the outer Spec of vP, which allows it to have its features checked and valued. In this proposal the feature checking is treated as a very local process that may only occur within the c-command domain of the probe (i.e. the feature that requires checking). This assumption is crucial in motivating the movement (by Attract) of the doubled XP to Spec, vP. Such an argument creates a local dependency between the clitic and verbal argument (object). This dependency relation is crucial with respect to the assignment of [+a] features. When [+a] features are assigned to the (anaphoric) clitic X by the pragmatic module at the end of the vP phase, Spec, vP also becomes [+a], as in (30). The schema is stated bellow:

(30)



The VP complement of X is then assigned [-a], which matches with information focus elements being non-anaphoric. Note that the Agree relationship between the clitic X and the clitic double is crucial for López's proposal, as elements that do not enter into such a relationship with the clitic X (e.g. fronted focus, which does not have a clitic double) cannot be marked [+a] by the pragmatics module. While on the one hand this prevents the external argument, elements that will be focus fronted, and non-D-linked phrases, which also stop in Spec, vP on their way to higher positions, from being marked with the [+a] feature, on the other hand it does not prevent the complement of Spec, vP from being marked [-a]. [+a]-marked elements then are peripheral elements which either remain in Spec, vP for clitic elements that later move higher in the structure for another interpretation. Since only constituents that move to Spec, FinP are assigned [+c] by the pragmatics module, let's examine how [±a]-assignment would work with contexts that would not involve structures in the higher, left-peripheral realm, and the sort of pragmatic predictions that it would make. Consider SVO (31a), VSO (31b) and VOS (31c) word orders.

(31) [Context: What did Sumon eat?]

a. [$_{TP}$ (Sumon) [$_{T'}$khelo [$_{vP}$ [$_{v'}$ [$_{VP}$ aekta aam$_{[-a]}$...]]]]]

Sumon eat.PST.3 a mango

'Sumon ate a mango'

b. [$_{T'}$khelo [$_{vP}$ (Sumon) [$_{v'}$ [$_{VP}$ aekta aam$_{[-a]}$...]]]]

c. #[$_{T'}$khelo [$_{vP}$ aekta aam [$_{v'}$ Sumon $_{[-a]}$ [$_{VP}$...]]]]

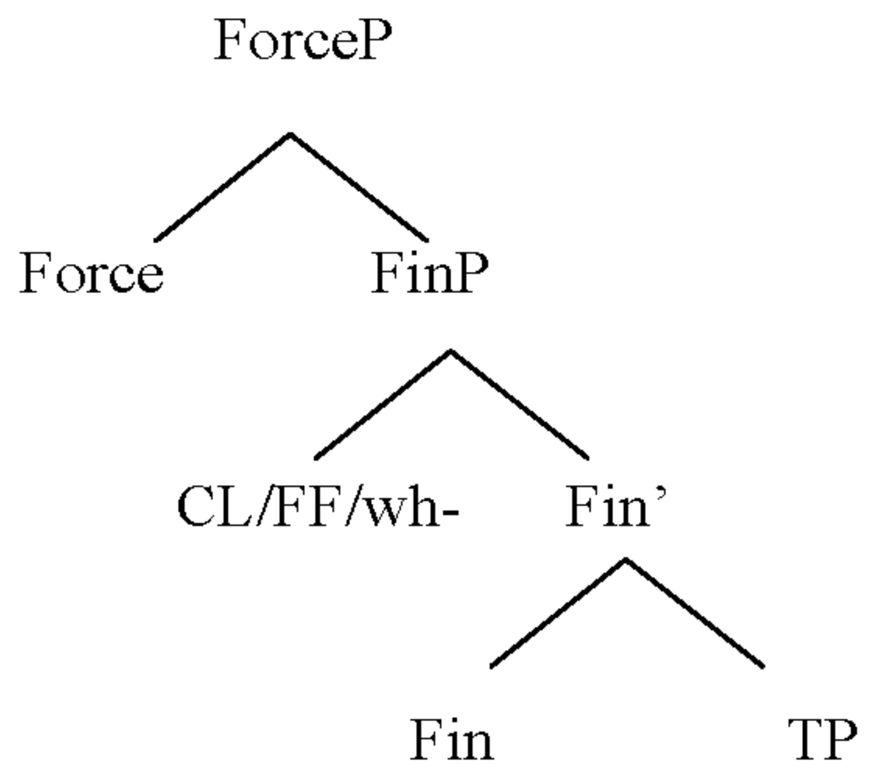
For the context provided in (31), both (31a) and (31b) are felicitous replies (with or without the subject *Sumon*) since in both of these *aekta aam* is marked [-a] for regular focus/rheme. In (31a) the external argument is not in an Agree relation with a clitic prior to moving on to Spec, TP, and therefore is unaffected by [+a]-marking at the end of the vP phase. The same applies in (31b), but *John* does not continue to move higher. In (31c), *Sumon* is marked [-a] thus correctly predicting its infelicitousness for this context. If I alter the question context to *Who ate a mango?*, however, then only (31c) is appropriate since it is the only configuration in which *Sumon* is marked [-a].

López's analysis runs into a bit of a problem for all-focus, or thetic sentences since in this sort of context, the whole sentence should be marked [-a]. To deal with this, he suggests that subjects can also bear an additional feature which he calls [*u*d]. The interpretable counterpart of this is [d], which appears on Fin. Unvalued φ-features on Fin allow it to probe and trigger movement of the subject DP to Spec, TP. This portion of the proposal is problematic, namely due to the [d] feature that he

proposes to initiate a new discourse. If [d] is a discourse feature like [±a] and [±c], it is unclear why this particular feature would be purely syntactic and not be involved with the pragmatic module.

Prior to continuing with an example of [±c]-assignment, let's examine the structure that López proposes for the left periphery. Since he does not assume topic and focus features, his left periphery consists of only ForceP and FinP (32).

(32)



López believes that wh-phrases, focus phrases and clitic phrases take up specifier positions of FinP. When more than one of these is available, they come out as stacked specifiers of FinP. Notably, preverbal subjects do not appear here. Now that I am familiar with López's left peripheral structure, let's take a look at how [+c] assignment is supposed to work if I change our context from (31).

(33) [Context: What did else Sumon eat?]

a. aam -o khelo Sumon

mango also eat.PST.3 Sumon

'The mango also, Sumon ate'

b. [$_{FinP}$ aam-o$_{[+a,+c]}$ [$_{Fin'}$[$_{TP}$[$_{T'}$ khelo [$_{vP}$ <aam-ti$_{[+a]}$> [$_{v'}$ Sumon $_{[-a]}$ [$_{VP}$...]]]]]

Following López, in a Clitic reply (33a) to the context in (33), *aam-o* first moves to Spec, vP, after which pragmatics marks it [+a] due to its agree relation with the clitic *o*, also marked [+a], while still in v. The direct object *aam-o* then later moves to Spec, FinP (43b). At the end of the phase, pragmatics [±c]-marks the element appearing in Spec, FinP, and the complement of FinP (i.e. the remaining structure) gets marked [-c]. Fronted focus elements are proposed to make the same movement steps as Clitics save for [+a] marking in Spec, vP, which does not apply due to the lack of a clitic-double dependency.

A particularly attractive aspect of this proposal is that it examines in detail how to determine the appropriateness of Clitic elements. López proposes that Clitic requires a particular type of discourse relation in order to be felicitous and appropriate. Analyses of discourse generally assume a hierarchical structure for discourse (see e.g. Hobbs (1985), Polanyi (1988), Grosz & Sidner (1986), Asher (1993) van Kuppevelt (1995) Asher & Vieu (2005)), working in the Segmented Discourse Representation Theory framework, distinguish between two particular types of discourse structure relations: coordination and subordination. The two are summarized as in (34).

(34) *Coordination:* narration, background, result, continuation, parallel, contrast, question-coordination, correction

Subordination: elaboration, instance, topic, explanation, precondition, commentary, question-answer pairs

According to these relations, both narrow- and wide-focus (López's *regular focus* or *rheme*) question-answer pairs would qualify as subordination contexts. Foci expressing contrast, however, would qualify as coordinating. This division matches López's proposed division between regular and contrastive focus. While thetic, "out of the blue" contexts can answer the question "What happened?", they may also initiate a discourse without such a question. Asher & Vieu provide only one example with such a question, and in that example, they define discourse relations as related to the first sentence of the reply. I therefore assume that thetic contexts are coordinating. López proposes that clitics requires discourse subordination as well as a discourse antecedent in the superordinate sentence in order to be felicitous and appropriate. The antecedent for clitics need not be immediately adjacent, but must be found within the preceding discourse in the appropriate subordinating relation.

López's proposal then provides not only a mechanism by which the pragmatic module interfaces with the syntax in order to assign discourse features to a variety of sentence elements, but also provides metrics for creating discourse appropriate clitic positions.

2.6 Implications: Discourse factors in the collection of speaker judgments

In order to determine focus-information structure interface, one must

generate discourse contexts and test them with native speakers of the target language. However, there are certain challenges present in creating discourse contexts and potential replies. In particular, subjects can be problematic in languages like Bengali in part because it is a null-subject language. The Bengali question-answer pair in (35) typifies the phenomenon I am referring to.

(35) a. Sumon ki randh-lo?

Sumon what cook.PST.3

'What did Sumon cook?'

b. (Sumon) bhat (randh-lo)

Sumon rice cook.PST.3

'Sumon cooked rice'

In (35b), the subject or the subject and verb may be lacking in a reply to the object- narrow-focus question in (35a). In fact, the two-word reply would likely be the most common reply to this question. However, in order to effectively establish word order preferences for the subject, verb, and object, all of these constituents must be present. Therefore, for all of the discourse contexts I examine, all of the possible replies are full sentences including a DP subject, verb and DP object.

As previously discussed, since there are many more possible word orders in a language like Bengali (as compared to a language like English or French), additional care must be taken since word orders like SOV, OSV and OVS are never ungrammatical, but may be *dispreferred*.

Therefore, the goal of eliciting judgments will be establishing *word order preferences* according to information structure context. The result may be that not all well-constructed syntactic structures are acceptable at the level of discourse structure. Determining these preferences then will assist in describing the syntax- information structure interface for Bengali.

2.7 Conclusion

In this chapter, I have reviewed many of the definitions of *topic* and *focus* in the literature. I have also discussed some analyses for the syntax-information structure interface in the literature. I follow many of López's (2009) assumptions in creating appropriate discourse contexts in my methodology in order to determine which word orders are most appropriate for a variety of discourse contexts in Bengali. Before proceeding with a description of the methodology that I employ, I briefly summarize the information structure assumptions that inform the following chapter.

The definition of *thetic contexts* that I assume in this dissertation follows Zubizarreta's (1998) characterization of these contexts as "out of the blue", or what many others have called "all focus". Such sentences may either initiate a discourse or provide replies to questions like "What happened?". As previously discussed, the basic definition of *topic* that I assume is based on the notions of "aboutness" or "discourse old" in the literature. However, since I examine cliticized topics in particular, I

assume López's (2009) analysis for clitics, as discussed in section 2.5. According to López's proposal, clitic placement requires discourse subordination, which consists of a discourse antecedent and a subordinating discourse context as defined by Asher & Vieu (2005). In the methodology described in Chapter 3, I design task conditions to test the validity of this proposal for Bengali. When I refer to topics in Chapters 3 and 4, I occasionally make use of the terms *subject old* and *object old*. In most cases, this is due to space constraints in tables. These terms refer to topical, discourse-old subject DPs and discourse-old object DPs, respectively.

The definition of *focus* that I assume follows López (2009). Recall that for him, the difference between regular focus (rheme) and contrastive focus depends on how the focus integrates into the existing discourse. While regular focus resolves an open variable in the discourse, contrastive focus simultaneously opens and resolves a variable. Regular focus provides an answer to an explicit or implicit wh-question, while contrastive focus cannot answer a wh-question. The regular focus contexts I will employ are what I refer to as *narrow focus*, whereby a wh- question elicits either the subject or object. I do not employ any contrastive focus or focus fronting contexts in the quantitative data that I gather.

In the following chapter, I detail more precisely how the above information structure assumptions are tested for in the quantitative and

qualitative tasks and conditions that I employ in my investigation of focus constructions in Bengali. The following questions will guide my investigation to get the answer of the main query of this research:

1. What is the preferred structure in Bengali for *out-of-the-blue* as well as thetic, sentences?
2. What is the preferred structure in Bengali for sentences in which the grammatical subject represents discourse-old information?
3. What is the preferred structure in Bengali for sentences in which the grammatical object represents discourse-old information?
4. What is the preferred structure in Bengali for sentences in which the grammatical subject is narrow-focused?
5. What is the preferred structure in Bengali for sentences in which the grammatical object is narrow-focused?
6. How does the data collected contribute to the overall analysis of information structure in Bengali?

Chapter Three

Methodology

By now, it has been studied that there are intricacies entailed in settling on the exact status of Bengali focus constructions and their relevant information structure. I presume that the syntax–information structure interface is a vital step in the constructing of an acceptable focus structure. Precisely explaining this interface necessitates organized and empirical data on what kinds of sentence structures are suitable in what types of discourse contexts to serve as a guide. Therefore, I have designed quantitative and qualitative tasks to collect such data for Bengali, which will then supply the part of the empirical base for my analysis of Bengali focus constructions.

3.1 Mixed Method: An Overview

Nowadays, mixed methods research has drawn particular attention in empirical research and a large quantity of studies have addressed theoretical and methodological aspects of incorporating qualitative and quantitative methods (e.g. (Caracelli, 1997); (Morgan, 2007); (Creswell et al, 2008)). In field linguistics research, even though the nature and prevalence of qualitative and quantitative methods have already been explored, a tiny amount of researches have been addressed the integration of the two methods (see (Do¨rnyei, 2007)). In this section, I would like to present a number of reflections on the utilization of mixed methods research in linguistics.

Mixed methods research has developed from mono-method research; and its recent history revisits the concept of 'triangulation'. As described

by (Bergman, 2008, p. 1), mixed methods research requires 'the combination of at least one qualitative and at least one quantitative component in a single research project or program.' Nevertheless, it would be simplistic to presuppose that the use of the two components in a study could be judged as utilizing mixed methods research (Bryman, 2008). High-quality mixed research needs high levels of integration at different phases of the study; that is, while forming research questions, during sampling, data collection and analysis, while making interpretations, and drawing conclusions (for more, see (Yin, 2006)).

In social and behavioral research, theoretical and methodological aspects of mixing methods have been explored. Theoretically, a number of studies have explored and re-examined epistemological and ontological issues of combining qualitative and quantitative methods (see e.g. (Morgan, 2007)). Methodologically, a number of studies have investigated questions related to mixed research designs (Creswell et al, 2008), mixed sampling designs (Collins et al, 2007), the nature of data collection and analysis, and the 'quality of inferences' in mixed methods research.

Several studies have dealt with mixed research designs ((Nastasi, B.K. et al, 2007); (Creswell et al, 2008)). Based on wide-ranging literature reviews and scrupulous content analysis, Creswell et al. (2008: 67–68) categorize the designs as either 'concurrent' (designs that draw on qualitative and quantitative research concurrently) or 'sequential' (designs

that are carried outed sequentially). According to Creswell et al. (2008: 67–70), these major designs can be additionally classified into:

(i) 'concurrent triangulation design' (QUAN and QUAL data are collected and analyzed in parallel and interpretations are drawn based on QUAN QUAL results); (ii) 'concurrent embedded design' (QUAL data are collected within the QUAN design, between pretests and posttests, and interpretations are based on QUAN QUAL data); (iii) 'sequential explanatory design' (first QUAN data are collected and analyzed; then, to further explain the results QUAL data are collected and analyzed; finally, interpretations are based on QUAN QUAL data); (iv) 'sequential exploratory design' (first QUAL data are collected and analyzed; then, to further explore the problem QUAN data are collected and analyzed; finally, interpretations are based on QUAL QUAN data); and (v) 'sequential embedded design' (this 'typically involves collecting qualitative data before an intervention begins or after it is complete' (p. 69); interpretations are then made based on data integration).

Performing research in linguistics, as Larsen-Freeman and Cameron (2008, p. 251) verify, engages investigating the changes in intricate systems with 'adaptive', 'dynamic', and 'nonlinear' procedures. In this regard, mixed methods research is a helpful device for exploring intricate systems, examining both the processes and the outcomes (see Yin (2006)). The qualitative exploration of the processes and quantitative measurement of the outcomes in a concurrent design would

supply a more holistic depiction of the phenomenon under study. In addition, longitudinal inquiry of a vibrant arrangement using qualitative methods coupled with the evaluation and measurement of the experimental features of the system at different points in time in a sequential.

As a result of the qualitative content study, there come out a array of functions and applications for mixed method research in linguistics, which replicate extensive potential. Concurrent triangulation designs, for example, appear to be used to provide supplementary data in multilevel discourse analysis, needs analysis, and program evaluation. Concurrent embedded designs materialize to be used to qualitatively discover the processes within an experimental design. In studies where a qualitative explanation of the findings from the quantitative phase is required, sequential explanatory designs may serve as an appropriate tool. It is mentionable to state that I have chosen sequential explanatory design for this present study. Since this design would add explanatory power to coding and rating procedures in language assessment, I have selected it as the most suitable way out for my research. In contrast, sequential exploratory designs can be utilized to expand the scope of the qualitative phase and further explore the problem quantitatively. The qualitative phase may also be used as a pilot study. Such designs would be used in developing and validating measurement instruments such as questionnaires, tests, and rating scales (see (Onwuegbuzie et al, 2010)). The exploratory designs could also be merely used to endow with

quantitative support for the qualitative findings, chiefly in discourse analysis. Like exploratory designs, sequential embedded designs could be utilized for test validation purposes, where the researcher requires to quantitatively looking at cross-sections of the test development and validation process within the research program. Also, these designs may supplement qualitative ethnographic studies for the purpose of developing patterns and evaluating models in a more precise way.

It comes into views that linguistics has immense potential for making use of mixed methods research. I strongly feel that by developing more carefully informed opinions about mixed methods research and by making creative use of a range of mixed-method designs the quality of research in the field of field linguistics could be further improved. Regarding the present study, my choice of mixed methods research is validated by the research rationale, the research questions, the economic and political aspects of this project, and the requirements of its research context. In this research, my qualitative data supports my quantitative findings. Therefore, I am able to get the measuring figures of Bengali native speakers' focus preferences and at the same time the reason behind their choice is also revealed. Needless to say, a deeper and more comprehensive understanding of mixing qualitative and quantitative components at different stages of the study would be helpful to researchers.

3.2 Research Design

The experimental measures detailed in this chapter seek to gather data

on acceptability judgments, preferences, and uses of a variety of word orders for a variety of information structure contexts. The verbs used to elicit judgments and preferences in these tasks are all agentive transitive verbs in order to guarantee the presence of a subject and object argument in the clausal structure, as opposed to intransitive verbs—either unaccusative or unergative—which have only one or the other. In section 3.3, I describe the subjects who participated in this investigation. In section 3.3.1, I describe the variables that I examined, and justify the modifications that I made in these variables. In section 3.3.2, I describe the data-gathering procedures that I used and the conditions involved in each task. In section 3.4, I describe Task 1 and each of the seven conditions that I examined in this task. In section 3.5, I explain Task 2 and the six conditions that I examined in that task. In section 3.6, I explain the methods I used in Task 3 as well as the participant groups and the experimental procedures involved in this task. I offer concluding remarks for this chapter in section 3.7.

3.3 Participants: Tasks 1 and 2

Prior to completing tasks 1 and 2, participants completed a linguistic history questionnaire, which included a maximum of 20 questions about age, place of birth, current place of residence, the language(s) that they speak at home, at school, at work, with friends and colleagues, and with family relations. The linguistic questionnaire also includes questions about experience with Bengali-language education at the primary, secondary and tertiary (university) level. I report on their responses to

these questions below in my description of the variables I originally planned to consider, and explain the decisions that I made to remove variables from consideration.

3.3.1 Participant variables

For this investigation, I recruited 114 male and female participants currently living in Dhaka district. Among the 114 who started the initial linguistic profile making task, the results from 34 of these participants were removed from statistical consideration. 28 of these did not complete the whole tasks and 6 belonged to an age group not under consideration in this dissertation (31-49 years old). Of the 80 remaining participants, 42 were male and 38 were female, all of whom reported Bengali native speakers. According to 2011 figures from the Bangladesh Bureau Statistics, the country has a population of 14,97,72,364 of which 50.06% is male and 49.94% female. The two age groups examined in this study are ages 18-30, and older than 50. Considering groups differing by one generation enables me to control for and to detect potential intergenerational grammar differences, which may provide evidence of linguistic changes in progress.

The next variable I consider is primary living environment. For the purposes of this investigation, I adopt the binary variable [±urban] to describe the primary living environment of Bengali-speaking participants. I define [+urban] as speakers who primarily reside in Dhaka city and in the peripheral locations attached to this city. I define

[-urban] as speakers who primarily reside outside of the city area; however all of them belong to the larger Dhaka district.

The remaining participant variable that I consider in this investigation is level of education. There is a notable difference in reported reading and writing abilities between those with and without a minimum of secondary education. For example, those with secondary education actually *completed* their secondary education. Speakers who were classified as having *primary* level studies, however, included speakers who had and had not completed their primary studies. Lacking such specifics in definitions, I define [+educated] as Bengali speakers who have completed at least some level of secondary education. I attempt to include a wide spectrum of [+educated] speakers. A summary of the participant variables under consideration to this point, as well as their numerical representation in the questionnaire results, appears in Table 1 below.

Categories	**Male**				**Female**			
	Age	**Number**	**Age**	**Number**	**Age**	**Number**	**Age**	**Number**
[+urban]/[+educated]	18-30	n=29	50+	n=3	18-30	n=26	50+	n=5
[+urban]/[-educated]	18-30		50+	n=1	18-30		50+	n=2
[-urban]/[+educated]	18-30	n=9	50+		18-30	n=4	50+	n=1
[-urban]/[-educated]	18-30		50+		18-30		50+	

Table 1. Summary of participant variables.

As it can be seen above, not all of the variables are represented in the

participant population that successfully completed the whole tasks. Over half of the participants came from [+urban], [+educated] backgrounds and the majority of those belong to the 18-30 year-old group. There were very few participants from [-urban] backgrounds, yet among those, there was no participant from [-urban], [-educated] backgrounds. As many of the above variables were not sufficiently represented in the results gathered, I only report on the variable age as it relates to participant preferences and ratings in Chapter 3.

3.3.2 Procedures: Tasks 1 and 2

Tasks 1 and 2 were quantitative data-gathering tasks. Participation was entirely voluntary and anonymous. Monetary compensation or otherwise was not provided. Participants are randomly divided into two groups, thus varying the order in which the quantitative linguistic tasks as well as the items within the tasks are presented in order to avoid a task effect. In the case of task 1, which includes audio files, the gender of the interlocutors in the task items is also varied between the two groups. For Task 1, there were seven different condition types, and three tokens for each condition, yielding a total of 21 tokens. There were five different condition types in Task 2, and three tokens for each condition type, netting a total of 15 tokens. As Task 2 was composed of five different extended discourse contexts, participants completed almost half of the discourse contexts per visit in order to avoid task-type fatigue as well as overall saturation fatigue. Context presentation order for Task 2 was also randomized, but the question items within each context were not, as

this would have adversely affected the construction of the information structure contexts within them.

3.4 Task 1: Scaled Pragmatic Appropriateness Task

Task 1 is based on the grammaticality judgment task (Bley-Vroman, Robert & Naoko Yoshinaga et al., 1992) used in Second Language Acquisition research to test the acceptability of sentences in learner grammars, results from which were then compared with a native speaker results. In this task however, participants read a conversational context and then provided an appropriateness rating for (syntactically) grammatical sentences that varied in focus structure. For each conversational context, a triad of possible responses with varying word orders was provided following Kallestinova's (2007) methodology employed for gathering data on clausal variants in Russian.

Participants were instructed to listen to each possible response in the triad and then to rate them on a scale of acceptability. As all focus word order options were grammatical, this task scale utilized an ordinal scale of 1 to 5, (1 meaning 'unacceptable', and 5 meaning 'preferred'), following recommendations in Schütze (1996) and White (2003), and thus differed from Bley-Vroman & Yoshinaga's (1992) original scale, which used a five-point ordinal scale from -2 to 2. This scale accompanied each task 1 triad for the participants' convenience.

1 = not acceptable (*grohonjoggo noy*)

2 = marginally acceptable (*kichuta grohonjoggo*)

3 = more or less acceptable (*moter opor grohonjoggo*)

4 = rather acceptable (*pray grohonjoggo*)

5 = totally acceptable (preferable) (*shompurno grohonjoggo*)

Instructional sample tokens preceded the task so that participants would know that there may be more than one acceptable manner in which to conclude the discourse situations in the task, (i.e. that two or more sentences may receive the same acceptability rating). Pragmatic conditions included thetic (*out-of-the-blue*) situations, subject narrow-focus or rheme, object narrow-focus or rheme, subject arguments as old information, and object arguments as old information. Audio clips accompanied both the conversational context and the possible responses in order to control for the intonation properties of the clausal structures presented. I discuss each condition type below.

3.4.1 Condition A

Condition A tokens sought to establish a word order preference for thetic contexts or, as Zubizarreta (1998) labels them, *out-of-the-blue* situations. These do not presuppose knowledge of the subject, verb, or object, but only presuppose that something occurred (i.e. they ask the basic question "What happened?") These questions examine three word orders: Subject-Object-Verb (SOV), Object-Subject-Verb (OSV), and Object-Verb-Subject (OVS), as in (1).

(1) Context: Sumon and Shobuj are friends. They are talking with each other.

Sumon– What are you doing tonight?

Shobuj– Why? What's up?

A. Sumon– Sho'jib amader-ke nimontron korte chay
Sho'jib us-Acc invite.INF do.nonFin want.PRS

B. Sumon– amader-ke Sho'jib nimontron korte chay
us-Acc Sho'jib invite.INF do.nonFin want.PRS

C. Sumon– amader-ke nimontron korte chay Sho'jib
us-Acc invite.INF do.nonFin want.PRS Sho'jib

Shajib wants to invite us.

3.4.2 Condition B

Condition B items sought to determine the preferential position of the subject when it is discourse-old information (i.e., it has already been introduced into the common ground of the discourse), as discussed in Chapter 2. To avoid a null subject as well as unnatural repetition of an overt subject in these items, a switch reference or paraphrase of the subject referent is used (e.g. in (2), the switch from "publisher" to "Muktodhara"). The word orders examined in item B are SOV, OSV, and OVS, as in example token (2).

(2) Context: Sumon and Shobuj are friends. They are talking about a book.

Shobuj– Who has published the book?

A. Sumon– Muktodhara boi-ti prokash koreche

Muktodhara book-the-Acc publish do.PRF

B. Sumon– boi-ti Muktodhara prokash koreche

book-the-Acc Muktodhara publish do.PRF

C. Sumon– boi-ti prokash koreche Muktodhara

book-the-Acc publish do.PRF Muktodhara

Muktodhara has published the book.

Condition B contexts were all Question-Answer (Q-A) pairs, which are classified as *subordination* contexts according to Asher & Vieu's (2005) analysis of discourse relations in the Segmented Discourse Representation Theory (SDRT) framework (Asher, 1993). According to López (2009), discourse subordination creates a context appropriate for focus of previously introduced referents. Since in this condition the subject has already been introduced within the discourse, the preferred word order for this condition will provide data to inform this question. Note, however, that in SOV word order it is impossible to detect whether a preverbal subject appears in a canonical or left-peripheral position. The availability of clitic is also of relevance for discourse-old objects, as in Condition C below.

3.4.3 Condition C

Condition C tokens were designed to establish preferred focus structure when an object is discourse-old information within the context presented. These contexts also involved Q-A pairs, thus serving as a counterbalance for the Condition B tokens above. These items seek to test the availability and relative appropriateness of clitics using in Bengali focus structure. The response triads in this condition have a bi-clausal structure, placing the first conjunct with focus in opposition to the second conjunct. Furthermore, the focus object in the first conjunct will create a strong preference for a parallel focus structure in the second one. The word order possibilities for the first conjunct in these triads are OclSV (clitic object followed by a preverbal subject and a verb), OclVS (clitic object followed by a verb and a postverbal subject), and SOV, as in (3).

(3) Context: Sumon and Shobuj are friends. Shobuj has moved in a new house. When they enter Sumon's new house, there is a large number of books inside.

Shobuj– Wow! Your family already brought in all books!

A. Sumon—boi-gulo-i baba eneche kintu book shelf -ti dokan-e
book-Pl-cl papa bring-PRF but book shelf- the shop- in (the)

B. Sumon— boi-gulo-i eneche baba kintu book shelf -ti dokan-e
book-Pl-cl bring-PRF papa but book shelf- the shop- in (the)

C. Sumon— baba boi-gulo eneche kintu book shelf -ti dokan-e
papa book-Pl bring-PRF but book shelf- the shop- in (the)

Papa has only brought the books but the book shelf is in the shop .

Word orders did not vary in the second conjunct in this condition, and all included an appropriately anaphoric focus object according to the prerequisites introduced above. The prediction is that the object would be preferred in a focus position in the first conjunct. Were it to appear in its canonical position, this would violate the parallel structure of focus between the two conjuncts. This item also sought to determine whether a preverbal or postverbal subject would be available or preferred in the first (focus) conjunct. Note that the second conjunct involves a null subject. All subjects in the response triads were subsets, recasts, or switch references of the subject in the elicitation context in order to avoid 1) repetitiveness, which would create an unnatural response, and 2) null subjects, since a null-subject response in the first conjunct would not indicate the preferential position of the subject.

3.4.4 Condition D

Condition D tokens were designed to determine preferred focus structure for contexts in which a subject is old information within the discourse context. This information structure condition serves as a counterbalance for Condition B (above) on the one hand, and for Condition E (below) on the other. This item is similar in nature and structure to Condition B; however, Condition D responses employed in this item are *continuation* and *result* in relation to the elicitation contexts, both of which are classified as coordination contexts by Asher & Vieu (2005). The word orders examined in this item set are SOV, OSV, and OVS, as in (4).

(4) Context: Sumon– Did you hear?

Shobuj– What?

Sumon– Shajib stood first in the class.

Shobuj– Really! That's so great!

A. Sumon– Sho'jib ejonno britti peyeche

Shajib for this scholarship get.PRF

B. Sumon– ejonno britti Sho'jib peyeche

for this scholarship Sho'jib get.PRF

C. Sumon– ejonno britti peyeche Sho'jib

for this scholarship get.PRF Sho'jib

For this, Shajib has got a scholarship.

The availability of focus in this discourse condition is crucial in regards to the preferred clausal position of the subject. Therefore, discourse-old preverbal subjects appear to be non-peripheral elements.

3.4.5 Condition E

Condition E contexts also sought to verify the availability and appropriateness of dislocation of objects for Bengali clausal structure. In these discourse contexts, objects represent discourse-old information, as in Condition C. However, the contexts employed in these response triads are *continuation* and *result*, as in Condition D above, thus providing a counterbalance on two fronts. The word orders examined in this item

were OclSV, OclVS, and SOV, as in (5) below. Three questionnaire items of this type included PP adjuncts, and two include Adv adjuncts, neither of which was predicted to affect the appropriateness of the replies in the triads.

(5) Context: Sumon saw Shajib at the playground on Monday.

A. budhbar- e ei chele-ti- ke- i Shobuj dekhe-che
Wednesday on this boy the-Acc-cl Shobuj see.3.PRF

B. budhbar- e ei chele-ti- ke- i dekhe-che Shobuj
Wednesday on this boy the-Acc-cl see.3.PRF Shobuj

C. budhbar- e Shobuj ei chele-ti- ke dekhe-che
Wednesday on Shobuj this boy the-Acc see.3.PRF

On Wednesday, the boy, Shobuj has only seen (him).

As with Condition D, these contexts were *coordination* contexts, and should therefore prohibit clitic construction (focus) of an object DP.

3.4.6 Condition F

Condition F tokens were designed to establish a preference for Bengali focus architecture when there is narrow focus on the subject (i.e. the subject is the rheme, or new information). According to the Zubizarreta's (1998) reformulation of the Nuclear Stress Rule, new information appears to the right of the VP *in situ*. By her analysis, this is the result of the object scrambling past the subject to avoid the projection of focus. This task seeks to establish this

preference/constraint for Bengali for subjects. Condition F is a counterbalance for Condition G below, in which objects are narrow focus. The word order permutations that appeared in triads of this type are SOVX, OVSX and OSVX.

Within an information structure context in which only the subject is unknown, frequently the verb and object complement(s) have already been introduced into the discourse. Three of the five tokens in this condition involved permutation of an adverbial XP adjunct instead of the object, as in (6).

(6) Context: Sumon and Shobuj are friends. Sumon asks Shobuj about a box in his room.

Sumon– What's that? How pretty?

Shobuj– That? Well, it's a gift that arrived in the mail for my birthday.

Sumon– Who sent it to you?

A. Shobuj—gato shaptahe amar bon eti ama-ke pathie-chilo kintu
last week my sister it-Acc me-Dat send.3.PST but
she amar onushthan-e ashte sakkhom habe na.
she my party to to come-INF able be.FUT not.NEG

B. Shobuj—gato shaptahe ama-ke amar bon eti pathie-chilo kintu
last week me-Dat my sister it-Acc send.3.PST but
she amar onushthan-e ashte sakkhom habe na.
she my party to to come-INF able be.FUT not.NEG

C. Shobuj—gato shaptahe ama-ke eti pathie-chilo amar bon kintu
last week me-Dat it-Acc send.3.PST my sister but

she amar onushthan-e ashte sakkhom habe na.

she my party to to come-INF able be.FUT not.NEG

My sister sent it to me last week but she won't be able to come to my party.

.

3.4.7 Condition G

Condition G tokens counterbalance those of Condition F, seeking data on the preferred focus structure for Bengali when the information structure involves the object as new information (rheme). The word orders examined in these object-focus triads were SOV, OVS and OSV, as in (7).

(7) Context: Sumon Shajib and Shobuj are friends. Sumon and Shajib are talking about the achievement of Shobuj.

Sumon– Did he win any prize?

Shajib– Well, yes.

Sumon– What did he win?

A. Shobuj ekti medal peye-che

Shobuj a medal get.3.PRF

B. ekti medal Shobuj peye-che

a medal Shobuj get.3.PRF

C. ekti medal peye-che Shobuj

a medal get.3.PRF Shobuj

Sobuj has got a medal.

As previously mentioned, if Zubizarreta (1998) is on the right track,

narrow-focus new information appears to the right of the VP *in situ*. If this is the case, experiment participants should rate SVO as the most appropriate word order in this condition.

3.5 Task 2: Word Order Preference Task

Since Task 1 allowed for the possibility that participants may rate differing focus structures with identical or very similar ratings, Task 2 was designed to encourage participants to choose a clear preference from two word order possibilities. If it turns out that two word orders are equally acceptable, this task will also provide confirmation of such a preference. Although Task 2 items did not include accompanying audio for the word order possibilities, they appeared in an extended pragmatic context of three to seven items in a connected, continuous discourse. Task two items involved the same information structure contexts as in task one, with the exception of subordination versus coordination items testing the availability of focus variations.

Three sample questions preceded the items in task two in order to provide instructions to participants on how the contexts will be presented and how they may be answered. These instructions highlighted to participants that while more than one word order may be possible, one or another may not be appropriate depending on the context. This was included to encourage them to pay particular attention to the discourse contexts provided in the tasks that follow. This was to highlight to participants that all three possible answers provided over the

course of the task were possible responses. Therefore, if it was the case that two word order options are equally possible and appropriate in a given context, participants should indicate this.

3.5.1 Condition A

Condition A involved thetic, or *out-of-the-blue,* contexts. Participants had to choose the appropriate word order, with the subject preceding or following the verb. Therefore the possible word orders in this item type were SOV or OVS:

(8) Context: Sumon– What happened?

Shajib– Nobody knows exactly.

A. buRo lok-ti- ke mrito pawa geche

old man the-Acc dead find.3.PRF.PAS

B. mrito pawa geche buRo lok-ti- ke

dead find.3.PRF.PAS old man the-Acc

C. A and B both are possible.

The old man has been found dead.

SOV and OVS orders are predicted to be the most acceptable clausal word orders for thetic contexts. This item was designed to elicit a preference between these two word orders in case it turned out that both of these responses in the triads received statistically similar ratings in the corresponding Task 1 condition.

3.5.2 Condition B

Condition B tokens involved a discourse old subject DP. The possible word orders in this context were SOV or OVS:

(9) Context: Father and mother are talking about their son Sobuj and his examination

Father– Who was the other fellow that I saw a little bit ago?

Mother– That was his friend. He came to play basketball this afternoon...

A. kintu she jane na je Sobuj-er agamikal porikkha ache
but he know not that Sobuj-Gen tomorrow exam has

B. kintu she jane na je agamikal porikkha ache Sobuj-er
but he know not that tomorrow exam has Sobuj-Gen

C. A and B both are possible.

But he does not know that Shobuj has a exam tomorrow

3.5.3 Condition C

In Condition C, the subject represented the rheme within the extended discourse provided. The word orders participants can choose from in this condition are SclOV or OSclV, as in (10).

(10) Context: Sumon– *Who recommended this business to you?*

A. Shobuj– amar bhai- i ama-ke ei paramorsho dieche
my brother-cl me-Acc it suggest give3.PRF

B. Shobuj– ama-ke amar bhai- i ei paramorsho dieche
me-Acc my brother-cl it suggest give3.PRF

C. A and B both are possible.

Shobuj– My brother suggested it to me

If Bengali clausal structure adheres to Zubizarreta's (1998) configuration of the Nuclear Stress Rule (NSR), I predict that participants will place narrow-focus (i.e. new information) subjects in a postverbal position.

3.5.4 Condition D

Condition D presented an object narrow-focus (rheme) information structure context. These tokens counterbalance Condition C, eliciting participants to choose between OVS and OSV order, as in (11).

(11) Context: Sumon– *Is your sister ready yet?*

Shajib– *Almost, almost...*

Sumon– *What is she looking for?*

A. Sumon– toale khuj-che she chul shukono-r jonno
towel look for.3PRS.PROG she hair (the) dry.INF for

B. Sumon– toale she khuj-che chul shukono-r jonno
towel she look for.3PRS.PROG hair (the) dry.INF for

C. A and B both are possible.

Sumon– She's looking for a towel to dry her hair (with).

If Bengali objects in OVS imply scrambling of the object to avoid the rightward projection of information focus as discussed above, the

prediction is that the object will remain *in situ*, therefore favoring a choice of OVS order in these discourse contexts. As the grammatical subjects in this condition represent old information, it was not assumed that SOV would be a preferred word order in the corresponding Task 1 condition. Yet, if object DPs remain in their thematic position, SOV should have been one of the word orders under consideration in this condition. This methodological shortcoming was remedied in the Task 2 follow-up condition, which gathered preference data for all three of the word orders examined in Task 1 for object narrow-focus conditions. I provide the methodological details of this follow-up as well as its results in section 4.3, following the results for Task 2.

3.5.5 Condition E

Condition E items are a counterbalance for Condition B. Within these contexts, the object represents discourse-old information, therefore satisfying the anaphoric requirements of focus, as previously introduced for Condition C in Task 1. The two possible word orders in this item type were OclSV and OclVS, with the discourse-old object, as in (12).

(12) Context: Shobuj– *And where did they leave the rest of the furniture?*

A. Sumon– sofa-ti-i amar bon khuje peyeche dorjar baire
sofa the-cl my sister find get.3.PRF door (the) outside

B. Sumon– sofa-ti-i khuje peyeche amar bon dorjar baire
sofa the-cl find get.3.PRF my sister door (the) outside

C. A and B both are possible.

Sumon – The sofa my sister found (it) on the balcony.

The response options for this token did not include SOV for two reasons: 1) it was not assumed that SOV would be a preferred word order when an object is discourse-old, and 2) I predicted that the focus word orders would receive higher ratings in Task 1.

These questions served not only as distractors, but also to ensure that participants completed the tasks properly. The word order options in conditions A through E did not incur sharp ungrammaticality depending on the choice of clausal word order; however, a choice of Adjective–Noun order, option B in item (13), would. Three different reply contexts were included: two in which only Adjective-Noun order was possible, two in which only Noun-Adjective order was possible, and one in which either Noun-Adjective or Adjective noun order was possible. If a participant chose an ungrammatical reponse, their questionnaire results were highly scrutinized, as it was taken as a potential indication that they did not complete the task(s) in a faithful or mentally-focused manner.

3.6 Task 3: Recorded field interview

Researchers in sociolinguistics have found that obtaining linguistic evidence from minority, non-prestige varieties present unique challenges. Cheshire and Stein (1997) claim that the "fluid" nature of

non-prestige varieties makes grammaticality judgments a difficult task, and that only standardization and the establishment of a grammar make such judgments possible. However, Henry (2005) points out that such 'standard forms' of a variety can also cloud the matter since speakers may consider certain standardized forms or uses to be 'incorrect', and therefore 'ungrammatical'. When such standardized forms are reinforced by an education system, it can lead to otherwise grammatical structures being highly stigmatized. Henry's (2005) examination of Belfast English found that follow-up interviews with participants uncovered subtleties in acceptability judgments from linguistic questionnaires that otherwise would have been overlooked. She suggests, therefore, that linguistic questionnaires alone often cannot provide a complete picture of a speaker's grammar. She also points out that paper-and-pencil questionnaires typically make use of a standard variety of a minority language which is often unfamiliar to older or less literate speakers of non-prestige languages. Such individuals are frequently not receptive to written questionnaires.

The third task consisted of 19 field interviews recorded with an Olympus DS-40 digital voice recorder. Although participants volunteered to be interviewed, a crown flat microphone was used as well to lessen potential recording anxiety. A small subset of the interviewees also participated in Tasks 1 and 2, but since anonymity was preserved, it is unknown which of the interviewees they were. A Bengali-speaking male and a Bengali-speaking female conducted all

interviews within the area of Dhaka city. Bengali-speaking individuals conducted the interviews in order to guarantee comprehension on the part of the interviewer, and to avoid simplification of linguistic structures for the benefit of the Primary Investigator and author of this dissertation. I was present for all of them and familiarized the interviewers with the questions and expectations of the interview beforehand. The interviewees were acquaintances, family or friends of all the interviewers, which greatly assisted in gathering more relaxed, informal data. The interview questions for Task 3 included questions about family, hometown, job-place, opinions about older and younger generations, opinions about the Bengali language, and thoughts about the future of Bengali. Participants are also asked to tell an anecdote from their youth or from their hometown. The older participants were asked about their experiences and views about the readymade garments sector in Bangladesh. Although Task 3 is ostensibly a spontaneous field interview, the interview questions asked of these speakers are quite structured. This task has been designed in this manner to attempt to ensure analogous, and therefore comparable, responses among interviewees. The goal of these short interviews is also to gather a qualitative sample of spontaneous speech with which to compare the results of the quantitative tasks (Tasks 1 and 2) in this experiment.

3.7 Conclusion

The methodology detailed in the above three tasks provide indications of preferred word orders in Bengali in a variety of pragmatic contexts. All

of the tasks detailed above were created with the consultation and advice of native Bengali speakers working in sociolinguistics and Bengali education in order to assure that these tasks would focus more on Bengali speakers' estimations of what is pragmatically acceptable and appropriate, and less on what is lexically appropriate.

Statistical results from Tasks 1 and 2, which appear in the following chapter, indicate the focus preferences of Bengali speakers for these contexts. The spontaneous production data gathered in Task 3 help to shed further light on how common such word orders are in the elicited speech of interviewees. Crucially, a lack of such word orders should not be taken as an indication that such word order(s) are lacking in their grammars. The combination of the indications in these task results provide vital qualitative and quantitative evidence guiding the syntax-information structure analysis that I propose for Bengali in Chapter 5.

Chapter Four

Analysis of Quantitative and Qualitative Measures

In this chapter, I present the results of the quantitative and qualitative tasks described in Chapter 3. In section 4.2, I present the statistical results for each Task 1 condition. In section 4.3, I present the statistical results for each Task 2 discourse condition. In section 4.4, I report on a follow-up quantitative task that I carried out for two conditions whose results in Task 2 did not indicate a clear focus word order preference. I describe the follow-up task and present the statistical results gathered for this task. I report on the focus word orders attested and their accompanying discourse contexts in the recorded field interviews in section 4.5. In section 4.6, I make closing comments on this chapter and the tasks reported on within.

4.1 Introduction

Initial data inspection for both tasks indicated that participant responses were not normally distributed. Rather, they were quite skewed. The five-point rating scale by which the word order options in Task 1 were rated for appropriateness is ordinal and not scalar (i.e. a rating of 4 is not necessarily twice as high as a rating of 2). The choices available in Task 2 were also ordinal, as they only provided a focus word order preference choice of a, b, or c. Therefore, when inter-group comparisons can be made, I have analyzed the data using the Friedman test, which is the non-parametric statistical alternative to either an ANOVA or a two-tailed t-test. In section 4.2 I detail the statistical results for Task 1. In sections 4.2.1 to 4.2.7; I provide descriptive statistics for the word order triads for each condition, and then provide statistical comparisons of the word

order responses for each triad. I summarize these results in 4.2.8. In section 4.2.9 I provide statistical comparisons for Task 1 conditions.

4.2 Task 1

Recall that Task 1 was the Scaled Appropriateness Task in which participants rated continuation/response triads to a variety of information structure contexts on a five-point scale. For each condition in Task 1, I present descriptive statistics as well as a prose description below. I follow the data descriptions with Friedman ranks of means measures to discover statistical differences. Additional comparisons follow when such differences are detected.

4.2.1 Condition A

For thetic contexts, SOV word order received a mean acceptability rating near ceiling (4.91), as in Table 1. In comparison with the mean ratings for OSV (3.14) and OVS (2.63), these statistics suggest that SOV is the preferred word order. The ratings for OSV and OVS display greater individual variation than SOV (minimum 4, maximum 5), as both received ratings between 1 and 5.

Word Order	N	Mean	Std. dev.	Min.	Max.
SOV	100	4.91	0.28762	4	5
OSV	100	3.14	1.08265	1	5
OVS	100	2.63	1.26055	1	5

Table 1. Word order triad ratings for Task 1 Condition A (thetic sentences)

Not surprisingly, the small range of variation in ratings for SOV is reflected in its small standard deviation (0.28762). Despite the wide range of ratings given for OSV and OVS, they do not exhibit very large standard variations (1.08265 and 1.26055, respectively).

WORD ORDERS	N	χ^2	DF	P VALUE
SOVv OSV v OVS	100	145.531	2	< .001
SOV v OSV	100	87.044	1	< .001
SOV v OVS	100	93.000	1	< .001
OVS v OSV	100	7.667	1	.006

Table 2. Friedman statistics for Task 1 Condition A

The p-value of the Friedman non-parametric test ranking the means for all three word orders (< .001) in the first row of Table 2 suggests a statistically significant difference between the distributions of the three scores. When the word orders are compared pair-wise, there are statistically significant differences between SOV and OSV (p< .001), SOV and OVS (p< .001), and OSV and OVS word orders (p=.006).

4.2.2 Condition B

In contexts in which the subject represents old information and discourse subordination contexts as instructed by Lopez (2009), mean ratings for SOV are near maximum (4.96). Mean ratings from for OSV (2.03) and OVS (3.01) are quite lower. As in Condition A, ratings for OSV and OVS for condition B received a wider range of scores (4 and 5,

respectively) than SOV did (range of 2, min. 4 and max. 5). SOV had a much smaller standard deviation (0.19695) than either OSV (0.93695) or OVS (1.23497).

WORD ORDER	N	MEAN	STD. DEV.	MIN.	MAX.
SOV	100	4.96	0.19695	4	5
OSV	100	2.03	0.93695	1	4
OVS	100	3.01	1.23497	1	5

Table 3. Descriptive statistics for Task 1 Condition B (subject old subordination)

When the mean rankings are compared for Condition B, the Friedman p value indicates a statistically significant difference between the distributions of focus word order ratings. When compared pair-wise, there are also statistically significant differences among each of the three focus word orders compared.

WORD ORDERS	N	χ^2	DF	P VALUE
SOVv OSVv OVS	100	159.314	2	< .001
SOV v OSV	100	99.000	1	< .001
SOV v OVS	100	84.045	1	< .001
OVS v OSV	100	27.597	1	< .001

Table 4. Friedman statistics for Condition B

4.2.3 Condition C

For subordination discourse contexts in which the object represents old information, mean ratings for SOV are 4.59. SOV for this discourse context received a wide range of ratings, and a standard deviation of 0.68306. OclSV received a mean rating of 1.87, a range of 5, and a standard deviation of 0.94980. OclVS received a mean rating of 3.70, a range of 5, but a larger standard deviation of 1.19342.

Word Order	N	Mean	Std. dev.	Min.	Max.
SOV	100	4.59	0.68306	2	5
OclSV	100	1.87	0.94980	1	5
OclVS	100	3.70	1.19342	1	5

Table 5. Descriptive statistics for Task 1 Condition C (object old subordination)

The Friedman test statistic for the distributions of the three word orders in Table 6 ($p < .001$) indicates statistically significant differences between the word orders. Pair-wise comparisons also indicate statistically significant differences among each word order, thus indicating that SOV is the preferred word order in subordination discourse contexts in which the object represents old information.

WORD ORDERS	N	χ^2	DF	P VALUE
SOV v OclSV v OclVS	100	136.622	2	< .001
SOV v OclSV	100	91.162	1	< .001
SOV v OclVS	100	21.278	1	< .001
OclSV v OclVS	100	64.205	1	< .001

Table 6. Friedman statistics for Condition C

4.2.4 Condition D

For coordination contexts in which the subject is old information within the discourse context provided, mean ratings for SOV are 4.79, means for OSV are 3.16, and means for OVS are 2.72. As with Condition C, all three word order possibilities received a wide range of ratings, yet despite such variation, all three word order ratings have relatively small standard deviations (0.6559 for SOV, 1.10755 for OSV and 1.13778 for OVS).

WORD ORDER	N	MEAN	STD. DEV.	MIN.	MAX.
SOV	100	4.79	0.6559	1	5
OSV	100	3.16	1.10755	1	5
OVS	100	2.72	1.13778	1	5

Table 7. Descriptive statistics for Task 1 Condition D (subject old coordination)

Friedman test results comparing the three word orders indicate

statistically significant differences between them (p< .001). There are also statistical differences between SOV and OSV (p< .001), SOV and OVS (p< .001), and between OVS and OSV (p< .001). These results suggest that SOV is the preferred word order in contexts in which dislocation is inappropriate and the subject represents old information within the discourse.

WORD ORDERS	N	χ^2	DF	P VALUE
SOVv OSV v OVS	100	130.08	2	< .001
SOV v OSV	100	73.179	1	< .001
SOV v OVS	100	85.172	1	< .001
OVS v OSV	100	10.889	1	.001

Table 8. Friedman statistics for Condition D

4.2.5 Condition E

In coordination contexts in which the (direct) object represents old information within the discourse, SOV word orders received the highest mean rating (4.75), followed by a mean rating of 2.70 for OclVS, and a mean rating of 1.55 for OclSV. For this task, all three word orders received a wide range of ratings, and all have fairly similar standard deviations.

WORD ORDER	N	MEAN	STD. DEV.	MIN.	MAX.
SOV	100	4.75	0.70173	1	5
OclSV	100	1.55	0.8333	1	4
OclVS	100	2.70	1.15907	1	5

Table 9. Descriptive statistics for Task 1 Condition E (object old coordination)

Friedman statistics for the mean rating distributions for the three possible word orders indicate a statistically significant difference between the three word orders ($p < .001$). Pair-wise Friedman comparisons also indicate statistically significant differences between each word order rating pairing. These results suggest that among the word order options given, OclSV is the preferred word order option for discourse condition E.

WORD ORDERS	N	χ^2	DF	P VALUE
SOV v OclSV v OclVS	100	163.107	2	< .001
SVcl v OclSV	100	95.040	1	< .001
SVcl v OclVS	100	86.170	1	< .001
OclSV v OclVS	100	43.556	1	< .001

Table 10. Friedman statistics for Condition E

4.2.6 Condition F

As previously discussed, Condition F sought to test word order preferences for subject narrow focus contexts (i.e. when the subject is

the rheme following López (2009)). For this discourse context, none of the word order options available approached the maximum possible rating, but both SOVX and OVSX word orders received mean ratings around the "rather acceptable" level of four (4.09 and 4.07, respectively). OSVX received a mean rating of 3.66, which also approaches the same level of acceptability. All word order options received the full range of ratings, and have very similar standard deviations.

Word Order	N	Mean	Std. dev.	Min.	Max.
SOVX	100	4.09	1.16424	1	5
OVSX	100	4.07	1.13933	1	5
OSVX	100	3.66	1.08451	1	5

Table 11. Descriptive statistics for Task 1 Condition F (subject narrow focus/rheme)

Friedman test results for this condition indicate the presence of statistically significant differences, but pair-wise comparisons only indicate significant differences between SVX and VXS ($p=.021$), and between VXS and VSX ($p< .001$), but no significant difference between SVX and VSX ($p=.448$), thus complicating the matter of determining a word order preference for this discourse context. Since determining a preference was problematic, I separated the X discourse response tokens according to their constituent identity for post-hoc analysis.

WORD ORDERS	N	χ^2	DF	P VALUE
SOVXvOVSXvOSVX	100	13.019	2	.001
SOVX v OVSX	100	0.576	1	.448
SOVX v OSVX	100	5.313	1	.021
OSVX v OVSX	100	16.000	1	< .001

Table 12. Friedman statistics for Condition F

Unlike the previous discourse conditions, only two of the five discourse context response tokens for Condition F included an argument direct object DP. The other three contexts included an adverbial XP adjunct. I examined these XP adjuncts in an attempt to determine if the ratings for these contexts are similar to the ratings for argument object DPs. For these possible replies, OVSA word orders received the highest mean rating (4.75), OSVA received the second highest mean (3.95), and SOVA received the lowest mean rating (3.583). OVSA displays a smaller range of ratings (range=2) compared to OSVA (range=4) and SOVA (range=5), as well as a smaller standard deviation (0.437) compared to the other word orders (0.852 for OSVA, 1.239 for SOVA)

WORD ORDER	N	MEAN	STD. DEV.	MIN.	MAX.
SOVA	60	3.583	1.239	1	5
OVSA	60	4.75	0.437	4	5
OSVA	60	3.95	0.852	2	5

Table 13. Descriptive statistics for Task 1 Condition F with adjunct XP

Friedman statistics for these word orders indicate a statistically significant difference ($p < .001$), and pair-wise comparisons display statistical differences for OVSA and SOVA ($p < .001$) and OVSA and VAS ($p < .001$), but not between OSVA and OSVA ($p = .238$), thus suggesting that OSVS is preferred to both SOVA and OSVA for these contexts.

WORD ORDERS	N	χ^2	DF	P VALUE
SOVAvOVSA	v60	31.387	2	< .001
SOVA v OVSA	60	17.163	1	< .001
SOVA v OSVA	60	1.391	1	.238
OSVA v OVSA	60	29.432	1	< .001

Table 14. Friedman statistics for Condition F with adjunct XP

SOV received the highest mean rating (4.85), quite near the "totally acceptable" ceiling of five points. OVS received the second highest mean rating (3.225), and OSV received the lowest mean rating (3.05), both near "more or less acceptable" levels. For these tokens, SOV received the lowest range of ratings (range=2) as well as the smallest standard deviation (0.36162). Both OVS and OSV received the full range of possible ratings and had comparatively higher standard deviations (1.25038 and 1.10824, respectively).

Word Order	N	Mean	Std. dev.	Min.	Max.
SOV	40	4.85	0.36162	4	5
OSV	40	3.05	1.10824	1	5
OVS	40	3.225	1.25038	1	5

Table 15. Descriptive statistics for Task 1 Condition F with DP object

Friedman analysis of the mean rating distributions for these three word orders indicates a statistically significant difference between the three word order options provided. Ratings for SOV were significantly higher than either OSV or OVS ($p < .001$ in each case), while the comparison of OSV and OVS shows no significant difference between these word orders. This suggests that SOV is the preferred word order for these discourse contexts.

Word Orders	N	χ^2	DF	P value
SOV v OSV v OVS	40	46.217	2	< .001
SOV v OSV	40	36.000	1	< .001
SOV v OVS	40	22.730	1	< .001
OVS v OSV	40	0.037	1	.847

Table 16. Friedman statistics for Condition F word orders with DP object

The difference between ratings of discourse contexts with adverbial adjuncts and those with argument direct object DPs therefore explains the lack of a clear word order preference in Condition F contexts as a whole.

4.2.7 Condition G

For direct object narrow focus contexts (i.e. rheme in López (2009)), SOV received a mean rating of 4.89, despite receiving ratings ranging from three to five and a rather small standard deviation (0.37322). OVS received a mean rating of 2.75, and OSV a mean of 2.18. Even though both of these word orders showed a wide range of variation in their ratings (range=5 for both), the standard deviation for OSV (0.95748) was smaller than it was for OVS (1.28216), thus displaying a lesser degree of variability in ratings for OSV.

Word Order	N	Mean	Std. dev.	Min.	Max.
SOV	100	4.89	0.37322	3	5
OSV	100	2.18	0.95748	1	5
OVS	100	2.75	1.28216	1	5

Table 17. Descriptive statistics for Task 1 Condition G (object narrow focus/rheme)

Friedman tests show statistically significant differences between the word order ratings for this discourse condition.

WORD ORDERS	N	χ^2	DF	P VALUE
SOV v OSV v OVS	100	144.005	2	< .001
SOV v OSV	100	98.000	1	< .001
SOV v OVS	100	78.681	1	< .001
OVS v OSV	100	8.333	1	.004

Table 18. Friedman statistics for Condition G

When pair-wise comparisons were analyzed with the Friedman test, all p-values were lower than the statistically significant level of (.05), thus indicating a statistical difference between the distributions of each pair-wise word order comparison in Table 18. The statistical results suggest that SOV is the preferred word order for contexts in which the direct object is narrowly focused.

4.2.8 Summary of Task 1 discourse conditions

In Table 19, I provide a summary of the preferred word orders for the discourse contexts provided in Task 1 based on participant ratings of possible word order triads on a five-point scale as discussed in Chapter 3. For the majority of the information structure contexts provided SOV is the preferred word order. In Condition E, the object appears as a direct object clitic since it is old information within the given discourse context. The presence of an adverbial adjunct in Condition F appears to affect the preferred position of the subject as OVSA is the preferred word order in these environments.

CONDITION	INFORMATION STRUCTURE CONTEXT	PREFERRED WORD ORDER
A	thetic, "out of the blue"	SOV
B	subject as old information subordination	SOV
C	object as old information subordination	SOV
D	subject as old information coordination	SOV
E	object as old information coordination	OclSV
F	subject narrow focus with argument direct object XP	SOV, OVSA
G	object narrow focus	SOV

Table 19. Summary of Task 1 word order preferences by discourse condition

I discuss the implications of these preferences further in the syntactic analysis. It has already been proved that such word order preferences are not at all stylistic. Therefore, I am able to reject the null hypothesis. Henceforth I shall try to establish the alternative hypothesis in a plausible manner. I propose for Bengali clausal structure in Chapter 5. In section 4.3., I present the results of the word order preference task, which was the second questionnaire tasks presented in Chapter 3.

4.2.9 Task 1 by sociolinguistic factors

As discussed in Chapter 3, due to variation in participant responses to language history and use questions in the linguistic history questionnaire, I separated them into four sociolinguistic groups. The groups are: +urban/+educated, +urban/-educated, -urban/+educated and –urban/-educated. However, 1 failed to collect data from –urban/-educated people.

Therefore, I have only three groups who have actively taken part in my experiment. I did this out of curiosity as a *post hoc* measure in order to determine if any differences would surface between those who speak Bengali to a greater or lesser degree. All of the participants reported exposure to and use of Bengali in their everyday lives. In an attempt to determine the presence of statistical differences between the three groups' Task 1 results, I compared each language groups' ratings for each possible word order for each condition in Task 1 using the Friedman Test.

Table 20 shows the descriptive statistics for the Condition A SOV word order option in the first three rows, OSV in the second three rows, and OVS in the final three rows. The Friedman test statistics obtained did not indicate a significant difference based on sociolinguistic factors for SOV (p=.549), or for OVS (p=.489), but did indicate a difference approaching significance for OSV (p=.07).

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.97	0.183	4	5
	+urban/-educated	30	4.90	0.305	4	5
	+urban/+educated	30	4.93	0.254	4	5
OSV	-urban/+educated	30	3.33	1.241	1	5
	+urban/-educated	30	2.90	1.094	1	5
	+urban/+educated	30	3.17	1.020	1	4
OVS	-urban/+educated	30	2.83	1.392	1	5
	+urban/-educated	30	2.50	1.253	1	5
	+urban/+educated	30	2.50	1.167	1	5

Table 20. Descriptive statistics for Condition A by sociolinguistic factors

Table 21 shows the descriptive statistics for Condition B word orders. For SOV, the Friedman test indicated no significant difference according to sociolinguistic factors for SOV (p=.223), but did indicate the presence of statistical differences for OSV (p=.01) and for OVS (p=.006). Follow-up pair-wise Friedman measures for OSV indicate a difference between -urban/+educated ratings and +urban/+educated ratings (p=.012), and between +urban/-educated ratings and +urban/+educated ratings (p=.003). No such difference was found between the –urban/+educated speaker group and the +urban/-educated speaker group for OSV (p=.827).

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.97	0.183	4	5
	+urban/-educated	30	5.00	0.000	5	5
	+urban/+educated	30	4.93	0.254	4	5
OSV	-urban/+educated	30	1.73	0.640	1	3
	+urban/-educated	30	1.77	0.935	1	4
	+urban/+educated	30	2.37	1.033	1	4
OVS	-urban/+educated	30	3.27	1.172	1	5
	+urban/-educated	30	2.60	1.329	1	5
	+urban/+educated	30	3.20	1.215	1	5

Table 21. Descriptive statistics for Condition B by sociolinguistic factors

For OVS, the Friedman statistics indicate a statistical difference between ratings for -urban/+educated speakers and +urban/-educated speakers (p=.014), and between +urban/-educated speakers and +urban/+educated

speakers (p=.005). No significant difference was found between -urban/+educated speakers and +urban/+educated speakers (p=.819).

For Condition C discourse contexts are described in the statistics in Table 22. Friedman test measures do not indicate the presence of statistical differences by sociolinguistic factors for SOV (p=.796), OclSV (p=.406), or for OclVS (p=.783). The table is stated bellow:

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.67	0.547	3	5
	+urban/-educated	30	4.60	0.622	3	5
	+urban/+educated	30	4.70	0.596	3	5
OclSV	-urban/+educated	30	1.93	0.944	1	4
	+urban/-educated	30	1.63	0.890	1	5
	+urban/+educated	30	1.77	0.898	1	4
OclVS	-urban/+educated	30	3.67	1.269	1	5
	+urban/-educated	30	3.80	1.243	1	5
	+urban/+educated	30	3.60	1.163	1	5

Table 22. Descriptive statistics for Condition C by sociolinguistic factors

Table 23 shows the descriptive statistics for Condition D information structure contexts by sociolinguistic factors and word order. The Friedman test statistic does not indicate a statistical difference according to sociolinguistic factors for SOV (p=.529). For OSV, however, there is a difference. Pair-wise Friedman measures uncover a statistical difference between ratings by +urban/-educated speakers and -urban/+educated speakers (p=.023), but not between -urban/+educated and

+urban/+educated speakers (p=.108), nor between +urban/-educated and +urban/+educated speakers (p=.102). The Friedman test also indicates a statistical difference by sociolinguistic factors for OVS (p=.039). Pair-wise comparisons show a statistical difference between +urban/-educated and -urban/+educated speakers (p=.022). The comparisons also show the minimum statistical difference at a 95% confidence interval between -urban/+educated ratings and +urban/+educated ratings (p=.05). No statistical difference was found between +urban/-educated speakers and +urban/+educated speakers (p=.835).

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.73	0.785	1	5
	+urban/-educated	30	4.77	0.774	1	5
	+urban/+educated	30	4.93	0.254	4	5
OSV	-urban/+educated	30	3.63	1.066	1	5
	+urban/-educated	30	2.67	1.124	1	5
	+urban/+educated	30	2.97	0.999	1	4
OVS	-urban/+educated	30	2.93	1.337	1	5
	+urban/-educated	30	2.60	1.329	1	5
	+urban/+educated	30	3.20	1.215	1	5

Table 23. Descriptive statistics for Condition D by sociolinguistic factors

The descriptive statistics for the information structure contexts in Condition E appear in Table 24 separated by sociolinguistic factors and by word order response options provided in Task 1. The Friedman test

does not indicate any presence of statistical differences based on self-reported sociolinguistic factors for SOV (p=.459), OclSV (p=.112), or for OclVS (p=.519).

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.73	0.640	3	5
	+urban/-educated	30	4.93	0.253	4	5
	+urban/+educated	30	4.77	0.817	1	5
OclSV	-urban/+educated	30	1.67	0.922	1	4
	+urban/-educated	30	1.37	0.669	1	3
	+urban/+educated	30	1.53	0.860	1	4
OclVS	-urban/+educated	30	2.80	1.297	1	5
	+urban/-educated	30	2.53	1.074	1	4
	+urban/+educated	30	2.77	1.104	1	5

Table 24. Descriptive statistics for Condition E by sociolinguistic factors

Recall that the response triads for Condition F information structure contexts varied due to constituent type, and for each subject narrow focus data set there were only two context tokens with an argument direct object. This would net a total of only twelve response tokens for each sociolinguistic factors group. Given that these are not sufficiently numerous to warrant statistical comparison, I do not consider Condition F results for comparison by sociolinguistic factors.

The descriptive statistics for object narrow focus information scenarios

in Condition G appear in Table 25. A Friedman comparison of mean ranks do not indicate statistical differences by sociolinguistic factors for SOV (p=.646) or for OVS (p=.322). For OSV however, the Friedman test statistic does indicate the presence of a statistical difference by sociolinguistic factors (p=.006). Pair-wise comparisons indicate a statistical difference between +urban/-educated speakers' and -urban/+educated speakers' ratings (p=.014), as well as between -urban/+educated speakers' and the +urban/+educated speakers' ratings (p=.005), but not between +urban/-educated speakers' ratings and those of +urban/+educated speakers (p=.322).

Word Order	Sociolinguistic Factors	N	Mean	Std. dev.	Min.	Max.
SOV	-urban/+educated	30	4.90	0.403	3	5
	+urban/-educated	30	4.93	0.365	3	5
	+urban/+educated	30	4.90	0.305	4	5
OSV	-urban/+educated	30	2.30	0.988	1	5
	+urban/-educated	30	1.67	0.802	1	4
	+urban/+educated	30	2.33	0.884	1	4
OVS	-urban/+educated	30	2.77	1.305	1	5
	+urban/-educated	30	2.60	1.404	1	5
	+urban/+educated	30	2.83	1.147	1	5

Table 25. Descriptive statistics for Condition G by sociolinguistic factors

In summary, significant differences were found for five different word orders in three different information structure conditions. These were conditions B, D, and G, which share a subject as old information. The

two word orders that display significant statistical differences are OSV and OVS which, for their respective discourse conditions, were not the most highly rated word order options in their triad. The +urban/-educated and +urban/+educated groups both rated OVS higher than their -urban/+educated counterparts for Conditions B and D, but only for Condition B was OVS rated higher than OSV. OSV word orders were not rated as uniformly with respect to sociolinguistic factors. Both +urban/-educated and +urban/+educated groups rated OSV significantly higher than the -urban/+educated group for Condition G. For Condition D, the +urban/-educated group rated OSV significantly higher than the -urban/+educated group, but not significantly higher than the +urban/+educated group. OSV received the lowest mean ratings for both Condition B and G triads, but for Condition B, the +urban/+educated group rated OSV significantly higher than either the +urban/-educated or -urban/+educated groups.

Although +urban/-educated and +urban/+educated speakers rated OSV and OVS word orders more highly than -urban/+educated speakers for information structure contexts with discourse-old subjects, there is a lurking variable that obscures any sort of sociolinguistic factors based conclusions that might otherwise be suggested: age. However, it is worth to mention here that age balance was not possible to maintain in this research. I just put emphasis on the selected age groups. Therefore, age could not be possible to be considered as an analyzing factor from sociolinguistic angle.

As discussed before, since the syntax-information structure interface involves an interface between syntax and the discourse, I should expect to see greater signs of instability and optionality in heritage bilinguals and second language learners. Such instability should manifest itself in this task as a wide(r) range of ratings for certain word orders depending on the discourse-pragmatic context. What I have seen in this section, however, is that the variation between groups is nearly identical – appropriate rating ranges and standard deviations differ only slightly from one sociolinguistic group to another. When there are differences, they are not unidirectional. At the level of individual variation, there are individuals in all three groups that gave the full range of ratings to certain word orders – even for preferred word orders. If there is residual interface optionality and instability in these groups then, the data in this chapter suggest that this comes as an all-or-nothing proposition: either all of them are showing signs of instability at this particular interface or none of them are. This is a difficult issue to address with the data gathered and presented here due to lack of a truly monolingual control group with which to make comparisons and draw conclusions. I return to this issue later in the chapter.

4.2.10 Task 1 by gender

In this section, I report on statistical differences detected according to gender. As there were very few statistical differences based on gender, I only report on these differences in this section. Note however, that in

many of these cases, the ratings in question are very low, thus indicating marginal acceptability. Although little, if anything, may be concluded based on these ratings, I report on these statistical differences for the sake of completeness.

The descriptive statistics for OSV word order in Condition D, coordination contexts in which subjects were discourse-old, appear in Table 26.

GENDER	N	MEAN	STD. DEV.	MIN.	MAX.
female	50	3.46	1.054	1	5
male	50	2.86	1.088	1	5

Table 26. Task 1, Condition D. Descriptive statistics for OSV word order by gender

Statistical differences by gender were found for OSV word orders in reply to Condition D (p=.002, χ^2=9.256). It is interesting that these word orders received the full range of ratings, and that they were rated as more or less acceptable by the participants. Statistical differences by gender are also present for OSV word orders for Condition G (p=.005, χ^2=7.811). Although this condition was an object narrow-focus (rheme) context, it also included a discourse old-subject DP, as previously discussed. The descriptive statistics for this condition appear in Table 27.

GENDER	N	MEAN	STD. DEV.	MIN.	MAX.
female	50	2.42	1.032	1	5
male	50	1.94	0.818	1	4

Table 27. Task 1, Condition G. Descriptive statistics for OSV word order by gender

Female participants gave higher maximum ratings for OSV than their male counterparts did, which led to a higher standard deviation for female ratings. However, the statistics above indicate that females tend to rate OSV word orders higher than males when a discourse-old subject is present. Despite the statistical difference, the low mean ratings for these word orders do not provide indication of any sort of meaningful conclusion.

Females rated OclSV word orders significantly higher than males for subordination discourse contexts with a discourse-old object (p=.005, χ^2=7.811). The descriptive statistics for Condition C appear in Table 28 below.

GENDER	N	MEAN	STD. DEV.	MIN.	MAX.
female	50	2.14	0.948	1	4
male	50	1.60	0.881	1	5

Table 28. Task 1, Condition C. Descriptive statistics for OclSV by gender

Although males assigned higher maximum ratings for OSV in this condition, their standard variation shows less variation. Even though female participants rated this word order higher overall than their male counterparts did, the low overall ratings of these word orders preclude any sort of meaningful conclusion.

The ratings of the discourse context counterbalance for Condition C displayed results approaching statistical significance (p=.072, χ^2=3.240). Condition E also involved a discourse-old object, but in a coordination

(i.e. continuation) context. The descriptive statistics for this condition appear in Table 29.

GENDER	N	MEAN	STD. DEV.	MIN.	MAX.
female	50	1.70	0.909	1	4
male	50	1.40	0.728	1	4

Table 29. Task 1, Condition E. Descriptive statistics for OclSV by gender

While it is interesting that females show a tendency to rate these word orders higher than males do, the mean ratings for this word order were still quite low on the five-point appropriateness scale, somewhere between marginally acceptable (2) and not acceptable (1). Therefore, no conclusion can or should be made based on these results.

4.3 Task 2

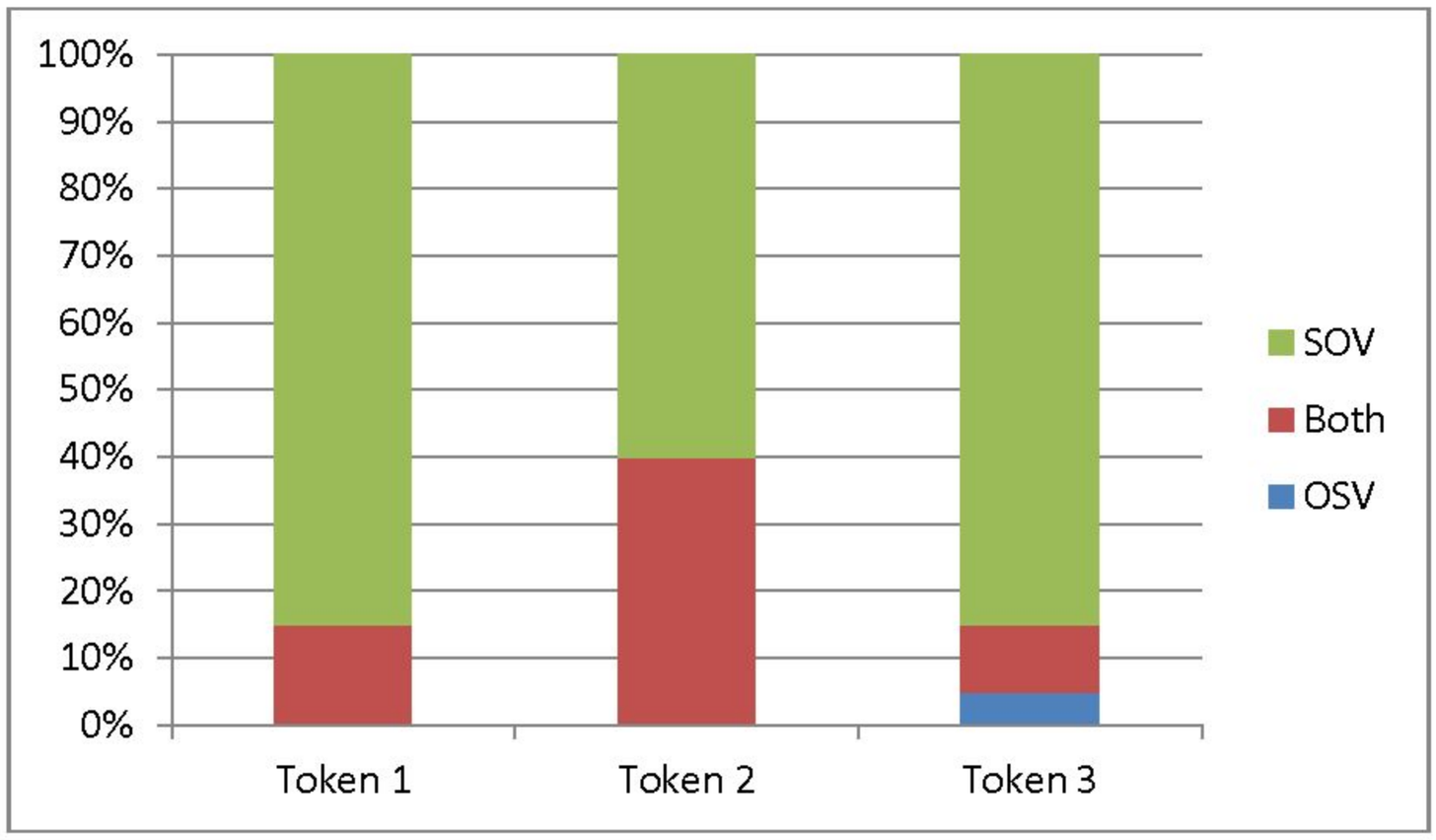
Recall that Task 2 was the Word Order Collocation Task, in which participants responded to an information structure context with a choice of clausal word order options, with the additional option for no preference. As discussed in Chapter 3, in Task 2

I sought to gather further data on the word order preferences indicated in the first task, the scaled pragmatic appropriateness task. In order to detect tendencies in the data from token to token, in initial data analysis, the two word order options presented were assigned a value of either 1

or 3, while the option for both word orders was assigned a value of 2. By this method, mean data preference values gravitating towards one extreme or another would appear at either vertical extreme of the figure, while a preference of "both" would gravitate towards the middle of the figure. Following initial analysis, outlying or unexpected results were examined in further detail. For each discourse context, I also report mean preference percentages for the condition as a whole.

4.3.1 Condition A

Condition A in Task 2 was a follow-up for Condition A in Task 1. Recall that there were three tokens for each information structure preference context in this task, and that the word orders tested in this condition were SOV and OSV. I report on the mean preference percentages for each questionnaire token for this condition in Figure 2 (following page). The results for Condition A indicate a strong preference for SOV order in thetic sentences, thus echoing the indications from Task 1 above. For this condition overall, SOV was preferred in 83% of the ratings, OSV was preferred in 1% of the ratings, and both word orders was chosen in 16% of the replies.

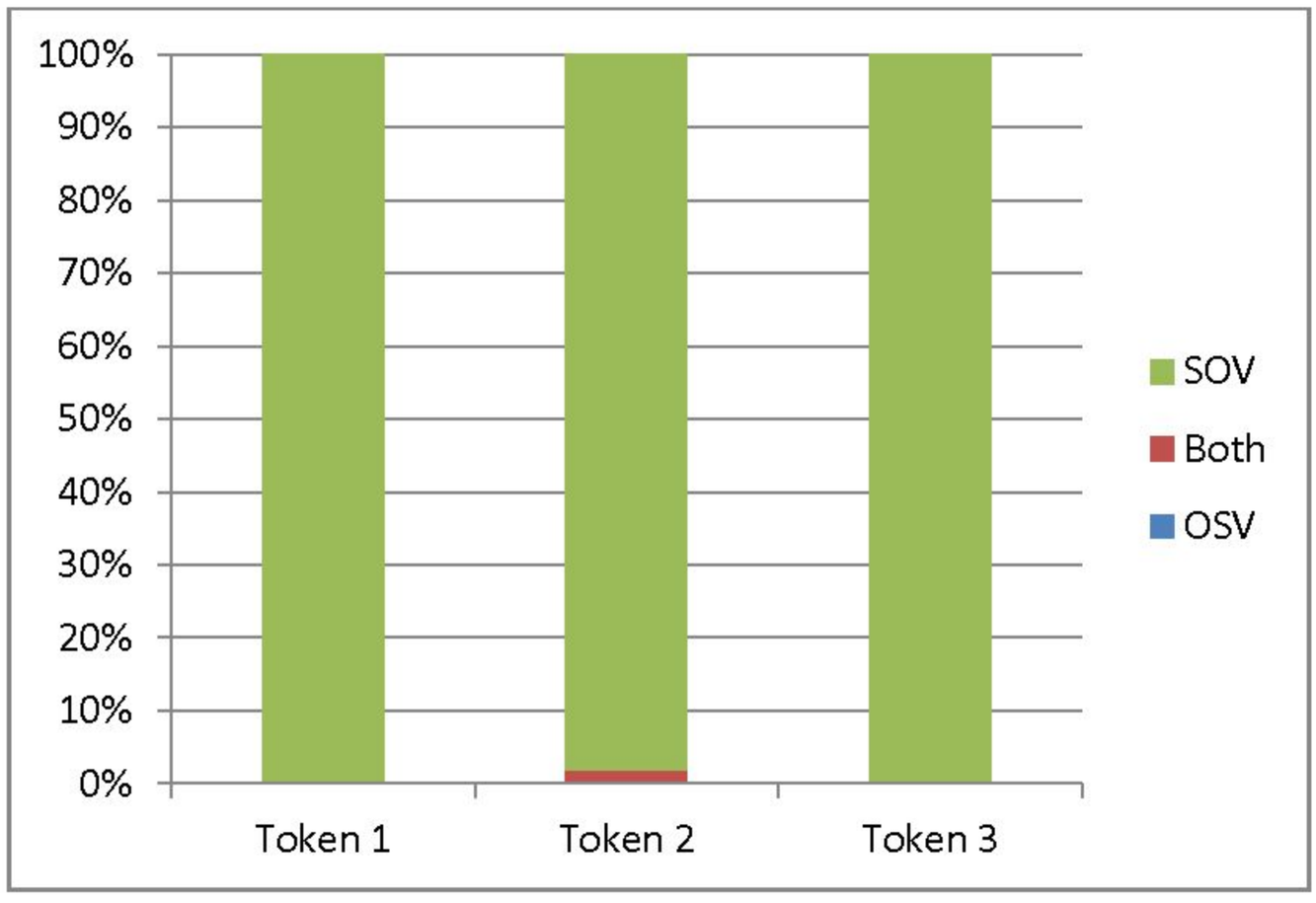


Condition A (thetic) mean ratings by token

Figure 2. Condition A mean word order preference percentages by token

4.3.2 Condition B

Condition B was a follow-up task for Condition B in Task 1. Although this task did not seek to test clitic position in general, all of the discourse contexts provided were subordination contexts. These information structure contexts involved a subject DP that was topical, or discourse-old, and tested the preference of SOV against OSV. The mean preference percentages for each token in Figure 3 show a strong preference for SOV order in this condition. For the condition as a whole, participants chose SOV in 98% of the contexts provided, OSV in 0%, and both word orders in 2% of the tokens. These results confirm the results of Condition B in Task 1.

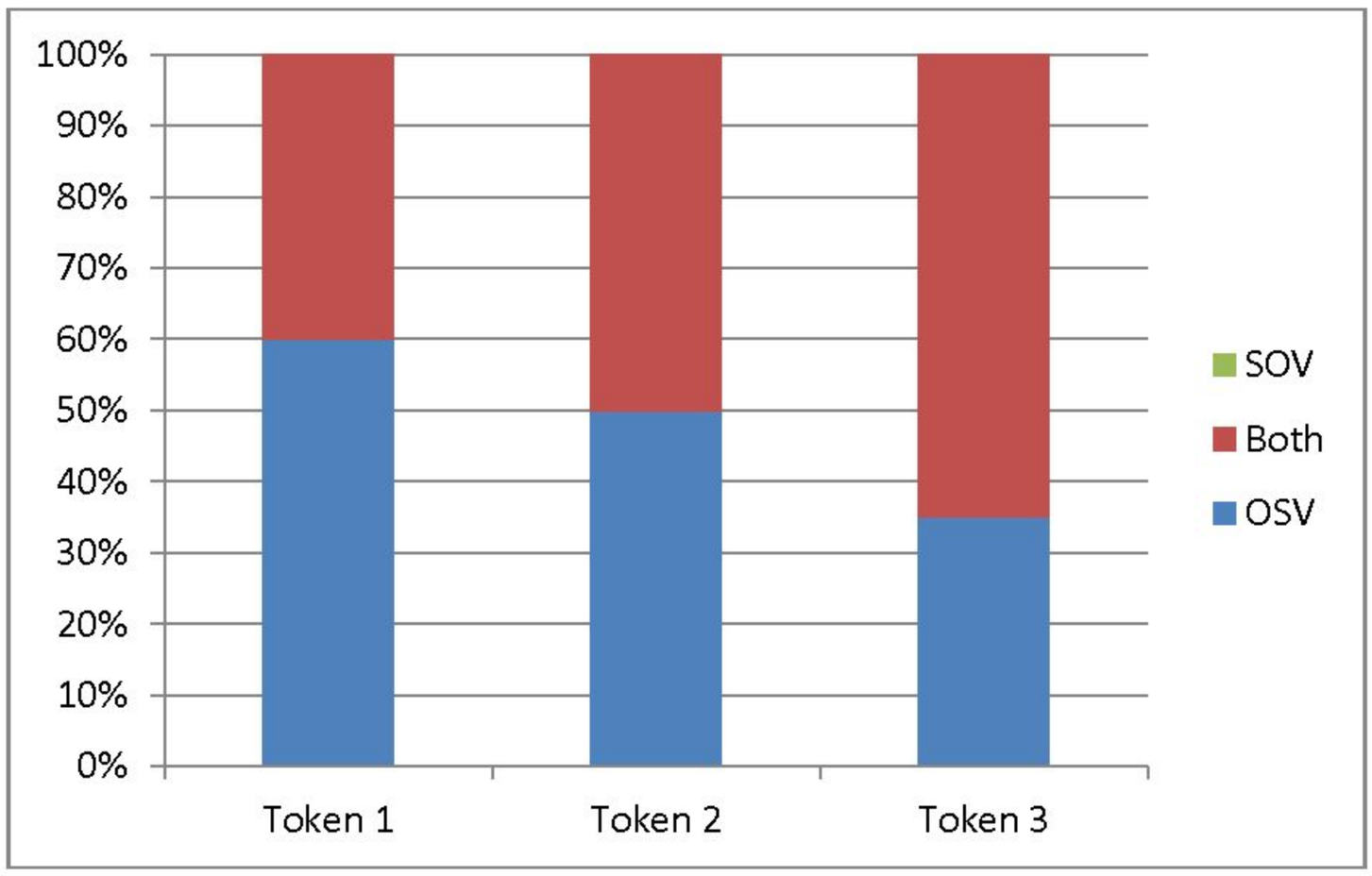


Condition B (topical subject) mean ratings by token

Figure 3. Condition B mean word order preference percentages by token

4.3.3 Condition C

Condition C sought to test word order preference in subject narrow focus (rheme) information structure contexts. This condition served as a follow-up to Task 1, Condition F and compared the relative preference of S(O)V and (O)SV word orders.



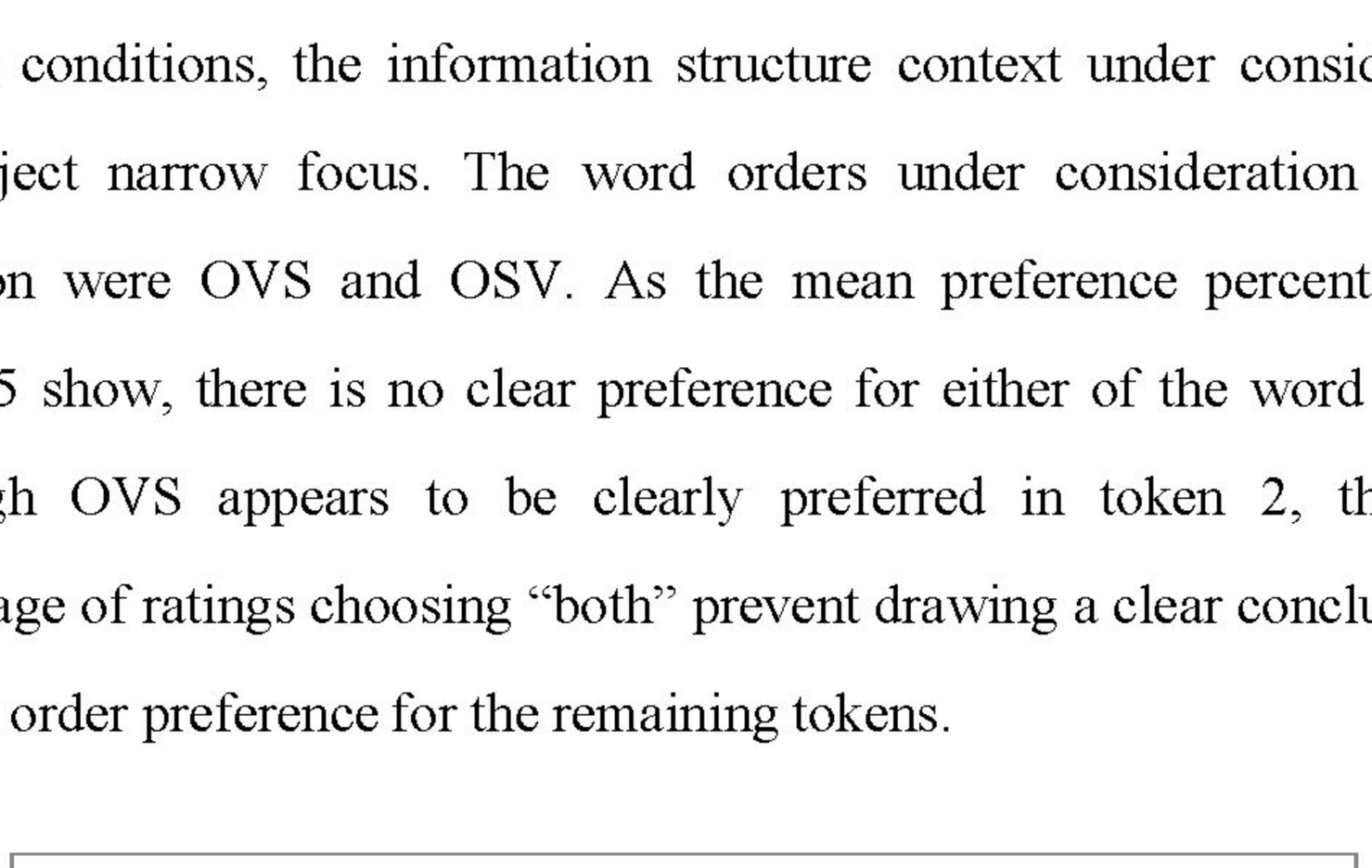
Condition C (subject rheme) mean ratings by token

Figure 4. Condition C mean word order preference percentages by token

Recall that the results for this condition in Task 1 indicated differing preferences depending on whether an argument direct object DP or an adverbial adjunct was present in the possible responses. For Condition C in Task 2, all of the context tokens involved a direct object clitic, and of these, only one token included an adverbial adjunct (C2).

4.3.4 Condition D

Condition D in Task 2 was a follow-up measure for Condition G in Task 1. In both conditions, the information structure context under consideration was object narrow focus. The word orders under consideration in this condition were OVS and OSV. As the mean preference percentages in Figure 5 show, there is no clear preference for either of the word orders. Although OVS appears to be clearly preferred in token 2, the high percentage of ratings choosing "both" prevent drawing a clear conclusion as to word order preference for the remaining tokens.



Condition D (object rheme) mean ratings by token

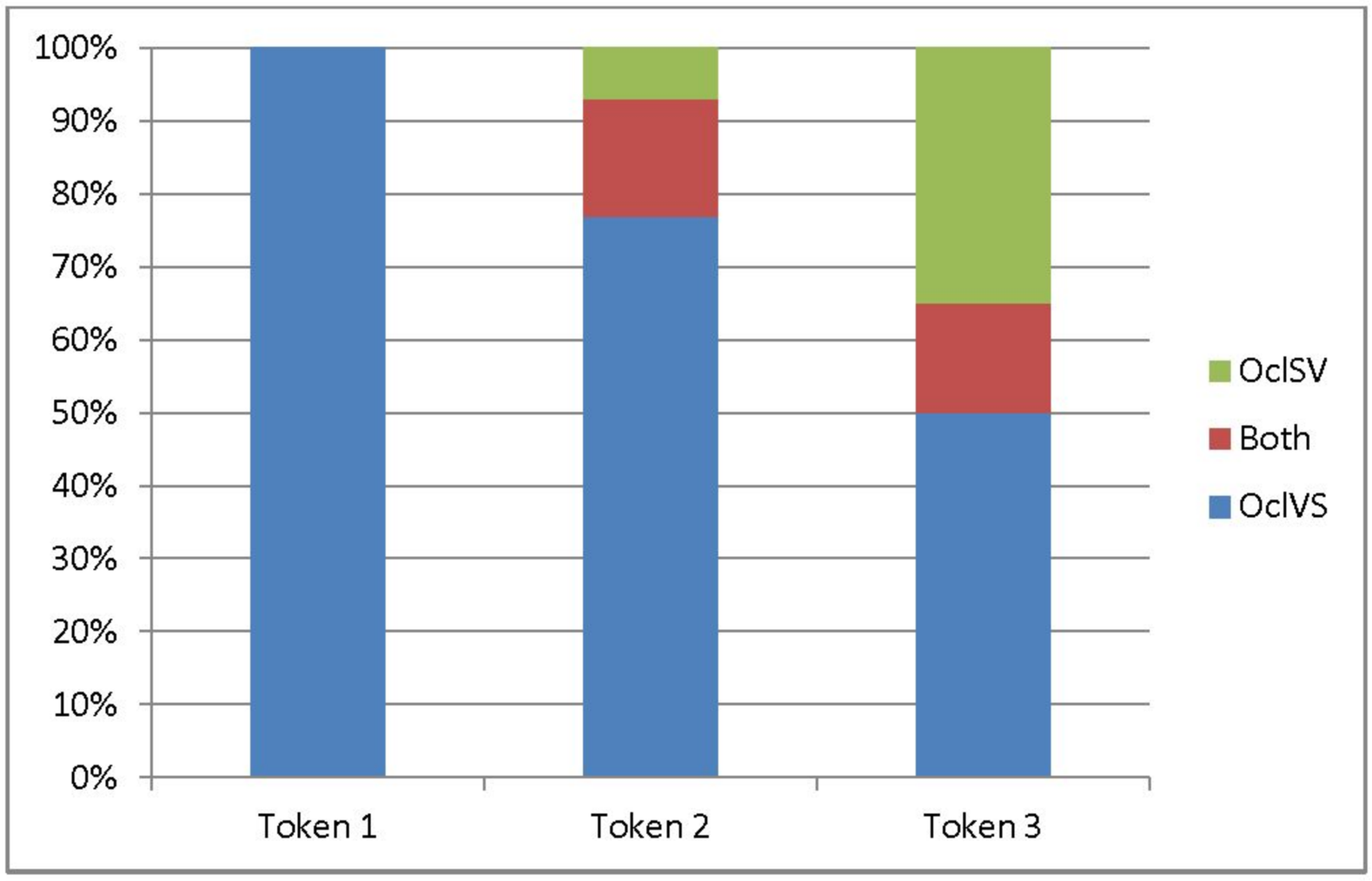
Figure 5. Condition D mean word order preference percentages by token

The overall percentage ratings for this condition suggest a preference for OVS, as this word order was chosen in 46% of replies in comparison with 27% of replies in favor of OSV. As with Condition C in the previous section, a large percentage of participants (27%) chose "both" for their preference. A complicating factor for declaring a preference based on these data lies in the fact that this large of a percentage could potentially sway a preference for either possible word order. Recall that in Task 1, Condition G, SOV was the preferred word order, not OSV or OVS (see Table 17), and that the mean rating for SOV (4.89) was significantly higher than that of OVS, the next highest rated word order (2.75). Therefore, in retrospect, the competing word orders in the current task and condition should have been SOV and OVS. Given the difficulty involved in determining a preference for object narrow focus based on the results gathered, I conducted a supplemental follow-up task for this condition as well. I describe this task in section 4.4.

4.3.5 Condition E

Condition E served as a follow-up to Condition C in Task 1. Recall that this task provided subordination discourse contexts involving an object DP that was topical, or discourse-old. This condition examined a preference between OclVS and OclSV word orders. With the exception of token E3 in Figure 6, most of the data suggests a preference for OclVS word order. The overall percentages for this condition suggest the same preference, as OclVS was preferred in 77% of the replies, OclSV

was preferred in 7% of the replies, and "both" was chosen in 16% of the replies. Removing token E3 from consideration results in a complete disappearance of preferences for OclSV word orders (84% OclVS, 16% "both").



Condition E (topical object) mean ratings by token

Figure 6. Condition E mean word order preference percentages by token

It merits mentioning that *post hoc* examination of token E3 (2 below) does not suggest any significant variation in information structure or constituent structure from the other Condition E tokens.

(1) Context: Shobuj– *And where did they leave the rest of the furniture?*

A. Sumon– sofa-ti-i amar bon khuje peyeche dorjar baire
sofa the-cl my sister find get.3.PRF door (the) outside
Sumon – (Sigh) The sofa my daughter found (it) on the balcony.

In the Task 1 results, mean ratings for SOV were significantly higher (4.59) than either OclSV (1.87) or OclVS (3.70). Although SOV was the preferred word order, it was left out of this condition in the early stages of design so that the verb and its enclitic direct object pronoun would not have to be modified. In the original task design, I had envisioned participants filling in one of two possible blank spaces within the provided continuation of the discourse context, in the spirit of a cloze test. For this particular condition, this would have involved the participant placing the subject to the left or the right of the verb. In the end, the limitations of the online survey system precluded such a response design, so full sentences were provided as response options. However, the provided reply context remained unchanged. I recognize that the design chosen limits the type of conclusions that I can make for this particular discourse context in comparison with the other conditions, but this should not diminish the clear preference for SOV as indicated in Task 1 for this information structure context.

4.3.6 Task 2 by sociolinguistic factors

As I conducted Friedman Tests for Task 1 to determine the presence of statistical differences between the three sociolinguistic factors dependent groups, I carried out the same comparisons for these groups' Task 2 results as well. Recall that the determination of sociolinguistic factors was made *post hoc* based on participant replies to questions on the linguistic history questionnaire. For these particular Friedman measures, I compare the overall ratings by the group for each condition. In order to

maintain consistency with the graphical analyses of each condition in the figures above, I have maintained the numbers assigned to each word order possibility as previously described. As with Task 1, the size of the groups was unequal, consisting of +urban/-educated individuals, -urban/+educated individuals, and +urban/+educated individuals. This examination determined two individuals whose data exhibited outlier characteristics as compared to the rest of the sociolinguistic groups.

The descriptive statistics for Condition A appear in Table 30. Each mean clusters toward 1, and a preference for SOV word order. The Friedman statistic for the three-way comparison of these groups does not indicate any statistical significance between the distributions of the mean scores (p=.704).

SOCIOLINGUISTIC FACTORS	N	MEAN	STD. DEV.
+urban/-educated	30	1.20	0.484
+urban/+educated	30	1.10	0.305
-urban/+educated	30	1.17	0.379

Table 30. Descriptive statistics for Condition A by dominant language

The Condition B means in Table 31 show a greater preference for SOV word order, and a lesser degree of variation, as seen in the standard

deviations. The Friedman statistic for this condition (p=.607) does not indicate the presence of any statistical differences between the sociolinguistic groups' preferences.

SOCIOLINGUISTIC FACTORS	N	MEAN	STD. DEV.
+urban/-educated	30	1.00	0.000
+urban/+educated	30	1.03	0.183
-urban/+educated	30	1.03	0.183

Table 31. Descriptive statistics for Condition B by sociolinguistic factors

The means for Condition C in Table 32 cluster toward 2, indicating a preference for both SOV and OVS for each sociolinguistic group. There is also a very larger degree of variation in the groups' ratings, as seen in the standard deviations for each group. The Friedman Test measure indicates a lack of statistical significance in the differences between the distributions of the mean ratings of the groups for this condition (p=.167).

SOCIOLINGUISTIC FACTORS	N	MEAN	STD. DEV.
+urban/-educated	30	2.00	0.643
+urban/+educated	30	1.80	0.761
-urban/+educated	30	1.77	0.728

Table 32. Descriptive statistics for Condition C by dominant language

The statistics for Condition D shows (not displayed in a table) mean scores slightly higher than 2, suggesting a mild preference for the word order option "+urban/+educated" in all of the groups, but more so for the +urban/-educated group. As with the results for Condition C, the standard deviations for each group indicate a fairly large degree of variation for this condition as well, hardly surprising when one considers that neither of the overall percentages for these two conditions pointed toward a clear word order preference.

The average means for Condition E in Table 33 show a rather strong preference for OclVS for each of the groups. The Friedman statistic for this condition (p=.622) does not indicate the presence of any statistical differences between the mean rating distributions for the three sociolinguistic factorsgroups.

SOCIOLINGUISTIC FACTORS	N	MEAN	STD. DEV.
+urban/-educated	30	2.70	0.596
+urban/+educated	30	2.73	0.583
-urban/+educated	30	2.63	0.669

Table 33. Descriptive statistics for Condition E by dominant language

Overall, none of the Friedman measures conducted for the conditions in Task 2 indicate significant differences based on the sociolinguistic

factors of the participants.

4.3.7 Task 2 statistical results by gender

When subjected to Friedman analysis for statistical differences, only one discourse condition indicated the presence of statistical significance among the numbers assigned to the word order preference options. This was Condition D, which was the object narrow-focus (rheme) information structure context. I provide the descriptive statistics for this condition in Table 34 below.

GENDER	N	MEAN	STD. DEV.	MIN.	MAX.
female	50	2.00	0.857	1	3
male	50	2.38	0.780	1	3

Table 34. Task 2, Condition D. Descriptive statistics for object narrow-focus (OSV vs. OVS)

Although a statistical difference was present (p=.012, χ^2=6.259), these results must be taken with caution. Recall that, as discussed in section 4.3.4, SOV was not a possible word order option for this particular preference condition. Also, as the means gravitate very near to two, the number that was assigned to "both are acceptable", little can be concluded with respect to word order preference by gender for the statistical difference present for this particular discourse condition.

4.3.8 Task 2 Summary

The word order preferences for Task 2 as a whole are reported in the first two columns of Table 35. In the third column, I show the corresponding word order condition from Task 1. In the fourth column, I show whether the results from Task 2 confirm the indications from Task 1. The word order preferences for conditions A and B in Task 1 are confirmed in Task 2, i.e. they show a preference for S(O)V word order.

TASK 2 CONDITION	DISCOURSE CONTEXT	PREFERRED WORD ORDER	CORRESPONDING TASK 1 CONDITION	CONFIRMATION OF TASK 1 RESULTS
A	thetic	SOV	A	Yes
B	subject old	OSV	B	Yes
C	subject focus	"both"	F	unknown
D	object focus	OVS	G	unknown
E	object old	OclVS	C	Yes?

Table 35. Summary of Task 2 conditions

As previously mentioned, the results of conditions C and D in Task 1 could not be confirmed in Task 2, and for that reason, a further follow-up task was conducted. I report on the results of this task in section 4.3. The results of Condition E confirm the findings of Condition C in Task 1 in showing a preference for OclVS over OclSV, but crucially did not include SOV, the preferred word order in Task 1.

4.4 Follow-up task for Task 2 (Task X)

A total of 54 subjects initially participated in the follow-up for Task 2, which attempted to clarify the results for subject and object narrow-focus information structure contexts. There were four discourse conditions. The format of the Task X was identical to the format of Task 2, but with a couple of exceptions. First, the option of "both word orders" was removed in order to arrive at clearer conclusions, thus leaving only two word order response options. Secondly, the close format of the prompts used in Task 2 was completely eliminated in favor of sentence-length replies to the narrow-focus questions in the priming contexts. Among the 54 subjects who participated in the questionnaire task, eight did not complete the task, and thus were removed from consideration. As the goal of this dissertation is to describe the word order preferences of native and habitual Bengali speakers, the results of 14 completed questionnaires were removed from statistical consideration based on participant responses to a brief revised linguistic history questionnaire that preceded the discourse conditions. The summary of the participants from Task X appear below in Table 36.

Age group	Gender		Total
	Male	Female	
18-30	5	6	11
31-49	7	11	18
50+	1	2	3
Total	13	19	32

Table 36. Bengali participant representation by age and gender in Task X

As very few significant statistical differences were detected among some of the non-preferred word orders as discussed in sections 4.2.10 and 4.3.8, I did not take into account the variable of age in Task X. Rather, this task focused on clarifying the previous results for subject and object narrow-focus discourse contexts, as stated at the beginning of this section. The results for the following four conditions are based on the remaining 32 participants who successfully completed the questionnaire.

Follow-up Condition X1 involved subject narrow focus information structure contexts. For this condition the possible word order response options were SV or VS, nine of which included an enclitic direct object pronoun, while one included a proclitic pronoun due to due to its appearance in a subordinate clause. Participants chose VS word order in 96.88% of the contexts provided and SV word order in 3.12% of the contexts, thus showing a clear preference for VS word order in subject narrow focus information structure situations.

Follow-up conditions X2, X3, and X4 were designed to clarify the results gathered in Tasks 1 and 2 for object narrow focus discourse contexts. Each condition compared two word orders in order to more clearly establish a word order preference for this pragmatic context. The results for these conditions are summarized on the following page in Table 37.

Condition X2	OSV	46.25%
	OVS	53.75%
Condition X3	SOV	83.13%
	OVS	16.17%
Condition X4	SOV	90%
	OSV	10%

Table 37. Summary of object narrow-focus follow-up conditions

In Condition X2, OVS was preferred to OSV only slightly. However, SOV was preferred to OVS in Condition X3, and was also preferred to OSV in Condition X4. Given the difference between the rating percentages of SOV to both OSV and OVS, these ratings strongly suggest a preference for SOV in object narrow focus discourse contexts.

4.5 Task 3

The 19 interviews recorded lasted between 12 minutes 51 seconds and one hour 34 minutes 58 seconds, totaling 14 hours 34 minutes 47 seconds of interviews. It should be pointed out that, in the interview corpus, there are numerous cases of SOV word order. There are also many null subjects and one-word replies, which are more typical in everyday conversation, thus following a Gricean model more faithfully than the quantitative tasks reported on in the other sections of this chapter. Recall, however, that such word order options and preferences had to be constructed in the quantitative tasks in order to be able to reach

conclusions about Bengali focus word order according to discourse context. Although many of the word orders may have appeared to be artificial and forced within their particular information structure context in the quantitative tasks, many of them did surface during the interview sessions. In this section, I report on word orders that departed from SOV, those that involved a post-verbal subject (OVS), and those that contained clitic elements.

For thetic sentences, the only word order found in the recordings was SOV, as indicated by participant preferences and ratings in Tasks 1 and 2. As these types of sentences were extremely numerous, and no other word orders were found for this sort of discourse context, I do not report on these sentences here.

For discourse contexts in which a subject was old information, SOV was preferred in the previous quantitative tasks. Despite such preferences, I found greater clausal word order variety in the Task 3 interviews. In over fourteen and a half hours of interview recordings, I found only four examples of VS word order, which represents an extremely small percentage. All of these appeared in subordinate clauses. The first example (2) followed a discourse in which the interviewee was recounting a childhood experience.

(2) takhon matro baRte shuru koreche dhaka shahor
that time just spread.PROG start have.PRS.3 dhaka city
"That time Dhaka city has just started spreading."

Within this parenthetical, the post-verbal subject serves to clarify the subject of the present-tense verb within the past-tense narration. The interviews also attest VS word order for topic-switch to a previously discourse active subject topic in a subjunctive subordinate clause (3).

(3) Interviewer- office-e bangla baebohar korte paren?
office in Bangla use do.PRS can
Are you able to use Bangla in office?
Participant- office-e ami karo karo shathe bangla boli, kintu
office in I a few with bengali speak but
bakider shathe poristhitir opor nirbhor kore.
rest of all with situation on depend do
Shadharonoto ami bangla boli jakhon keu
normally I bengali speak when someone
amar shathe bangla bole. tai ami tader
me with bengali speak.PRS.3 so I them
onujaii bhasha poriborton kori
according to language switch do.PRS.1

(Participant) – "In the office I speak Bengali with a few people. And with others, it depends on situation. Normally, I speak Bengali when someone speaks Bengali with me. So, I switch language according to them"

Within this discourse, the interviewee is speaking about her university roommates and their linguistic preferences when speaking to one

another. In (3), the subject of the subjunctive subordinate clause is the same subject from the interviewer question.

At a bare minimum, however, OVS word order caused no communication breakdown or problems otherwise in this particular conversation, thus suggesting that OVS forms part of the clausal word order inventory of these speakers.

When direct objects were discourse-old information in the interviews they tended to surface as clitic pronouns and were not frequently accompanied by an overt subject. In this type of sentences a discourse-old direct object appears in a clitic position.

According to Asher, Nicholas & Marie Vieu (2005) this would be considered either *elaboration* or *explanation*, both of which fall under the classification of *subordination*, which for López (2009) is required for clitic to be appropriate. The first-person singular subject was also already discourse-active, and remains clear due to the inflection of the verbal suffix.

Sentences with post-verbal subjects represent new information regarding the house, and within a clause that elaborates on the topical, discourse-old elements, however, although in a clitic position, are not accompanied by a resumptive clitic. There are various other cases in which a post-verbal subject represents new information without elicitation by a

subject-focus context (i.e. rheme).

Indefinite DPs are generally considered to represent new information when they appear. In such a sentence, the only discourse-old information is available, the direct object from the previous sentence, which does not appear as a clitic-doubled DP.

Post-verbal subjects also appear as contrastive or emphatic elements in the interview corpus. While attempts were made to elicit narrow-focus subjects (i.e. rheme) during the interviews, the answers given typically lacked a verb. There was only one case in which a verb accompanied a subject response to such a question. While it was not very common in this corpus, clearly it is an attested word order. Despite its low rate of appearance, it should be pointed out that such field interviews are not typically the most natural of venues for questions of this sort, especially considering that object narrow-focus replies including a verb are non-existent within this interview corpus. This surely does not indicate that such contexts are non-existent, but rather that replies that would help to shed light on clausal word order were in short supply.

To summarize this section, SOV word orders were well attested for thetic sentences as well as numerous other discourse contexts. In fact, SOV was the most common word order in this corpus. As exemplified in the data collected in this research, a discourse-old subject may appear in either OSV, or OVS word orders. It may also appear pre-verbally in a

dislocated, left-peripheral position. Discourse-old direct objects appeared in a clitic position. While not entirely common within the corpus, clearly these structures also form part of these speakers' grammars. Subjects that represent new information may appear post-verbally and in a variety of ways. Narrow-focus (rheme) subjects also appear post-verbally. Various examples of canonical-divergent word order share is a post-verbal subject DP, and a left periphery that has been activated either by the presence of a clitic element or a post-positional adjunct. In a number of collected samples, post-verbal subjects appear in subordinate clauses, all of which contain a clitic pronoun of some sort. I discuss the importance of the left-periphery and cliticization for the syntactic analysis of Bengali in the following chapter.

4.6 Summary and discussion: quantitative measures

In this section I summarize the results and indications of the above quantitative tasks and discourse conditions. For convenience, these results are also represented in Table 38. For the majority of the information structure contexts provided, the participants in this investigation showed a preference for SOV word order. In four of these discourse contexts, this preference was confirmed by participant results in two separate task conditions.

INFORMATION STRUCTURE CONTEXT	TASK 1 CONDITION	TASK 2 CONDITION	FOLLOW-UP TASK CONDITION	WORD ORDER PREFERENCE
thetic	A	A	N/A	SOV
subject old (subordination)	B	B	N/A	SOV
object old (subordination)	C	E	N/A	SOV
subject old (coordination)	D	N/A	N/A	SOV
object old (coordination)	E	N/A	N/A	OclSV
subject narrow focus	F	C	A	OclVS
object narrow focus	G	D	B, C, D	SOV

Table 38. Summary of quantitative data measures

In the two remaining discourse conditions, SOV was not the preferred word order. In both of these situations, the direct object DP appeared as a clitic pronoun. The direct object had been previously entered into the common ground of the discourse in these contexts and could not be subordinated (i.e. repeated) in each of these conditions without causing an inappropriateness violation. Participant ratings displayed a preference for OclVS word order in only one of these two conditions: the subject narrow focus context. I discuss the syntax-information structure ramifications if this preference in Chapter 5.

4.6.1 A note on clitic response ratings

Coordination contexts involving a discourse-old direct object only

appeared in one task condition (Task 1, Condition E), and merit some discussion. This was one of the two discourse conditions that offered clitic response alternatives. Although SOV was the preferred word order (μ=4.75), I would like to offer comments on participant ratings of the clitic response options for this condition. Participants gave OclVS word order a mean rating of 2.70, which neared on the "more or less acceptable" level defined by a rating of 3 in this task. OclSV word order however, received a significantly lower rating (μ=1.55), which places this mean near a midpoint between the ratings "not acceptable" = (1) and "marginally acceptable" = (2). These results suggest a rather strong dislike for SV order when preceded by clitic-left position.

The only other information structure condition with one or more response options involving clitic position also involved a discourse-old direct object, but these were subordination discourse contexts. These appeared in Tasks 1 and 2. In Task 1 Condition C, SOV was the preferred word order receiving the highest score (μ=4.59). Despite the fact that the Friedman test measure determined this score to be significantly higher than the mean ratings for either OclVS or OclSV, the former still received a mean (μ=3.70) nearing on the level of "rather acceptable" = (4), thus suggesting this word order as a valid response option to this discourse context. In contrast, OclSV only received a mean rating (μ=1.87) approaching the level of "marginally acceptable" = (2). While Task 2 did not offer SOV as a response option (for reasons discussed above), the response percentages for OclVS (77%) were more

than ten times that of OclSV (7%), and five times that of "both" (16%) as preferred response options. These indications, in conjunction with the results for OclSV above, therefore strongly suggest a distinct dislike for a preverbal subject when preceded by a clitic element. I return to this dispreference for OclSV word order and the ramifications it holds for the syntactic analysis I propose for Bengali in Chapter 5.

4.7 Methodological considerations

The results of the research presented in this chapter would have benefited from a larger number of participants in the quantitative tasks. This would have allowed for more powerful statistical comparisons, and for more varied comparisons based on education level, living environment [± urban], and age. I do not doubt that the length of the linguistic history questionnaire completed by participants who participated in tasks 1 and 2 led to the high levels of attrition discussed in Chapter 3. With higher participant numbers, shorter questionnaires overall may be used, thus precluding the need for all participants to complete all of the tokens. This could be remedied by random participant assignment to groups, each of which would complete a subset of the conditions from each task type. This strategy can be useful in a setting where sociolinguists in the host country discourage cash payments to survey participants, as was the case during the data-gathering stage of this dissertation research.

Although (follow-up) Task X employed a shorter, more precise linguistic

questionnaire, it did not provide the data that led to dividing the participants who did complete the questions in tasks 1 and 2 into sociolinguistic factorsgroups. However, as I saw in section 4.2.9, dividing the data into groups based on sociolinguistic factors only uncovered restricted inter-group differences by condition. As these were multi-directional in nature, I concluded that none of these statistical differences suggested overall differences among the sociolinguistic groups.

Looking toward the future, these results would benefit from comparison with results from participants. In light of the appropriate and inappropriate Bengali word orders discussed in the chapter 2, one would predict differences in focus word order preferences depending on the language involved in the discourse context. Another possibility, however, is that there has been a sort of *interface leveling* due to the extended exposure that these languages have had to each other in the minds of speakers for centuries now.

With respect to diachronic changes, the data reported in this chapter does not appear to indicate any significant changes in progress related to the syntax-information structure interface among Bengali speakers. However, since only six speakers over 50 years of age participated, the above conclusion is strictly limited in scope to the data presented above. Clearly, this conclusion would be more definitive had a larger population of speakers from this age group participated in the quantitative tasks.

Future research related to diachronic differences should also consider speakers from the 31-49 year old age group – an age group not considered in this dissertation. As I excluded 6 speakers from this age group in the results reported in this research, I will analyze the data gathered from this age group in follow-up research to this dissertation.

For many of the quantitative tasks, the data displayed a large amount of variation, or gradience. In other words, for Task 1 (which used a 5-point scale), many of the response sentences provided in the triads received the full range of possible ratings. One participant in particular whom I spoke with described her '4' ratings as scores that would have been '5' ratings had it not been for other elements in the discourse context that she did not feel comfortable with. These were mostly individual preference issues related to word choice, and not word order. She also indicated that her ratings of '1' were very similar to her ratings of '2', which perhaps would have warranted collapsing the two extreme point values.

The issue of gradience in speaker grammars with respect to grammaticality remains an issue of debate. Wasow (2007, p. 261) entertains the possibility that grammars may be gradient, citing three key factors that may lead to such gradiency.

First, it is possible that at least some cases of apparent gradience in the choice of grammatical form are actually cases in which different

meanings are being expressed. Second, some cases of gradience may be the result averaging over the usage of multiple individuals, who have different internalized grammars. Third, there are gradient factors affecting language use that are not part of linguistic competence.

If gradience may exist for grammaticality judgments, a greater level of gradience is expected for pragmatic acceptability judgments, where categorical judgments are harder to come by. As an alternative to ordinal scales, Bard et al. (1996) suggest magnitude estimation methodology for linguistic research, and Adli (2005) discusses a (somewhat) practical manner in which to put such methodology into practice. Despite the benefits of magnitude estimation data as opposed to ordinal data (e.g. a judgment of '4' in the former carries a statistical value equivalent to double the value of '2', but not in the latter), research using such methodology may be limited not only by methodological inertia, but by practical concerns such as participant fatigue/patience and (in the case of this dissertation) compatibility with statistical survey packages.

4.8 Conclusion

Although it is not the norm in theoretical generative syntax to gather experimental data (at least not quantitative) to arrive at theoretical conclusions, there are those who have expressed the need for quantitative data (e.g. (Zubizarreta, 1998)), and those who have collected it outside of SLA research. As noted previously, in the case of

Bengali, there are few extant works on information structure as related to syntax. Therefore, the data that I have presented on Bengali represents an important first step in describing Bengali on the one hand, and on the other, in gathering more varied data to inform a theoretical issue such as the information structure-syntax interface in languages in general. This type of research has the potential to be a useful tool in other languages for describing dialect differences and perhaps even diachronic change in progress.

Chapter Five

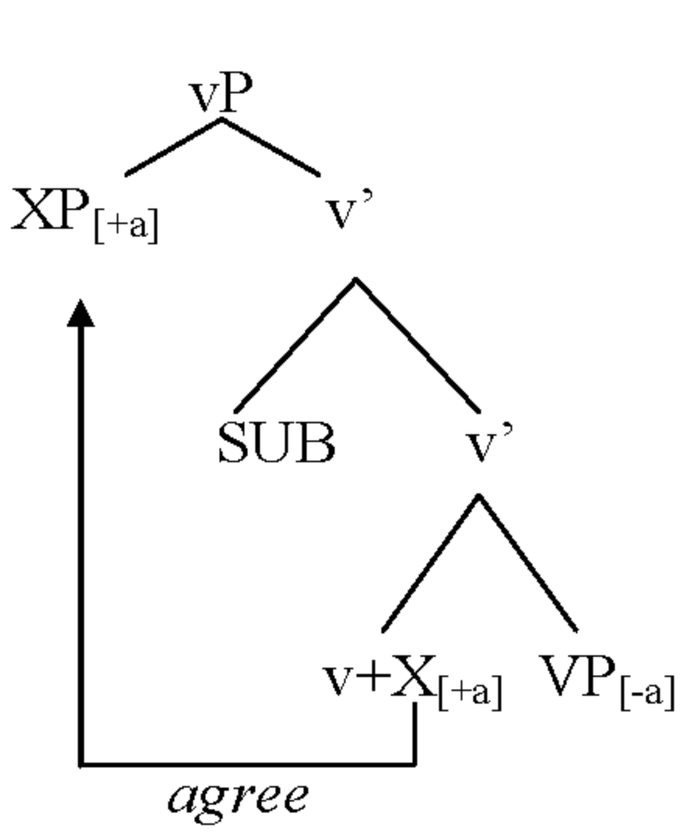
Focus and Information-structure in Bengali

In this chapter, I examine the significance of the data discussed in Chapter 4 in relation to López's (2009) proposal. In particular, I study his analysis of v-realm components in this section and whether such [±a] feature assignment can explain the characteristics of subjects and objects positions in Bengali. Finally, I attempt to find designated positions for Bengali focus elements. Before doing this I may recall López.

5.1 The Syntactic Behavior of Bengali Focus

I may start out with the v-realm notion. The v-realm is where [±a] (anaphoric) assignment is executed by the module pragmatics, which inspects the syntactic structure and the end of the first phase (1).

(1)



Elements that form an agree relationship with a clitic that move to Spec, vP are assigned the pragmatic feature [+a], while the complement of V is assigned [-a]. Now, since any (remaining) *in situ* object would come into view in the complement of V, it would be assigned [-a]. As I have seen in the Bengali data, a discourse new subject XP is preferred either at the

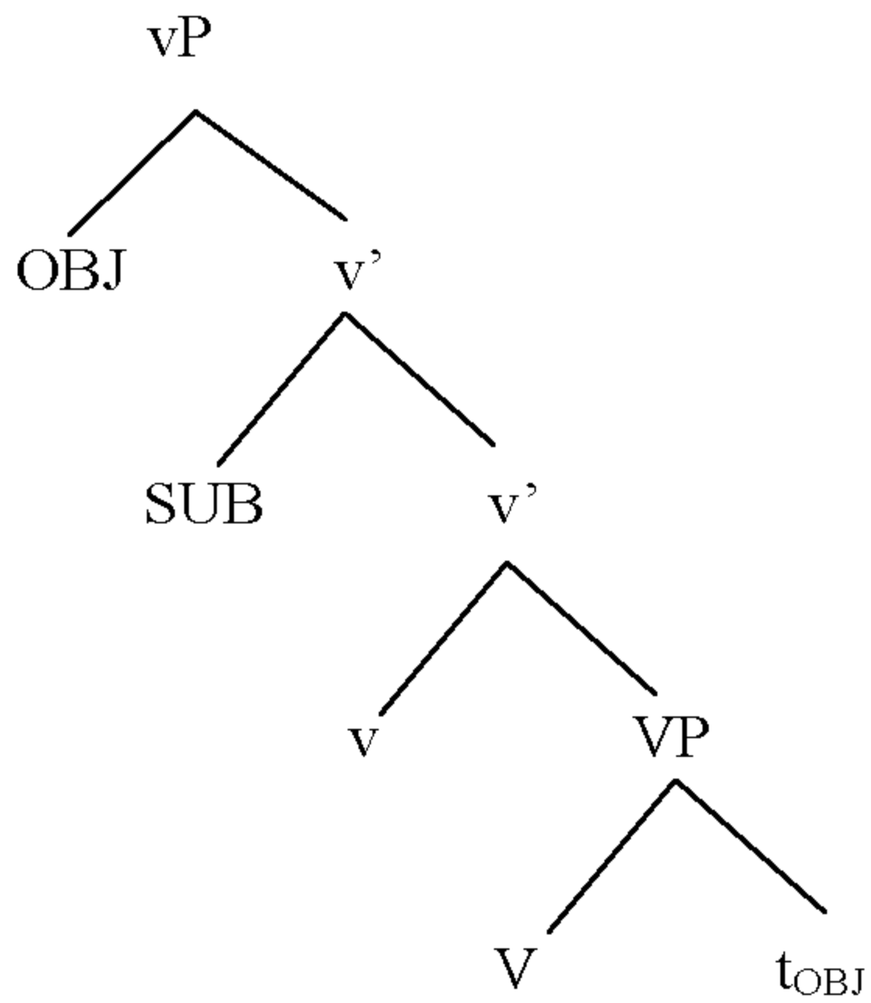
right edge of the clause in reply to subject narrow-focus (rheme) contexts, or in a preverbal position for thetic sentences. However, discourse-old subjects may also appear in a preverbal position. For López, elements appearing *in situ* are assigned [-a] by default by Pragmatics. Since preverbal subjects (once they move on from Spec, vP) may be either discourse-old or new, there are two directions I can take from here: I can assume that they are not marked either way, or I can assume that they are marked [-a] by default owing to their *in situ* position at the end of the first phase. López does not opt for either in his analysis. His analysis also clarifies preverbal subjects in thetic sentences, which move on from Spec, vP to Spec, TP. The disadvantage to assignment of [-a] in Spec, vP is that it cannot explain the discourse anaphoricity of discourse-old preverbal subjects.

As I have observed, however, López's definition of [+a] strictly applies to elements that have a discourse antecedent and enter an appropriate structural relation with that antecedent. Since it appears that there is no simple solution to this issue, for the moment I will follow López in assuming that the external argument in Spec, vP is not assigned any pragmatic feature in the v-phase. Of greater interest with respect to [±a] assignment for the moment are objects, and how they interact with v-realm subjects at the end of the first phase.

Recall from Chapter 4 that two tasks were carried out to gather data on

participant judgments of narrow focus (rheme) contexts. In two separate object narrow-focus conditions, my Bengali subjects displayed a preference for SOV word order as compared to OVS or OSV. In Task 1, Condition G, SOV received a mean rating of 4.89, while OSV received a mean of 2.75, and OVS a mean of 2.18. This preference does not cause any significant conflict for López's proposal since rheme constituents remain *in situ* and are appropriately marked [-a] by Pragmatics. Following López's proposal, an object moving to Spec, vP over the external argument should be assigned [+a]. Recall however, that *only* elements that establish an agree relation with a clitic are those that get marked [+a] by the pragmatic module at the end of the first derivational phase. Let me assume that OSV is the result of "object shift" to Spec, vP. As in (2).

(2)



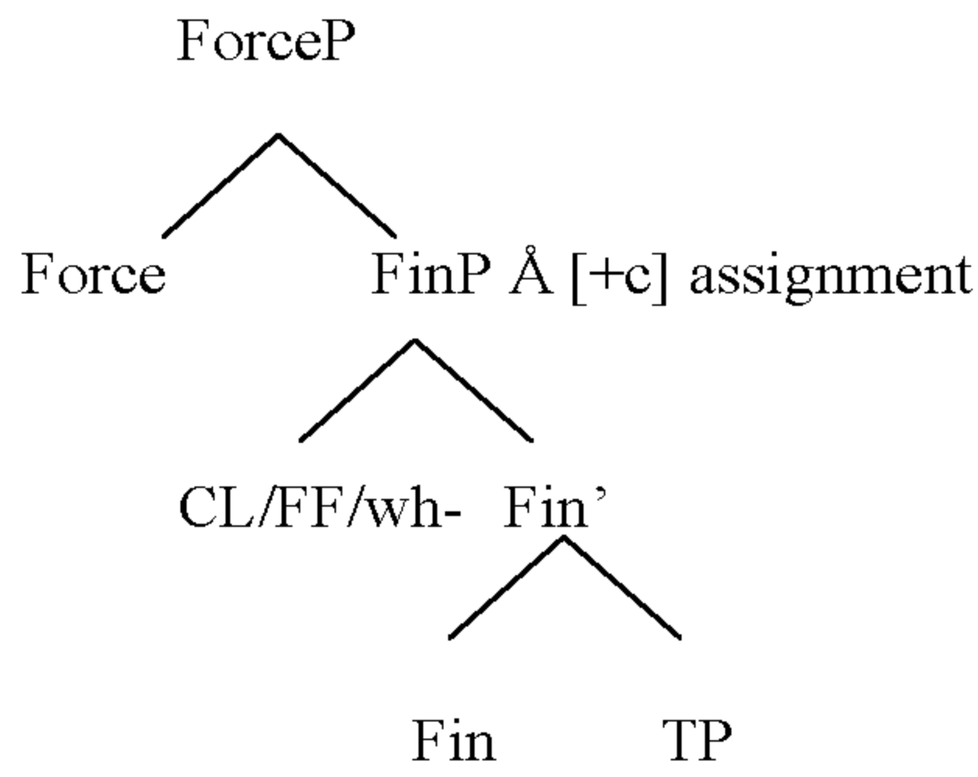
With this sort of movement, the scrambled object in OVS does not enter

into a relation with a clitic in v since it is not marked [+a], as in (1). However, since it has moved from its *in situ* position as complement of v, it should not get marked [-a] either, thus preventing it from being regular focus, which is underived in López's proposal. Movement for reasons of interpretation is in line with current minimalist discussions of scrambling or object shift (e.g. Chomsky (2005)), by which movement must be motivated by some feature having an influence on the eventual derivational outcome (Bhattacharya T et al, 2007).

While both of these analyses explain why OSV is not preferred in object narrow-focus contexts, López's analysis does not predict that OVS should not be preferred in the same context. In OVS the underived object gets correctly marked [-a] while the *in situ* postverbal subject should remain unmarked. By Zubizarreta's analysis, the subject in OVS also falls within the scope of focus, thus correctly predicting that OVS would not be preferred for object narrow-focus contexts.

So, how does López's proposal fare in the left periphery? Recall that he proposes a reduced, syncretic left periphery in which a variety of elements appear in Spec, Fin (3).

(3)



For López, if more than one FF/wh- or clitic element appears the result is stacked specifiers of Fin. At the end of the second phase in López's proposal, the Pragmatics module assigns [+c] to Spec, Fin, and [-c] to the complement of Fin (i.e. TP).

Now, I have been assuming that topicalized elements are attracted to Spec, DFceP due to the fact that topics appear between two complementizers without triggering proclisis. Now if the DFce projection is the lexicalization of topic features it is also possible that topical preverbal subjects are attracted to the same Spec, DFceP position. If topical preverbal subjects appear in Spec, DFceP, when *f* is projected (as I have been assuming for all preverbal subjects), enclisis will still obtain in main clauses. Since preverbal subjects in thetic sentences lack topical properties, when *f* is projected in main clauses they will not be attracted to Spec, FceP. Recall also that because of the enclisis facts, they may not appear in Spec, FP in main clauses, either.

This could explain the appearance of preverbal subjects in Spec, SubjP. In main clauses without a clitic (i.e. when *f* is not projected), a preverbal subject in Spec, DFceP would be practically indistinguishable from a preverbal subject in Spec, TP in a main clause. In subordinate clauses lacking a projection of *f*, a preverbal subject appears in either Spec, TP or Spec, DFceP. In subordinate clauses when *f* is projected, any sort of preverbal subject – except for truly topicalized ones – will appear in Spec, FP since it can provide a leftward clitic fusion site.

The remaining problem in the clausal architecture above (with SubjP) in the context of López's Pragmatics module is that it still has to possess some sort of mechanism for determining which elements should be [+c]-marked (affective phrases) and which elements should not be [+c]-marked (preverbal subject) in the Spec, FP position. If Pragmatics can determine which elements to mark [+c] in Spec, FP, I see no compelling reason why it would not also be able to do the same in Spec, DFceP. Therefore, the alternative clausal architecture with an additional dedicated position for preverbal subjects is no more advantageous than the architecture in with respect to [+c] assignment.

I have seen in this section that the left periphery in Bengali requires a greater number of projections in order to account for the cliticization data. A crucial issue for López (2009) in light of the clausal architecture I have proposed for Bengali has to do with what "counts" as the edge

with respect to pragmatic feature assignment. Even if more than one projection may count as an edge, López's Pragmatics module would need to be modified in order to correctly assign [+c] discourse features to topical(ized) elements and affective phrases on the one hand, and not assign them to non-topical preverbal subjects on the other. Even putting aside this complication, edge calculation for [EPP] checking is even more problematic. Following e.g. Chomsky (2005), the [EPP] feature is inherited by T from C. Even if [EPP] inheritance can be "intercepted" by a C-realm head higher than T, given the variety of possible preverbal subject positions described above, such interception would have to be arbitrary and *ad hoc* to fit the purposes of the preverbal element in question. This is clearly an undesirable proposition.

5.2 Position of the FocP in Bengali

5.2.1 Semantic Division of the Clause Structure

I now move onto the issue of focus comprehensively, and it's insinuations in the semantic division of the clause structure. The semantic division of the clause structure that is credibly proved to be true for Bengali, militates against the semantic division proposed in Kamp (1981), , Heim (1982), and with revisions in Diesing (1992). In Diesing's theory, the clause is divided into restrictive clause and nuclear scope. Since only 'discourse concepts' are encoded in the Bengali syntax, it is possible to divide the clause broadly into two parts:

(4) a) Presupposed—Non-Presupposed

or

(4) b) Top(field)—Focus—Background(vP Phase)

In these terms, the 'non-presupposed' part or the 'background' in (4a) and (4b) would constitute the rheme, or the 'new information'. The combination of the topic field and the FocP constitute the 'presupposed part'. The 'topic field' encompasses all the positions above FocP. In Abraham (2002) it was argued that German is discourse prominent in the following sense: There is a basic word order in German (and Dutch and West Frisian) with rhematic (informationally new) material in VP. Something similar is happening in Bengali.

5.2.2 Proof that Bengali has a Clause-int(ernal) Focposition between TP and vP

In this study, I am going to establish a proposition that a clause-internal position is designated for Bengali focus and which is placed in between TP and vP. A focused phrase must move overtly to SpecFoc. The verb is constantly right next to the focused particle. It means, no other phrase can be placed between the focused phrase and the verb. By following Cinque's (1997) suggestion, it is possible to state that the verb is in the Foc head.

The most clear-cut substantiation of the above claims comes from the reality that the subject can be placed on both sides of the focused phrase, and they have diverse elucidations. Whereas the subject in Spec TP

must be 'presupposed', the subject in Spec vP is 'not presupposed'. By 'not presupposed', I mean that it has not been prominent in the discourse—in advance. In a way, then, I can say that the EPP feature of 'T' (Chomsky, 1995), in Bengali is the feature [+N, +presupposed], since only presupposed NPs can be allowed to be placed in Spec TP. This feature on TP will be [-int(erpretable)], though, it will be [+int] on the NP in its Spec. The non-overt 'Expl(etive) pro' that I assume is merged in Spec TP when there is no subject in Spec TP, will have a feature [+N, +presupposed]. In the following sections I am going to discuss on this 'presupposition effect' which is also opposed to 'Specificity Efect'. The 'Presupposition effect' has its effect on an 'entire range' of heads, which include the Topic(s) and TP.

5.2.3 Specificity Effect

Bengali also allows 'Specificity Effect', that is, when the focus within the clause is 'Contrastive', only a specific or a definite NP can be in Spec TP. The Specificity eifect however, is to be analyzed in a different manner from the 'presupposition effect' stated above. That is, the 'Specificity effect' is not a result of 'T' having an EPP feature (=[D] feature), such that only specifics and definites can be in Spec TP (Chomsky, 1995), as I shall explain below, see (5a). For example, if the focus is of 'Contrastive type', the subject in Spec TP can only be a specific. Specifics are always a part of a presupposed D-set, so, it is easy

to associate presupposition and specificity as part of the same phenomenon. But they are not as can be seen in (5b).

(5) a) 'Specificity Effect is directly a consequence of the type of focus in the clause.

b) Specifics can be viewed as follows:

A They are part of a set

B That set is presupposed.

Contrastive focus within the clause structure entails that the subject in Spec TP be element of a set, and the EPP feature of 'T' needs that the phrase in Spec TP be 'presupposed'. As a result it follows specificity of the subject in Spec TP. The subject in Spec vP quite the opposite, will not only be an indefinite, but will not be part of the discourse either. That is, even under contrastive focus, Spec vP will host a 'non-presupposed' indefinite when the subject in Spec TP will be specific. The following examples demonstrate this:

(I) When Focus is CONTRASTIVE FOCUS

(6) a) [$_{TP}$ aekjon keu (Specific) [$_{FocP}$ OI boi-ta t_1 [$_{Foc}$' poRchilo$_2$ [$_{vP}$ t [$_{VP}$ t_1,

Someone (Spcific) THAT book-cl$_1$ reading-was

go'tokal t_2]]] (ei boi-ta noy)

yesterday (this book-cl not)

'Someone(spcific) was reading THAT book yesterday' (not this book)'

b) [$_{Top}$go'tokal$_2$ [$_{TP}$ expl(pro) [$_{FocP}$OI boi-ta$_1$ [$_{Foc}$' PoRchilo$_3$ [$_{vP}$ aekjon

yesterday expl(pro) THAT book-cl reading-was

keu [$_{VP}$ t $_1$ t$_2$ t$_3$]]]] indef(non-presupp)

someone indef(non-presupp)

'Yesterday, it was THAT book that someone (indef—nonpresupposed) was reading.'

In (6a), the subject which is found in Spec TP position is a specific, whereas in (6b), the subject in Spec vP is licensed as a non-presupposed indefinite. If the subject bears a proper name (DP), then, under the contrastive focus frame work in the clause, it will be able to have the following readings in the two positions, Spec vP and Spec TP:

(7) a) Spec vP—Non-presupposed DP

b) Spec TP—DP will be part of a presupposed D-set

For instance,

(8) a) [$_{TP}$ Sumon-to [$_{FocP}$ OI boi-Ta $_1$ [$_{Foc}$' poRchilo$_2$ [$_{vP}$ t [$_{VP}$ t$_1$ go'tokal t$_2$]]]]

Sumon-top THAT book-cl reading-was yesterday

(ei boita noy)

(this book-cl not) (presupposed D-set)

b) [$_{Top}$ go'tokal$_1$ [$_{TP}$ expl(pro) [$_{FocP}$ OI boiTa$_2$ [$_{Foc}$' poRchilo$_3$

yesterday THAT book-cl reading-was

[$_{vP}$ Sumon [$_{VP}$ t$_1$ t$_2$ t$_3$]]]]]] (non-presupp)

Sumon

Since both Spec TP and Spec vP are bearing a proper name, the disparity in interpretations are hard to identify. Comparison with the data in (6) facilitates to obtain the right type of elucidation.

(II) When Focus is INFORMATION FOCUS (Kiss, 1998):

(9) a) [$_{TP}$ aekjon keu$_1$[$_{FocP}$ OI boiTa $_2$ [$_{Foc}$' poRchilo$_3$ [$_{vp}$ t$_1$[$_{vp}$ t$_2$go'tokal t$_3$]]]]]

Someone THAT book-cl reading-was yesterday

'Someone (indef-presupp) was readingTHAT book yesterday.'

indef -presupposed

b) [$_{Top}$ go'tokal $_1$ [$_{VP}$ Expl (pro) [$_{FocP}$ OI boiTa $_2$ [$_{Foc}$' poRchilo$_3$

yesterday THAT book-cl reading-was

[$_{vP}$ aekjon keu [$_{VP}$ t$_1$ t$_2$ t$_3$]]]]] indef (non-presupp)

someone

Unlike in (6a), the indefinite *aekjon keu* 'someone' in (11a), no longer has a Specific comprehension. There is a dissimilarity between the elucidation of *aekjon keu* 'someone' in Spec TP, as in (9a), and that in Spec vP in (9b), nevertheless both are indefinites. Where *aekjon keu* 'someone (indef)' in (9a) is PRESUPPOSED, and for this reason exists in the discourse, *Ekfon keu* 'someone (indef)' in (9b) is NOT. This 'someone' did not subsist in the discourse in advance.

It is this data which substantiates our argument that 'Presupposition effect' is dissimilar from the 'Specificity Effect'. Whereas 'Presupposition effect' on the subject in Spec TP can be found in both

types of focus; specificity effect on the subject in Spec TP is available only in case of contrastive focus. Structurally this is clarified by acknowledging that the EPP feature of TP is [+N, +presupposed], which can however be only a [-interpretable] feature. This is a 'discourse' principle predetermined in the computational system. It is true for Bengali, and most other Indian languages (Dasgupta, 2007). It is then feasible, that it is a universal principle, previously not taken note of in the literature. The feature 'presupposed' should not be jumbled with the feature 'existential', for the reason that even such a kind can be either 'presupposed' or 'not presupposed' in the discourse.

This difference is present in all varieties of structures. Even generics show this distinction. In the following, the focus is information focus on the 'late' as a part of 'late night':

(10) a) $[_{TP}$Sumon$_1$ $[_{FocP}$ khub RAtE$_2[_{foc}$'kaj kore$[_{vP}t_1[_{VP}t_2t_3]]]]$

Sumon late NIGHT work do

'Sumon works at late night'

b)$[_{TP}$Expl(pro)$[_{FocP}$khubRATE$[_{Foc}$'kaj kore $[_{vP}$Sumon$[_{VP}$tt]]]]]

late NIGHT work do Sumon

5.2.3.1 Specificity explained

I have claimed that specificity is a result of contrastive focus in the clause. I now mention examples with other types of focus: (III) *Sudhu* 'Only' type:

(11) a) [$_{TP}$ aekjonkeu [$_{FocP}$ Sudhu OI chobi-tai [$_{Foc}$'dekhchilo
someone(spcfc) only THAT movie-cl-foc watching-was
[$_{vP}$ t [$_{VP}$ t go'tokal t]]] spcific
yesterday
'Someone(spcfc) was watching only THAT movie yesterday.'

b) [$_{Top}$go'tokal [$_{Tp}$expl(pro)[$_{FocP}$ Sudhu OI Chobi-tai [$_{Foc}$'dekhchilo
yesterday only THAT movie-cl-foc watching-was
[$_{vP}$ aekjon keu [$_{VP}$ t t t]]]]
someone (indef- nonpresupp)

The subject in Spec TP is in (11a) is specific, but the subject in Spec vP in (11b) is a non-presupposed indefinite.

(IV) *aemon ki -o* 'even-also' type:

(12) a) [$_{TP}$ aekjon keu [$_{FocP}$aemon ki chobi-ta-o [$_{Foc'}$dekhchilo
someone even THAT movie-cl-also watching-was
[$_{vP}$t [$_{VP}$ t go'tokal t]]] spcfc
yesterday
'Someone (spcfc) was watching even THAT movie yesterday.'

b) [$_{Top}$ go'tokal [$_{TP}$expl (pro) [$_{FocP}$aemon ki OI chobi-ta-o

yesterday even THAT movir-cl-also

[$_{Foc'}$ dekhchilo [$_{vP}$ aekjon keu [$_{VP}$ t t t]]]] indef(non-presp)

watching-was someone indef (non-presupp)

The subject in Spec TP is in (12a) is Specific, but the subject in Spec vP in (12b) is a non-presupposed indefinite.

(V) X *nOeto* 'either or' type: 'X or Y'

(13) a) [$_{TP}$ aekjon keu (spcfc) [$_{FocP}$ ei chobi-ta noyto oi chobi-ta

Someone(spcfc) [this movie-cl or that movie-cl]

[$_{Foc'}$ dekhechilo [$_{vP}$ t [$_{VP}$ t go'tokal t]]]

watch yesterday

'Someone(spcfc) watched either this movie or that movie yesterday.'

b) [$_{Top}$ go'tokal [$_{TP}$ expl(pro) [$_{FocP}$ ei chobi-ta noyto oi chobi-ta

Yesterday this movie-cl or that movie-cl]

[$_{Foc'}$ dekhechilo [$_{vP}$ aekjon keu indef (non-presp) [$_{VP}$ t t t]]]]

watch someone(indef-non-presupp)

It is again found that in the example (13a), the subject in Spec TP is specific. In contrast, the example (13b) shows that the subject in Spec vP is a non-presupposed indefinite.

I therefore study a specificity effect (specifics are elements of a presupposed set by definition) in all the three cases stated above, where each clause has a different type of focus. The 'presupposition' segment of the specific NP is because of the EPP feature of 'T', and the fact that it is part of a 'set' is due to the focus. All types of focuses above are working like the contrastive focus, vis-a-vis generating a specificity effect. Information focus, which is the most common category of focus, is the only exclusion. Information focus does not give rise to specificity effect.

I explicate the phenomenon as follows:

(14) I assert that all kinds of focus can be segregated into two types:

(A) BINARY FOCUS:

Ones determining Binary sets: contrastive, 'only' type, 'even' type, 'either-or' type etc. Even when there are multiple alternatives, out of which a focused constituent is chosen, for which the proposition is true, the given set would be binary in the following way:

{a, not however {b, c, d ...}}.

(B) INFORMATION FOCUS:

The same specificity effect is found with all the focus types listed as 'BINARY' Focus in (A) above. It is possible to raise a question: why should BINARY Focus permit a Specific NP in Spec TP? How are the two related? Instinctively, the clarification would be along the

subsequent lines: specifics split a given set into two parts. One part contains the member(s) of the specific NP, and the other (silent, but understood) part consists of the rest of the members of the given set. Thus, the Specific *aekjon keu* 'someone' in the examples above is a member of a following BINARY set (presupposed in the discourse).

(15) {{aekjon keu}, {b.c.d...}}—{{Someone}, {b.c.d...}}

Now I can consider the set associated with a focused phrase. A contrastive focus too splits the given set into a binary set. The disparity between the specific set and the set belonging to the contrastively focused set is that the second member of the contrastively focused set is a negation of the given set. See e.g. the focused *Oi chobita* 'that movie' would be part of a following kind of set.

(16) {{OI chobita}, Not{ei chobita}}—{{that movie}, Not{this movie}}

The negation of the second member of the focused set describes the focused phrase uniquely in a prearranged clause. An analogous binary splitting up of the focused set can be imagined for all other binary focus types, where the second set consisting of one or more members will be negated to define the focused phrase uniquely in a given clause. Consider the elucidation of (6a) repeated below, in terms of specificity and focus:

(17) [$_{Tp}$ aekjon keu [$_{FocP}$ OI chobi-ta [$_{Foc}$' dekhchilo [$_{vP}$ t [$_{VP}$ t go'tokal t]]]

Someone (spcfc) THAT movie-cl watching-was yesterday

(ei chobita noy)

(this movie-cl not)

'Someone (spcfc) was watching THAT movie yesterday.'

Due to contrastiveness, there are two accounts embedded in (19). The precise reading of (17) would be:

(18) a) Someone from the given presupposed set watched THAT movie yesterday, And the implicit reading associated with it would be:

(18) b) And the REST, watched THIS movie.

With added types of binary focus, instead of 'THIS MOVIE' in (18b), I can have other options like 'OTHER MOVIES', 'also other movies' etc.

Despite the fact that focus negates the other components of its set in each clause, so that the focused phrase can be labeled exclusively for that particular clause, the other component of the focused set plays a part in the other clause, afresh, with a distinctive reading. As a result, in both (18a) and (18b), the focused phrase is exclusively described by not permitting its other member to take part in that particular event (of the clause it is part of). Syntactically, then, there is an agreement relation

holding between the focus head and the 'T' head with reference to their 'set theoretic' properties. If focus is BINARY, so is the 'T' head. Specificity effects are an end result of this agreement relation. See (19).

(19) Focus head agrees with 'T' head with reference to its 'set theoretic' features.

I shall perceive that this agreement relation expands to the whole CP system.

5.2.3.2 Explaining Focus

If contrastive focus is able to allocate quantificational properties to the 'T' head and also to the entire CP system as I shall demonstrate, the question naturally arises regarding the subject in Spec vP. The fact that it doesn't, justifies our division of the clause as in (4a) or (4b). It is the Focused phrase and the verb which defines the 'event' of the clause, and the 'background' part is more like the 'description' of the 'event'. (Chomsky, 1971, p. 102) defines focus as the 'predicate' of the dominant sentence of the 'deep structure'. Our definition of Focus comes close to that.

5.3 Encoding discourse features in the lexicon

My study above has depicted that the subject in Spec TP must allow the feature [+presupposed], indicating that 'T' holds an EPP feature=[+N, +presupposed]. This feature must then be predetermined in the lexical items. Reinhart (1995) proposes, along the lines recommended by

Chomsky (1995), that interface economy should be created into the numeration. In accordance with Chomsky's (1995, p. 76) economy principle, "a enters the numeration only if it has an effect on the Output". Alongside these lines, my proposal is that both [+Focus] and '[+presupposed]' should be marked lexically. According to Reinhart (1995, p. 14): "Interface economy, then, determines the shape of the numeration: The underlying Intuition may be that it is at the stage of choosing the 'stone blocks' that Speakers pay attention to what it is they want to say. (Theoretically, this line resembles earlier views, that all aspects of meaning are determined in Deep Structure)."

However, Noam Chomsky (p.c.) points out that: "Internal computational operations are 'dumb'. They can't know whether something's presupposed in a discourse. What must be happening is that NPs are moving freely (at least, not with this constraint), and once they end up in Spec-TP they are interpreted as presupposed."

This turns into complicated to ascertain if the whole Topic field (or the CP field), (which comprises Topics, Wh-phrases in Spec CP, NPs in Spec TP and even phrases in SpecFoc), is part of the 'presupposition' in the discourse, as I am asserting for the semantic splitting up of the clause. Thus, I presume that 'presupposition' is an attribute which is predetermined in the lexical items when they enter the numeration. This does not collide with the idea that "Internal computational operations

are dumb". The NPs are without a doubt shifting freely, but only those NPs continue to exist in Spec TP which has this feature 'presupposed', under Spec head agreement with the 'T' head which too carries the feature 'presupposed'. This feature on 'T' must then be [-interpretable].

My proposal that discourse features like 'presupposition' and 'Focus' are inserted in the lexical items is admitting that meaning is determined in the Deep structure itself in Chomsky's (1971) sense, and according to what Reinhart (1995, p. 14) suggests above. Perhaps I should make a diiference between the function of LF in case of the 'Speaker' and that of the 'hearer'. It is only in case of the 'hearer' that meaning is determined by the 'Spell Out' of the speaker's computational system (via PF) to LF of the hearer. In case of the 'Speaker', there are LF activities 'before' he speaks, i.e., when the lexical items are chosen in the 'numeration'. If 'binding' is an independent module, and scope effects are also dependent on Topic and Focus, as this work shows, it is a plausible idea. The speaker's 'intention' decides 'what he is going to speak'. (It is also true that children learn to construct sentences mechanically, but they do have a vague idea that one word order has a certain kind of effect different from that of the other. The finer semantics comes later.)

This analysis establishes that all Focused phrases move overtly to SpecFoc. It also establishes—'syntactically' at least—that every clause must have a FocP.

5.4 Topic Focus and Object Shift

In this section, I am going to show that there is no structural case position for the DO in Bengali. The DO either moves to SpecFoc, when it has a Focus feature, or moves to the Topic position, or remains in-situ in the vP Phase.

5.4.1 Topic Field

I am also going to study the Topic position(s) in Bengali Clause structure. Until now, I have assumed and shown that TopP exists in the Standard Clause-external position above TP. I am going to show that in addition to the clause-ext TopP, there can be Top phrase(s) inside the clause too, between TP and FocP, arranged hierarchically, in the following manner:

(20) a)[$_{\text{Top-ext}}$ **[TP** [$_{\text{Top-int}}$ (Top.int) **[FocP [vP [VP…**

Let's identify this space which holds this string of hierarchically set Topic phrases including TP, the Topic field. It resembles the series of hierarchically arranged specifiers in Bengali DP structure. It is significant to recognize what I mean by 'hierarchically arranged', in case of Topics. Despite the fact that they are all D(iscourse)-linked, the

order of Topics is decided by the way an 'event' is viewed by the speaker. This would become clearer as I proceed.

Like this, the position of all the phrases within a clause can be explained, and all word order possibilities too, without resorting to arbitrary scrambling operations as the 'adjunction theories' of scrambling postulate, with the help of the above clause structure. The commonality of Topic phrases is that all Topics are by definition 'presupposed' in the discourse. Focus (binary) too is part of the 'presupposition'. That is, the focused vP is selected from a set salient in the discourse. It is only the material present in the vP Phase that is not selected from the discourse. I have already proved in section 5.2.3 that there can only be [-presupposed] NP in Spec vP. In fact, it is seen that the entire vP Phase is the Non-Presupposed part. That is, whatever is there in the vP Phase at the end of the derivation, is outside discourse. It is not presupposed in advance. Hence, I call the vP Phase the 'Background'. Thus, the clause structure can hence be semantically divided into two parts:

(20) b) Presupposed—Non-Presupposed/Background

The presupposed part (/field) comprises Top Field and Focus, encompassing the verb, and the components in the vP phrase is the Non-Presupposed part or Background. Is it then unexpected that the verb is always in the presupposed field, in the Foc head? The fact that it is

never in the background, points out that a deeper semantics is involved here. Our data reveals that except in exceptional cases, which I shall describe, not all the phrases can move out of the vP Phase, i.e., from the 'background'. There must be some phrase always present in the 'background'. In all the cases I have considered till now, except in case of unaccusatives and statives, that is indeed the case. Here is a repeat.

(21) a) [$_{TP}$ Sumon-to [$_{FocP}$ OI boiTa [$_{Foc}$' diechilo [$_{vP}$ t [$_{VP}$ Shojibke t t]]]

Sumon-top THAT book-cl gave Shojib-dat

If all the phrases are in the presupposed part, the construction is unusual, particularly if the focus is information focus. For instance, if some speakers find the data acceptable, it is because a time adverb is 'non-overtly' present in the background. The construction would be very unusual if instead of *diechilo* 'gave' I would use *dieche* 'has given'.

(21) b) ??[$_{Top}$Shojibke-to [$_{TP}$ Sumon [$_{FocP}$ OI boiTa [$_{Foc}$' diechilo [$_{vP}$t [$_{VP}$ t t t]]]]]

Shojib-dat-top Sumon THAT book-cl gave

If I add one more phrase, e.g., a time adverb, then some phrase or the other must be in the background or the vP Phase, and thus, here are some of the possibilities:

(21) c) *[$_{Top}$ Shojibke-to [$_{TP}$ Sumon [$_{Top}$ go'tokal [$_{FocP}$ OI boiTa [$_{Foc'}$ diechilo [$_{vP}$ t [$_{VP}$ t t t t]]]]]]

Shojib-dat Sumon yesterday THAT book-cl gave

d) [$_{Top}$ Shojibke-to [$_{TP}$ Sumon [$_{FocP}$ OI boiTa [$_{Foc}$' diechilo [$_{vP}$ t

Shajib-dat Sumon THAT book-cl gave

[$_{VP}$ t t go'tokal t]]]]]

yesterday

e) [$_{Top\text{-}ext}$ go'tokal-to [$_{TP}$ Sumon [$_{FocP}$ OI boiTa [$_{Foc}$' diechilo

yesterday-top Sumon THAT book-cl gave

[$_{vp}$ t [w Shojibke t t t]]]]]]

Shojib-dat

(21c) is ungrammatical because all the phrases are in the presupposed part, with the time adverb in Top-int position. (21d) and (21e) are acceptable because some phrase is in the vP phase, in the background. It can be tested that the hierarchy of phrases in the Topic field is flexible, and is not rigidly fixed, unless the verb has a Completive aspect. If the verb has Completive aspect, firstly, all the phrases can be in the presupposed part, with the background empty. Secondly, the order of phrases is fixed now. In the following, there are three phrases in the Top-int positions, all arranged in strict hierarchy, and there is none in the vP phase.

(21) f) [$_{Tp}$ Sumon-to [$_{Top\text{-}int}$ 1 Shojibke [$_{Top\text{-}int\ 2}$ go'tokal [$_{Top\text{-}int}$ 3 boiTa

Sumon-top Shajib-dat yesterday that book-cl

[$_{Focp}$ [$_{Foc}$'DIE DIECHILO [$_{vP}$ t [$_{VP}$ t t t t]]]]]

[$_{Focp}$ [$_{Foc}$'GAVEAWAY[$_{vP}$ t [$_{VP}$ t t t t]]]]]

'Sumon, to Shpjib, yesterday, that book GAVE AWAY.'

These are some inquiring facts about Bengali word order can only be clarified in terms of how an 'event' is analyzed. A tentative presumption is that a completive aspect has by now described an event to the last bit, and so the relative hierarchy amongst the phrases in terms of their topicality has already been defined, and is thus inviolable, as in (21f), where the focus can only be on the verb. In contrast, in (21d) and (21e), the verb does not have a completive aspect. The event is defined as it is described by the word order hierarchy presented in these clauses (representing the speaker's perspective). Hence, the order is relatively free. The fact that some phrase or the other must remain in the background or the non-presupposed part only confirms the fact that in case of 'non-completive' aspect, the event is not totally defined. It is defined only when all the phrases are in the 'presupposed part' of the clause, as in (21f). I understand that there hasn't been any discussion along these lines in the literature. As I present more data, it will become clearer that there is indeed a semantic division of the clause along the lines suggested above. This division has sufficient 'syntactic effects' to suggest that it is not a separate level of interface, but is very much part of the narrow syntax.

5.4.2 Object shift

I have already observed that, when the aspect of the verb is 'non-completive' and which is the most frequent case, apart from for certain constraints, viz.5

(22) a) Having the Focused phrase in SpecFoc and verb in Foc head

b) Having some phrase or the other in the vP Phase (non-presupposed)

The word order is more or less free. It is verified by the way Topical hierarchy is perceived by the Speaker.

The identical rule applies to the DO, as to any other argument or adjunct in the clause. Thus, a DO, whether definite, specific or indefinite, that is, irrespective of it's type, can either be in the vP Phase, in-situ, when it would be 'non-presupposed', or in SpecFoc, or in the Topic field, Top-ext or Top-int. That way, even an indefinite can be in Top-ext, where it would be a 'presupposed' indefinite—one which has taken place in the discourse before.

5.4.2.1 Position of an Indefinite

I demonstrate below, that DOs in Bengali can only be in Spec Top, or SpecFoc, or in-situ in the vP Phase. In the following, I explain that for indefinites:

(A) AN INDEFINITE IN SPEC TOPIC

(23) [$_{Top.ext}$ Ekta kono chobi [$_{TP}$Sumon [$_{FocP}$ SHOJIB-ke [$_{Foc}$' diechilo [$_{vp}$ t

some movie (presupp indef) Sumon SHOJIB-dat gave

[$_{Vp}$ t t go'tokal t]]]]].

yesterday Presupp indef

For the NP in the topic position to have an indefinite interpretation, the focus must be information focus and not one of BINARY focus. That is, the specificity effect I studied earlier extends to the topic position. The focus type not only influences the subject in Spec TP, it affects the Topic phrase too. This indicates that the following agreement relation holds amongst all the heads in the Top-field.

(24) Top—T—(Top)—Foc

should all agree vis-a-vis their set theoretic or quantificational property.

If focus is BINARY, all the heads in the Top-field should be part of a presupposed set. In other words, they should be quantificational. And if focus is information focus, then topics simply need to be 'presupposed' without being quantificational.

(B) AN INDEFINITE IN SPEC FOC:

In the following, the focus (any type) is on the N-head *chobi* "movie" of the indefinite.

(25) a) [$_{TP}$ Sumon [$_{FocP}$ EkTa chobi [$_{Foc}$' dieche [$_{vP}$ t [$_{VP}$ Shojibke t t]]]]]

Sumon one-cl MOVIE given Shojib-dat

'Sumon has given a MOVIE to Shajib.'

In the following, the Focus is on the numeral of the indefinite.

(25) b) [$_{TP}$[Sumon[$_{FocP}$ Sudhu EKTAI khelna [$_{Foc}$' dieche [$_{vP}$ t [$_{VP}$ Shojibke t t]]]]

Sumon only ONE-cl-foc toy given Shojib-dat

(C) AN INDEFINITE IN THE BACKGROUND OR IN-SITU:

(26) [$_{TP}$ Sumon-to [$_{FocP}$ SHOJIBKE [$_{Foc}$' diechilo [$_{vP}$ t [t kichu boi t]]]]]

Sumon-top SHOJIB-dat gave some books

(non-presupp indef)

It is essential to note that, for the 'non-presupposed indefinite' to be able to take place in the vP phase in-situ, the focus in the clause must be 'Information Focus'. If the focus is contrastive, the indefinite can only occur in SpecFoc, itself, carrying the contrastive focus feature. A contrastive focus does not allow an indefinite DO anywhere eise in the clause, except in SpecFoc. This is a surprising result. How come focus type is affecting the Interpretation of the DO in-situ, when it doesn't, if the subject is in-situ? The subject in Spec vP can be an indefinite, but the DO in-situ cannot, under Contrastive Focus. Why? This is a very fine semantics. I have the following explanation to offer.

The specificity effect enforced on the indefinite in-situ (under contrastive focus) in (26) will not be due to a D-linked set, but will be due to an implicit set linked to the subject *Sumon.* A phrase in the 'background' cannot be 'presupposed' or D-linked according to our

theory. Thus, the set associated with the specific indefinite in (26) cannot come from discourse. It can come from a 'possessor' set in the Spec of the indefinite, anaphoric with the subject Sumon. Thus *kichu boi* 'some books' in (26) (under contrastive focus) is to be interpreted as:

(26) (Sumober) kichu boi

'(Sumon's) some books'

Here *Sumoner* 'Sumon's' is phonologically non-overt.

5.5 Conclusion

In this chapter, it has been systematically observed how focus is the 'driving force' of Bengali Syntax. The clause structure of Bengali is more like a 'discourse structure', which is surprisingly, part of the computational system of Bengali. The word order and scope effects are thus the corollaries of the semantic partition of the clause structure that I have proposed.

Chapter Six

Conclusion

This dissertation thus far has dealt with the syntax-information structure interface as it pertains to theoretical issues regarding focus constructions in Bengali. Within, I have examined the syntactic issues related to preverbal subjects and clausal word order on a theoretical level, and have sought to inform the issues and debate by examining Bengali. Sociolinguistic issues have also been noted. I have discussed and established the notions related to information structure in developing a quantitative data-gathering methodology, and have detailed this methodology. This methodology was counterbalanced and supplemented by qualitative data gathering. I have presented the results that I gathered using this methodology and I have described how the variety of experimental data that I gathered for Bengali informs this issue. I have discussed the implications that these experimental data have for analyses and proposals related to the interface between syntax and information structure (Lopez, 2009).

Prior to concluding this dissertation, I would like to return to the research queries that guided this dissertation and, to the best of my ability, provide answers for these questions. As discussed in Chapter 4, according to the data gathered in this dissertation, the preferred word order for is SOV. For a subset of the data, discourse subordination contexts with a discourse-old object, OclVS was also rated very highly – much higher than OclSV. As I saw in Chapter 5, Bengali allows a Clause-internal Focus position between TP and vP. A Focused phrase

must move overtly to SpecFoc. The subject in Spec TP is required to be 'presupposed', the subject in Spec vP is 'not presupposed'. The EPP feature of 'T' in Bengali is the feature [+N, +presupposed]. The reason is, only presupposed NPs can be licensed in Spec TP. The non-overt 'Expl(etive) pro' that I assume is merged in Spec TP when there is no subject in Spec TP, will have a feature [+N, +presupposed].

Moreover I saw, however, the syntactic analysis of a language like Bengali must take into account the clitic pronoun system. The data examined in Chapter 5 also suggest the existence of additional left-peripheral architecture to which preverbal constituents are attracted when a clitic is present. I have pointed out, however, that there are complications related to motivating such movement within current Minimalist assumptions.

Finally, it is worth to state that further researches in this specific field of study will be able to discover many issues related to the syntax-information structure interface of Bengali.

Appendix

PARTICIPANT'S PROFILE DETERMINING QUESTIONERS

১. আপনার নামঃ .....

২. আপনার শিক্ষাগত যোগ্যতাঃ .....

৩. আপনার বয়সঃ ..... লিঙ্গঃ.....

৪. আপনার পেশাঃ .....

৫. আপনি কোথায় থাকেন? ক) শহরে খ) গ্রামে

৬. আপনার ছেলেবেলা কোথায় কেটেছে? ক) শহরে খ) গ্রামে

৭. বাংলা কি আপনার মাতৃভাষা? ক) হ্যাঁ খ) না

৮. বাংলা ব্যতীত অন্য কী কী ভাষা ব্যবহার করে থাকেন? .....

৯. পরিবারের সদস্যদের সঙ্গে বাংলা ভাষার কোন রূপ ব্যবহার করেন? ক) প্রমিত খ) আঞ্চলিক

১০. পরিবারের বাইরে অনানুষ্ঠানিক পরিবেশে কোন ভাষারূপ ব্যবহার করেন? ক) প্রমিত খ) আঞ্চলিক

১১. আনুষ্ঠানিক পরিবেশে কোন ভাষারূপ ব্যবহার করেন? ক) প্রমিত খ) আঞ্চলিক

১২. বাংলা ভাষা ব্যবহারে আপনি কেমন দক্ষ? ক) ভাল খ) মাঝারি গ) ভাল নয়

১৩. বাংলা লিখতে গেলে অসুবিধার সম্মুখীন হন কি? ক) হ্যাঁ খ) না

১৪. উত্তর 'হ্যাঁ' হলে তা কী ধরনের? ক) বানানগত খ) বাক্যগঠনগত গ) অন্যান্য

১৫. বাংলা বলতে গেলে অসুবিধার সম্মুখীন হন কি? ক) হ্যাঁ খ) না

১৬. উত্তর 'হ্যাঁ' হলে তা কী ধরনের? ক) উচ্চারণগত খ) বাক্যগঠনগত গ) অন্যান্য

১৭. বাংলা ভাষার বাক্যে পদ অদল-বদল করা যায় কি? ক) হ্যাঁ খ) না

১৮. বাংলা বাক্যে সাধারণত কর্তা কোথায় বসে? ক) প্রথমে খ) মাঝে গ) শেষে

১৯. বাংলা বাক্যে সাধারণত কর্ম কোথায় বসে? ক) প্রথমে খ) মাঝে গ) শেষে

২০. বাংলা বাক্যে সাধারণত ক্রিয়া কোথায় বসে? ক) প্রথমে খ) মাঝে গ) শেষে

TASK 1 & TASK 2

Both have been illustrated in Chapter Three.

TASK 3

RECORDED FIELD INTERVIEW

Interview A: Questions for young people

১. আপনার নিজের সম্পর্কে বলুন।

২. আপনার পরিবার সম্পর্কে বলুন।

৩. আপনার গ্রাম/শহর সম্পর্কে বলুন।

৪. আপনার ছেলেবেলার কোনো গল্প বলুন।

৫. আপনার প্রজন্মের সঙ্গে আগের প্রজন্মের পার্থক্য কোথায় আছে বলে মনে করেন?

৬. এখন যারা শিশু, তাদের সঙ্গে আপনার প্রজন্মের পার্থক্য কোথায় বলে মনে করেন?

৭. আপনার শিক্ষাপ্রতিষ্ঠান /কর্মক্ষেত্র সম্পর্কে বলুন।

৮. বাংলা ভাষার বর্তমান অবস্থা কীরূপ বলে মনে করেন?

৯. ভবিষ্যতে বাংলা ভাষার ভাষা-পরিস্থিতি কেমন হবে বলে মনে করেন?

১০. আপনার জীবনের লক্ষ্য সম্পর্কে বলুন।

Interview B: Questions for older people

১. আপনার নিজের সম্পর্কে বলুন।

২. আপনার পরিবার সম্পর্কে বলুন।

৩. আপনার গ্রাম/শহর সম্পর্কে বলুন।

৪. আপনার ছেলেবেলার কোনো গল্প বলুন।

৫. মুক্তিযুদ্ধের সময়ের কোনো অভিজ্ঞতা সম্পর্কে বলুন।

৬. আপনার প্রজন্মের সঙ্গে আগের প্রজন্মের পার্থক্য কোথায় আছে বলে মনে করেন?

৭. এখন যারা শিশু, তাদের সঙ্গে আপনার প্রজন্মের পার্থক্য কোথায় বলে মনে করেন?

৮. আপনার শিক্ষাপ্রতিষ্ঠান /কর্মক্ষেত্র সম্পর্কে বলুন।

৯. বাংলাদেশে গণতন্ত্রের বর্তমান অবস্থা কেমন বলে মনে করেন?

১০. বাংলা ভাষার বর্তমান অবস্থা কীরূপ বলে মনে করেন?

১১. ভবিষ্যতে বাংলা ভাষার ভাষা-পরিস্থিতি কেমন হবে বলে মনে করেন?

References

Aboh, E. K. (2007). The interaction of focus and grammar in Niger-Congo and Afro-Asiatic. In E. K. Aboh, *Focus strategies in African Languages.* Berlin: Mouton de Gruyter.

Abraham Werner & Laszlo Molnárf. (2002). The German Clause under discourse functional weight: Focus and Antifocus . In I. W.-W. (eds.), *Issues in formal German(ic) typology* (pp. 1-44). Amsterdam: J. Benjamins.

Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford University Press.

Adli, A. (2005). Gradedness and Consistency in Grammaticality Judgments . In S. &. In Kepser, *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives* (pp. 7-25). Berlin: Mouton de Gruyter.

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht: Kluwer.

Asher, N. (1993). *Reference to abstract objects in discourse*. . Dordrecht: Kluwer.

Asher, Nicholas & Marie Vieu. . (2005). Subordinating and coordinating discourse relations. *Lingua 115*, 591-610.

Azad, H. (1994). *Bakkyatatto.* Dhaka: University of Dhaka.

Bard, E. G. (1996). Magnitude Estimation of Linguistic Acceptibility. *Language*(72(1)), 32-68.

Bayer, J. &. (2011). Bengali emphatic clitics in the lexicon-syntax interface. In W. U. Ed. by Dressler, *Contemporary Morphology, TiLSM* (pp. 3-16). Mouton De Gruyter.

Bergman, M. M. (2008). Introduction: Whither mixed methods? In M. M. (ed.), *Advances in Mixed Methods Research* (pp. 1–7). Sage.

Bhattacharya T et al. (2007). *Argument Structure*. John Benjamins.

Bley-Vroman, Robert & Naoko Yoshinaga et al. (1992). *Broad and narrow constraints on the English dative alternation: some fundamental differences between native speakers and foreign language learners.* University of Hawai'i.

Brunetti, L. (2004). *A unification of focus.* Padova: Unipress.

Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/merge/ fuse quantitative and qualitative research? . In M. M. (ed.), *Advances in Mixed Methods Research* (pp. 87–100). Sage.

Caracelli, V. J. (1997). Crafting mixed-method evaluation designs . In J. C. (eds), *: Advances in Mixed Methods Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms* (pp. 19-32). Jossy-Bass.

Cardinaletti, A. (2004). Toward a cartography of subject positions. In I. L. (ed.), *The Structure of CP and IP, The Cartography of Syntactic Structures* (Vol. Vol. 2, pp. 115-165). Oxford: Oxford University Press.

Casielles-Suárez, E. (2004). *The Syntax-Information Structure Interface: Evidence from Spanish and English.* New York: Routledge.

Cheshire, Jenny & Dieter Stein. (1997). The syntax of spoken language. In Cheshire, Jenny. In J. &. In Cheshire, *Taming the Vernacular:From Dialect to Written Standard Language.* London: Longman.

Chomsky, N. (1971). Deep Structure, Surface Structure, and semantic interpretation. In D. D. (ed.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (pp. 183-216). New York: CUP.

Chomsky, N. (1995). *The Minimalist Program.* . Cambridge, MA: MIT Press.

Chomsky, N. (2005). *On Phases.* Ms: MIT.

Cinque, G. (1990). *Types of A' dependencies.* Cambridge, MA: MIT Press.

Cinque, G. (1997). *Adverbs and Functional Heads.* Oxford: Oxford University Press.

Collins et al. (2007). A mixed methods investigation of mixed methods sampling designs in social and health science research . *Journal of Mixed Methods Research, 1/3*.

Contreras, H. (1976). *A theory of word order with a special reference to Spanish.* Amsterdam: North-Holland.

Creswell et al. (2008). Methodological issues in conducting mixed methods research designs. In M. M. (ed.), *Advances in Mixed Methods Research* (pp. 66–83). Sage.

Dahl, O. (1974). Topic-comment structure revisited . In I. O. (ed.), *Topic and comment, contextual boundness and focus* (pp. 1-24). Hamburg: Helmut Buske.

Dasgupta, P. (2007). A dependency syntax of Bangla. *Indian Journal of Linguistics, 25-26*, 15-97.

Diesing, M. (1992). *Indefinites.* Cambridge, Mass: M.I.T. Press.

Do¨rnyei. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies.* Oxford University Press.

Emonds, J. (2004). Unspecified categories as the key to root projections. In D. C. In Adger, *Peripheries: Syntactic edges and their effects* (pp. 75-120). Dordrecht: Kluwer Academic.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface.* New York: OUP.

Firbas, J. (1964). On defining the theme in functional sentence perspective. *Travaux linguistiques de Prague 2*, 267-280.

Grosz, Barbara & Candace Sidner. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12 (3)*, 175-204.

Gundel, J. (1988 [1974]). *The role of topic and comment in linguistic theory.* New York: Garland.

Gundel, J. K. (1994). On different kinds of focus. In I. P. eds. (Ed.), *Focus and Natural Language Processing. Vol. 3*, pp. 457-466. Heidelberg: Working Papers of the Institute for Logic and Linguistics. IBM Deutschland.

Hajičová, E. (1987). Focusing – a meeting point of linguistics and artificial intelligence. In P. &. In Jorrand, *Artificial Intelligence II: Methodology, systems and applications* (pp. 311-321). Amsterdam: Elsevier/North-Holland.

Halliday, M. (1967). Notes on transitivity and theme in English. *Journal of Linguistics 3*, 37-81, 199-244.

Heim, I. (1982). *The Semantics of Defmite and Indefinite Noun Phrases.* Amherst: University of Massachusetts.

Henry, A. (2005). Non-Standard Dialects and Linguistic Data. *Lingua, 115 (11)*.

Hobbs, J. R. (1985). *On the coherence and structure of discourse.* Center for Study of Language and Information.

Hockett, C. (1958). *A Course in Modern Linguistics.* New York: Macmillan.

Horn, L. R. (1981). Exhaustiveness and the semantics of clefts. *NELS 11*, 125-142.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar.* Cambridge, MA: MIT Press.

Kallestinova, E. (2007). *Syntactic aspects of word order in Russian.* Iowa City: University of Iowa.

Kamp, J. A. (1981). A Theory of Truth and Semantic Representation. In T. J. In: J. Groenendijk, *Formal Methods in the Study of Language* (pp. 277-321). Amsterdam: Mathematical Centre.

Kiss, K. E. (1998). Identificational focus versus information focus. *Language, 74*, 245-273.

Koster, J. (1986). *The Relation between Pro-drop, Scrambling, and Verb Movements.* . Groningen: University of Groningen.

Kratzer, A. (1989). Stage-level and individual-level predicates . In E. g. AnIn Bach, *Papers on quantification.* Amherst: University of Massachusetts.

Kratzer, Angelika & Elizabeth Selkirk. (2007). Phase theory and prosodic spellout: the case of verbs. *The Linguistic Review, 24*, 93-135.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

Larsen-Freeman, D. and L. Cameron. (2008). *Complex Systems and Applied Linguistics*. Oxford University Press.

Lopez, L. (2009). *A derivational syntax for Information Structure*. Oxford: Oxfrod University Press.

Majumdar, B. (1920). *The History of the Bengali Language*. New Delhi: Asian Educational Service.

Mathesius, V. (1975 [1961]). A functional analysis of present day English on a general linguistic basis. In J. V. Libuše Dušková (transl.). Prague: Academia.

May, R. (1977). *The Grammar of Quantification*. MIT.

Miyagawa, S. (1997). Against optional scrambling. *Linguistic Inquiry, 28*, 1-26.

Morgan, D. L. (2007). Paradigms lost and prag¬matism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1/1*, 48–76.

Nash, L. (1995). *Argument scope and case marking in SOV and in ergative languages:the case of Georgian*. Paris: Université de Paris.

Nastasi, B.K. et al. (2007). Mixed methods in intervention research: Theory to adaptation. *Journal of Mixed Methods Research, 1/2*.

Notes on Transitivity and Theme in English. (1967). *Journal of Linguistics, 3*(Part 2), 199-244.

Onwuegbuzie et al. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4/1*.

Parafita Couto, M. d. (2005). *Focus at the Interface*. Lawrence, KA: University of Kansas.

Partee, B. (1992). Topic, Focus, and Quantification. *SALT I Proceedings*, 159–189.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics, 12*, 601-638.

Pollock, J.-Y. (1989). Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry, 20 (3)*, 365-424.

Prince, E. F. (1986). On the syntactic marking of presupposed open propositions. . *CLS 22*, 208-222.

Reinhart, T. (1995). Interface strategies. *OTS Working Papers in Theoretical Linguistics 02-001*. Utrecht: OTS, Utrecht University.

Reinhart, T. (2006). *Interface Strategies*. Cambridge, MA: MIT Press.

Rizzi, L. (1997). The fine structure of the left periphery. In I. L. (ed.), *Elements of Grammar.* Kluwer.

Rizzi, L. (2004). On the Form of Chains: Criterial Positions and ECP Effects. In L.-S. &. In Cheng, *Wh-movement: moving on*. Cambridge, MA: MIT Press.

Rochemont, M. (1986). *Focus in generative grammar.* Amsterdam: Benjamins.

Rooth, M. (1985). *Association with focus*. USA: University of Massachusetts, Amherst.

Rooth, M. (1985). *Association with focus*. Amherst: University of Massachusetts.

Sarkar, P. (2006). *Bangla Byakaran Prasanga*. Kolkata: Dey's .

Sasse, H.-J. (1987). The Thetic/Categorical distinction revisited. *Linguistics, 25 (3)*, 511-580.

Schütze, C. (1996). The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. Chicago: University of Chicago Press.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. . Cambridge/London: MIT Press.

Selkirk, E. (1996). Sentence prosody: Intonation, stress, and phrasing. In J. G. (ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge: Wiley-Blackwell.

Sgall, Peter, Eva Hajičová, & Jarmila Panenová. (1986). *The meaning of the sentence in its semantic and pragmatic aspects.* Dordrecht: Reidel.

Spyropoulos, V. &. (2007). Subject chains in Greek and PF processing. *Paper presented at the Workshop on Greek Syntax and Semantics*. Cambridge: MIT.

Steedman, M. (2000). *The Syntactic Process.* MIT Press: Cambridge, MA.

Stoyanova, M. (2008). *Unique Focus: Languages withot Multiple wh-questions.* USA: John Benjamin Publishing Co.

Sudhoff, S. (2010). *Focus Particles in German: Syntax, Prosody, and Information Structure.* Amsterdam/Philadelphia: John Benjamins.

Szendrői, K. (2001). *Focus and the syntax-phonology interface.* London: University College.

Szendrői, K. (2004). Focus and the interaction between syntax and pragmatics. *Lingua , 114*, 229-254.

Uriagereka, J. (1999). Multiple spell-out. . In S. D. In Epstein, *Working Minimalism* (pp. 251-282). Cambridge, MA: MIT Press.

Vallduví, E. (1990). *The Informational Component.* Philadelphia: University of Philadelphia.

Vallduví, E. (1992). *The Informational component.* New York: Garland.

Vallduvi, Enric; Elisabet Engdahl. (1996). The linguistic realization of information packaging. *Linguistics 34*, 459-519.

van Kuppevelt, J. (1995). Main structure and side structure in discourse. *Linguistics, 33*, 809-833.

Ward, G. (1988). *The semantics and pragmatics of preposing.* New York: Garland.

Wasow, T. (2007). Gradient data and gradient grammars. In I. M. (ed.) (Ed.), *CLS 43: Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, (pp. 255-271).

Wasow, T. (2007.). Gradient data and gradient grammars. In I. M. (ed.) (Ed.), *CLS 43: Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society* (pp. 255-271). Chicago: Chicago Linguistics Society.

White, L. (2003). *Second Language Acquisition and Universal Grammar.* Cambridge: Cambridge University Press.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools, 13(1)*, 41-47.

Zubizarreta, M. L. (1998). *Prosody, Focus, and Word Order.* Cambridge, MA: MIT Press.

= = =