

Towards an Alternative Description of
Incomplete Sentences in Agglutinative Languages

S. Ido

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

2001

University of Sydney

I declare that this thesis is all my own work. I have acknowledged in formal citation the sources of any reference I have made to the work of others.

Shinji Ido

Date

Title:

**Towards an Alternative Description of
Incomplete Sentences in Agglutinative Languages**

Abstract:

This thesis analyses ‘incomplete sentences’ in languages which utilise distinctively agglutinative components in their morphology.

In the grammars of the languages dealt with in this thesis, there are certain types of sentences which are variously referred to as ‘elliptical sentences’ (Turkish *eksiltili cümleler*), ‘incomplete sentences’ (Uzbek *to‘liqsiz gaplar*), ‘cut-off sentences’ (Turkish *kesik cümleler*), etc., for which the grammarians provide elaborated semantic and syntactic analyses.

The current work attempts to present an alternative approach for the analysis of such sentences. The distribution of morphemes in incomplete sentences is examined closely, based on which a system of analysis that can handle a variety of incomplete sentences in an integrated manner is proposed from a morphological point of view. It aims to aid grammarians as well as researchers in area studies by providing a simple description of incomplete sentences in agglutinative languages.

The linguistic data are taken from Turkish, Uzbek, and Japanese, with special reference to (Bukharan) Tajik.

CONTENTS

Acknowledgment	v
Preface	vii
Notes on the Text	ix
1 INTRODUCTION	1
1.1 Introduction	1
1.1.1 Information Structure and Agglutinative Languages	3
1.1.2 Morpheme-based Analysis and Observability of Information Structure	5
1.1.3 Definition of Information Structure for This Thesis	11
1.1.4 Summary	12
1.2 Agglutination	12
1.2.1 Agglutinative Morphology	12
1.2.2 Agglutination and Fusion	17
1.2.3 Remarks on the Languages Used in This Thesis	22
1.2.3.1 Turkish	24
1.2.3.2 Uzbek	26
1.2.3.2.1 General Remarks	26
1.2.3.2.2 Orthography	27
1.2.3.3 Bukharan Tajik	29
1.2.3.3.1 Tajik	29
1.2.3.3.2 Agglutination in Tajik	30
1.2.3.3.3 Tajik Orthography	32
1.2.3.3.4 Bukharan Tajik	33
1.2.3.3.4.1 Tajik and Bukharan Tajik	33
1.2.3.3.4.2 Characteristics	34
1.2.3.3.4.3 Agglutination in Bukharan Tajik	36
1.2.3.4 Japanese	37
1.2.3.5 Summary	39
1.3 Summary of the Chapter	39
2 MORPHEME-BASED SYSTEM OF INFORMATION STRUCTURE ANALYSIS	41
2.1 Introduction	41
2.2 Preliminary Notes	41
2.3 ‘Old/New Information’: A Review of previous works	42
2.3.1 Old/New Information	42

2.3.2	Units Associated with ‘Old/New Information’	45
2.3.3	Summary	48
2.4	Morpheme-based system	49
2.4.1	Parameters and Constants	49
2.4.2	Variability and Values	50
2.4.3	Variables and Non-Variables	55
2.4.3.1	Allosentences	56
2.4.3.2	Variables and Non-Variables	57
2.4.3.3	Summary	61
2.4.4	Analysis Using Variables and Non-variables	61
2.4.4.1	‘Ellipsis’ or ‘Inaudible Elements’	63
2.4.4.1.1	Rule 1	68
2.4.4.1.2	Variable = Free Morpheme	69
2.4.4.1.3	Variable = Bound Morpheme	75
2.4.4.1.4	Multiple Variables	81
2.4.4.1.5	Notes on “Variable = Phrase”	93
2.4.4.1.6	Notes on Variable = Absence	94
2.4.4.1.7	Summary for the ‘Ellipsis’ Subsection	96
2.4.4.2	Potential (Partial) Applicability of Rule 1 in Fusional Morphology	97
2.4.4.3	Language Specificity in the Operation of Rule 1-2	99
2.4.4.3.1	Orthographic Words	100
2.4.4.3.2	Summary	103
2.4.4.4	Echoing and Unattended Parameters	103
2.4.4.4.1	Echoing	103
2.4.4.4.2	Unattended Parameters	105
2.4.4.4.2.1	Parameters Unattended/Attended by Intention	107
2.4.4.4.3	Summary	109
2.4.4.5	Pre-grammatical Sentences	109
2.4.4.5.1	Pre-Grammatical but not Un-grammatical	110
2.4.4.5.2	Child Language	112
2.4.4.5.2.1	Language Learning	112
2.4.4.5.2.2	Children’s Pre-grammatical Sentences	118
2.4.4.5.2.3	Adults’ Pre-grammatical Sentences	120
2.4.4.5.2.4	Pre-grammatical Sentences with ‘Dropped’ Case Markers	126
2.4.4.5.3	Summary of ‘Pre-grammatical Sentences’	132

2.4.5	Summary of ‘Analysis Using Variables and Non-variables’	133
2.5	Summary of the Chapter	133
3	PROMINENCE	135
3.1	Introduction	135
3.2	Information Structure and Prominence	136
3.3	Data	138
3.4	Terminology	138
3.5	Prominence	140
3.6	Overview of Possible Phonetico-Prosodic Nature of Prominence	141
3.7	Phonetic Nature of Prominence	142
3.8	Review of Previous Works	146
3.9	Basis of Parsing	149
3.10	Unit with which Prominence is Associated	153
3.11	Summary	156
3.12	Prominence – Morpheme	157
3.13	Prominence – Variables	158
3.13.1	Variable = Free Morpheme	159
3.13.2	Variable = Bound Morpheme	160
3.13.3	Summary	164
3.14	Cases Where Information-Structural Constraints Rule Over Phonological Constraints	164
3.15	Variable-Prominence Correlation Overriding Phonological Rules	166
3.16	Summary	169
3.17	Summary of the Chapter	170
4	SUMMARY AND CONCLUSION	171
	Appendix 1: Glossary	174
	Appendix 2: Transliteration	175
	Appendix 3: A Short Bibliography of Languages in Bukhara	178
	REFERENCES	180

ACKNOWLEDGEMENTS

To two of my best friends, Leyla Alioğlu and Sarvar Rahmatullaev, my deep gratitude.

I would like to express sincere thanks to the following friends (who were also patient informants) and people without whose kindness and support this study could not have been completed:

Sherzod Abdurauf, Abdolsalam Adinehzadeh, Haluk L. Aka, Elvan Alptekin, Tilla Deniz Baykuzu, İlhame Caferova, Sveta Caferova, Ildar Davletchine, Tunay Ersez, Linar Gubaidulin, Zuhra Halimova, Muqaddas Hikmatova, Shohida Istanova, F. Jamolova, Oğuz Kuşhanoğlu, Mahmood Mahmoodian, the late M. M. Mirzaev's children, Necla Münevveroğlu, Barış Müstecaplıoğlu, İpek Müstecaplıoğlu, Oybek Namazov, Shukurjon Nazarov, Erkan Nazlı, Oya Nazlı, Zekiye Nazlı, Nodira Nuritdinova, Shōko Ōno, Ayling Rubin, Mubin Rahmatullaev, Muazzam Rahmatullaeva, Cemal Sarıcı, Nurettin Dino Şentürk, Moukim Temourov, and Rustam Xushvaqtoʻv.

I am very grateful to Hugh Clarke, William Foley, Nerida Jarkey, Aet Lees, Christopher Manning, Sumru Özsoy, and Jane Simpson for their helpful comments on proposals and drafts of this thesis and for their guidance and encouragement. I also thank Hamid Ne'matov for hours of pleasant discussion on information structure in Uzbek.

The present thesis has greatly benefited from Tajik and Uzbek books, articles, and microfilm which the staff of the Interlibrary Loans Section of Fisher Library spent much time and effort to obtain. My thanks go also to my Tajik teacher Qahramon

Tūxsanov at Bukhara State University who gave me books which proved to be essential in writing the present thesis.

Finally, I would like to thank the Australian Commonwealth Department of Employment, Education, Training, and Youth Affairs (DEETYA) and the University of Sydney. DEETYA and the University of Sydney helped me by awarding me the Overseas Postgraduate Research Scholarship and International Postgraduate Award, respectively.

PREFACE

This thesis is the result of my attempt to explain some aspects of agglutinative languages in a simple and psychologically convincing manner.

The present research started with studying certain types of sentences which are variably referred to as ‘elliptical sentences’ (Turkish *eksiltili cümleler*), ‘incomplete sentences’ (Uzbek *to‘liqsiz gaplar*, Tajik *jumlahoi nopurra*), ‘cut-off sentences’ (Turkish *kesik cümleler*), etc. in grammars of the languages dealt with in this thesis.

Such sentences, which I call collectively ‘incomplete sentences’ here, are analysed by the authors of the grammars into a number of disparate phenomena such as pro-drop and null subjects, case marker drop, deletion, etc., for each of which an elaborated syntactic or semantic analysis existed.

After a study of existing analyses of such sentences, analyses that make little reference to one another, I began to think of the possibility of an integrated account for these apparently diverse phenomena. This led me to a somewhat unconventional approach to incomplete sentences – I started to examine them from a morphological point of view.

It seemed to me that such phenomena which have been analysed in terms of syntax or semantics could be given simpler accounts if the distribution of morphemes in sentences in which the phenomena take place was examined closely. For any analysis, simplicity is a positive attribute – I therefore embarked on setting up a simple system of analysis that can handle a variety of phenomena yielding incomplete sentences in an integrated manner.

As the reader will immediately notice, simplicity and general accessibility are the key concepts in this system. Particular attention was paid to make the system conceptually tractible and independent of theoretical complications. The parsimonious nature of the system is obvious from the fact that a very limited number of principles have been posited. For example, all the linguistic data that this thesis contains are analysed in terms of the operation of only two principles.

The importance attached to simplicity is due partly to the fact that the accounts derived from the simple system seemed to comply with the psychological reality of many of my informants as well as myself. It is also due to another purpose of this thesis, namely to devise a system of analysis that is accessible to individuals whose areas of specialty are not within linguistics, e.g. researchers in area studies (of whom I am one), language teachers, etc.

I see the main contribution of this thesis as being not so much in the area of theoretical linguistics, in spite of some reference to works on pragmatics in the literature review and the employment of the term ‘information structure’, but more within the tradition of descriptive linguistics.

The system which has emerged from my analysis of the data was not constructed to fit any particular established theoretical framework. However, it benefits a great deal from previous linguistic work on information structure and I believe that it will be as much of interest to general linguists as it will be to teachers and students of Japanese, Turkish and other agglutinative languages.

NOTES ON THE TEXT

The editorial plural is used throughout the text, except when the use of it may cause confusion regarding the exact identity of its referent. My use of ‘we’ rather than ‘I’ for self-reference has more to do with editorial reasons than with stylistic considerations.ⁱ

Explanations of the transliteration systems for Tajik and Bukharan Tajik used in this thesis are found in the subsection 1.2.3.3. and appendix 2.

All examples in the present thesis are taken from my field notes and e-mails from native speakers, unless indicated otherwise. A large majority of Bukharan Tajik examples are taken from the language used by young Bukharans in their twenties who have had no formal education in standard (literary) Tajik.

ⁱ I had written a fair amount of text before I attempted to convert the editorial plural to the first person singular, which conversion proved to be difficult because the text contains citations as well as English translations of sentences in which first person plural pronouns and agreement morphology are used.

1. INTRODUCTION

1.1. Introduction

This thesis aims to propose a practical system of description of incomplete sentences in agglutinative languages. The system will be based on observation of data from several languages which utilise distinctively agglutinative morphology.

The current work, as will be clarified in the following discussion, attempts to aid grammarians of agglutinative languages by presenting an alternative approach for the analysis of incomplete sentences (i.e. sentences which have fewer components than is necessary according to the classical sentence grammar; see § 2.4.4). Therefore, the focus is primarily on practical description and only secondarily on existing theories of information structure, even though we refer to certain linguistic theories in this introduction and in chapter two.

The system of description of incomplete sentences that we propose in this thesis has several features that distinguish it from the systems found in many grammars.

Firstly, our system has its basis in an analysis of Turkish, Uzbek, and Japanese, with special reference to (Bukharan) Tajik. Concentrating on these languages, all of which have agglutination as a prominent component of their morphology (see § 1.2), brings about an interesting consequence. It prompts us to approach incomplete sentences in a radically different way from that used by many grammarians and attempt to explain them from a morphological point of view. This is the second feature of the system.

While existing grammars and linguistic theories typically explain incomplete sentences in terms of the lack of certain grammatical, syntactic, or lexical components (e.g. predicates, noun phrases, words), our system identifies non-occurrence of certain morphemes in these types of sentences. The feasibility of this morphological approach is in fact conditional on the first feature of our system, namely its concentration on languages with prominent agglutinative morphology. This is part of the reason why, as will be explained later, we restrict the applicability of the system to agglutinative languages.

The system is based on the assumption that there is one-to-one correspondence between morphemes and pieces of information in agglutinative languages (see § 2.4.3.2) and the

discussion in this thesis involves contrasting our system with linguistic theories of information structure.

For the purposes of the present work, the term ‘information structure’ (hereafter referred to in its abbreviated form as IS) is used very loosely to refer to any line of thought that relates, directly or indirectly, sentence form and information. Accordingly, we call the analysis which this thesis presents, a system of information structure. The use of this term for our system may be misleading, considering its non-theoretical, practical nature (indeed, the basic concept of our system is adopted from the work of Greenfield, a psychologist¹) and also, because it does not attempt to contest current theories of information structure in general linguistics. Wherever a distinction has to be made between our system and the prevailing theories of information structure, we will refer to the latter as ‘linguistic systems/analyses/theories’ of information structure.

In the languages from which we take data, morphemes coincide with pieces of information to a considerable degree. This may not be surprising since morphemes are usually defined as the smallest meaningful units. However, as the existence of so-called portmanteau morphs or words like ‘ran’ and ‘stood’ (as opposed to ‘run’ and ‘stand’, respectively) suggests, one-to-one correspondence between pieces of information/meaning and morphemes is not a salient feature in many languages, particularly in languages which have morphology characterised by its fusional (or replacive) nature.

On the other hand, in the morphology of agglutinative languages, this correspondence is relatively consistent. This enables us to establish direct links between morphemes and pieces of information, which, in turn, brings in observability of information in our analysis. The observability of distribution of pieces of information in sentences is the third of the distinctive features of our system of IS analysis.

In linguistic works on IS, pieces of information are typically ‘propositions’ which can often be expressed only with clumsy long sentences (see § 2.4.1). This is a natural consequence of linking syntactic phrases with pieces of information. On the other hand, in our analysis where pieces of information are linked with morphemes, pieces of information are usually identical with the ‘meanings’ of morphemes such as [negative],

¹ The term as it is used in our system is given a definition in section 2.4.3.2, which is not compatible with the concept used in the linguistic literature.

[plural], etc. As will be discussed later, this approach diminishes the descriptive ambiguity which seems almost inherent in IS analyses.

To sum up, the IS analysis which we propose in this thesis differs from most other IS analyses in the following aspects: It is based 1) on data taken from **agglutinative languages** and 2) on **morphemes**. As it will become clearer in the discussion in the next section, these features of our system of IS analysis are not independent from each other – the latter feature, which forms the backbone of what we will call ‘morpheme-based IS analysis’ (section 2.4.), would not have arisen in the absence of the former.

In this thesis, we will analyse data gathered from natural conversation among native speakers of the languages concerned to provide an explanation of the framework of our system. We aim to make our exposition clear and conceptually tractible, allowing the reader to reason about the implications of the system.

Accordingly, chapter two concentrates mainly on the major issue that is vital for the exposition of how the system works, following which it proposes a set of principles pertaining to the system. The chapter integrates the principles with the analysis of incomplete sentences. It also makes references to children’s incomplete sentences.

The third chapter extends the framework to prominence and discusses the relation between perceptive prominence and information structure. It discusses the abstraction of the prominence-information correlation as much as it explains the correlation itself.

In the remaining part of this chapter, the features of our IS analysis mentioned above are further explained in detail, which draw a distinguishing line between our system of analysis with those of many existing IS analyses. Laying out the system involves putting together pieces of supporting evidence from different disciplines and different branches of linguistics.

1.1.1. Information Structure and Agglutinative Languages

Most of the theories of information structure written in English are based mainly on analyses of major Indo-European languages (e.g. Lambrecht 1994), English, French, and German being the most popular. While there are exceptions to this tendency for major European languages to be the sole source of data in constructing theories of information structure, and there exist theories which are based primarily on analyses of non-European

language(s), such as Kamio and Takami (1998), their number is few. As a result, English (or some other languages which the theoreticians happen to know in addition) dominates the western and western-influenced study of information structure.

We think, in linguistics, this concentration on data from the ‘big’ languages has biased the study of information structure in such a way that the study deals mainly with information structure in major Indo-European languages, whose morphology is known for its relatively prominent fusional (or inflectional) nature. This in turn has made linguistic theories of information structure look, at least in part, analogous to theories of the information structure of languages with salient fusional morphology.

The present thesis, which is not in theoretical linguistics, goes contrary to this trend in the linguistic study of IS in setting its basis on languages with distinctively agglutinative morphology. Two of them are not even ‘major’. As we will see later, agglutinative morphology provides a unique perspective from which to observe information structure in sentences. More importantly for us, it makes a morpheme-based analysis of information structure feasible. This perspective leads us to tailor a system of IS analysis specifically for agglutinative morphology.

The concentration of IS analyses on few major European languages, particularly English, seems to have had an influence also on the way the study of IS in general is perceived by linguists. Despite the fact that many linguistic works on information structure are in most cases synonymous with works on information structure in English, some authors of such works do not hesitate to claim cross-linguistic applicability of their theories or systems of IS analysis. The following is an example of such a claim:

The book contains relatively few analyses of linguistic data, and most of these data are from English. However, the principles discussed have wide crosslinguistic applicability. (Lambrecht 1994:xv)

Perhaps unlike many linguists, we believe it is unrealistic to claim the wide cross-linguistic applicability of principles elicited from data taken from only a few languages. We do not know whether the principles proposed in Lambrecht’s book have wide cross-linguistic applicability. However, we do suspect that the principles would have been different from what they are, if, say, the data had been from the Turkic languages.

The present thesis contains analyses of Turkish, Japanese, Uzbek, and Bukharan Tajik data. In other words, our analysis will be based almost entirely on data from languages

other than major European languages. It is almost certain that our concentration on data from these languages will yield a bias which is not dissimilar to the one which the concentration on English data brought about in linguistic work on information structure. We therefore refrain from claiming wide cross-linguistic applicability of the principles which will be discussed in this thesis. Accordingly, instead of insisting on the wide cross-linguistic applicability of our principles, we aim to tailor a system of analysis applicable specifically to languages in which one morphological characteristic is prominent – we posit principles for languages with a strong agglutinative nature, which comprise the languages from which we take examples. No universality of the principles set out in this thesis will be claimed.

As we shall see in the next chapter, while agglutinative languages are the source of data on which we posit a system of information structure analysis, their agglutinative morphology is conversely what defines our system of information structure analysis. Agglutination and the consequent segmentability of words into readily identifiable morphemes provide us with the possibility of analysing information structure of sentences on the basis of morphology. Thus, agglutinative morphology is itself a factor which inspires us to posit a system of information structure analysis distinct from linguistic IS analyses.

1.1.2. Morpheme-Based Analysis and Observability of Information Structure

Concentration on data from agglutinative languages is thus one chief feature of our IS analysis. This feature, combined with ready segmentability of words which is characteristic of agglutinative morphology (see the section ‘Agglutination’), gives rise to another crucial feature that characterises our system of IS analysis. This feature, which is also another feature that distinguishes our system from many other systems, is ‘observability’ of information structure.

In this thesis, we attach particular importance to observability of information structure in sentences. How, though, can information structure be observable? Information does not have physical representation in the same way as, say, tangible ‘things’ such as a pen and table do. If information itself does not have an observable attribute, how do we observe the structure of information in sentences? The inevitable answer to this question would be ‘by assigning pieces of information to tangible or observable things’, because

observability of pieces of information can only be obtained by giving ‘tangible’ bodies to them.

In sentences, the observable ‘things’ to which information is assigned are parts of sentences. Assignment of information to parts of sentences is therefore inevitable in discussing information structure and linking pieces of information with parts/segments of sentences is an approach employed, in one form or another, in virtually all work on information structure. Examples of segments of speech or sentences being assigned information are easy to find in any IS analyses. For example, in the following examples taken from Kamio and Takami (1998:120-121), subjects and predicates are assigned ‘old’ and ‘new’ information (This example will be revisited in chapter two):

- | | | |
|----|-------------------|-----------------|
| A: | What did John do? | [Kamio’s (22)] |
| B: | He hit Mary. | |
| | old information | new information |
| A: | Who hit Mary? | [Kamio’s (23)] |
| B: | John hit her. | |
| | new information | old information |

In our system too, pieces of information are assigned to parts of sentences. In this respect, our analysis is not different from analyses in linguistics. The uniqueness of our system lies in the specific parts of sentences to which pieces of information are assigned. We assign pieces of information not to grammatical categories (as in the analysis cited above) or syntactic phrases (as in many works in linguistics), but to morphemes. By linking morphemes with pieces of information, our system contrasts with many analyses of information structure which typically associate syntactic or grammatical units with pieces of information.

There are understandable reasons for the preference of syntactic units to morphemes as units to which information is directly associated. One reason for not associating pieces of information to morphemes can be the existence of such words as ‘run’ and ‘ran’ – predictably, an attempt to identify a morpheme with which, say, the information of [past] is associated in ‘ran’ has to be faced with difficulty. These sorts of words which are typical of fusional (or replacive) morphology naturally discourage the linkage of pieces of information with morphemes.

Another reason could be the existence of function morphemes which some think are often not capable of providing the kind of information which content morphemes manages to provide (see Lambrecht 1994:215-216). We will discuss these issues at length in the next chapter where we will see that some of the factors that prevent linking of pieces of information and morphemes are unique to languages with highly fusional morphology and hence do not cause major problems in our analysis of agglutinative languages.

What is, though, the benefit of positing morpheme-based analysis of IS? There is at least one obvious benefit, namely reduction of ambiguity of explanations which is present in a number of studies of IS. How a morpheme-based analysis reduces the ambiguity will be explained in the next few paragraphs.

Analysis of IS usually involves segmentation of information into sizeable pieces. Some works also label segmented pieces of information as ‘new’ or ‘old’. The pieces of information are, then, assigned to parts of sentences – this is where a potential problem (at least for the aim of the present thesis) lies.

As was stated above, many of linguistic IS analyses assign pieces of information to syntactic or grammatical units such as noun phrases or predicates. Such analyses either deny or neglect the possibility of pieces of information being assigned to units which are ‘smaller’ than syntactic/grammatical units, such as words and morphemes. Naturally, in such analyses, information is not segmentable to the degree that pieces of information can be assigned to morphemes. This insegmentability provokes ambiguity as to, say, whether the ‘-ed’ [past] in the following dialogue taken from the previously cited Kamio’s examples with modification conveys a ‘new’ piece of information despite its being in the syntactic unit conveying ‘new’ information:

[1] What did John do?
He [killed Mary].

The information [past] does not appear to be ‘new’, unless ‘newness’ of information is defined in a very counter-intuitive manner (‘new/oldness’ of information will be discussed in 2.3). This in turn diminishes the reasonableness of identifying ‘killed Mary’ as a unit to which a chunk of information classified as ‘new’ is assigned, because the unit appears to comprise a part of the sentence that conveys ‘old’ information – ‘-ed’ which seems to be ‘old’ information appears in the unit which is claimed to be ‘new’. Is ‘-ed’,

then, ‘new’ or ‘old’? This kind of ambiguity seems unavoidable in IS analyses which associate syntactic units with pieces of information.

On the other hand, in our analysis which will be proposed in the next chapter, breaking down of information to pieces encoded in individual morphemes allows us to distribute pieces of information which are more or less in direct correspondence with glosses for morphemes in a sentence. This allows us to say, for instance, that in the Turkish equivalent of ‘He killed Mary’ in the example above *Mary-’i öl-dür-dü* [Mary-acc die-caus-past.3sg] ‘(he) hit Mary’, *-dü* [past.3sg] is not an encoder of a ‘new’ piece of information. In other words, the kind of ambiguity exemplified in the example above cannot occur if morphemes are units linked to pieces of information.

IS analyses based on syntactic units also have the tendency of devising layers of abstract representations of information which intervene between information and its linguistic representations. Such abstract representations of information are meant to eliminate the kind of ambiguity explained above, but, to the contrary, they often seem to contribute to the ambiguity.

Observe, for example, Lambrecht’s (1994:226) analysis² of what he terms the ‘predicate-focus structure’. The following is presented as an answer to the question ‘What happened to your car?’:

Sentence:	<i>My car broke DOWN.</i>
Presupposition:	“speaker’s car is a topic for comment x”
Assertion:	“x = broke down”
Focus:	“broke down”
Focus domain:	VP

² Lambrecht organises linguistic IS theories based on the traditional ‘old-new’ view of information in a unique way and, with the addition of a number of original insights, proposes an original theory of information structure. A brief review of his theory is a good starting point for our discussion, because 1) it comprises many of the previously posited ideas of information against which our approach will be contrasted and 2) our approach necessitates postulation of ideas which Lambrecht goes to a great length to invalidate. He emphasises throughout the book that information is not conveyed by lexical items or individual sentence constituents (Lambrecht 1994:207). On the other hand, in our approach, morphemes are agents of ‘value-specification’ in our analysis (see next chapter) – they may not be ‘conveyers’ of information in the exact sense of the word, but are certainly units with which pieces of

Lambrecht's 'focus' is (ibid:213):

The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition.

Presupposition and assertion, on the other hand, are defined thus (ibid:52):

PRAGMATIC PRESUPPOSITION: The set of propositions lexicographically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered.

PRAGMATIC ASSERTION: The proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered.

It is easy to notice that there is a great deal of abstraction³ involved in Lambrecht's system of IS analysis and that only 'focus' is directly associated with specific parts of sentences, which he calls 'focus domain'.

He links the 'semantic component' of 'focus' to the syntactic unit VP. In doing so, he naturally, and indeed necessarily, acknowledges the piece of information 'broke down'. However, identifying this piece of information as the 'focus' appears to be problematic, because it seems to be certain that at least the information that the tense in which something happened to the speaker's car [past] is in the 'presupposition'.

In other words, the 'presupposition' seems to consist not only of "speaker's car is a topic for comment x", but includes at least the information of tense (lexicographically) evoked by 'broke' as opposed to 'break'. This seems to entail that the 'focus' should not contain the information [past] (see Lambrecht's definition of 'focus' above). However, Lambrecht's 'focus' does contain the information [past] as it is clear from Lambrecht's presentation of the 'focus': 'broke down'. This fact does not seem to comply with his definition of focus.

Here we find the same problem Kamio's analysis has – a part of a sentence which is claimed to be associated with 'new' information has an 'old' part in it. We can see that positing abstract representations of information, e.g. focus and topic, does not free a theory of IS from the kind of problem explained above with the example 'He killed

information are associated. Naturally, this makes our approach incompatible with Lambrecht's and, indeed, most of the previously proposed theories of information structure.

Mary'. While it allows one to assign pieces of information to syntactic units and avert the difficulty of dealing with the IS of units 'smaller' than the syntactic ones, it seems to necessarily abstract IS at lexical and morphological levels.

It appears typical of IS analyses involving assignment of pieces of information to syntactic units to have this kind of disregard of IS at sub-syntactic levels. Many IS analyses seem to have a general tendency of abstracting information at the morphological level, showing general unwillingness to recognise information that words/morphemes may encode in their own right.

For example, Lambrecht's 'focus' is defined as a 'semantic component', which we take as some sort of piece of information. However, information that may exist at the lexical or morphological level is paid little attention in his book. His reasoning for his exclusion of sub-syntactic units from his discussion is as follows:

It follows from my definition of "focus" that focus domains must be constituents whose denotata are capable of producing assertions when added to presuppositions. As we shall see, such denotata are either predicates or arguments (including adjuncts), or else complete propositions. This entails that focus domains must be phrasal categories (verb or adjective phrases, noun phrases, prepositional phrases, adverbial phrases, and sentences. Focus domains cannot be lexical categories.

In the following chapter we will introduce a morpheme-based analysis of IS which contradicts this statement⁴, but that is not the point here. Our concern here is what makes him maintain that 'lexical categories' cannot be units with which pieces of information are associated. He goes on to say:

This is so because information structure is not concerned with words and their meanings, nor with the relations between the meanings of words and those of phrases and sentences,

³ These notions are, by nature, not associated directly to observable structures of language, but considered to affect linguistic expressions, i.e. observable structures of language.

⁴ In arguing against identification of lexical focus domains, he claims that the preposition 'to' in the following cannot be a focus domain, claiming that the focus domain is the syntactic phrase 'to the pigs', not the morpheme 'to':

And then, when we'd finished talking about pigs, we started talking TO the pigs.

However, as Polinsky (1999:575) points out in her review of Lambrecht's book:

[...] it is unclear why the presupposition 'We talked x' is superior to 'We talked x the pigs' or alternatively, 'Our talking and the pigs were related as x'.

but with the pragmatic construal of the relations between entities and states of affairs in given discourse situations. Entities and states of affairs are syntactically expressed in phrasal categories, not in lexical items.

It is clear from the above passage that his view on how IS should be defined is the basis of his exclusion of lexical (and also morphological) units from his discussion of IS. This means that, since Lambrecht himself provides his own definition of IS, the restriction of his discussion to syntactic units is a self-imposed one, which in turn means that there is no inherent characteristic of information that prevents us from assigning pieces of it to lexical or morphological items.

As explained before, many linguistic theories of IS assign pieces of information to syntactic units, but there are usually no attempts to examine the IS that may exist within the syntactic units. We can call this a ‘top-down’ approach to IS where syntactic phrases are assigned with pieces of information, abstracting the pieces of information the morphemes in the phrases may encode.

On the other hand, our system of analysis employs a ‘bottom-up’ approach – it focuses on pieces of information encoded by morphemes. No superimposition of pieces of information (such as ‘foci’ on syntactic units) is assumed in our analysis. This is the reason why we will rely on the defining character of agglutinative morphology, namely ready segmentability of words into morphemes, in proposing our system of IS analysis. It is also the reason why we restrict the assumed applicability of our system of analysis to highly agglutinative languages.

In summary, the system of IS analysis which we will develop in the following chapter is most saliently distinguished from many analyses of IS by being morpheme-based, thus resulting in observability.

1.1.3. Definition of Information Structure for This Thesis

The term ‘information structure’ has gained wide currency relatively recently. As a result of the newness of the term, the referential domain which the term covers intersects domains which are covered by such terms as ‘pragmatics’ and ‘discourse analysis’. There is also a rather novel term whose use is difficult to distinguish from that of IS, namely ‘information packaging’. We do not review the use of the terms or discuss the history of

the terminological confusion, since they are of little importance in our discussion. The reader is referred to Lambrecht (1994:section 1.1.) for a detailed explanation of the use of such terms. (As was stated in the introduction (§ 1.1), the term ‘information structure’ is used very loosely in this thesis to refer broadly to any line of thought that relates sentence form and information, whether it is psychological, philosophical, or linguistic.) The definition of IS for the system proposed in this thesis, which probably explicates little at this stage where we have not introduced notions employed in our system of IS analysis, is simple:

Information Structure: Distribution of variables and non-variables in sentences.

This definition should be self-explanatory when we proceed to the discussion in the next chapter.

1.1.4. Summary

In this section, we have explained the differences between our system of IS analysis and other (mainly linguistic) theories of IS. The differences can be summarised in a simple sentence: **our system is morpheme-based** while many others are phrase- or word-based. Our system will be proposed and explained with examples in the next chapter. However, before we proceed to the discussion of our system, explanations of agglutinative morphology and the languages which we use as sources of data are in order.

1.2. Agglutination

1.2.1. Agglutinative Morphology

As we have stated in the introductory section, this thesis concerns IS of languages with distinctively agglutinative morphology. Since the whole of our proposition made in this thesis will be in regard to such languages, we need a clear understanding of ‘agglutination’. Accordingly, we devote this entire section to clarification of agglutination as it is understood here and an explanation of agglutination in the languages which are the subject of this thesis. We also explain how agglutination is relevant to the system of analysis of IS that we will propose in the subsequent chapter. In doing so, we will focus on agglutination, keeping explanation of morphological language

typology which yielded the notion of agglutination to the minimum. The reader is referred to Shibatani and Bynon (1995) for an overview of various language typologies including morphological language typology.

The notion of agglutination has its root in the morphological typology of languages suggested by August von Schlegel (1818) who outlined three types of languages: ‘language without grammatical structure, affixing and inflectional languages’ (Schlegel 1818:559, cited in Greenberg 1974:37). These are, in later terms, three canonical types of language: isolating, agglutinative, and inflectional (or fusional)⁵.

The isolating type is often called analytic and contrasted with the other types which are called, as opposed to analytic, synthetic or polysynthetic. Languages of this type are characterised by some as languages which have no affixes (Isaev 1978:72). In Comrie’s (1989:43) description, an isolating language is ‘one which has no morphology, i.e. at least ideally, a language where there is one-to-one correspondence between words and morphemes’.

On the other hand, a fusional language ‘encodes relational meaning by modifying the lexical base by ‘true’ (that is to say, internal) inflection (as in English *sang*)’ (Shibatani and Bynon 1995:5). English ‘went’ as the past tense of ‘go’ may serve as an example of absolute fusional morphology. Typically, fusional languages have words which are not readily segmentable into morphemes, as can be seen in the following declension table of the Russian noun *den* ‘day’:

	Singular	Plural
Nominative	<i>den</i> ’	<i>dni</i>
Accusative	<i>den</i> ’	<i>dni</i>
Genitive	<i>dnja</i>	<i>dnej</i>
Dative	<i>dnju</i>	<i>dnjam</i>
Instrumental	<i>dnjom</i>	<i>dnjami</i>
Prepositional	<i>dne</i>	<i>dnjax</i>

Unlike in Turkish which we will observe below, we cannot segment, for instance, the genitive plural *dnej* [day.pl.gen] into the [day] morpheme, plural morpheme, and genitive morpheme. Moreover, the ‘fleeting vowel’ *e* in *den*’ eliminates the formal/morphological

⁵ To these, often the fourth type, incorporating (or polysynthetic), which originates in Wilhelm von Humboldt’s works, is added.

invariability of the root morpheme and thus undermines the ready segmentability of words in the table. (The ‘fleeting vowel’ is explained by Wade (1992:55): ‘[t]he vowel in the final syllable of many nouns which end in a hard consonant or soft sign does *not* appear in oblique cases’)

Lastly, there is the agglutinative type of morphology. Schlegel (1818:14ff, cited in Ramat 1987:204) calls agglutinative (or affixing) languages simply ‘languages which use affixes’. (This characterisation has problems as we will see below.) Turkish is probably the most frequently cited language in explaining agglutinative morphology. An example of agglutinative morphology where various suffixes such as the plural *-lar* and accusative *-ı* follow the noun *adam* ‘man’ is cited from Comrie (1989:44):

	Singular	Plural
Nominative	<i>adam</i>	<i>adam-lar</i>
Accusative	<i>adam-ı</i>	<i>adam-lar-ı</i>
Genitive	<i>adam-ın</i>	<i>adam-lar-ın</i>
Dative	<i>adam-a</i>	<i>adam-lar-a</i>
Locative	<i>adam-da</i>	<i>adam-lar-da</i>
Ablative	<i>adam-dan</i>	<i>adam-lar-dan</i>

When we compare this table with the Russian declension table previously cited, the difference between Turkish morphology and Russian morphology becomes apparent. Turkish words appearing in the table are characterised by their ready segmentability into morphemes which in turn is made possible by the invariability of the forms of the morphemes. As we will discuss in the paragraphs below, the ready segmentability is the defining character of agglutinative morphology.

Obviously, the canonical types of languages explained above are ideal and perhaps also conceptual types – ideally isolating/fusional/agglutinative languages probably do not exist. Since a language can and does exploit more than one of the isolating, fusional, and agglutinative morphologies, ‘the majority (perhaps all) of the world’s languages do not correspond exactly to one or other of these types’ (Comrie 1989:47). Shibatani and Bynon write (1995:5):

For example, English shows its isolating character in the encoding of modal meanings by independent words such as *will* and *may*, its agglutinative character in the regular plural

formation (e.g. *books*), and its inflectional character in the irregular plural and past tense formation (e.g. *feet, sang*).

Thus, the morphological typology explained above is not a system of strict categorisation where each category is mutually exclusive. Following Comrie (1989:46), we assume the morphological typology is designed to operate with two parameters, namely the ‘synthesis’ parameter:

[...] the number of morphemes per word, and its two extremes will be isolating and polysynthetic

and the ‘fusion’ or ‘segmentability’ parameter:

[...] the extent to which morphemes within the word are readily segmentable, its two extremes being agglutination (where segmentation is straightforward) and fusion (where there is no segmentability)

The first of these parameters, i.e. synthesis parameter concerns the isolation-polysynthesis axis, and hence does not concern agglutination. Only the latter of these parameters concerns us in this thesis, because segmentability is the defining character of agglutinative morphology. As is clear from the explanation above, agglutinative languages are most characteristically languages with ready segmentability of words into morphemes, which is explicitly observable in the table of Turkish inflection above. To put this in a different way, languages with ready segmentability are characterised as being agglutinative. It is in this sense that we call the languages which we deal with in this thesis ‘agglutinative languages’, though Japanese and Bukharan Tajik’s place in the ‘segmentability’ axis may not as high as the others’ (see § 1.2.3).

Thus, agglutinative languages are languages in which words are readily or easily segmentable. The ready segmentability of words which characterises agglutinative languages entails that, in an agglutinative language, each morpheme has its own ‘meaning’, in the sense that ‘individual exponents of relational categories are attached one by one to the lexical basis’ (Shibatani and Bynon 1995:5). Indeed, the idea that affixes ‘have an independent meaning’ in agglutinative language was present even in the original version of morphological typology by Schlegel (Greenberg 1960:180).

This ‘meaning’-morpheme correspondence can be exemplified by attaching a gloss i.e. ‘meaning’ to one of the sequences of morphemes appearing in the table of Turkish inflection presented before:

- [2] *adam-lar -a*
 man -pl -dat
 ‘to (the) men’

Note that the pieces of information [man] and [pl] are encoded with two separate morphemes in Turkish whereas in English, ‘men’ alone encodes both. This morphologically different encoding of the same information overtly contrasts the difference between segmentability which characterises agglutination and insegmentability which characterises fusion. (It may be possible, however, to consider the change of the ‘a’ in ‘man’ to the ‘e’ in ‘men’ to be a non-morphological segment that corresponds to the morphological segment, i.e. the morpheme, *-lar* [pl] in Turkish. This possibility of non-morphological segmentability will be discussed in § 2.4.4.2.)

The clearly observable one-to-one correspondence between ‘meaning’ and morphemes in the example above inspires us to suppose that information structure also somehow corresponds with morphemes in agglutinative morphology.

Assuming that ‘meaning’ or functions and morphemes correspond to a great extent (or completely, if segmentability is maximally high) in agglutinative languages, could we associate pieces of information with certain morphemes in sentences? – if we could, we could also ascribe changes in information that take place as a result of the utterance of sentences to the occurrence of certain morphemes.

For example, in the following question-answer pair, which will be discussed extensively in the next chapter, would it be possible to say that, not the entirety of (b), but only the negative morpheme *-ma* is responsible for narrowing down the possibilities of the answerer B having met Ali from [positive/negative] to [negative], despite the appearance of the verb stem *tanış-* [meet⁶], past tense marker *-dı*, and personal suffix *-m* in the sentence?:

- [3] a. *Ali-yile tanış-tı-n mi?*
 Ali-com meet-past-2sg Q
 ‘Did you meet Ali?’
- b. *tanış-ma-dı-m.*
 meet-**neg**-past-1sg

⁶ ‘Become acquainted’ is probably a more appropriate gloss for the verb *tanış-*, but we use in this thesis ‘meet’ for simplicity’s sake.

‘I didn’t’

These are some of the questions which prompted us to propose a system of analysis designed for languages with distinctively agglutinative morphology. Our answers to these questions is positive. The remaining part of this thesis is devoted essentially to link informational changes or lack of them directly to certain morphemes. This involves some reviews of previous theories of IS and, of course, detailed explanation of our system of IS analysis for agglutinative languages.

However, before we start setting up our system of IS analysis, some additional notes on fusional morphology and its segmentability, as well as brief descriptions of the languages from which examples are taken are in order.

1.2.2. Agglutination and Fusion

Morphological segmentability which is characteristic of agglutinative languages is indeed the factor that prompts us to devise a system of analysing information structure of sentences on the basis of morphemes, and we tailor in the next chapter a system specifically for agglutinative languages. However, even highly agglutinative languages can exploit fusional morphology just as highly fusional languages can exploit agglutinative morphology. Consequently, the distinction between the two types of languages is often subtle. We therefore adopt the parametric view of the agglutination-fusion distinction based on the segmentability of words into morphemes.

See, for example, Estonian and Finnish noun inflection paradigms from Comrie (1989:50-51) where segmentability is much higher in the Finnish paradigm, the Estonian paradigm exhibiting fusional morphology, despite the languages’ very close genetic relationship (*jalka* is ‘leg’ and *lippu* is ‘flag’):

	Finnish		Estonian	
Nominative singular	<i>jalka</i>	<i>lippu</i>	<i>jalg</i>	<i>lipp</i>
Genitive singular	<i>jala-n</i>	<i>lipu-n</i>	<i>jala</i>	<i>lipu</i>
Partitive singular	<i>jalka-a</i>	<i>lippu-a</i>	<i>jalga</i>	<i>lippu</i>
Partitive plural	<i>jalko-j-a</i>	<i>lippu-j-a</i>	<i>jalgu</i>	<i>lippe</i>

Finding examples of a language’s exploitation of both fusional and agglutinative types of morphology is also never difficult. English irregular and regular nouns (e.g. cup/cups but

man/men) and verbs (e.g. kill/killed but go/went) are apt examples. Even in Turkish, a textbook example of agglutinative languages, fusional morphology is easy to spot. Let us observe below how insegmentability of words in a language with highly agglutinative morphology causes disagreement among grammarians regarding the way the insegmentability is treated.

Turkish morphology is not entirely agglutinative and exhibits a fusional character in some components of its grammar. This fact has posed a problem for any attempts to account for Turkish morphology exclusively in terms of agglutination. This problem surfaces most typically in segmentation of verbs in the negative aorist.

Observe the following aorist verb paradigms. The words in the chart are segmented into morphemes in accordance with Kornfilt's (1997) analysis and glosses:

	Positive		Negative	
1sg	<i>oyna-r-im</i>	'I play'	<i>oyna-ma-m</i>	'I do not play'
2sg	<i>oyna-r-sin</i>	'you play'	<i>oyna-ma-z-sin</i>	'you do not play'
3sg	<i>oyna-r</i>	's/he plays'	<i>oyna-ma-z</i>	's/he does not play'
1pl	<i>oyna-r-iz</i>	'we play'	<i>oyna-ma-y-iz</i>	'we do not play'
2pl	<i>oyna-r-siniz</i>	'you play'	<i>oyna-ma-z-siniz</i>	'you do not play'
3pl	<i>oyna-r-lar</i>	'they play'	<i>oyna-ma-z-lar</i>	'they do not play'

The positive paradigm in the left column is straightforward, systematic, and, above all, distinctively agglutinative – the verb stem *oyna-* 'play' is followed by the aorist *-(I)r*, which is in turn followed by an agreement suffix. However, the negative paradigm where the negative morpheme *-mA* is inserted after the verb stem is obviously not as systematic as its positive counterpart. According to Kornfilt's (1997:337) analysis:

The liquid consonant of the aorist turns into a **z**, which is dropped in the first person singular, and which itself turns into a **y** in the first person plural [...]⁷

⁷ Ergin (1958/1997:295) notes that, from a diachronic point of view, *-mAz* can be divided into the negative *-mA* and de-verbal nominalising *-z*. Kornfilt's identification of *z* as an alternate of *-(A)r* has its predecessors in the grammars by Türeli (1968:35) and Swift (1963:85). According to Deny (1995:32), the *z* is a product of the sound change of *r* into *z*. There are some pieces of supporting evidence for this claim. For example, there are cognates of *-maz* ending with *r* in other Turkic languages spoken in the neighbouring areas of Turkey. Özkan (1996:146) notes in his Gagauz grammar that the aorist *-r* either retains its form or is turned into *-z* in the negative aorist in Gagauz, e.g. *al-ma-r-sin* or *al-ma-z-sin*

This analysis, which has its predecessor in Underhill’s textbook of Turkish (1976:147), has a support of diachronic evidence⁸. However, there are a number of different analyses for the negative aorist paradigm. The fact that there are a number of analyses of the paradigm, which do not comply with each other, is interesting, particularly because they all attempt to analyse the negative aorist paradigm entirely in terms of agglutination – this implies that it is not agglutinative morphology that is in operation in the paradigm.

Some of these analyses resort to the ‘morpheme-drop’ analysis in order to account for the apparent absence of the aorist morpheme in the negative first person singular, which is exactly what Kornfilt does. Some others try to explain the absence and the disturbing appearance of *y* in the first person plural by phonological rules. Let us observe below some of the mutually disagreeing analyses.

Kornfilt’s glossing is consistent with her analysis. Note that, in the following, the morpheme *-mA* is presented simply as a negative morpheme, not as a negative aorist morpheme (my bold-face):

oku-yá + **ma**-m (ibid:375)
 read-Abil.+**Neg.**-1.sg.
 “I am unable to / not permitted to read”

Lewis, on the other hand, refrains from segmentation and seems to regard *-mAz* as a single (portmanteau) morpheme encoding [neg] and [aor]. Interestingly, Kornfilt too, despite her analysis cited above, employs this segmentation in a gloss (my bold-face):

[...] gör-müş ol-a-**maz**-sın (ibid:352)
 see-Ppart be-Abil-**Neg.Aor.**-2sg.
 “You can’t have seen [...]”

These are not the only analyses of the negative aorist paradigm in Turkish. Türeli (1968:35), like Banguoğlu (1959:453), identifies the aorist morpheme as *-Iz* rather than *z*

[take-neg-aor-2sg]. According to Deny (1995:32), the Azerbaijani negative aorist retains the old form *-mer* of Turkish *-mez*.

⁸ In Turkish in the first half of the 15th century, where the negative aorist suffix is *-maz/-mez*, the suffix retains its form (i.e. is not *-ma/-me*) in the first person singular, as in the following examples from Türk Dil Kurumu (1942:83,127) cited in Timurtaş (1977:125): *iç-mez-em*, *sor-maz-am*. This of course suggests that the *-mAs* in the Modern Turkish translations of these, *iç-me-m* and *sor-ma-m*, encode both [neg] and [aor] as *-maz/-mez* in the 15th century Turkish does.

in the paradigm. As for the *y* in the first person plural, Türeli considers it to be a buffer consonant inserted to prevent the occurrence of succession of two vowels. Swift's analysis (1962:85) which posits the morpheme *-z/-Ø* as an 'alternate' of the morpheme *-r/-ir/-ir* is also different from all of these analyses. On the other hand, Lees (1961:34-5) proposes yet another analysis of the negative aorist which involves rather complex phonetic reduction and assimilation rules, which we do not review in detail here.

In the following, we have schematised these mutually contradicting analyses. Sounds which are identified as aorist markers are bold-faced where they are clearly segmented. The 'DROP' in (1) signifies the assumed 'drop' of the aorist marking morpheme in the analysis:

	1	2	3	4	
					<i>oyna-</i>
1sg	<i>-ma-DROP-m</i>	<i>-ma-m</i>	<i>-ma-Ø-m</i>	<i>-mA-Iz-yIm > -mA-I-Im > -ma-m</i>	<i>> -mam</i>
2sg	<i>-ma-z-sin</i>	<i>-maz-sin</i>	<i>-ma-z-sin</i>	<i>-mA-Iz-sIn > -ma-z-sin</i>	<i>> -mazsin</i>
3sg	<i>-ma-z(-Ø)⁹</i>	<i>-maz</i>	<i>-ma-z</i>	<i>-mA-Iz-Ø > -ma-z-Ø</i>	<i>> -maz</i>
1pl	<i>-ma-y-iz</i>	<i>-ma-yız</i>	<i>-ma-Ø-yız</i>	<i>-mA-Iz-yIz > -mA-Iy-yız > -ma-y-iz</i>	<i>> -mayız</i>
2pl	<i>-ma-z-siniz</i>	<i>-maz-siniz</i>	<i>-ma-z-siniz</i>	<i>-mA-Iz-sIn-Iz > -ma-z-sin-iz</i>	<i>> -mazsınız</i>
3pl	<i>-ma-z-lar</i>	<i>-maz-lar</i>	<i>-ma-z-lar</i>	<i>-mA-Iz-lAr-Ø > -ma-z-lar-Ø</i>	<i>> -mazlar</i>

Obviously, all of these analyses, except for the second one, strive to make sense of the paradigm (shown in the rightmost column) within the framework of agglutinative morphology which explains quite neatly many other paradigms in Turkish grammar, e.g. the positive counterpart of the above paradigm. The apparent absence of an aorist marking morpheme in the first person singular is the issue for which the grammarians have much difficulty providing accounts. It is explained as a result of the 'drop' of the morpheme in the analysis (1), while (3) posits an aorist marking zero morpheme and (4) proposes phonological rules that account for the apparent absence. This lack of consensus among linguists regarding how *oynamam* is yielded through suffixation points to the lack of ready segmentability and hence also involvement of fusional morphology in the word.

⁹ Kornfilt acknowledges the existence of the 'zero' morpheme (1997:338), but notes that 'The third person singular agreement suffix is null for finite verbs; I have not represented this null morpheme in the examples, nor have I glossed it' (ibid:xxiv).

An implication which the existence of fusional morphology in a highly agglutinative language such as Turkish has for the system of IS analysis which we will propose for agglutinative languages, is that it cannot dismiss fusional morphology completely.

If we take these facts into consideration, a system of analysis of sentences must have at least potential explanatory power for cases where both types of morphology are present in a sentence. There is also a factor that encourages this attempt to, though partially, incorporate fusional morphology in our system of analysis of IS: simplicity. Obviously, devising another system in order to account for sporadic occurrence of fusional morphology in agglutinative languages would introduce unwanted complexity into our analysis. This would also contradict our aim to devise a simple system. The measure we take to (partially) incorporate fusional morphology into our system of analysis is not to rule out the possibility that morphologically unsegmentable entities ‘encode’ information as we will assume morphemes do.

Let us take the English *man/men* pair for example. We have mentioned in the previous subsection the possibility of identifying the phonetic changes that takes place in the pluralisation of ‘man’, which yields ‘men’, as what corresponds to a [pl] morpheme. This would mean that, in this English example, the changes of the distinctive phonetic features of [+low] > [-low] and [-ATR] > [+ATR] (or the change of the phoneme /æ/ to /e/) in the pluralisation of ‘man’ correspond to the occurrence of the [pl] morpheme ‘-(e)s’ in regular pluralisation. (Our pursuit of simplicity and the unified treatment of regular and irregular, i.e. agglutinative and fusional morphology, are inspired by connectionists’ (see the next chapter) treatment of English irregular and regular verbs and nouns (e.g. Rumelhart and McClelland 1986, Plunkett and Juola 1999)¹⁰.)

Although we will largely confine our discussion to IS in relation to agglutinative morphology in the next chapter, in the chapter following the next one, the possibility of our system’s applicability to some components of Turkish grammars which exhibit fusional morphology will be investigated. There, we will explain in some detail how our system may treat fusional morphology in a predominantly agglutinative language. It is also worth mentioning that, as Comrie (1982:94) points out:

[...] many languages come close to exhibiting canonical agglutination, but none comes anywhere close to canonical fusion: a language exhibiting canonical fusion would have to

¹⁰ For an overview of morphology in connectionism, see Christiansen and Chater (1999:423-425).

lack segmentability and invariance completely, i.e. every ‘sentence’ would simply differ holistically from every other, with no possibility of formal analysis.

Thus, we do not sacrifice much of the usefulness of our system of analysis by restricting its applicability to agglutinative languages.

1.2.3. Remarks on the Languages Used in This Thesis

In this section, we give a brief description of each of the languages from which we take examples in the present thesis, namely Turkish, Japanese, Uzbek, and Bukharan Tajik. In the course of description, we will give special emphasis to a description of agglutinative morphology in these languages, which has the highest relevance with our discussion in the subsequent chapters.

We limit ourselves to synchronic descriptions of the languages, though we do provide explanations from a diachronic perspective where we think such explanations are beneficial in comprehending how agglutinative morphology came into existence in particular domains of the grammars of the languages.

As we will see below, the morphology of every language we discuss in this thesis exhibits a certain degree of agglutination. This is a prerequisite for our chief object, i.e. the proposal of a system of analysis of information structure in agglutinative languages. However, it is worth mentioning that the defining factor of agglutination, namely morphological segmentability of words, in the languages is conversely what prompted us to set up a method of analysis of information structure – as we will see in the next chapter, our system relies much on the visible one-to-one correspondence of ‘meanings’ and morphemes, such as the one exemplified with a Turkish example in the preceding section. If there were no such correspondence, devising a system as simple as ours would be an impossibility.

Before we proceed to the description of the languages, brief comments on each of them are in order. Turkish is very often cited as a textbook example of agglutinative languages. Uzbek is another of the Turkic languages which are known for their distinctively agglutinative morphology. Japanese too is generally considered to be an agglutinative language (see e.g. Szabo 1994, Katada 1995:115). On the other hand, Tajik is usually not considered to be typically agglutinative although the Iranian languages to

which Tajik belongs have some distinctively agglutinative components in their grammars¹¹ – as we will see in this subsection, the strength of agglutinative character of Tajik may be comparable with that of Japanese. As for Bukharan Tajik, it is stronger in agglutination than literary Tajik is in some components of the grammar, while in some other components fusional morphology is more prominent than it is in their literary Tajik counterparts. We will observe them in detail later.

Restricting our source of data to Turkic languages would certainly have made it possible for us to set up principles applicable (exclusively) to the information structure in Turkic languages in general. However, we are aware of the possibility of a wider applicability of the principles which we will propose in the next chapter, which prompts us to include Japanese and (Bukharan) Tajik in our discussion. The inclusion of Japanese and (Bukharan) Tajik is inspired by several different reasons. Japanese is indispensable for our argument in the next chapter where we contrast Turkish and Tajik both of which utilise subject-verb agreement morphology with Japanese which lacks the agreement morphology altogether. (Bukharan) Tajik is of special usefulness for our argument regarding language-specificity in the operation of some of the information-structural principles which will also be proposed in the next chapter. These issues will be discussed at length in the following chapter.

Since Turkish and Japanese have been subjects of unusually intensive linguistic studies and very well documented, our accounts of the two languages will be minimal. For general descriptions of their grammars, the reader is referred to such books as Lewis (1967) and Shibatani (1990).

Uzbek is also a relatively well documented language with several grammars and dictionaries published in different languages, though the intensity of research of Uzbek is obviously lower than that of Turkish or Japanese. Accordingly, remarks on Uzbek will be in a little more detail than those on Turkish or Japanese.

As for Bukharan Tajik, it is added to the list of languages from which we take examples to broaden the scope of our discussion. It is not considered a typically agglutinative language, but where it does show agglutinative characteristics, it bears a curious resemblance to Uzbek. The Tajik language, particularly its dialects are extensively

¹¹ Hence the Osset (a North-East Iranian) declension chart used to exemplify agglutinative morphology in Isaev (1978:73).

studied mainly by local scholars. Nevertheless, to most western readers, Tajik is arguably the least accessible among the languages from which data are taken in this thesis.

Moreover, our source of Tajik data is Bukharan Tajik, which may be even esoteric to some readers. We will, therefore, assuming little previous knowledge of Tajik by the reader, give relatively detailed accounts of a range of issues in Tajik linguistics and Bukharan Tajik.

1.2.3.1. Turkish

Let us start our description with Turkish, a textbook example of an agglutinative language. First, general facts about the language: Turkish is the official language of the Republic of Turkey and is the largest of the Turkic languages in terms of the number of speakers. Turkish is based on the dialect of Istanbul (although the orthography's divergence from spoken Istanbul Turkish is easily noticeable¹²).

Turkish has attracted European scholars' attention by its agglutinative morphology, which, Lewis (1967:xx) writes, 'English-speakers find most alien'. Segmentability of words is surprisingly high in Turkish as we have already seen the systematic suffixation in the paradigm of a verb in the positive aorist. The very systematic way in which suffixes are added to stems are so clearly observable in the current orthography of Turkish that Turkish has been cited frequently in explaining agglutinative morphology (e.g. Comrie 1989:44, Fujii et al 1998:27), which is large part of the reason why Turkish is used in this thesis as a source of data.

Agglutination in Turkish has been exemplified and discussed in the previous section and we do not redundantly exemplify it here. However, we would like to note that the high degree of agglutination is very much a characteristic of written Turkish. In spoken Turkish, particularly in rapid speech, invariability of morphemes' forms gives way to morphological fusion, undermining high morphological segmentability. (This type of

¹² There seem to be also some undercurrents that show the signs of some users' desire to narrow the divergence between the written language and spoken language or to write Turkish more 'phonetically'. This trend manifests itself in a variety of ways, e.g. as catch lines in advertisements (*kokuyo* for *kokuyor*), words in cartoons (*kusucam* for *kusacağım*), in e-mail messages (*n'aber?* for *ne haber?*) etc. There was even a newspaper columnist who would write his column 'as he speaks it' (*di mi?* for *değil mi?*).

morphological fusion has yielded a number of new morphemes in the history of Turkish¹³. See, e.g., Deny (1995:125-127), Ergin (1997)) Observe the following examples from Ergenç (1995:40) in both of which the suffix *-(y)AcAk* (i.e. *-(y)acak ~ -(y)ecek*)¹⁴ occurs:

Spoken	Written
<i>ari:cak ~</i>	<i>ara-yacak</i>
	search-Fut.3sg
	‘s/he will search’
<i>olucak ~</i>	<i>ol-acak</i>
	become-Fut.3sg
	‘s/he will become’

We can see that morphological segmentability is markedly lower in spoken Turkish than in written Turkish. Perhaps due to the remarkably phonetic orthography of modern Turkish as well as many Turks’ belief that Turkish is written as it is spoken, some Turks appear to be tempted to reflect their pronunciation to writing at the expense of orthographic correctness. Such writing is salient, not surprisingly, in informal writing, such as e-mail messages, cartoons, and words in pop songs, but appears occasionally also in catch lines in advertisements and a newspaper columnist’s writings. Some examples are shown below: *diyiceeniz*¹⁵ (*di-yeceğ-iniz* in written English) *edicen*¹⁶ (*ed-ecek-sin*), *sorucam*¹⁷ (*sor-acağ-ım*), and *yatcaz*¹⁸ (*yat-acağ-ız*). Such writing indicates the discrepancy between writing and speech, and also the lower morphological segmentability in speech in comparison with writing in Turkish.

¹³ As it has in other Turkic languages, e.g. , *işlep yatip turur* which turned into *işlā(v)yätiptü>işlâyätipti>işlâyä(ti)pti* in Uzbek and *işlev(y)atıptu>işlevatıptu>işlevatidu* in Uyghur (Öztürk 1997:129).

¹⁴ According to Korkmaz (1960:179), etymologically speaking, this suffix itself consists of two morphemes, *-a (<-ğā)* and *+cağ*.

¹⁵ ‘that you (would like to) say’. From Savaş Ay’s writing cited in Hepçilingirler (1997:179).

¹⁶ ‘you will do’. From pop singer Sezen Aksu’s ‘Onu Alma Beni Al’.

¹⁷ ‘I will ask’. From Erdil Yaşaroğlu’s *Komikaze 4*.

¹⁸ ‘we will lie’ From pop singer Nazan Öncel’s ‘Erkekler de Yanar’.

However, as it may be clear from the above discussion, such large discrepancy between speech and writing which involves the issue of morphological segmentability is confined largely to cases where the morpheme *-(y)AcAk* (the *k* in the suffix turns into *ğ* when followed by a suffix beginning with a vowel) or the *y* phoneme appears. All the Turkish examples which we will examine in this thesis do not involve such cases. Accordingly, as far as our examples are concerned, there is no problem in complying with orthographic rules in their presentation, because the language's unusually high morphological segmentability is reflected in the orthography with relatively high precision.

1.2.3.2. Uzbek

1.2.3.2.1. General Remarks

Uzbek is another of highly agglutinative languages with possibly even higher segmentability of words than Turkish.

Uzbek, a Turkic language, is the official language¹⁹ of the republic of Uzbekistan which has Chaghatay²⁰ as its literary predecessor (Boeschoten 1998:357). According to the 1959 census, Uzbeks live in Uzbekistan, and its neighbouring countries such as Kazakhstan, Tajikistan, Turkmenistan, and Kyrgyzstan²¹. China is home to 12,433 Uzbeks (1982 census) where they live in Yinin, Kashgar, etc. (Cheng and A-pu-tu-je-ho-man 1987:1). There are Uzbeks also in Northern Afghanistan, Saudi Arabia, Pakistan, Turkey, Iraq, and other middle eastern countries. According to Rajabov (1996:19), there are also large groups of Uzbeks who moved to such cities as Sibir', Orenburg, and Astraxan in the 16-17 centuries to promote commercial ties between Russia and Uzbek khanates. We have no data of the situation of language maintenance/shift in all these groups (see, however, Akiner (1989) for the situation in Uzbekistan prior to its independence).

¹⁹ Or *Davlat tili* 'state language' as used in some texts, including the constitutional law and Abdurahmonov (1996:5).

²⁰ This term has different referents according to different authors. See Eckmann (1988:vii-xvi) and Boeschoten and Vandamme (1998:166-169).

²¹ See Faensen (1983, cited in Öztürk 1997:3) or Devlet (1990, cited in Öztürk 1997:3) for the figures of 1979 census.

The basis of standard Uzbek is not a single dialect. Yaman and Mahmud (1998:iv) note that Uzbek's phonetic basis is the Tashkent dialect whereas its morphologic basis is Ferghana valley dialect²². Also, Sjoberg (1962:5) writes:

‘Today the Tashkent dialect alone serves as the phonological basis for standard written Uzbek; however, the standard grammar and vocabulary are also based in part on certain South Uzbek dialects of the Ferghana region’

1.2.3.2.2. Orthography

In this thesis, Uzbek sentences, terms, etc. which are originally in Cyrillic writing are transliterated into Latin-based writing in accordance with the new orthography.

The current writing system of Uzbek is based on a Latin alphabet which uses only ASCII symbols. There was a consideration for digitalisation of Uzbek behind the decision to restrict the letters used in the Latin-based Uzbek alphabet to the ASCII symbols (Ne'matov 1999:private communication). The restriction naturally necessitates exclusion of letters with diacritics, such as ‘ö’ and ‘ç’, from the alphabet. This lead to devising such representations of phonemes as *g* ‘[ɣ], *o* ‘ [o], *sh* [ʃ], *ch* [tʃ], and *ng* [ŋ], which correspond to *g*, *o*, *sh*, *ch*, and *ng*, respectively in the previous Cyrillic-based alphabet. Thus, the following title of the book ‘the orthographic dictionary of the Uzbek language’ (Rahmatullayev and Hojiyev 1995) in the Cyrillic-based orthography:

Ўзбек тилиниг имло луғати

is transliterated into the Latin-based orthography as follows:

O'zbek tilining imlo lug'ati

Many Uzbek orthographic rules are based on morphology rather than phonology or phonetics.²³ These rules which abstract inter-morphemic phonetic assimilation/dissimilation naturally enhance direct reflection of (agglutinative) morphology onto writing and makes morphemes, in writing, immune to phonological or phonetic variation. For example, the orthographic dictionary resolves that the past tense

²² For a description of Uzbek dialects, see Rajabov (1996).

²³ There are also a number of orthographic rules based on phonetics/phonology (see G'ulomov, Tixonov, and Qo'ng'urov 1977:365-398, Rahmatullayev and Hojiyev 1995:7-27).

marker *-di* is always written with *d*, even when the consonant at the beginning of it is pronounced as *t* (ibid:15):

[4]	<i>ket-di</i>	<i>o't-di</i>	<i>bor-di</i>
	leave-past	pass-past	go-past
	's/he left'	's/he passed'	's/he went'

Compare this with Turkish's more phonological orthography where the devoicing of *d* is reflected in writing:

[5]	<i>git-ti</i>	<i>geç-ti</i>	<i>kal-dı</i>	<i>dön-dü</i>
	go/leave-past	pass-past	stay-past	return-past
	's/he went/left'	's/he passed'	's/he stayed'	's/he returned'

The invariance of the morpheme *-di* [past] in Uzbek orthography helps easy visual recognition of morphemes. This in turn 'visualises' very high ready segmentability of words in Uzbek, which is exemplified in the following Uzbek paradigm (*o'yna-* 'play' + *-y* [nonpast²⁴] + personal suffix):

	Positive		Negative	
1sg	<i>o'yna-y-man</i>	'I play'	<i>o'yna-ma-y-man</i>	'I play'
2sg	<i>o'yna-y-san</i>	'you play'	<i>o'yna-ma-y-san</i>	'you play'
3sg	<i>o'yna-y-di</i>	's/he plays'	<i>o'yna-ma-y-di</i>	's/he plays'
1pl	<i>o'yna-y-miz</i>	'we play'	<i>o'yna-ma-y-miz</i>	'we play'
2pl	<i>o'yna-y-siz</i>	'you play'	<i>o'yna-ma-y-siz</i>	'you play'
3pl	<i>o'yna-y-di(-lar)</i>	'they play'	<i>o'yna-ma-y-di(-lar)</i>	'they play'

As is clear from the paradigm shown above, segmentation of words is, in general, a straightforward process in Uzbek. (However, this is not necessarily the case in spoken Uzbek, where, for instance, such an utterance as *opket*, which is *ol-ib ket* [take-ger leave.imp] (O'rinboev 1982:133) in written Uzbek, is observed.) This high segmentability is favourable to our discussion of agglutinative morphology, which is the chief reason why Uzbek is used as a source of data in the present thesis.

²⁴ This glossing is a tentative one adopted from the Japanese tense system. In Uzbek, it is called 'present-future tense' (*hozirgi-kelasi zamon*).

1.2.3.3. Bukharan Tajik

1.2.3.3.1. Tajik

Tajik is a South-West Iranian language which is genetically closely related to Persian. Tajik is heavily influenced lexically, morphologically, and syntactically by Uzbek. Most Tajik speakers are in Tajikistan and Uzbekistan²⁵. Within Uzbekistan, Samarkand and Bukhara are particularly densely populated by Tajik speakers.

In the beginning of the twentieth century, Tajik was considered by a number of scholars to be not an independent language but a mere variety of Persian (Halimov 1974:30-31). The popularity of this conception of Tajik as a (less prestigious) variety of Persian was such that, during the period when Tajik intellectuals were trying to establish the Tajik language separate from Persian, Sadriiddin Aynī, the most prominent figure in the history of the Republic of Tajikistan, had to make a statement that ‘Tajik is not a bastardised dialect of Persian’ (cited in Halimov 1974:31) in 1928²⁶. The issue of whether Tajik and Persian are to be considered two dialects of a single languages or two separate languages has political sides to it (see Perry 1996) with which we are not concerned in this thesis.

1.2.3.3.2. Agglutination in Tajik

Tajik is not known for its agglutinative morphology. However, Tajik has many morphemes whose forms possess high invariability. These formally invariable morphemes contribute to the high segmentability of words, which in turn enhances the language’s agglutinative nature. For example, see the following systematic paradigm where possessive suffixes²⁷ (or pronominal enclitics) and the plural suffix *-ho* are suffixed to *kitob* ‘book’:

²⁵ A 1970 census also records 15.000 ethnic Tajiks in Kyrgyzstan (Nawada 1989:614). The 26,500 (according to Kao 1985:1) or approximated 33,000 (according to the 1990 census in *Zhongguo dabai ke quan shu zong bian ji wei yuan hui* 1993:331) ‘Tajiks’ in the Tašqorghān autonomous prefecture of the Tajiks, Qāshqār, and Hotan in China are, despite being called *Ta-chi-ko* ‘Tajik’ in Mandarin, speakers of Pamir (mainly Sariqul and Wux dialects) (Kao 1985:1-4. ‘Sariqul’ and ‘Wux’ are written as ‘Sarikol’ and ‘Wakhi’, respectively, in Schwarz 1984).

²⁶ In his article ‘About the Persian language and Tajik language’ (*Dar atrofī zaboni forsi va tojikī*) which appeared in the journal *Rahbari doniṣ*, 1928, No 4-5.

²⁷ *Bandakjonišinhoi šaxsī-sohibī* ‘personal-possessive suffixal pronouns’.

	Single	Plural	
	<i>kitob</i>	<i>kitob-ho</i>	‘book/s’
1sg	<i>kitob-am</i>	<i>kitob-ho-yam</i> ²⁸	‘my book/s’
2sg	<i>kitob-at</i>	<i>kitob-ho-yat</i>	‘your book/s’
3sg	<i>kitob-aš</i>	<i>kitob-ho-yaš</i>	‘his/her book/s’
1pl	<i>kitob-amon</i>	<i>kitob-ho-yamon</i>	‘our book/s’
2pl	<i>kitob-aton</i>	<i>kitob-ho-yaton</i>	‘your book/s’
3pl	<i>kitob-ašon</i>	<i>kitob-ho-yašon</i>	‘their book/s’

(An argument can be made for further segmentation of the personal suffixes *-(y)amon*, *-(y)aton*, and *-(y)ašon* – they can be analysed as their singular counterparts *-(y)am*, *-(y)at*, and *-(y)aš* to which the morpheme *-on*, whose form is invariable, is suffixed. However, we follow the traditional grammarians’ segmentation here.) The invariability of the forms of all the morphemes in this table is very high, which ensures high segmentability of the words and hence also their high degree of agglutination.

On the other hand, Tajik verb paradigms usually exhibit both fusional and agglutinative natures. For example, there are a large number of verbs whose present tense and past tense forms differ radically from each other (see Rustamov and Ghafforov 1985:173-174). The difference between the present and past stems of a few of such verbs can be as big as that between English ‘go’ and ‘went’, e.g., the present stem of *didan* ‘to see’ is *bin-*, which shows little resemblance to the verb’s past stem *did-*²⁹.

However, aside from the variability of the forms of verbs, basic verb paradigms³⁰ are fairly systematic, as can be seen in the following chart where conjugations of the verb *xondan* ‘to read’ is tabulated (copied from Rastorgueva 1992:57-58):

	Non-past (subjunctive) ³¹	Past	
1sg	<i>xon-am</i>	<i>xond-am</i>	‘I read’
2sg	<i>xon-ī</i>	<i>xond-ī</i>	‘you read’

²⁸ The buffer sound *y* is inserted before a personal suffix when the suffix is preceded by either of the vowels *o*, *ū*, and *ī* (see Nemenova 1955:552).

²⁹ The present and past stems of *didan* are etymologically different stems.

³⁰ We count out here verb paradigms that involve participle/gerund forming suffixes such as *-a* and *-gī*.

³¹ We use in this table our (tentative) terminology of tenses and aspects based on Kerimova’s (1997:108-109) terminology.

3sg	<i>xon-ad</i>	<i>xond</i>	‘s/he read(s)’
1pl	<i>xon-em</i>	<i>xond-em</i>	‘we read’
2pl	<i>xon-ed/-eton</i>	<i>xond-ed/-eton</i>	‘you read’
3pl	<i>xon-and</i>	<i>xond-and</i>	‘they read’
	Non-past	Durative past	
1sg	<i>me-xon-am</i>	<i>me-xond-am</i>	‘I (would) read’
2sg	<i>me-xon-ī</i>	<i>me-xond-ī</i>	‘you (would) read’
3sg	<i>me-xon-ad</i>	<i>me-xond</i>	‘s/he (would) read(s)’
1pl	<i>me-xon-em</i>	<i>me-xond-em</i>	‘we (would) read’
2pl	<i>me-xon-ed/-eton</i>	<i>me-xond-ed/-eton</i>	‘you (would) read’
3pl	<i>me-xon-and</i>	<i>me-xond-and</i>	‘they (would) read’

Thus, relatively high segmentability of words is observed in at least some domains of the grammar in standard Tajik. However, it should be noted that, in colloquial Tajik, ready segmentability is much less prominent than it is in literary standard Tajik. Observe, for example, the following literary Tajik example (the auxiliary verb *istodan* ‘stand’ is used as the progressive aspect marker):

[6] *xond-a istod-a-ast*
 read.past-ptcpl stood-ptcpl-cop.3sg
 ‘(s/he) is reading’

In this literary Tajik example, the succession of *xond-a istod-a-ast* can be segmented into morphemes with relative ease. However, this is not the case with colloquial Tajik. Buzurgzoda (1940:66) writes that, in conversation, many syllables and sounds drop from this, yielding the following:

xɔnda istɔdaast → *xɔnsɔdast* → *xɔndsɔas* → *xɔnsɔs*

This form is attested in Xujand-Koni Bodom dialect and takes place in the following paradigm (Rastorgueva 1956:58):

1sg	<i>xonsodiyam</i>	‘I am reading’
2sg	<i>xonsodi</i>	‘you are reading’
3sg	<i>xonsos</i>	‘s/he is reading’

Clearly, ready segmentability does not exist in this colloquial Tajik paradigm. Thus, we cannot assume that colloquial Tajik always shares the agglutinative morphology which literary Tajik exploits (which is probably the case with any agglutinative languages).

In sum, Tajik exploits agglutinative morphology perhaps not to the extent that Turkish or Uzbek does. However, the language does have distinctively agglutinative components, although fusional morphology is perhaps equally conspicuous in its grammar, particularly in its colloquial variations.

1.2.3.3.3. Tajik Orthography

Following the foundation of the Soviet republic of Tajikistan, Tajik (or Central Asian Persian) was elaborated to the status of the national language. One of the most important tasks Tajik intellectuals were to perform in the first years of the foundation of the republic was standardisation of the Tajik language. As Lazard (1970:71) writes, ‘In the twenties and the thirties, linguistic activity in Tajikistan was mainly devoted to the practical work of standardising the language and its orthography first in Latin, then in Cyrillic script’³².

A close examination of the process of standardisation of Tajik orthography in Latin script provides us with an insight into what ‘word’ is, which will be discussed in detail later. Since we need to come back to the standardisation of Tajik orthography in 2.4.4.3.1, in this subsection, we content ourselves with only explaining the transliteration we employ in this thesis.

For transliterating the current Tajik alphabet in Cyrillic script, we use the transliteration system for the Russian alphabet by the International Organisation for Standardisation with modification. The Cyrillic *x* remains *x* and is not transliterated into *h* or *ch* as the ISO system suggests. Cyrillic *Ээ* will be *Éé*. The characters unique to Tajik are transliterated as follows: *к* > *q*, *ғ* > *gh*, *ӯ* > *ī*, *ӯ* > *ū*, *ҳ* > *h*, and *ҷ* > *j*.

1.2.3.3.4. Bukharan Tajik

³² See the 1929 issues of *Rahbari doniş* for the Latin-based orthography in transition. See also Vahhobov (1980) and Perry (1997) for an account on how the writing system of Tajik changed.

1.2.3.3.4.1. Tajik and Bukharan Tajik

The aforementioned process of standardisation of Tajik languages involved many interrelated issues, and Bukharan Tajik is in fact the dialect which played an important role in the standardisation.

Some of the issues which intellectuals brought up in the course of the standardisation of Tajik included the following: 1) creation of a literary language simple and accessible to the masses (A'lo-zoda 1930:4, Jahangiri 1997:24), 2) determining the dialect or (Iranian) language on which the language is to be based, and 3) making the language a component of the specific identity of Tajikistan.

The second was the issue which provoked much debate, as depicted in A'lo-zoda (1930:3). One of the languages we benefit from in this thesis, namely the Bukharan dialect of Tajik (referred to simply as Bukharan Tajik in this thesis) was one focus of this controversy.

The dialects of the centres of Tajik culture such as Bukhara and Samarkand were natural candidates for the basis of standard Tajik, but there was a problem. The dialects were highly Uzbekicized, which was not a welcome feature for a language that was to constitute the identity of the newly founded republic of Tajikistan. This led to such a claim as 'adopting the language of city-dwellers and, among them, that of Bukharans as the basis of the Tajik language does not appear to be acceptable'³³ by Azizī (1929:5), who also writes that the language of city-dwellers would need 'cleaning (of Uzbek influence)'.

However, the odds were in Bukharan Tajik's favour and it eventually became the basis of standard Tajik (A'la zoda 1930:5). This is due at least partly to the fact that the aforementioned Tajik educator/writer Sadriiddin Ainī, who dominated the academia of Tajikistan since the establishment of the republic, was from Suktare in the province of Ghižduvon, which is some 30 km west from Bukhara, and was educated in Bukhara. A dialectological work by Melex (1968:22 cited in Éšniězov 1977:20) also suggests the influence of Ainī's own dialect on standard Tajik: 'a comparison between Ghižduvon dialect and literary language shows that, both in the domain of phonetics and in the

³³ [...] *saboni* [sic] *šahrijon va az ončumla Buxoraijon* [sic] *ro asosi zaboni točikī karda griftan ma'qul namenamojad.*

domain of morphology and vocabulary, they possess an immense unity' (see also Jahangiri 1997:24-25).

1.2.3.3.4.2. Characteristics

Despite its status as the basis of standard Tajik, Bukharan Tajik of today exhibits radical phonological, morphological, and syntactic divergence from standard Tajik in various ways. This is partly because standard Tajik incorporates many classical Persian(-Tajik) elements which do not exist in colloquial Northern Tajik (Soper 1996:58) to which Bukharan Tajik belongs, and perhaps also because Bukharan Tajik itself has undergone changes since the beginning of the twentieth century.

Divergence of Bukharan Tajik from literary Tajik is not limited to lexical borrowings from Uzbek but include syntactic and morphological peculiarities³⁴ many of which are documented in the literature of dialectology (e.g. Kerimova 1959). Such peculiarities of Bukharan Tajik are usually ascribed, probably rightly, to Uzbek influence on the language³⁵.

Let us see why it is tempting to account for Bukharan Tajik's peculiarities in relation to Uzbek. Examine the following Bukharan Tajik and Uzbek examples (modifier phrases

³⁴ For example, many young Bukharans (like the illiterate Persians in Chodzko's (1852:59, cited in Windfuhr 1979:106) description) do not know of the *-a šudan* passive construction, which they invariably translate into Uzbek as *-(i)b bo'ymoq* 'finish -ing'. This seems to be because Uzbek *-(i)b* corresponds to *-a* in Bukharan Tajik in a number of other cases, e.g. in the converb construction (see Perry 1979). Young Bukharans, probably unlike their preceding generations (see data in Kerimova 1959:129-139), also seem to use pre-nominal modifier phrases introduced with the suffix *-(a)gī* very extensively (the comments on this matter in Ido (2001) may be somewhat misleading), whereas in literary language the norm is that modifier phrases are post-nominal (Rustamov and Ghafforov 1986:257-258; See, however, Zehnī's (1987:154) comment on adjectives in Tajik and Darī.) and pre-nominal modifier phrases that are not in a simple mono-lexical form but in an extended form with 'explanatory/commentary' parts of speech are considered to be peculiar to colloquial speech (Rustamov and Ghafforov 1986:258).

³⁵ Tajik-Uzbek bilingualism has persisted for centuries (Mirzaev 1969:25) and in Bukhara, virtually all Tajik speakers have a certain command of Uzbek (see Rastorgueva et al 1970:717).

are square-bracketed) which show one-to-one morphological correspondence with each other³⁶:

[7] B. Tajik [*man navist-agi*] *xat*
 I wrote-ptcpl letter
 ‘the letter I write/wrote’

Uzbek [*men yoz-gan*] *xat*
 I write-ptcpl letter
 ‘the letter which I write/wrote’

Compare these with the following (colloquial) examples taken from Rastorgueva’s Tajik grammar (1992:80-81, glosses and translations are mine) where, unlike in the Bukharan Tajik and Uzbek examples above, modifier phrases follow modified nouns to which the izafet *-i* is attached (as they do in Persian):

maktub-i [*Hasan navišt-agī*]³⁷
 letter-iz³⁸ Hasan wrote-ptcpl
 ‘the letter that Hasan wrote’

kitob-i [*man ovar-d-agī*]
 book-iz I brought-ptcpl
 ‘the book that I brought’

³⁶ Some other examples of pre-nominal modifier phrases with *-agi* in Bukharan Tajik are shown below. (Taken from e-mails. Personal names are altered. Modified nouns are underlined): [...] *550 megriftagi darajaba meomdagis ku?* ‘[...] she will be at (lit. come to) the level where she will gain 550 points, won’t she?’, *Muhsin guftagi bača* [...] ‘a young man called Muhsin’, *Man da boli asp budagi rasm* ‘the photo of me being on the top of a horse’, *Az Ūzbekiston gde to yak ma peš omadagi bača guftiki* [...] ‘a guy who came from somewhere in Uzbekistan said that [...]’, and *Firmomoba 2 sol boza korkardaštagi Fozil* [...] ‘Fozil, who has worked for (lit. in/at) our company for 2 years [...]’.

³⁷ Most grammars and dictionaries analyse *-agī* as the combination of the (gerund/participle forming) suffix *-a* and past participle forming *-gī* (e.g. Bobomurodov and Mūminov 1983:97, Rastorgueva 1992:79-80). However, as far as colloquial Tajik where the past participle is always formed with *-agi* (or *-agi* in Bukharan Tajik) is concerned, segmentation that contrasts *navišt-a* with *navišt-agī* is not always very practical in morphological analyses. This is the reason why we treat *-agī/-agi* as a single morpheme here (Niězmuhammadov et al (1955), on page 95, take a somewhat similar approach to ours to *-agī*).

Pre-nominal modifier phrases with *-agi* is prevalent in the Tajik language used by young Bukharans in their twenties who have had no formal education in standard Tajik³⁹. Some examples are shown below:

[8] *Čunki man ham xohiš kard-agi payt-am bromad-a me-ton-am.*
 because I also wish did-ptcpl time-1sg leave-ger impf-can-1sg
 ‘Because I can also leave when I want to.’

[9] *me-raft-agi maktab-am-a web adres-aš [...]*
 impf-went-ptcpl school-1sg-gen web address-3sg
 ‘The web address of the school to which I will go / at which I will be enrolled is ...’

To sum up, Bukharan Tajik is a dialect with a number of non-standard features, many of which can be ascribed to Uzbek influence on the dialect.

1.2.3.3.4.3. Agglutination in Bukharan Tajik

Observe the following chart provided (in writing) by a Bukharan Tajik speaker and compare it with its literary Tajik equivalent previously cited:

	Single	Plural	
	<i>kitob</i>	<i>kitob-o</i>	‘book/s’
1sg	<i>kitob-am</i>	<i>kitob-om</i>	‘my book/s’
2sg	<i>kitob-at/d</i>	<i>kitob-ot/d</i>	‘your book/s’
3sg	<i>kitob-aš</i>	<i>kitob-oš</i>	‘his/her book/s’
1pl	<i>kitob-amo</i> ⁴⁰	<i>kitob-omo</i>	‘our book/s’
2pl	<i>kitob-aton</i>	<i>kitob-oton</i>	‘your book/s’
3pl	<i>kitob-ašon</i>	<i>kitob-ošon</i>	‘their book/s’

Ready segmentability in the succession of morphemes following *kitob* in the plural forms which is retained in the literary Tajik paradigm is absent in the Bukharan Tajik paradigm.

³⁸ Iz stands for izafet.

³⁹ It should be noted that the use of pre-nominal modifier phrases is not unique to Bukharan Tajik (see Rustamov and Ghafforov 1986:258) and that young Bukharans do use post-nominal modifier phrases (alongside pre-nominal modifier phrases).

⁴⁰ An informant provided both *kitobamo* and *kitobomo* for [1pl].

We can also see that portmanteau morphs are formed at the expense of morphological segmentability. Now look at the Bukharan Tajik equivalent of the other of the previously cited Tajik example:

1sg	<i>xond-ašt-am</i>	‘I am reading’
2sg	<i>xond-ašt-i</i>	‘you are reading’
3sg	<i>xond-ašt-as</i>	‘s/he is reading’

It is unclear whether these words have their etymologies in their literary counterparts, namely *xonda istodaam*, *xonda istodaī*, and *xonda istodaast*. (Bukharans claim that the *št* sound in them might be a remnant of the verb *šistan* ‘to sit’ used as an auxiliary verb instead of *istodan* ‘to stand’.) However, what is interesting about this paradigm is that, despite sound changes or/and unclear etymology, it somehow manages to form a system where ready segmentability is not sacrificed. In the paradigm above, if we assume *-ašt* to be a continuous aspect marker, we get a set of highly segmentable words. This implies that Bukharan Tajik, despite its status as an almost exclusively colloquial language, may not necessarily be less agglutinative than literary Tajik in all components of the grammar.

1.2.3.4. Japanese

Japanese is the official language of Japan and is considered by some scholars (e.g. Miller 1971) to be a member of ‘the Altaic languages’ which comprise subgroups of languages, one of which is the Turkic languages. The existence of genetic relationships between such a huge group of languages is a matter of debate (see Unger 1990) and there are some studies whose discussions are based on dubious data exist. Some characteristics that are claimed to be largely unique to the languages have been put forward by a number of Japanese linguists (see Hattori 1959:Chapters 1,2,11). Characteristics of Japanese that have prompted the scholars to assume the language’s remote genetic relationship with such languages as Turkish, Mongolian, etc. include the existence of partial vowel harmony in Old Japanese and agglutination.

Agglutinative morphology is indeed widespread in the grammar of Japanese, as can be observed in the following examples where the morphemes *-(a)na*, *-i*, *-(s)ase*, and *-(r)are* are suffixed to the verb stem without altering their forms:

[10] *hanas-ana-i*

talk-neg-nonpast

‘does not talk’

hanas-ase-na-i

talk-caus-neg-nonpast

‘cause not to talk’

hanas-ase-rare-na-i

talk-caus-potential-neg-nonpast

‘cannot cause to talk’

However, despite the distinctive agglutinative morphology observed in these examples and the general agreement among Japanese grammarians that Japanese is an agglutinative language, segmenting words into morphemes is no more straightforward than it is in Tajik. We only need to look at the past tense formation of verbs in Japanese to realise this. Observe the following chart taken from Kiyose (1995:48) with modification:

Nonpast	Past		
<i>kaku</i>	<i>kaita</i>	‘write’	‘wrote’
<i>oyogu</i>	<i>oyoida</i>	‘swim’	‘swam’
<i>tatsu</i>	<i>tatta</i>	‘stand’	‘stood’
<i>kiru</i>	<i>kitta</i>	‘cut’	‘cut’
<i>kau</i>	<i>katta</i>	‘buy’	‘bought’
<i>shinu</i>	<i>shinda</i>	‘die’	‘died’
<i>yomu</i>	<i>yonda</i>	‘read’	‘read’
<i>tobu</i>	<i>tonda</i>	‘fly’	‘flew’

A number of different ways of segmentation for the examples above have been proposed (see Vance 1987:186-7,192-3), much in a similar way different segmentation patterns were proposed for the first person plural forms of verbs in the aorist in Turkish.

However, the variability of the proposed past tense (or perfective aspect) marking morphemes such as *-ta* (Bloch 1970:9) or *-(i)ta* (Kiyose 1995:48) is evident. This variability undermines ready segmentability of words and hence high level of agglutination of Japanese morphology.

To sum up, Japanese morphology shows distinctively agglutinative characters, but is probably not much more agglutinative than languages which are not known for their agglutination (e.g. Tajik).

1.2.3.5. Summary

In this section, we have provided an overview of the morphological language typology and explained the parameter according to which the degree of agglutination of a language's morphology is determined. The parameter is the ready segmentability of words into individual morphemes which is based on the invariability of the forms of the morphemes.

As a preliminary for the discussion in the following chapter, we have also provided brief descriptions of Turkish, Uzbek, Bukharan Tajik, and Japanese, with an emphasis on agglutination in their morphology. The descriptions revealed that agglutination is prominent in all of these languages, but to varying degrees. Agglutination is particularly salient in Turkish and Uzbek and less so in Japanese and Bukharan Tajik.

1.3. Summary of the Chapter

In the first part of this chapter, we have explained the factors which motivated us to set up a system of IS analysis. One of the factors is our understanding that existing systems of IS analysis are not necessarily appropriate for analysing IS in agglutinative languages. It seems to us that some IS analyses are moving towards more abstraction by creating layers of abstract representations of information between information and linguistic expressions, which in turn undermines immediate observability of IS.

We think that the lack of direct observability of IS undermines the *raison d'être* of such theories, because, after all, their aim is to link information with observable linguistic expressions. We attempt to achieve the observability of IS by linking pieces of information directly to lexical and morphological units, which generally have 'tangible' bodies and hence are observable.

At the same time, as has been explained there, our aim is to tailor a system for analysing IS in agglutinative languages. Agglutinative languages are chosen as the source of data in this thesis because they are characterised by their ready morphological segmentability.

As we will see in the following chapter, this defining character of agglutinative morphology is a prerequisite for our direct linking of pieces of information to morphological units. In the next chapter where we explain our system of IS analysis, a unique insight into IS is gained by restricting our source of data largely to agglutinative languages.

2. MORPHEME-BASED SYSTEM OF INFORMATION STRUCTURE ANALYSIS

2.1. Introduction

As was stated in the introductory chapter, the main objective of the present thesis is to propose a system of IS analysis which links information structure and morphology in agglutinative languages. The present chapter attempts to achieve this objective by presenting the most immediate consequence which linking of information structure with morphology would bring about, namely a direct connection between pieces of information and morphemes.

In our analysis of information structure, it will be not predicates or phrases, but morphemes that are linked to pieces of information. Consequently, our method of analysis will contrast sharply with those adopted in linguistics which see stretches of multiple morphemes as bearers/conveyers of information. It differs from many previous approaches also in not using the popular ‘old information’ - ‘new information’ dichotomy in explaining information structure of sentences. Our system of analysis sees pieces of information not as bearing characteristics of ‘new’ or ‘old’, but as entities which may be varied or remain constant.

In our analysis, for example, if one says ‘two’ in response to the utterance ‘so you have a cat’, the utterance ‘two’ alters the information about the singularity of the object ‘cat’ – the information which ‘two’ encodes is not given any judgement as to whether it is ‘new’ or ‘old’. That is, in our analysis, pieces of information themselves are neutral in terms of ‘new/oldness’ – there are only pieces of information which are altered/varied and those which remain unchanged. This issue will be discussed at length later.

In the immediately following section, we will analytically review some of the previous analyses of information structure, following which our analysis that links pieces of information and morphemes will be devised.

2.2. Preliminary Notes

As we have briefly stated in the preceding section, we will link pieces of information not to stretches of morphemes or words but to morphological units, i.e. morphemes. In accordance with the pieces of information they encode, morphemes are categorised into two groups. In our analysis, morphemes are classified into two categories, namely ‘variables’ and ‘non-variables’, according to whether there is any increase or alteration of pieces of information.

The distinction between the alteration and non-alteration of pieces of information, which will be explained in detail later, contrasts with the popular classification of information: ‘old’ vs. ‘new’. We do not employ in this thesis the ‘old/new’ classification of information and hence do not assume pieces of information bearing old/newness. In the next section, we review briefly the old/new dichotomy of information and contrast it with our way of grouping pieces of information. The review of the old/new dichotomy is given, so that the reader will have something to compare our approach with and, hopefully, the explanation of our approach gains more clarity than it would if no ‘control’ were provided.

2.3. ‘Old/New Information’: A Review of previous works

2.3.1. Old/New Information

In an overwhelming number of publications on information structure, information has been assumed by many to have a dual character. Typically, scholars put pieces of information into some sort of dichotomous opposition, the opposing categories being termed most commonly ‘old (or given) information’ and ‘new information’. This classification of information into two ‘information statuses’ is particularly popular in the English literature on information structure (see below), but a similar classification is found also in Japanese (e.g. Szatrowski 1993:79) and Tajik works (Rustamov and Ghafforov 1985:78)⁴¹.

Aside from the widely used term ‘old/new information’, a number of other terms, whose referential domains partly or largely coincide with those of ‘old information’ and ‘new information’, are in use. ‘Presupposition’, which appeared in Lambrecht’s theory

⁴¹ E.g. *shinjōhō* ‘new information’, *kyūjōhō* ‘old information’ (Szatrowski 1993:79), *xabari nav* ‘new information’, *xabari kūhna* ‘old information’ (Rustamov and Ghafforov 1985:78).

reviewed in the preceding chapter, is one such term. Other terms for ‘old information’ include, for example, ‘presupposed information’, ‘known (given) information’, ‘less (or un-) important information’, etc. There are various terms for ‘new information’ too, e.g. ‘(more) important information’, ‘focus’, etc. (see Halliday 1966, Selkirk 1984 cited in König 1991:32, Mikami 1963:106-108, Rustamov and Ghafforov 1985:60, Firbas 1994:118, Duszak 1994:120, and Kamio and Takami 1998:121).

What, then, does the term old or new information represent? ‘Old’ and ‘new’ seem to be recognised generally as statuses into which any given piece of information is assumed to fall, as the following simple definitions of the terms show:

Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance. So-called new information is what the speaker assumes he is introducing into the addressee’s consciousness by what he says. (Chafe 1976:30)

[...] **new information**, which is information that the addresser believes is not known to the addressee, and **given information** which the addresser believes is known to the addressee [...] (Halliday 1967 introduced in Brown and Yule 1983:153)

Thus, according to these definitions, old information may be summarised as ‘information which the speaker assumes to be already in some way in the consciousness of the listener’ (Cruttenden 1997:81), and new information is basically the information which old information is not. Most definitions of old/new information coincide at this point, but there are some aspects in which they differ from each other.

One such difference is the subject who judges the old/newness of pieces of information. In both of the definitions cited above, old/newness is determined from the speaker/addresser’s point of view, as it is clear in the expressions ‘(information which) the speaker assumes to be ... ’ and ‘the addresser believes is ... ’ in the definitions cited above.

There are, on the other hand, definitions of old information which are independent of the speaker’s judgement⁴². This alternative view of old/new information, which is employed

⁴² See, e.g., Venneman (1975: 314, cited in Brown & Yule 1983:79). See also Ochs (1979:1-5) and Birner (1994:241) for the ambiguity such definitions tend to have.

by such a scholar as Venneman (1975: 314, cited in Brown and Yule 1983:79)⁴³, is to consider old information as the entirety of information shared by the interlocutors. Old information according to this view is, then, ‘predisposing factors’ comprising the entire life history of the speaker and hearer up to the time of the speaker’s utterance, which can be paraphrased plainly as the entirety of the information the speaker and hearer retain at the time of utterance.

Naturally, shared information comprises context, which ‘includes, minimally, language users’ beliefs and assumptions about temporal, spatial, and social settings’ (Ochs 1979:1-5) Shared information can also be characterised by its undetectability and unobservability, as Birner (1994:241) aptly notes: ‘only an omniscient observer can truly know what knowledge is in fact shared between interlocutors, while actual language users must operate on the basis of what they assume to be familiar to their interlocutors’. This undetectability stemming from the identification of old information as ‘shared information’ is probably what makes some scholars think the addresser-based old/new information is somewhat preferable to the ‘omniscient observer-based’ old/new information. However, the two in fact do not contradict each other because they are essentially different notions which need to be labelled with different terms, e.g. 1) information assumed by the addresser to be shared and 2) actual shared information. This difference⁴⁴, as it will become clear in the discussion that follows, in our system of IS analysis, loses much of the importance it may have in these theories. This is because whatever ‘old information’ or ‘shared information’ may be, most of it remains unchanged in a dialogue (which is what we are concerned with in this thesis) and we need not and indeed, as Birner points out, cannot describe the entity of the ‘old’ or ‘shared’ information to analyse the IS of sentences which take place in dialogues.

Taking the dialogue about ‘two’ cats presented in the first section of this chapter for example, the information regarding the number of cats undergoes alteration, but other pieces of information which may be involved in or even indispensable for the dialogue to

⁴³ Venneman proposes a ‘presupposition pool’ which contains information ‘constituted from general knowledge, from the situative context of the discourse, and from the completed part of the discourse itself’ which Brown and Yule regard as having much in common with what they call ‘topic framework’. These notions, though different from one another in terminology, essentially point to information shared between the speaker and the hearer.

⁴⁴ See Tomlin et al (1997:77-80) for a discussion of these differences.

take place (such as ‘cat is an animal’, ‘cats eat’, ‘cats may be owned by a human’, etc.) all remain constant. This naturally prompts an analyser of IS in dialogues to focus on alternating rather than non-alternating pieces of information, which is what we do in this chapter. We do not have to be, and are not concerned with the entirety of ‘old information’ – we are interested only in pieces of information which take part in the process of formation of information structure in a given segment of dialogue.

The distinction between ‘old’ information and ‘shared’ information is thus not important for us. Nevertheless, the issue of what units pieces of information are associated with in sentences, which is seldom a subject of discussion, does concern us. What units do the definitions associate or assign old/new information to? What units are considered by them to bear information (statuses)? Are they syntactic phrases like NPs or VPs, or grammatical categories such as subjects and predicates? This is examined in the next section.

2.3.2. Units Associated with ‘Old/New Information’

We have seen in the preceding section a point where the definitions differ from one another, i.e. the identity of the criterion of ‘oldness’ they assume. There is another point at which definitions of old/new information do not conform. This point, which is the subject of this section, concerns units with which pieces of information are linked.

For some scholars, old/new information is an abstract (semantic) notion which has no ‘tangible’ referent such as morphemes, grammatical categories, and syntactic phrases (e.g. Cruttenden 1997:43-50). Some other scholars (e.g. Kamio and Takami 1998:121), however, use the term as synonymous with constituents/elements/phrases/etc. that contain old/new information. The others are not particular about the distinction between information and its representation or manifestation.⁴⁵

Nevertheless, regardless of the differences between the scholars’ views about the representation or lack of representation of information, virtually all theories of IS link (pieces of) information to certain units in sentences. This is an inevitable consequence which the nature of the study of IS brings about – if one studies involvement of language in the structure of information, and if information is segmented in categories such as

⁴⁵ E.g. Wade’s (1992:528-530) Russian grammar.

‘old’ and ‘new’, linguistic expressions also have to be segmented in some way in accordance with the information segmented into pieces or categorised in separate groups. Even in Lambrecht’s (1994:214) theory cited in the preceding chapter, in which the semantic nature of information and its lack of physical/linguistic representation is emphasised, there are ‘foci’ and ‘focus domains’ which ‘express’ them.

Thus, study of IS assumes an association between parts of speech and pieces of information. This thesis is not an exception and, in our system of IS analysis, pieces of information are linked with morphemes. However, morphemes are not the type of units which are linked with pieces of information in the majority of works on IS. In many theories of IS, the units are strings of morphemes, e.g. syntactico-grammatical units or lexical units.

This general lack of precedence of using morphemes as units to which pieces of information are associated necessitates us to discuss in some detail why morphemes are preferable to larger units as units to which pieces of information are linked for our analysis of IS in agglutinative languages. Accordingly, in the remaining part of this section, we review some previous works in which multi-morphemic units are associated with (pieces of) information, following which our system of IS analysis will be introduced.

In works on old/new information, particularly in those in linguistics, the units with which pieces of information are associated are typically syntactic (e.g. noun phrases) or grammatical (e.g. subjects) or lexical (e.g. noun). One such example taken from Kamio and Takami (1998:120-121), where the term ‘old/new information’ is used to refer to both a category for information and its representation in sentences, is shown below:

What did John do?	[Kamio’s (22)]
He hit Mary.	
old information new information	

Who hit Mary?	[Kamio’s (23)]
John hit her.	
new information old information	

In this example, pieces of information are equated apparently with subjects and predicates. On the face of it, this assignment of ‘old/new information’ to the grammatical

categories looks feasible. However, there is ambiguity which is concealed from analytic view by the use of the irregular verb ‘hit’ in its past tense form. If we alter the verb or tense in the above example, we would have, for instance, a question-answer pairs such as this:

[11] What did John do?

He bashed Mary.

In this example, the ‘bashed Mary’ part which, by analogy with Kamio’s original example, may be identified as ‘new information’, has a past tense marking ‘-ed’ in it. How is this to be analysed when the tense’s being ‘past’ does not appear to be a ‘new’ piece of information (as the use of ‘did’ in (a) indicates)? In this example, we think, the logical basis for assigning ‘new information’ to the whole of the ‘bashed Mary’ part would be, to say the least, ambiguous. This sort of example naturally prompts us to identify morphemes rather than grammatical units as the units to which pieces of information are linked.

Let us look also at the following passage from Chafe (1970:30). It seems to be a syntactic phrase, an NP, that is equated with ‘new information’ in this passage:

[...] a speaker who says I saw your father yesterday is unlikely to assume that the addressee had no previous knowledge of his father, even though by the usual criteria your father would be considered new information. [...]⁴⁶

Reading the above, we are inclined to think the NP ‘your father’ is the unit through which the ‘new information’ manifests itself. This analysis too, however, is not without

⁴⁶ Chafe does not call *your father* ‘new information’ but does not discuss in detail what units are associated to old/new information either. A fuller text of the passage is shown below: What is it? The key to this distinction is the notion of consciousness (Chafe 1974). Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of utterance. So-called new information is what the speaker assumes he is introducing into the addressee’s consciousness by what he says. [...] The terminology has been and continues to be misleading to linguists and psychologists who use it. Calling something “old information” suggests it is “what the listener is expected to know already” and “new information” is “what the listener is not expected to know already,” something that is being introduced into the addressee’s knowledge for the first time (the quotes are from Haviland and Clark 1974). But a speaker who says I saw your father yesterday is unlikely to assume that the addressee had no previous knowledge of his father, even though by the usual criteria your father would be considered new information. The point is that the speaker has assumed that the addressee was not thinking about his father at the moment.

ambiguity. Chafe, for some reason which he does not explain (ibid:27), confines his discussion to nouns, leaving the parts in the sentence aside from ‘your father’ out of the scope of discussion. Consequently, the question of whether, say, ‘yesterday’ is ‘old’ or ‘new’ is left unanswered. This is the case also in the following example:

The next day **he** discovered *the lake*.

According to Tomlin et al (1997:78), in this sentence from E. Rutherford’s (1988:17) novel, ‘the bold-faced NPs are generally taken to be given information and the italicised NPs new information’. However, since they confine their analysis to NPs here, we are left with no clue as to whether we should consider, say, ‘next’ or ‘discovered’ is ‘old’ or ‘new’.

In short, these works aim to identify ‘old/newness’ of some parts of sentences, but not of every part. This is a problem for us, because our purpose is to assign and link pieces of information not to arbitrarily chosen phrases or sequences of morphemes but to all parts of sentences. Their approach does not provide us with a means to exhaustively analyse information structure of sentences. Nor does it have morphology in their framework. This constitutes another factor which inspires us to link pieces of information to morphemes rather than to stretches of words.

2.3.3. Summary

We have seen in this section that identifying grammatical or syntactic units as units corresponding to pieces of information poses problems for our purpose here, namely establishing a direct connection between pieces of information and morphemes.

Constructing a theory based on the abstract distinction of information statuses, i.e. the ‘old-new’ (or ‘presupposition-focus’) dichotomy, then, does not seem to be a promising way to go in order to provide a method of analysis useful for our purpose. We need an alternative approach for analysing information structure of sentences – a system of IS analysis which does not rely on abstract categorisation of information and does not use syntactic or grammatical units as units linked to pieces of information.

In the next chapter we will attempt to devise a system of IS analysis designed to enable linking of information structure and morphology. It will naturally be very different from the ones which have been reviewed in this section. The differences between our system

and others will become immediately apparent when the reader proceeds to read the next chapter where the basis of our system is explained.

2.4. Morpheme-based System

2.4.1. Parameters and Constants

As has been stated a few times before, we do not rely on the abstract distinction of ‘old’ and ‘new’ categories of information. On the other hand, we do segment information into smaller pieces of information. That is, in our system of IS analysis, we divide information into pieces but the pieces do not constitute the categories of ‘old’ and ‘new’ and remain neutral in regard to ‘old/newness’. In other words, we assume that pieces of information bear no inherent characteristics such as ‘old’ and ‘new’.

The ‘contents’ of pieces of information in our system is also different from those in many linguistic theories. Unlike pieces of information in such theories which typically call for lengthy descriptions, our pieces of information are rarely more than such fragments as [past], [pl], [father], etc. This difference derives from the fact that many of such theories link pieces of information with syntactic phrases/units whereas our system links them with morphemes. Assigning pieces of information to syntactic units entails that the pieces of information are ‘bigger’ than ones which single morphemes manage to encode e.g. [pl], [3sg], etc. The natural consequence of this is that a piece of information identified in such analyses can very often be expressed only in the form of a long sentence. The following comment of Lambrecht regarding the information conveyed by the answer in the question-answer pair: ‘When did you move to Switzerland?’ ‘When I was seventeen.’ (1994:48) may serve as an example of such a sentence:

The conveyed information is not “when I was seventeen” but (clumsily expressed) “The time when I moved to Switzerland is the time when I was seventeen.”

In this section, we explain the way our system handles pieces of information, contrasting it with the way other theories handle them. The explanation will serve also as the introductory part to our system of IS analysis.

2.4.2. Variability and Values

Our system of analysis of information structure will be based essentially on ‘variability’, a concept used by psychologist Greenfield (1982) in analysing children’s language. She uses the notion of ‘variability’ basically to account for children’s choice of words at the stage of one- and two-word utterances, but we find the notion useful also in analysing adults’ speech.

According to Greenfield, variability ‘exists to the extent there are alternatives in a given referential situation’ and ‘the alternatives may exist across either space or time’. She gives the following examples for variability (ibid:2):

For example, the mother might tell the child to take off a series of items of clothing, e.g. *hat, jacket*. Here action is constant, while object of action varies. Or, in the opposite case, she might tell the child to first put *on* a hat and then take it *off*. Now action is variable while object remains constant.

The ‘variable’ and ‘constant’ in the first example in this statement where the mother tells the child to take off items of clothing, then, may be schematised as follows:

Action:	Constant	taking off
Object of the action:	‘Variable’	hat/jacket/etc.

Although Greenfield does not mention any other ‘constants’ or ‘variables’, the action is clearly not the only constant in this situation. Certainly, the performer of the action of taking off also remains constant, thus:

Performer of the action:	Constant	child
--------------------------	----------	-------

So does probably the number of the performer of the action (Greenfield writes ‘the child’, not ‘(the) children’, in her comment):

Number of the performer:	Constant	one
--------------------------	----------	-----

So does, say, the tense in which the action takes place: [present]. Thus, a countless number of constants can be added to this list of constants. Accordingly, the constants and variables in this case may be shown as a chart as in the following:

Action:	Constant	taking off
Object of the action:	‘Variable’	hat/jacket/etc.
Performer of the action:	Constant	child
Number of the performer:	Constant	one

Tense:	Constant	present
etc.	Constants	

Note that, in Greenfield's analysis, pieces of information are not given inherent 'old' or 'new' characteristics. For example, the information about the 'performer of the action' maybe constant, but not 'old' as many other theories may hold that it is.

In her analysis, there are pieces of information which remain constant before and after an utterance or linguistic expression. There are, on the other hand, pieces of information which are specified as a result of the utterance/linguistic expression. However, this difference between these two sorts of pieces of information does not derive from inherent 'statuses' (such as 'new' and 'old') of the individual pieces of information. As Greenfield exemplifies by alternating 'item of clothing' and 'action' as 'variables', the same piece of information, e.g. 'hat' or 'taking off', can belong to either of these two groups of pieces of information depending on what is 'variable' in a given situation.

Greenfield's analysis of information suggests that there is an alternative to the IS analysis based on 'old-new' distinction of information and that IS analysis of sentences does not have to involve the assignment of the 'old/new' status to pieces of information. In Greenfield's works, the potential of her analysis to be an antithesis to the popular analysis based on the 'old/new' distinction of information is not fully exploited by her, perhaps because her chief concern was to give a working account on the mechanism behind children's choice of words at the stage of one- or two-word utterances and she did not have much interest in IS as it is discussed by grammarians and linguists. However, we think her analysis deserves attention for its capability of providing an alternative to the style of analysis which a number of scholars have uncritically accepted as the only way to analyse IS. It is basically on Greenfield's approach that our analysis of IS in sentences is based.

A more elaborated version of the idea of variability, which is apparently developed independently from Greenfield's, is found in Clark (1996:296-7). He cites Wittgenstein's (1958:3) example where a builder A and an assistant B communicate in what Wittgenstein calls 'primitive language':

A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a

language consisting of the words “block,” “pillar,” “slab,” “beam.” A calls them out; - B brings the stone which he has learnt to bring at such-and-such a call. [...]

Clark calls the situation where Wittgenstein’s ‘primitive language’ takes place a ‘closed situation’, ‘a social situation’ which is ‘tightly circumscribed, fixed’ and writes the following:

[...] A and B take it as common ground that there is one parameter (type of block) with four possible values (block, pillar, slab, beam), and A’s utterance specifies the intended value.

Then, if A utters not ‘pillar’ or ‘slab’ or ‘block’ but:

Beam.

he specifies the value of the parameter, which corresponds to Greenfield’s term ‘variable’, ‘type of building-stone’, whose possible values are [block/pillar/slab/beam], as ‘beam’. If we symbolise the parameter with x , the utterance yields $x = [\text{beam}]$.

That ‘type of building-stone’ is the only parameter naturally entails that there are no other x s in this ‘closed situation’. What Clark calls a ‘closed situation’, then, can be interpreted as paucity of parameters. Fewness of parameters, in turn, seems to be interpretable as abundance of constants. That is, ‘action’ (which is [passing]), ‘performer of the action’ (which is [B]), etc. have to be constants for ‘type of building-stone’ to be the only parameter.

These infinite number of constants and a single parameter, then, can be schematised thus, with the specified values shown in the far right column:

Action:	Constant	passing
Object of the action:	Constant	building-stone
Number of the object:	Constant	one
Type of building-stone:	Parameter	block/pillar/slab/beam >beam
Performer of the action:	Constant	B
etc.	Constants	

Note that in the above table of parameter and constants⁴⁷, only those that may directly concern our analysis are enlisted. By adding ‘etc’ at the end of the above list, we allow

⁴⁷ Labels attached to constants and parameters in this thesis admittedly lack rigour definitions. Thus, the parameter in Clark’s example was ‘type of block’, which is also Clark’s own expression, and constants were ‘performer of the action’, ‘number of object’, etc. However, these labels secure intuitive

for the existence of an enormous number of constants that make the joint activity by the interlocutors, i.e. discourse possible. The constants in the ‘et cetera’ include, for example, ‘the language of communication’ and ‘B is hearing-impaired’ whose values are pre-specified as [English] and [Negative], respectively. Without an immense number of constants, communication could easily be an impossibility. It should be also noted that the table is constantly updated every moment as the values of parameters are specified.

Constants, or rather, specified values for constants, then, seem to constitute something equivalent to ‘old information’ discussed in the previous section. This in turn means that the above analysis resolves ‘old information’ into an immense number of (pre-)specified values.

Why, then, do we want to break down the single entity of ‘old information’ into specified values? There is an advantage in having many sizeable (constant) pieces of information or specified values rather than a huge collective whole of ‘old information’. This segmentation of ‘old information’ into such values as [one], [beam], etc. is, in fact, instrumental in claiming a correspondence between information structure and morphology which we aim to present, because each one of pieces of information which we link to morphemes can be only as ‘big’ as a piece of information encoded by a morpheme. This is to say that they are typically such pieces of information as [pl(ural)], [past], [neg(ative)], [beam], [you], etc. However, we are not adopting Greenfield’s and

understanding of what constants and parameters represent and we continue to use them in the following discussion. It has to be emphasised that constants and parameters are directly linked to morphological units, i.e. variables and non-variables. This link naturally presupposes that each parameter/constant is representative of (a) piece(s) of information to which morphemes correspond. Consequently, many of the labels of parameters/constants reflect the functions of the morphemes which they have links with. For example, the label of a parameter/constant whose value is specified by a passive morpheme may be represented as ‘passivity’ or simply ‘passive’. Similarly, parameters/constants may be labelled as ‘causality’, ‘potentiality’, ‘action performed’ (or ‘sentence negation’), and ‘tense’, in accordance with the types of morphemes which specifies their values.

A label may contain different types of information (semantic, grammatical, functional etc.) and will often be represented as a mixture of them. This is because morphemes are rarely only semantic or grammatical or functional and it is not our purpose here to be able to identify exactly which one of the types of information encoded by a given morpheme specifies (a) value(s) for a parameter or constant. It is unnecessary for the present discussion to classify the pieces of information which each morpheme

Clark's analytical framework simply because it is convenient for our analysis. Their system of IS analysis has at least one big advantage over many other systems of IS analysis.

The segmentation of information into sizeable values allows us to identify very specifically values which are specified by a sentence, something a number of theories previously proposed overlook. Recall, for example, Lambrecht's example cited in the preceding chapter (copied below for convenience) where he claims that the 'focus' of the answer is 'broke down':

What happened to your car?

My car broke down.

As was stated before, the piece of information 'broke down' which, according to Lambrecht's definition of 'focus', must not contain any information (or semantic component) which the speaker assumes the hearer already knows, does contain at least one piece of information which the speaker assumes the hearer already knows, namely [past].

This piece of information can be rendered as the pre-specified value [past] for the constant 'tense' in the analytical framework of Clark's and can be isolated from values specified by the sentence 'my car broke down'. This is an advantage which segmentation of information that Clark's and Greenfield's systems of analysis have over many other systems of IS analysis – they allow one to distinguish more rigidly between what Lambrecht calls 'presupposition' and 'focus'.

Now, let us reexamine Clark's example above. In our system of IS analysis, the sentence/morpheme /beam/ does not convey information such as 'the type of block A wants B to pass to him is beam', as many other systems may claim it does. /Beam/ occurs to specify the value [beam] for the parameter 'type of building-stone'. The morpheme encodes the piece of information [beam], but it is only instrumental in specifying the value for the parameter and does not have an inherent 'old' or 'new' status.

In Clark's example, the only parameter is 'type of building-stone', whose value is specified as [beam] by A's utterance of the morpheme /beam/. Is this apparent

encodes according to their types and make a list of them. More rigid formalisation of this labelling convention is desirable, but casual labelling suffices for the discussion here.

correspondence between the specified value [beam] and A's choice of the only morpheme to be uttered, which is /beam/, a coincidence? We think there is a reason why specifically the occurrence of the morpheme /beam/ is called for here: the morpheme has to occur because only 'beam' is needed to specify the value for the parameter 'type of building-stone'.

Note that in many other systems of IS analysis, an account for the apparently obligatory occurrence of the morpheme /beam/ cannot be straightforward. If we assume that /beam/ conveys the information of, say, 'the type of block A wants B to pass to him is beam', we would be at a loss to explain why specifically the morpheme (or sentence) /beam/ has to occur.

Accordingly, the first claim that we make regarding the relation between information structure and morphology will be the correspondence between values and morphemes. This will be the topic of the next section and should enable us not to leave any morpheme or constituent in a sentence ambiguous as to the value it specifies or pre-specifies. This will be an advantage of our approach, because, as we have seen before, a number of other works (e.g. Prince 1992, Takami 1995) typically neglect constituents/morphemes they do not specifically take up for discussion.

2.4.3. Variables and Non-variables

In this section, we will classify morphemes in sentences into two categories: morphemes that specify values for parameters and those that specify values for constants, which will be called variables and non-variables, respectively.

However, before we go into discussion, the notion of 'allosentences' which will be used extensively in the following discussion needs to be explained.

2.4.3.1. Allosentences

In explaining variable and non-variables, we will benefit from the notion of **allosentence**, a term introduced by Daneš (1966) and used extensively in Lambrecht (1994). Lambrecht uses the term in the sense of:

[...] semantically equivalent but formally and pragmatically divergent sentence pairs, such as active vs. passive, canonical vs. topicalized, canonical vs. clefted or dislocated, subject-accented vs. predicate-accented sentences, etc. (ibid:6)

He gives the following sentences as an example of a pair of allosentences (ibid:17):

She likes GERMANS.

It is GERMANS that she likes.

Put simply, Lambrecht's allosentences are 'semantically equivalent but formally and pragmatically divergent surface manifestations of given propositions' (ibid:35).

Lambrecht does not give a conclusive definition of 'proposition', but since he says 'we may refer to this sum of propositions [...] as the hearer's knowledge' (ibid:44), it is probably reasonable to think it as something like (a) piece(s) of information being conveyed by the speaker.

I adopt Lambrecht's definition of allosentence with a modification, which is in appearance minor, but in effect major. The modification is replacement of his 'proposition' with 'specification of values for parameters'. Accordingly, our allosentences are 'semantically equivalent but formally and pragmatically divergent surface manifestations of given specification of values for parameters'. Allosentences in this thesis, then, will mean **sentences which, despite their morphological and syntactic differences, specify the same value for a given parameter**. For example, 'Ali' and 'I met Ali' as two possible answers to the question 'Who did you meet Ali?' are allosentences because they both specify the value of the parameter 'object of meeting' as 'Ali', despite their formal differences. Allosentences are explained in more detail in the next subsection 2.4.3.2. In short, as long as the value they specify for the parameter is the same and the constants have the same values, they are a set of allosentences.

As a result of this modification, some sets of sentences Lambrecht identifies as sets of allosentences fall outside our version of category of allosentence or become unclear as to whether they are allosentences or not.

For example, we cannot tell with any certainty whether the 'she likes Germans' pair cited above are allosentences or not, because what the parameter(s) is/are is not certain. If the parameter or parameters for which the sentences specify value(s) is/are unknown, naturally, whether they specify the same value(s) for the same parameter(s) is also unknown. As a result, we cannot be sure whether the sentences are allosentences or not.

The following is also a set of sentences which are allosentences by Lambrecht's (ibid:121) definition but not by ours, because, as the first three of these explicitly show, the parameters in (a)-(d) are different from one another:

- a. (What did the children do next?) The children went to SCHOOL.
- b. (Who went to school?) The CHILDREN went to school.
- c. (What happened?) The CHILDREN went to SCHOOL!
- d. (John was very busy that morning.) After the children went to SCHOOL, he had to clean the house and go shopping for the party.

The sentence (a) specifies a value for the parameter 'action', whereas (b) does for the parameter 'performer of action'. (c), on the other hand, specifies values for multiple parameters. As for (d), we cannot say with any certainty what are constants/parameters. Thus, these are not allosentences by our definition, because they do not share the same parameter(s). They hence do not specify the same value for a parameter (or values for parameters) and therefore are not allosentences as they are defined here.

Having introduced and defined the notion of allosentences, in the following subsection, we will explain what variables and non-variables are, using comparisons between allosentences.

2.4.3.2. Variables and Non-variables

Clark's analysis reviewed before neatly explains the linguistic interaction in what he calls closed situations where there are few parameters. However, we think the notion of variability is applicable also to less 'closed' situations. We will discuss below question-answer pairs in such less-closed situations. In doing so, we introduce and explain the notions of **variables** and **non-variables** with examples. The discussion will also familiarise us with the notions of parameters, constants, and allosentences.

Though these notions have been introduced to analyse the languages that are the subjects of this thesis, in this preliminary section we use English sentences for ease of exposition. First, observe the following constructed question-answer pair:

- [12] a Who did you meet?
 b Ali.

Let us schematise the constants and parameter as well as their values as we did with the previously cited examples:

		Values
Action:	Constant	meeting
Performer of the action:	Constant	answerer
Object of the action:	Parameter	Jamila/Ali/Ayshamgul/etc. > Ali
Tense:	Constant	past
etc.	Constant	

Now that the parameter and the value the answer specifies is clear, we can devise an allosentence of (b). As is explained in the preceding subsection, allosentences have to specify the same value for the parameter. In this case, then, an allosentence of (b) has to specify the value of the parameter ‘object of the action’ as [Ali]. Accordingly, we devise the following as (b)’s allosentence:

b’ I met Ali.

This sentence is formally and probably also pragmatically divergent from (b). For instance, the ‘I met Ali’ answer may induce the questioner to expect the answerer to go on to talk about someone who did not meet Ali, or his having met also Ali’s wife, which expectation (b) does not arouse. Nevertheless, the sentence (b’) specifies the same value ‘Ali’ for the parameter ‘object of action’ as (b) does. The sentences (b) and (b’) are therefore allosentences by our definition.

Having devised allosentences, now we turn to ‘variables’ and ‘non-variables’. What are they and how do we classify morphemes in sentences into variables and non-variables? These will be explained in the next few paragraphs, using an English example. Our use of an English example here is entirely expedient – it is certainly not ideal for explaining our system of IS analysis for agglutinative languages and induces some unwanted complexity in the explanation. However, the benefit of the reader’s understanding of the notions of ‘variables’ and ‘non-variables’ which our use of an example in English may facilitate may outweigh the benefit of presenting a more appropriate example.

Accordingly, we opt to use an English example, but only in this preliminary section.

Variables are morphemes that specify values for parameters. As we will see below, allosentences prove to be instrumental in distinguishing variables from non-variables.

First, let us compare the allosentences (b) and (b’):

[13] a Who did you meet?

b Ali.

b' I met Ali.

The morpheme /Ali/, which is the value-specifier for the parameter in both (b) and (b'), is an example of what we call variables in this thesis.

Note that both (b) and (b') contain the variable /Ali/, which is the only morpheme which the allosentences have in common – it appears as if the occurrence of the variable is essential in providing an answer to (a), regardless of the allosentences' formal and pragmatic properties. When we consider the fact that the single-word utterance we have seen in the preceding section also consists of one variable /beam/, there seems to be some sort of an information-structural constraint involved in the production of these sentences. The constraint will be discussed at length in the following subsection.

Non-variables, on the other hand, are morphemes that redundantly specify values for constants. Non-variables in the above example are, therefore, 'I' and 'met' in (b') which redundantly specify the values for the constants such as 'performer of the action', 'action', and 'tense'⁴⁸.

To sum up, **variables are morphemes that take part in the specification of the values of parameters**. As for non-variables, they are morphemes other than variables.

According to these definitions of variables and non-variables, not syntactic phrases or grammatical categories but morphemes are the units responsible for altering or retaining the information structure of sentences. This is to say that, in our system of IS analysis, the information structure of a sentence equals the distribution of variables and non-variables in the sentence.

By postulating the 'morpho-informational' units of variables and non-variables, we have, in effect, asserted the existence of a direct connection between morphemes and pieces of information and proposed a new theory to information structure based on morphemes. Accordingly, our definition of morpheme-based IS for the present system is much simpler than most other definitions of IS:

Information Structure: Distribution of variables and non-variables in sentences.

⁴⁸ We consider that the 'met' morpheme (redundantly) specifies values for the constants 'action' and 'tense'.

Needless to say, this morpheme-based system of IS analysis could not have been derived from the popular line of IS analyses which are based on syntactic units, some of which were reviewed in the preceding chapter. There is a reason why in the present study which aims to provide a system of IS analysis specifically for agglutinative languages the morpheme-based approach to IS is preferred over the by far more popular phrase-based approach. The preference has to do with the defining characteristic of agglutination, i.e. ready segmentability of words into morphemes.

The lexicon and inventory of morphemes of an agglutinative language, unlike those of a fusional language, consist predominantly of readily identifiable morphemes, whose forms are largely invariable. We have already seen the robust invariability of morphemes in agglutinative languages in the preceding chapter in such examples as *adam* [man], *adam-lar* [man-pl], *adam-lar-ı* [man-pl-acc]. Ready identifiability of morphemes in agglutinative languages, which is supported by the invariability of forms of morphemes, enables us to use pieces of information which morphemes necessarily encode as those that form IS of sentences. This feature of agglutinative languages is both what prompts us to propose a morpheme-based analysis of IS of sentences and a prerequisite for this proposal. Our assumption is that, if each morpheme can encode a piece of information, they can individually take part in specifying values for parameters and constants.

In the following section where we analyse examples from agglutinative languages using the notions of variables and non-variables, this assumption that each morpheme can be a value-specifier for parameters (and also constants) will prove valid.

2.4.3.3. Summary

In this subsection, we have introduced some concepts which are essential for the morpheme-based system of IS analysis. In the next subsection, we will embark on analysis of example sentences from agglutinative languages, using these concepts of parameters, constants, allosentences, variables, and non-variables.

2.4.4. Analysis Using Variables and Non-variables

In the first subsection in this section, we examine a linguistic feature which the languages dealt with in this thesis have in common. The feature, which is referred to as ‘incomplete

sentences’, ‘cut(-off) sentences’ etc.⁴⁹ by grammarians of the languages, is given an alternative account here based on the notions introduced in the preceding section. This feature is most commonly analysed to be a result of ‘ellipsis’ or ‘pro-drop’ in the linguistic literature.

Observe the following Tajik passage from Hakim Qarim cited in Rustamov and Ghafforov (1986:336. The glosses and translations are mine) According to them, it is common for such an answer, to take the form of an ‘incomplete’ short sentence, which is referred to as *jumlai nopurra* ‘incomplete sentence’:

- [14] a *Tu ki-st-ī?*
you who-cop-2sg
‘Who are you?’
b *Odam.*
human being
‘(a) human being.’

They say the answer can appear also as a ‘complete’ sentence, though the ‘complete’ version is not suggested. In the first subsection, it will be shown that the use of the notion of variables will facilitate a better and more consistent explanation for so-called ‘ellipsis’ and ‘phonetically null subjects’ than previous ones do.

In the second subsection, we look at what we term ‘echoing’ as well as ‘unattended parameters’.

Finally, in the third subsection, we deal with sentences which are ungrammatical in the sense that they do not comply with rules set by classical sentence grammar. Such sentences, despite their diversion from the classical grammatical rules, often make very good sense, as can be observed in the answer of the following dialogue taken from van Dijk (1981:77):

- (6) A: With what has the postman been murdered?
B: John thinks with a knife.

The notion of variables proves useful in providing an account for the mechanism behind the generation of these sorts of sentences which have what van Dijk calls ‘relative grammaticality’.

⁴⁹ For example: Turkish *kesik cümle* ‘cut(-off) sentence’, *eksiltili cümle* ‘sentence with reduction’, Uzbek

Within the third subsection, a brief discussion about a possible relation between so-called ‘case marker drop’ and IS will also be provided. ‘Case marker drop’ is found in all the languages from which we take data in this thesis. For example:

[15] *Xoni Šohida-ba⁵⁰ telefon kard-am*
house.iz Shohida-to/dat telephone did-1sg
‘I phoned Shohida’s house’

The ‘dative’ case marker *-ba* does not have to appear in this Bukharan Tajik sentence. Similarly, the following sentence without *-ba* is also acceptable:

Xoni Šohida telefon kardam

However, in the following example, *-ba* has to appear:

[16] *Oni Šohida-ba telefon kard-am*
mother.iz Shohida-to/dat telephone did-1sg
‘I phoned Shohida’s mother’

There have been attempts to explain this kind of apparent absence of case markers in terms of definiteness or animacy of nominals to which case markers are attached. We will attempt to explain this phenomenon using the notions introduced in the preceding section.

Advantages which the morpheme-based system has in analysing IS in agglutinative languages will become apparent through the simplicity of the accounts which we will provide for the phenomena listed above.

2.4.4.1. ‘Ellipsis’ or ‘Inaudible Elements’

In this subsection, we examine the phenomena which are often referred to by grammarians and linguists as ‘ellipsis’ (and ‘gaps’) and ‘phonetically null’ elements, e.g. ‘pro’. We provide, using variables and non-variables, a unified way of analysing these

to ‘liqsiz gap ‘incomplete sentence’, Tajik *jumlai nopurra* ‘incomplete sentence’.

⁵⁰ We present *ba* ‘to’, which is a preposition in literary Tajik, with a hyphen here, following the convention in the literature of Tajik dialectology (e.g. Mahmudov and Berdiev 1989, Kerimova 1959, Rastorgueva 1956, 1964). A native speaker of Bukharan Tajik writes it suffixed to words preceding it in his writing, which probably reflects accurately how native speakers, who all know Uzbek, identify the morpheme.

phenomena which are treated separately by many scholars. Firstly, some accounts for ‘ellipsis’ or ‘elision’ in general will be reviewed, following which an analysis of the phenomenon using the morpheme-based system of IS analysis will be provided.

Let us start with looking at a simple account for ‘ellipsis’ in a book aimed at general readers. Palmer (1981:38) writes:

It is simple enough to envisage a situation in which someone might simply say *Horses*. This could be in reply to a question such as *What are those animals in that field?* Although some scholars have talked of ‘one word sentences’ in describing such expressions, it seems more helpful to treat *Horses* as a sentence fragment and as an incomplete version of *They are horses*; certainly we should need to reconstruct the complete sentence in this way to talk about its meaning.

He further states that *Coming?* or *Coming!* may be used instead of *Are you coming?* or *I’m coming!*. The extreme simplicity of this view on ‘elliptical’ sentences, which appears to be based on intuition, is probably justified provided that the book is geared towards general readers rather than to specialists. However, in fact, this ‘ellipsis’ view forms the basic assumption which prevalent theories in the literature of linguistics adopt, although they typically make use of more complicated linguistic notions. For example, Kornfilt (1997:128) gives a similar explanation to Palmer’s for a Turkish example:

[...] if the pragmatic situation makes the reference of a constituent clear, that constituent can be elided. For example, suppose a setting where speaker A is looking for his glasses. Speaker B finds them, holds them up and says:

bul -du -m!
find -Past -1.sg.
“I found (them)!”

Here, the direct object has been elided.

Obviously, the basic assumption behind this explanation is the same as the one employed in Palmer’s explanation. The assumption is, put simply, that ‘**sentences which have fewer components than they can (or should) have according to the classical sentence grammar miss some components**’. (‘Classical sentence grammar’ will be hereafter abbreviated as ‘CS grammar’.) The same assumption is shared by some Turkish and Tajik grammarians. For example, the following Turkish definition of *eksilti* ‘reduction’ is a paraphrase of the assumption above: ‘composition which misses some components compared to the form under normal conditions (i.e. some components that ‘normal’

composition has) [...]’ (Berke et al. 1988:95 cited in Çokuğurel 1996:2). In Tajik linguistics, the same assumption was present in as early as 1930, when the standardization of the Tajik language was still in progress. A textbook of Tajik language published in 1930 (Fitrat 1930:103) gives some question-answer pairs such as *Ki omad?* – *Şarif.* (*Şarifomad*) and explains that items that have appeared in the interrogative sentence fall off from the ‘answer-giving’ sentence. This assumption, however, yields more confusion than it does explication.

How can one confirm what is ‘elided’? In Kornfilt’s example where the ‘elided constituent’ is considered to be ‘them’, the ‘reconstruction’ may appear to be feasible. However, the ‘elision’ explanation faces difficulty once we look at a somewhat wider variety of examples. For example, in the following utterance by a man who came into a bookshop⁵¹, it is unlikely that one can tell with certainty what constituent is ‘elided’:

[17] *Balkan-lar-in Tarih-i, Georges Castellan.*
 Balkan-pl-gen history-3sg Georges Castellan
 ‘The History of the Balkans, Georges Castellan.’

The owner of the bookshop replied with the following to the man who is presumably from another bookshop in the alley:

[18] *Yok!*
 non-existent
 ‘There is not (in effect: we do not stock that book).’

The man’s utterance, in effect, is for asking whether the bookshop stocks the book the title of which is ‘The history of the Balkans’ and the author of which is ‘Georges Castellan’. Note that no utterance or dialogue took place prior to the man’s utterance; the succession of a book title and author was the very first utterance after his entrance to the shop. It seems to be difficult to tell with any certainty what is ‘elided’ in the utterance, if we should subscribe to the ‘elision’ explanation. As is clear in this example, the ‘elision’ explanation has at least one difficulty – talking about ‘elided constituents’ without being able to tell what is the elided constituent.

The assumption that sentences with fewer components than they should have according to CS grammar miss some components has a reflection also on syntactic theories. There is an explanation for elements which appear to many syntacticians to be unexpressed (or

For example, in Turkish linguistics, some grammarians, following the intuitively satisfying generalisation that ‘subjects can “drop” when verbal inflection is rich enough to identify at least some features of the missing subject’ (Franks 1995:288-9), ascribes the ‘drop of subject’ in Turkish to its verbal agreement suffixes and claims that subjects are used to obtain opacity, to draw attention to the persons (who perform the action denoted by the verb), or to ‘intensify’ the subject (Gencan 1979:140). In Tajik linguistics, Rustamov and Ghafforov adopt the same view as Gencan’s and write the following regarding the sentence *nadoram*. [neg-have-1sg] ‘(I) don’t have (it/one).’⁵³:

[...] mentioning of the subject is not inevitable [*hatmī*], because it is evident from the form of the verb that the person is first person singular⁵⁴ (1986:335 free translation by me).

However, this reasoning is not always applicable, because, as has been repeatedly pointed out by linguists (including Franks himself), ‘[t]here is a large class of languages [...] that drop subjects despite the complete absence of any subject-verb agreement whatsoever (Franks 1995:288-9)’. Moreover, it does not explain the obligatory occurrence of subjects in the presence of their agreement suffixes in examples we will examine later. These facts, together with the previously addressed problem which the ‘elision’ explanation has, we think, leave little good reason to assume the existence of ‘inaudible’ or ‘dropped’ elements in sentences.

To sum up, despite the different terminologies (‘ellipsis’, ‘elision’, ‘inaudible’, ‘dropped’, and ‘omission’) and the theoretical frameworks which they employ, all the explanations cited above have one assumption in common, i.e. ‘sentences with fewer components than they can have according to CS grammar miss some components’. However, as we have seen in the paragraphs above, the assumption that ‘complete’ sentences exist for all ‘incomplete’ sentences/utterances does not seem to be generally applicable. This is probably the reason why ‘ellipsis’ or ‘inaudible elements’ have attracted such criticism as ‘[an attempt] to “explain” a grammatical structure not by what is in the text but by what the analyst wished were there instead’ (Miller 1970:xxxiii). We are inclined to agree with this criticism, at least to a certain extent, which is why we prefer Greenfield’s bottom-up approach to the top-down approach used by many.

⁵³ As an answer to the following question: —*Metini tu kanī, hoy, pisaram?* ‘Where’s your pickaxe, my son?’.

⁵⁴ [...] *zikri muftado hatmī nest, čunki az šakli fe’l ma’lum ast, ki on šaxsi yakumi tanhost.*

Accordingly, we do not subscribe to these explanations which ascribe absence of constituents to agreement morphology. In this thesis, we put aside this assumption of ‘ellipsis/elision/omission’ of constituents (or components or elements) in sentences and see whether we can provide an alternative, and hopefully also functional, analysis of the kinds of sentences observed above.

Mahmudov and Nurmonov (1995:95) claim that not all ‘elements of existence’ (*borliq unsurlari*) have to be expressed with the help of linguistic means and that the elements which are not directly expressed are understood from the situation or context. While this idea itself may not sound particularly new – for example, in the last century, Vygotsky (1934 reprinted in 1962 and cited in Greenfield and Zukow 1978:294) noted expressions which are “abbreviated” in certain situations – it is closer to the one employed here than the ones employed in the explanations reviewed above. Our view on sentences which have fewer components than they can or should have according to CS grammar will be explained in the next subsection.

2.4.4.1.1. Rule 1

We will propose several rules based on the notions of variables and non-variables. Successful accounts for some linguistic phenomena by the rules, naturally, will support the feasibility of the postulation of the morpho-informational units. The rules are the following:

Rule 1: In informative communication in an agglutinative language, the following exist:

Rule 1-1: Variables.

Rule 1-2: Non-variables grammatically required⁵⁵ to accompany variables.⁵⁶

Rule 1 lists two groups of morphemes which are essential for informative communication and hence have to appear in sentences. What, then, about morphemes which are neither variables nor their accompaniments? Our claim is that they are not essential in

⁵⁵ What ‘grammatically required’ means will be discussed in detail in the subsection ‘language specificity in the operation of rule 1-2’.

⁵⁶ See, however, the analysis of the first example in the subsection ‘Pre-grammatical Utterances in Adult Speech’.

informative communication and hence may not appear even when their appearance is predicted by CS grammar.

This rule, then, may be capable of accounting for the phenomenon which is variously called ‘ellipsis’, ‘omission’, etc., because these terms are used to refer to cases where many people think some elements in a sentence are ‘missing’. While the analyses reviewed in the preceding subsection invariably contrast such sentences with sentences in which no constituent is ‘missed’ or ‘inaudible’, our analysis involves no such contrast of ‘incomplete’ and ‘complete’ sentences, because rule 1 can yield only self-contained, i.e. ‘complete’, sentences consisting only of morphemes that need to appear.

The difference between our analysis and those many others may be better explained by looking at how they differ in identifying self-contained sentences. Many previous analyses see sentences consisting of the most constituents they can possibly have as ‘complete’. According to these analyses, sentences which have a less number of constituents than they can have in them have to be considered ‘incomplete’ or to have ‘inaudible’ constituents. In other words, this is a top-down approach to such sentences – in this approach, constituents in ‘complete’ sentences are mapped onto actual sentences which may not contain all the constituents that appear in the assumed ‘complete’ sentences. Hence Palmer’s explanation that ‘horses’ is a fragment of ‘they are horses’.

On the other hand, the system of analysis employed here takes a bottom-up approach where morphemes appear as they are called for by rule 1. Sentences in our analysis are, then, assembled morphemes rather than (fragments of) sentences as the term is used in a large number of analyses. Naturally, in our analysis, there are no ‘complete’ sentences to be mapped onto the assembled morphemes (= sentences). Our analysis will be exemplified in the next subsection.

If the rule accounts for ‘ellipsis’ in a convincing way, the claimed feasibility of the rule, and hence also the validity of our postulation of variables and non-variables would be reinforced. In order to validate the notions of variables/non-variables, in the remaining part of this section we will examine the rule against examples from agglutinative languages.

The following discussion is divided into two sections. The first section deals with sentences in each of which there is only one variable. The second analyses sentences with multiple variables.

2.4.4.1.2. Variable = Free Morpheme

‘Elliptic’ sentences have been analysed by a number of scholars (e.g. Kuno 1978, Boboeva 1978) including the ones whose explanations for the phenomenon have been cited in the preceding subsection. However, their largely separate efforts have not established a unified way of analysing the phenomenon of ‘ellipsis’ (or ‘omission’ or ‘incompletion’) in its entirety. We attempt to provide a new standpoint from which to approach the phenomenon as a whole, using variables and non-variables. It will be shown that rule 1 has an explanatory power for this phenomenon (at least in agglutinative languages), which will then support the validity of our system of analysis.

Let us start our discussion with a simple Turkish example. The kind of ‘ellipsis’ explanation which we have reviewed in the preceding section is adopted by the anonymous author of the Turkish grammar from which the following example is taken (Turska gramatika:255, glosses and translations are mine):

- [19] a *Ahmet ne zaman gel-miş?*
 Ahmet what time come-IE
 ‘When did Ahmet come?’
- b *Dün.*
 yesterday
 ‘Yesterday.’

The grammar calls the answer *dün* ‘yesterday’ ‘incomplete (elliptic)’ and explains that the answer represents the following:

- b’ *Ahmet dün gel-miş.*
 Ahmet yesterday come-IE
 ‘Ahmet came yesterday.’

Explanations similar to these ones are found also in Japanese works some of which prefer the term ‘omission’ to ‘elliptic’ or ‘incomplete’ sentences (e.g. Kuno 1978:Chapter 1).

As was stated before, in our analysis in which we use the notion of variables, the explanation for this example becomes radically different from these ones.

On the face of it, the explanation reviewed above may seem to account in sufficient detail account the occurrence of the sentence *dün* ‘yesterday’. However, this intuitively

satisfying account gives us no clues as to why *dün*, and not, say, *Ahmet* or *gelmiş*, has to occur in (b). If (b) is an incomplete version of (b'), what makes only *dün* but not *Ahmet* or *gelmiş*, which are constituents in (b') like *dün* is, appear in (b)? Note that *dün* is the only morpheme the allosentences (b) and (b') have in common – does the appearance of only *dün* ‘yesterday’, for some reason, have an exclusive importance in answering the question?

In fact, if the answer consisting of every morpheme which (b') has in it but *dün* ‘yesterday’, the answer would be unacceptable despite its grammatical correctness, as can be seen below:

[20] a ‘When did Ahmet come?’
 b’## *Ahmet gel-miş*
 Ahmet come-IE
 ‘Ahmet came.’

The sentence (b'') is grammatically correct, but is unacceptable as an answer to (a). What, then, causes the unacceptability of (b'') as an answer to (a), which unacceptability becomes obvious even by looking at the English translations of (a) and (b'')? As has been implied before, this is probably because the appearance of only *dün* ‘yesterday’ (and no other morphemes) has an exclusive importance in determining the acceptability of possible answers to (a). The reason why the appearance of *dün* alone is essential can be explained by rule 1-1 – according to our analysis of the sentence, which will be explained in the following few paragraphs, the morpheme *dün* has to appear because it is a variable.

Let us analyse (b) and (b') to see the value(s) specified by them. Observe the following table of constants and parameter for the example above:

Action:	Constant	coming
Tense:	Constant	past
Mood:	Constant	indirect experience (inference)
Performer of the action:	Constant	Ahmet
When the action took place:	Parameter	yesterday/a week ago/etc. > yesterday etc.
	Constants	

Obviously, the sentences (b) and (b') specify the same value [yesterday] for the same parameter ‘when the action took place’ although they are different in terms of linguistic

expression. This makes (b) and (b') each other's allosentences. Which morpheme, then, is the variable, i.e. the morpheme that specifies the value? The variable is necessarily *dün* 'yesterday', because there are no morphemes but *dün* in (b). If *dün* was not a variable, (b), which consists only of *dün* would not be able to specify the value for the parameter.

Thus, the morpheme *dün* 'yesterday' is a variable. According to rule 1-1, variables are morphemes whose occurrence is required for informative communication. This, in turn, explains the acceptability of (b) and (b') as well as the unacceptability of (b''). The variable *dün* occurs in both (b) and (b'). On the other hand, (b'') has all the morphemes which the acceptable (b') has except the variable *dün*. In other words, the morphemes aside from the variable may or may not appear in an answer to (a) – their occurrence does not contribute crucially to informative communication. In other words, the non-variables present in (b') accompany the variable but their presence is not essential in informative communication.

This suggests the exclusive importance which the occurrence of the variable has in determining the acceptability of sentences. Note that the variable appears in isolation in (b) – the acceptability (and also naturalness) of (b) is evidence that rule 1-1 is in operation in the production of the mono-morphemic sentence.

Let us examine some other examples to see whether rule 1-1 works in other examples as it does in the example above. Gencan (1979:141-142, the glosses and translations are mine) presents the following question-answer pair as an example where the predicate 'drops':

- [21] a *Kim uy-uyor?*
 who sleep-pr.prog.3sg
 'Who is sleeping?'
- b *Turgut.*
 Turgut
 'Turgut is.'

Gencan claims that the answer *Turgut* is in fact *Turgut uyuyor* 'Turgut is sleeping' from which the predicate *uyuyor* 'sleeping' has been 'dropped', thus:

- [22] a 'Who is sleeping?'
- b *Turgut.*
- b' *Turgut uyuyor.*

This explanation is along the same lines as the one given in the Turkish grammar previously cited. The deficiency of Gencan’s explanation is in its not explaining why specifically *uyuyor* ‘sleeping’ is ‘dropped’. Similarly, it does not explain why *Turgut* cannot be ‘dropped’. Why is the following sentence from which *Turgut*, instead of *uyuyor*, is ‘dropped’ not informative or even acceptable, despite its grammatical correctness?:

- [23] a ‘Who is sleeping?’
 b” ## *Uyuyor.*
 sleep-pr.prog.3sg
 ‘(it/he/she) is sleeping’

With rule 1-1, *Turgut*’s obligatory presence can be explained. Let us repeat the process of analysis we have gone through when we analysed the previous example. The parameter and constants of Gencan’s example are as follows:

Action:	Constant	sleeping	
Tense:	Constant	present	
Aspect:	Constant	progressive	
Performer of the action:	Parameter	Ayşe/Ahmet/Turgut/etc.	> Turgut
etc.	Constants		

The value for the parameter ‘performer of the action’ is to be specified, for which, obviously, the morpheme *Turgut* is responsible. That is, *Turgut* is a variable, the obligatory presence of which is predicted by rule 1-1. On the other hand, the non-obligatory occurrence of any other morphemes (which are non-variables) is also predicted by rule 1-1. Thus, rule 1-1 provides an explanation for the reason why anything but *Turgut* cannot be absent in the answer in Gencan’s example.

Thus, rule 1-1 explains the obligatory appearance of *dün* in the previous example and *Turgut* in Gencan’s example in a unified manner. This is a piece of evidence which increases the reliability of rule 1 and hence also the notion of variable/non-variable on which the rule is based.

Note also that the single principle expressed as rule 1-1 accounts for the occurrence and non-occurrence of morphemes with various different functional loads. Perhaps more importantly, rule 1-1 accounts for the non-occurrence of different syntactic units – in the *dün* example, non-variables which are present in (b’) but not in (b) constitute an NP and

VP, while in Gencan's example, non-variables in *Turgut uyuyor* make up a VP. Rule 1-1, then, is capable of handling different types of apparent non-occurrence of syntactic units without classifying them into pro, ellipsis, etc. This is one of the reasons why we think rule 1-1 is preferable to many other principles proposed in previous works explaining apparent non-occurrence of constituents in sentences in agglutinative languages.

The examples cited above are both from Turkish, but the applicability of rule 1-1 is by no means limited to Turkish, as is clear in the following Tajik 'incomplete sentence' (*jumlai nopurra*) taken from Niëzmuhammadov et al. (1955:50, my glosses and translations) where the parameter is B's name, which is specified as [Pūlod⁵⁷] by the variable *Pūlod*:

Nom-i tu chi-st?
 name-izf you what-cop
 'What is your name?'

Pūlod.
 Pūlod
 'Pūlod.'

Rule 1-1 operates also in the following Uzbek example of 'incomplete'⁵⁸ sentence (*to 'liqsiz gap*), where the parameter 'how Ljudmila's life is' is specified as 'very good' by the set of variables *juda yaxshi* (Gor'kij 1955:197 cited in Boboeva 1978:42, the glosses and translations are mine):

Rašel'. *Turmush-ing qalay?*
 life-2sg how
 'How's your life?'

Ljudmila. *Juda yaxshi.*
 very good
 'Very good.'

We have seen cases where only variables have to appear, which, in turn, supports the feasibility of rule 1-1: 'for informative communication, occurrence of variables is

⁵⁷ As a common noun, *pūlod* means 'steel'.

⁵⁸ It should be noted that, despite the term, some Uzbek grammarians do not necessarily consider them 'incomplete' (see Boboeva 1978: particularly p12).

essential'. The production of the sentences cited above which are explained to be 'incomplete' are thus simply and uniformly accounted for by rule 1-1.

In fact, the sentences cited above (*dün, Turgut, Pülod, juda yaxshi*) share a property other than consisting exclusively of variables. In all of them, the variables are free morphemes. Rule 1 has proved to be capable of accounting for such sentences, but can it explain cases where variables are function or bound morphemes which do not appear in isolation, i.e. appear attached/accompanying other (free) morphemes? It can, and it is where rule 1-2 is in operation.

2.4.4.1.3. Variable = Bound Morpheme

In the previous subsection, we have looked at cases where variables are free morphemes. The restriction of examples to ones in which variables are always free morphemes allowed us to observe the operation of rule 1-1 in its simplest form. However, it also concealed the potential problem the rule has – occurrence of variables is called for by rule 1-1, but what happens if the variables are morphemes that cannot appear in isolation?

In the examples examined so far, variables are all free morphemes which can form sentences (*dün, Turgut, Pülod, juda yaxshi*) by themselves. However, if variables are bound morphemes such as affixes and suffixes, their occurrence which rule 1-1 predicts, would be unacceptable. For example, if, say, the prefix 'un-' is a variable, the occurrence of the morpheme in isolation yields an unacceptable 'sentence'/utterance: 'un-'. This is the problem attended to in this subsection where the problem is resolved by the application of rule 1-2.

First, observe the following constructed Turkish example:

- [24] a. *Ali-yle tanış-tı-n mı?*
 Ali-com meet⁵⁹-past-2sg Q
 'Did you meet Ali?'
- b. *tanış-ma-dı-m.*
 meet-neg-past-1sg
 'I didn't'

⁵⁹ Hereafter, the translation of the verb *tanış-* will be shown as 'meet' in glosses for simplicity's sake.

What is the parameter in this example? The identity of the possible performer of the action of meeting is not – its value is constant: [B]. The action which may have been performed is not – the value is [meeting]. The tense in which the action is performed is not a piece of new information either – the value is [past]. The value specified by the answer is, then, the negativeness in regard to whether the action specified in the question is performed.

What part in the answer, then, is the variable? It is not the verb stem *tanış-* which specifies the action being carried out, because the value it specifies, i.e. [meeting] is constant. It is not *-di* [past] either since the tense [past] also is a pre-specified value (see *-ti* [past] in (a)). The personal suffix *-m* [1sg] does not bear information yet to be specified either because the identity of the performer of meeting has already been specified in the question. We are left with only one candidate for the variable, namely the negative morpheme *-ma*.

Thus, whether the action is fulfilled, i.e. ‘action performed’, is the parameter, the value of which is specified as [negative] by the morpheme *-ma* [neg]. The only variable in (b) is, then, *-ma*. However, in (b), other morphemes, namely *tanış-*, *-di*, and *-m* are also present, despite their status as non-variables. This contradicts rule 1-1. Why are the non-variables present?

The answer is in the latter half of rule 1. According to rule 1-2, morphemes grammatically required to accompany variables have to appear. (See, however, the subsection ‘language specificity in the operation of rule 1-2’.) This explains why the non-variables *tanış-*, *-di*, and *-m* are present in (b). They are there not to specify values, but to accompany the sole variable in (b), namely *-ma* [neg] and secure the grammatical acceptability of (b) as an utterance in Turkish.

The assumption that the postulation of rule 1-1 and rule 1-2 entails is that, if B could specify the value as ‘negative’ without breaching the grammatical restrictions of Turkish, s/he would have. In fact, Turks often do breach grammatical rules in order to specify values only for parameters (and not to redundantly specify values for constants). We will look at such examples in the section ‘pre-grammatical sentences’.

Let us observe a possible process of operation of rule 1-1 and 1-2 below. Rule 1-1 calls for the occurrence of variables. Accordingly, at least *-ma* has to occur:

-ma

neg

However, in Turkish one cannot say *-ma* in isolation, even if the speaker wanted to. Indeed, the morpheme *-ma* is the only variable, and any addition of a morpheme would be an addition of a non-variable which redundantly specifies a pre-specified value. However, the grammar does not permit an utterance consisting only of *-ma* [neg]. Accordingly, though superfluous in terms of information or value-specification, the variable *-ma* needs to appear with some other morphemes to be grammatically acceptable. Above all, it needs a verb stem:

taniş-ma

meet-neg

This is grammatically correct, but means ‘do not meet’ in the imperative mood, which is not what the answerer means⁶⁰. It is grammatically correct as it stands in isolation, but (grammatically) deficient as a sentence which is intended to specify the value for the parameter as [negative]. In order to avoid misapprehension, it needs more non-variables. Accordingly, in the following, *-dı* [past] is suffixed to the above:

taniş-ma-dı

meet-neg-past

This too is grammatically correct. However, this utterance does not fit the truth condition, because it means ‘s/he did not meet’. This is, then, another sentence whose grammar is incorrect for a sentence used to specify the value as [negative]. Accordingly, another morpheme is required for the value-specification by *-ma* to be executed without the deviation of the utterance from the truth:

taniş-ma-dı-m

meet-neg-past-1sg

With the addition of *-m* [1sg], the grammatical correctness of the sentence is secured. The variable *-ma* needs no more non-variables accompanying it. Consequently, this sequence of non-variables and a variable constitutes the answer to (a).

⁶⁰ *Taniş-ma* may mean ‘becoming acquainted’ if the morpheme *-ma* is not a negative but a nominaliser. If it is a nominaliser, however, the location of the word accent shifts from *niş* to *ma*.

What the above analysis reveals is that only a fragment of (b) which rule 1-1 calls for, i.e. the [neg] morpheme, is important/indispensable in terms of information/value-specification. All the other morphemes are non-variables which redundantly specify values for constants. A point that is worth being pointed out in relation to this analysis is that the variable does not require the accompaniment of the subject pronoun *ben* ‘I’, whose occurrence would have yielded the following allosentence of (b):

- a. ‘Did you meet Ali?’
- b’. *ben tanış-ma-dı-m.*
I meet-neg-past-1sg
‘I didn’t’

This is because the performer of the action is a constant the value for which is already specified as [B] prior to the utterance of (b) and the variable *-ma* does not need *ben* to be grammatically acceptable. Note also that this analysis differs radically from Frank’s analysis reviewed before. In the paragraph cited from Franks (1995:288), he writes ‘[p]ro is just like an overt pronoun except that it lacks phonetic content’. Our analysis which do not assume anything where there is nothing contrasts sharply with this analysis.

Returning to the discussion of (b), according to rule 1-1, only *-ma* needs to appear for the value-specification of the parameter, but rule 1-2 requires the three non-variables *tanış-*, *-dı*, and *-m* to accompany the variable. This, then, allows the hypothesis that non-variables which need to be present in one language may not have to be present in another language with a different grammar. Taking (b) for example, while the [neg] morpheme requires the morphemes for [become acquainted], [past], and [1sg] in Turkish, the grammar of another language may require different accompaniments for the variable or perhaps none at all. To see whether this is the case, let us look at Japanese translation of the Turkish *tanışmadım* example:

- [25] a. *Ari ni at-ta?*
Ali with meet-past
‘Did you meet Ali?’
- b. *aw-ana-katta*
meet-neg-past
‘I didn’t’

The parameter and constants for this example are the same as the Turkish ones. Naturally, as rule 1-1 predicts, appearance of the variable *-ana* [neg] is obligatory in the (b) above as the variable *-ma* [neg] is in the Turkish example. However, the Turkish and Japanese examples differ in the number of non-variables that accompany the variables. In the Turkish (b), the variable *-ma* has three accompanying non-variables, whereas its Japanese counterpart has only two. The non-variable that is present in the Turkish example and absent in the Japanese example is the personal (agreement) suffix *-m*.

The answer (b) does not contain any personal suffixes, which Japanese does not have. Despite the lack of a personal suffix whose occurrence is grammatically required in the Turkish (b), the Japanese (b) is a perfectly natural translation of the Turkish (b). This is in accordance with our assumption that the personal suffix *-m* [1sg] in the Turkish (b) is present only to secure the grammatical correctness of (b).

This analysis of ours contrasts with the intuitively appealing analysis which considers that personal suffixes supply ‘the features of person and number of the referent of the Subject’ (Napoli 1993:85) and hence the subject can be ‘omitted’. In our analysis, personal suffix appear only to secure grammatical correctness – they may be capable of providing ‘the features of person and number of the referent of the Subject’, but at least in this example, the personal suffix does not appear for the sake of providing any information or specifying any value. (Perhaps this allows the assumption that personal suffixes may be of only secondary importance in information structure and usually do not participate in value-specification for parameters. That is, personal suffixes may be constantly non-variables. Whether personal suffixes are an indispensable part of information structure or not is central to the discussion in the next subsection.)

Thus, the occurrence of the personal suffix has proved to be not essential in answering (a), i.e. in specifying the value [neg] for the parameter ‘action performed’. On the other hand, the occurrence of the negative morpheme *-ana*, as a variable, is essential. What about, then, appearance of the morphemes other than the [neg] and [1sg] ones?

Comparing the Turkish and Japanese (b)s, we notice that morphemes specifying values for the constants ‘action’ and ‘tense’ as [meeting] and [past] (i.e. *tanış-*, *-di*, *aw-*, *-katta*), respectively, are present in both of the Turkish (b) and Japanese (b). If they are non-variables whose appearance is determined by language-specific rules, why do they

appear in both the Turkish and Japanese (b)s, as if they are variables whose appearance is called for by rule 1-1?

There are two possible reasons for this unexpected coincidence between Japanese and Turkish (b)s: 1) The rules that we set up are wrong; 2) Turkish and Japanese grammars happen to coincide in the rules concerning the sorts of morphemes by which a negative morpheme must be accompanied. We will claim in the next few paragraphs that the reason is the second. To validate our claim, we have to use an example from another language. Accordingly, we use the Bukharan Tajik translation of the above example:

- [26] a. *Ali kati šinos*⁶¹ *šud-i-mi?*
 Ali with acquaintance became-2sg-Q
 ‘Did you meet Ali?’
- b. *Na-šud-am.*
 neg-became-1sg
 ‘(I) didn’t.’

In this Bukharan Tajik example, the verbal compound *šinos šud-* ‘become acquainted’ does not have to appear in its entirety in (b). (Recall that Turkish *tanış-* is [become acquainted], though we gloss it as [meet] for simplicity’s sake.) *Šinos šudan* ‘to become acquainted’ is a verbal compound. Note that the answer could contain in it the first component of the verbal compound *šinos šud-*, namely *šinos*, which would yield *šinos našudam* ((b’) in the following), an allosentence of (b):

- b. *Na-šud-am.*
 neg-became-1sg
 ‘I didn’t.’
- b’. *Šinos na-šud-am.*
 acquaintance neg-became-1sg
 ‘(I) didn’t.’

The variable in (b) is *na-* [neg]. This morpheme, like Turkish *-ma* [neg] and Japanese *-ana* [neg], cannot appear in isolation and calls for accompanying non-variables *šud-* [became] as rule 1-2 requires:

⁶¹ A Bukharan informant says the pronunciation of this word has two variations in Bukharan Tajik; either *šinos* or *šunos*. This is an excerpt from an e-mail from the informant: *...yakta zanak kati šunos šudas* ‘(she) met a woman’.

na-šud

neg-became

This is grammatically correct as it stands but attracts a third person singular interpretation, which does not conform with the truth. Therefore rule 1-2 calls for the presence of the [1sg] morpheme, i.e. *-am*:

na-šud-am

neg-became-1sg

However, the variable does not call for the appearance of another non-variable, i.e. the first half of the compound *šinos*, as its accompaniment. This is to say that in the above (b), there is no [meet/become acquainted] morpheme that specifies the value for the ‘action’, which has already been value-specified as ‘meeting’. Thus, the Bukharan example in which no [meet] morpheme has to appear shows that the presence of Turkish *tanış-* in (24b) and Japanese *aw-* in (25b) is called for not by rule 1-1, but merely to secure the grammatical acceptability of the sentences.

Accordingly, we can now safely say that the validity of rule 1 is evident as far as the above examples are concerned. Significantly, the rule has accounted for the non-occurrence of subject pronouns without assuming any of the notions of ‘ellipsis’ or ‘inaudible elements’. This is particularly important because rule 1 manages to avoid the problematic argument that subject-verb agreement morphology somehow relates to the non-occurrence of subjects.

In the next subsection, we will examine cases where both the subject and agreement suffix must occur in a sentence. Such cases naturally pose another problem to the view which associates occurrence or non-occurrence of the subject with subject-verb agreement morphology. On the other hand, rule 1 proves to have explanatory power for such cases.

2.4.4.1.4. Multiple Variables

In the preceding subsection, we have claimed that subject pronouns do not have to occur in Turkish, Japanese, and Bukharan Tajik examples because rule 1 does not require them to occur. This analysis of subjectless sentences contrasts sharply with Gencan’s analysis in not associating the non-occurrence of subjects with agreement morphology. Our

analysis reveals that subject pronouns do not have to occur unless rule 1 calls for their occurrence, whether or not there is agreement morphology.

This analysis entails that, if rule 1 calls for their occurrence, they have to occur regardless of whether there are agreement suffixes. According to our analysis, for example, if the subject pronoun is a variable in a sentence, it has to occur even in the presence of an agreement suffix attached to the verb. This type of obligatory co-occurrence of the subject pronoun and agreement suffix, which cannot be accounted for by Gencan's analysis, is the topic of this subsection. Needless to say, providing a feasible account for the obligatory co-occurrence of the subject pronoun and agreement suffix will strongly reinforce the validity of our system of IS analysis.

Obligatory co-occurrence of a subject pronoun and its agreement suffix in Turkish has been analysed by Enç (1986:195) from a semantic-pragmatic point of view. She claims that 'the pronominal subject [as opposed to 'null subjects'] signals topic change'. An examination of her claim will serve as a good starting point to our discussion. Let us start with examining a Turkish example from Enç (1986:205, the glosses are slightly modified by me):

- [27] a. *Herkes Ali-'yle tanış-tı mı?*
everybody Ali-com meet-past q
'Did everybody meet Ali?'
- b. *Ben tanış-ma-dı-m.*
I meet-neg-past-1sg
'I didn't'

A peculiarity of the above example is the obligatory occurrence of the pronoun *ben* 'I' in (b). As Enç claims, the subject pronoun must be used in (b). If the pronoun is removed from (b), the sentence would be awkward as a response to (a):

- b'.# *Tanış-ma-dı-m.*
meet-neg-past-1sg
'I didn't'

Why does *ben* have to appear in (27b)? Clearly, the intuitively satisfying generalisation that 'subjects can "drop" when verbal inflection is rich enough to identify at least some features of the missing subject' (Franks 1995:288-9) does not work here as it did not in the preceding subsection, because in (b), the subject has to occur despite the occurrence

of the agreement suffix. Enç, instead of looking for a syntactic solution to this problem, turns to a semantico-pragmatic notion which she calls ‘(yes/no) contrast’.

As the reason for the obligatory appearance of the pronoun in (b), Enç points to ‘contrast’ which she claims (b) involves. Regarding the example, she writes (20a and 20b in the following comment signify (a) and (b), respectively, in the above example):

(20a) does not make an assertion, but implicates that the speaker was expecting everybody to have met Ali. (20b) [...] provides a counterexample

According to Enç, the response (b) involves provision of counterexample or ‘yes/no contrast’ which she claims trigger the use of the subject pronoun *ben* in the example. As an explanation of ‘yes/no contrast’ she cites a dialogue from Liberman and Sag (1974) and writes:

- (23) a. All presidents are immortal.
b. Well, Kennedy died.

(23b), uttered with slightly rising intonation at the end, does not contrast Kennedy with anybody else. It is used to provide a counterexample to (23a), [...] I will call this ‘yes/no contrast’.

She presents some other Turkish examples where she thinks ‘yes/no contrast’ is involved. For instance:

- a. *Ekmeğ-imiz yok.*
bread-1PL:POSS non-existent
‘We don’t have any bread.’
- b. *Ben gid-ip al-ır-ım.*
I go GER buy AOR 1SG
‘I’ll go and buy some.’

She notes that the subject pronoun must be used in the (b) in this example. Certainly, (b) would be awkward without the *bens*, and the awkwardness could be ascribed to B’s providing counterexamples as Enç claims to be the case. However, while Enç’s theory is adequate for explaining the occurrence of the first person singular pronoun *ben* in this example, this example seems to be a partial manifestation of a rule pertaining to the entirety of the IS of Turkish as well as other agglutinative languages, namely rule 1.

In the remaining part of this subsection, I attempt to locate Enç’s rule in a rule whose application is not limited to the analysis of the occurrence of *ben*, but open to

agglutinative languages in general (of which, of course, Turkish is one). Firstly, we shall analyse the ‘meeting Ali’ sentence, following which the other will be analysed. This separation of examples is not without a reason. Though they are both presented as different exemplifications of a single phenomenon in Enç’s article, they are very different from each other in terms of information structure. One of the (b)s consists exclusively of variables, whereas the other has two variables each of which co-occurs with non-variables. The ‘meeting Ali’ sentence, which is repeated here for convenience, is the latter.

Observe the following (the translation the verb *tanış-* in the gloss is changed from ‘become acquainted’ to ‘meet’ for simplicity’s sake):

- a. *Herkes Ali-’yle tanış-tı mı?*
 everybody Ali-com meet-past q
 ‘Did everybody meet Ali?’
- b. *Ben tanış-ma-dı-m.*
 I meet-neg-past-1sg
 ‘I didn’t’

Enç, in her analysis of the example, affirms that the questioner ‘was expecting everybody to have met Ali’. In fact, in spite of her affirmation, the expectation does not become evident from the form of (a), since even without that expectation, the form can be the same, i.e. *herkes Ali’yle tanıştı mı?* However, we simply assume the existence of A’s expectation which Enç claims to be there, instead of modifying her example sentence, since, after all, the assumption of the existence of such expectation with the utterance is highly feasible in, say, a classroom situation. (The same expectation will be assumed also for the example’s Japanese and Bukharan Tajik interpretations appearing later.)

This, then, makes the utterance (a) resemble a soliloquy rather than a question, because A does not presume there is a parameter for which a value has to be specified. In other words, in the existence of the assumption of A’s assumption that everyone met Ali, his uttering *herkes Ali’yle tanıştı mı?* in the above is an equivalent of saying the affirmative sentence *herkes Ali’yle tanıştı* ‘everybody met Ali’. Accordingly, A’s constants can be charted thus:

Action:	Constant	Meeting
Tense:	Constant	Past

Performer of the action:	Constant	Everybody
Action performed:	Constant	Positive
Object of the action:	Constant	Ali
etc.	Constants	

B, however, does not comply with A's constants. For B, there are two parameters the values of which are yet to be specified. A thinks 'everybody met Ali', whereas B thinks 'B did not meet Ali'. Accordingly, B's constants and parameters can be charted thus:

Specification		
Action:	Constant	Meeting
Tense:	Constant	Past
Performer of the action:	Parameter	B/Everybody/etc.> B
Action performed:	Parameter	Negative/Positive > Negative
Object of the action:	Constant	Ali
etc.	Constants	

Note that, unlike in the examples observed in preceding subsections where there was only one parameter 'action is performed', in this example, there are multiple parameters – In addition to the 'action is performed' parameter, there is also the 'performer of the action' parameter. B's intention is to specify the values for his parameters, because otherwise they would remain constant. A's constants and B's constants (and parameters) differ and a couple of A's pre-specified values need to be re-specified by B as follows:

	A's		B's	Value-specification:
Action:	Constant	✓	Constant	
Tense:	Constant	✓	Constant	
Action performed:	Constant	×	Parameter 1	Positive > Negative
Performer of the action:	Constant	×	Parameter 2	Everybody > B
Object of the action:	Constant	✓	Constant	
etc.	Constants	✓	Constants	

Having identified B's constants and parameters, we now must be able to tell which morphemes are the variables in (b). First, let us identify the variable which specifies the value of parameter 1 to [negative] (see above chart). The variable which specifies the value for parameter 1, then, must be the only [negative] morpheme in (b), namely *-ma* shown in bold letters below:

[28] *Ben tanış-ma-dı-m.*

I meet-neg-past-1sg

‘I didn’t’

What about, then, the variable for the other parameter, namely ‘performer of action’? The variable which specifies the value for parameter 2 has to be a morpheme that specifies the performer of meeting. There are two candidates for the variable for parameter 2 in (b). One is, of course, the first person singular pronoun, *ben* [I]. The other candidate is the first person singular suffix *-m* [1sg]. We are left with two morphemes with the potential of being variables. Both of *ben* [I] or *-m* [1sg] appear to be capable of specifying the identity of the performer of the action. Do *ben* and *-m* jointly specify the value, or is only one of them is value-specifier, the other one being a morpheme appearing merely to secure the grammatical acceptability of (b)?

The idea of them jointly specifying the value is unfeasible, because if they do, subjects and agreement suffixes must appear in pairs in all occasions. This is not the case since agreement suffixes can appear without subjects, as they do in the examples observed in the preceding subsections. Does this mean, then, that the agreement suffix specifies the value by itself? In other words, does *-m* [1sg] specify the value on its own, and hence is a variable?

Such value-specification by agreement suffixes seems to be found in some languages⁶². However, agreement suffixes in the Turkic languages do not appear to be capable of specifying values. If *-m* [1sg] was the variable for parameter 2, the sentence without *ben* [I], i.e. *tanışmadım*, has to be acceptable, because then all variables (the presence of which rule 1-1 calls for) are present and accompanied by the morphemes that secure the grammatical correctness of the sentence. However, this is not the case. As Enç notes, a (b) without *ben*, i.e. *tanışmadım*, is awkward as a response to (a), as shown below:

[29] a. *Herkes Ali-yle tanış-tı mı?*
everybody Ali-com meet-past q

⁶² For instance, judging from Bresnan and Mchombo’s (1987:745) comments, Chechewâ, a Bantu language, seems to allow pronouns and agreement affixes to jointly specify values when they are both present, but allows only agreement affixes to specify values on their own. They take the latter case to be the pronominal subject interpretation of agreement affixes which are called S[ubject] M[arker]s in their paper.

‘Did everybody meet Ali?’

b'.# *Tanış-ma-dı-m.*

meet-neg-past-1sg

‘(I) didn’t’

Thus, the presence of *ben* is required, which in turn suggests that *ben* is a variable – if *-m* [1sg] was a variable which specified the value for the parameter ‘performer of action’, rule 1 would not call for the appearance of *ben*. This is to say that the value-specification of the parameter ‘the identity of the performer’ is executed not with *-m* but with *ben*.

The occurrence of the morpheme *-m* is called for not by the information structure of the utterance, but by the occurrence of the morpheme *-ma*. In other words, *-m* occurs only to accompany the variable *-ma* [neg], so that the utterance’s conformity with the truth as well as its grammatical correctness are secured. Judging from the above discussion, agreement suffixes are, despite their often very obvious pronominal etymology in the Turkic languages⁶³, not value-specifiers.

The other piece of evidence which confirms that agreement suffixes are not variables in Turkish is obtained by modifying the above to construct an example where the only parameter in (b) is the identity of the performer of the action of meeting, thus:

[30] a. *Kim Ali-’yle tanış-ma-dı?*
Who Ali-com meet-neg-past
‘Who did not meet Ali?’

b. ***Ben.***
I
‘I (didn’t)’

If the agreement suffix *-m* was the variable and the pronoun *ben* was not, the answer in the immediately above example would have consisted of *-m* with morphemes it requires for grammatical correctness, making the presence of *ben* unnecessary:

[31] a. *Kim Ali ’yle tanışmadı?*
‘Who did not meet Ali?’

b'.## *Tanış-ma-dı-m.*
meet-neg-past-1sg

⁶³ E.g. Uyghur *män barimän* ‘I go’ and Uzbek *men yozaman* ‘I write’ (Kononov 1960:173) etc. A simple list of the agreement suffixes in other Turkic languages is found in Zeynalov (1993:224).

‘(I) didn’t’

However, (b’) is clearly awkward as an answer to (a). This supports our analysis that variables are personal pronouns, not agreement suffixes.

The conclusion we can draw from the above discussion is that the morpheme *-m* is called for by rule 1-2 while the occurrence of *ben* is required by rule 1-1. Accordingly, the information structure of (b) can be schematised thus:

Values specified by (b):	[I]		[neg]	
Variables (rule 1-1):	<i>ben</i>		<i>-ma</i>	
Non-variables (rule 1-2):		<i>tanış-</i>		<i>-dı -m</i>
Utterance:	<i>ben</i>	<i>tanış-</i>	<i>-ma</i>	<i>-dı -m</i>

Accordingly, the information structure of (b) can be represented as ***ben tanışmadım***, but not *ben tanışmadım*. Compare this the IS of the allosentence in the example previously analysed, one of which is structurally identical with the this sentence:

- [32] a. *Ali’yle tanıştın mı?*
‘Did you meet Ali?’
- b. *Tanış-ma-dı-m.*
meet-neg-past-1sg
‘(I) didn’t’
- b’. *Ben tanış-ma-dı-m.*
I meet-neg-past-1sg
‘I didn’t’

The latter of these allosentences is structurally identical with the response to *herkes Ali’yle tanıştı mı?* However, unlike that sentence where *ben* has to appear, *ben* in the (b’) above does not have to appear. This difference can be explained by the difference between the information structures of the two sentences: the former is ***ben tanışmadım*** which contrasts with the latter: *ben tanışmadım*.

Values specified by (b):			[neg]	
Variables (rule 1-1):			<i>-ma</i>	
Non-variables (rule 1-2):		<i>tanış-</i>		<i>-dı -m</i>
Non-variable (optional):	<i>ben</i>			
Utterance:	<i>ben</i>	<i>tanış-</i>	<i>-ma</i>	<i>-dı -m</i>

Thus, rule 1 neatly explains obligatory and non-obligatory occurrences of morphemes in these Turkish examples.

Now, let us see whether rule 1 is identifiable in another of the languages discussed in this thesis, Bukharan Tajik. Observe the following Bukharan Tajik counterpart of the Turkish ‘everyone met Ali’ example:

- [33] a. *Hamma*⁶⁴ *Ali kati šinos šud(-mi?)*
 ‘Everybody met Ali(?)’
- b. *Man šinos na-šud-am.*
 I acquaintance neg-became-1sg
 ‘I did not.’

The constants and parameters for this example is identical with those for its Turkish counterpart, that is, the parameters are ‘action is performed’ and ‘performer of action’, the values for which are [neg] and [I], respectively. Here again, despite the occurrence of the agreement suffix *-am* [1sg], the pronoun *man* [I] has to be present, as the awkwardness of the following sentences suggests:

- b’ .## *Šinos na-šud-am.*
 acquaintance neg-became-1sg
 ‘(I) did not.’
- b’’ .## *Na-šud-am.*
 neg-became-1sg
 ‘(I) did not.’

All five native Bukharan and one Samarkandi Tajik speakers rejected (b’) for its being strange or even utterly wrong and unanimously insisted on the obligatory appearance of *man* [I]. This is of course a judgment which would not take place if the agreement suffix *-am* [1sg] was a specifier of the value [I] for the parameter ‘performer of action’. This suggests that the pronoun *man* is a variable. Accordingly, the IS of (b) can be tabulated thus:

Values specified by (b):	[I]	[neg]
Variables (rule 1-1):	<i>man</i>	<i>na-</i>
Non-variables (rule 1-2):		<i>šud- -am</i>

⁶⁴ An informant provided not *hamma* ‘all’ (*hama* in literary Tajik and *hamma* in Uzbek) but *har kas* ‘everybody’.

Non-variables (optional): *šinos*
 Utterance: *man šinos na- šud- -am*

Thus, operation of rule 1 is clearly identifiable in this Bukharan Tajik example as it is in its Turkish example. Lastly, another piece of evidence verifying that agreement suffixes are not variables comes from a comparison of a Japanese translation of the Turkish example.

Japanese is of great usefulness in examining whether the presence of the agreement suffix in the Turkish example is required by rule 1-1 or only to satisfy the requirement set by rule 1-2. If the former proves to be the case, the Japanese translation of Turkish (b) which lacks personal suffixes could not be acceptable whatsoever. This is not the case, as the following Japanese example shows:

- [34] a. ‘Everybody met Ali’
 b. ***Boku*** (*wa*) *aw-ana-katta*
 I (top) meet-neg-past
 ‘I didn’t’

This example shows that the absence of the agreement suffix causes no deficiency in communication. On the other hand, the variable *boku* ‘I’, whose appearance is predicted by rule 1-1, must occur:

b’ ## *Aw-ana-katta*
 meet-neg-past
 ‘(I) didn’t’

Accordingly, the IS of Japanese (b) can be shown as a chart, thus:

Values specified by (b):	[I]	[neg]
Variables (rule 1-1):	<i>boku</i>	<i>ana-</i>
Non-variables (rule 1-2):	<i>aw-</i>	<i>-katta</i>
Non-variable (Optional):	(<i>wa</i>)	
Utterance:	<i>boku</i> (<i>wa</i>) <i>aw-</i>	<i>ana-</i> <i>-katta</i>

This chart, as well as those presented before suggest that agreement suffixes are of secondary importance in information structure at least in these languages. The awkwardness of (b)s without pronouns such as *ben* (Turkish) ‘I’ and *boku* (Japanese) ‘I’ also supports the view that the morpheme involved in specifying the value for the parameter ‘performer of action’ is the first person singular pronoun in all of these

languages. Taking into consideration this supporting evidence from Japanese and Bukharan Tajik, we now can safely acknowledge the analysis of the information structure of (24b) and (25b) presented at the end of the previous section, 2.4.4.1.3.

Our analysis is radically different from syntactic analyses, most of which seem to take for granted the existence of inaudible or dropped elements. It is also more useful than the analyses that assume that ‘subjects can “drop” when verbal inflection is rich enough to identify at least some features of the missing subject’, as far as the examples reviewed above are concerned. Such analyses deduce the existence of the inaudible/dropped subject from the audible/present agreement suffix. However, in the present analysis, in the above example for instance, there is no *ben* [I] dropped off from (b), while many analyses assume the existence of an inaudible *ben* [I].

The present analysis diverges from most other studies also in ascribing the presence of the whole succession of morphemes of *tanışmadım* to the single morpheme *-ma* [neg]. This refutes the information-structural necessity of the appearance of *-m* [1sg] on the basis of which the existence of the ‘inaudible’ *ben* [I] is hypothesised in many other works. The present information-structural analysis ascribes the appearances of *ben* and *-m* to two different reasons. *Ben* is there because it specifies the value for the parameter, i.e. because it is a variable. On the other hand, *-m* is there only to accompany the other variable *-ma* so that the grammatical correctness of the utterance is secured.

According to this analysis, agreement suffixes are of secondary importance in information structure of the languages discussed in this thesis. This lack of information-structural importance of agreement morphology may be the reason why it often attracts such comments as ‘it is strange that inflectional morphology is so widespread among natural languages⁶⁵, given that it benefits nobody, and makes a language more difficult to learn’ (Hudson 1980:63 cited in Romaine 1988:25).

This analysis brings about another interesting consequence: the appearance of the personal ending is not called for by ‘inaudible’ *ben* but by the appearance of *-ma*. It takes the form of *-m* [1sg] (not, say, *-k* [1pl] or *-nız* [2pl]) agreeing with the subject [I]. That is, ‘subject-verb agreement’ does operate in the process of the determination of the form in which the personal ending appears. However, from our information-structural point of view, the appearance of the personal ending is not necessitated by the subject, but by the

appearance of *-ma*, which needs to be accompanied by a personal ending (which is in this case *-m*).

This means that, in our analysis, the existence of the subject, whether ‘audible’ or not, is not deducible from the appearance of a personal ending, because the appearance of the subject and that of the personal ending are necessitated independently from each other. Taking (33b) for instance, the appearance of the former is indispensable for its status as a variable, while the latter is demanded by that of another variable.

Before we proceed to the next section, let us briefly observe the other of Enç’s examples cited before:

- a. *Ekmeğ-imiz* *yok.*
bread-1PL:POSS non-existent
‘We don’t have any bread.’
- b. *Ben gid-ip* *al-ır-im.*
I go GER buy AOR 1SG
‘I’ll go and buy some.’

In this example, there are no pre-specified values among the values which morphemes in (b) specify. This lack of redundant specification makes all the morphemes in (b) variables. Accordingly, the production of (b) according to our analysis can be schematised as follows:

Values specified by (b):	[I]	[go]	[ger]	[buy]	[aor]	[1sg] ⁶⁶
Variables (rule 1-1):	<i>ben</i>	<i>gid</i>	<i>-ip</i>	<i>al</i>	<i>-ır</i>	<i>-im</i>
Utterance:	<i>ben</i>	<i>gid</i>	<i>-ip</i>	<i>al</i>	<i>-ır</i>	<i>-im</i>

Unlike in some of the examples previously examined, (b) does not specify any value which has already been specified by (a). The sentence (b), then, consists only of variables as most of the examples cited in the subsection ‘variable = free morpheme’ do. Our analysis suggests that the pronoun *ben*’s occurrence in Enç’s examples has nothing to do with semantic contrast but is a result of the information structural requirement which is summarised in rule 1-1.

Thus, rule 1 has not only given an account to the obligatory occurrence of the first person singular pronoun *ben* in the first of Enç’s examples (27b), but also provided an account

⁶⁵ As opposed to pidgins.

⁶⁶ The verb *al-* means ‘take’ or ‘get’. We adapt Enç’s translation here.

of why so-called ‘pro-drop’ occurs freely in a language with no verb agreement morphology.

2.4.4.1.5. Notes on “Variable = Phrase”

In this brief subsection, we look at an example which may appear to some as an instance where not morphemes but phrases are equated with variables. The “variable = phrase” view is of course incompatible with the definition of variables as ‘morphemes which specify values for parameters’ and we will show below that such an equation is explanatorily superfluous in our system.

Observe the following constructed Japanese answer to the question ‘who came?’:

<i>ōki-i</i>	<i>otoko</i>
big-nonpast	man
‘(a) big man’	

This answer also has a lot of possible alternatives such as *Ari* ‘Ali’, *otoko* ‘(a) man’, *onaji otoko* ‘(the) same man’, and *kinpatsu no onna no hito* ‘(a) blonde woman’, all of which are noun phrases. Does this mean, then, that the NP (i.e. the answer), which consists of multiple morphemes, is to be analyzed as a single variable? If we assume that the answer specifies the value of the single parameter of ‘performer of action’ as the chart below shows, it needs to be counted as a variable:

Performer of the action:	big man
--------------------------	---------

However, as we will see below, this assumption has a problem: there is no reason to assume ‘big man’ to be a single value. Note that the answer could be, say, *otoko* ‘man’, in which case the value specified for the parameter may be represented thus:

Performer of the action:	man
--------------------------	-----

Here, the morpheme *otoko* is a variable and specifies the value for the parameter by itself. If ‘man’ alone constitutes a value, we should expect ‘big man’ to consist of multiple values. This naturally implies the possibility of decomposition of the composite of values: ‘big man’. The value-specification that the answer executes, then, may be represented in the following way:

Values	Value-specifiers
--------	------------------

Performer of action:	man	<i>otoko</i>
Characteristic of performer:	big	<i>ōki-</i>
Tense of characteristic:	non-past	<i>-i</i>

Which ones of these are parameters, or whether any of these are constants for the questioner is unknown. For example, the questioner may believe that the comer is(/was) big, i.e. have ‘characteristic of performer’ as a constant, in which case *ōki-* ‘big’ redundantly specifies the value for the constant and is a non-variable. Similarly, if the questioner believes the comer to be someone who was big in the past (but does not know whether s/he is big in the present), the value of the second of the three parameters/constant above is prespecified at the time of B’s answering, which makes *otoko* ‘man’ and *-i* ‘non-past’ the only variables in the answer (*ōki-i otoko*). On the other hand, if the questioner has the last two of the parameter/constants listed above as constants, only the value of the first needs to be specified by the answer. In such case, of course, the utterance of only *otoko* ‘man’ would have been sufficient as an answer for the question (and *ōkii otoko* would have been its allosentence).

Of course, all of the ‘performer’, ‘characteristic’, ‘tense’ listed above can be parameters. If they are, the answer *ōkii otoko* is not a single variable *ōkiiotoko* ‘big+nonpast+man’, but a combination of three variables, *otoko* ‘man’, *ōki-* ‘big’, and *-i* ‘nonpast’.

Whatever the information structure of the example above is, what is apparent from the analysis above is that, in our system, there is no need to set up phrases as variables.

2.4.4.1.6. Notes on Variable = Absence

Before we proceed to the next subsection, some notes on an issue which we will not discuss in this thesis are in order. As the observant reader may have noticed, in the preceding subsections, we have not mentioned the possibility that absence specifies values for parameters. However, it is not difficult to find examples where absence is most likely the specifier of a value. An example from Uzbek is shown below:

- [35] a. *to ‘p o ‘yna-y-sen-mi?*
ball play-impf-2sg-Q
‘Will you play ball?’
- b. *o ‘yna-ma-y-man.*

play-neg-impf-1sg

‘I won’t play.’

b’. *o’yna-y-man.*

play-impf-1sg

‘I’ll play.’

Note that the negative morpheme *-ma* specifies the value of the parameter ‘action performed’ as [negative] in (b), while in (b’) there seems to be no particular morpheme that specifies the value of the parameter as [positive]. This, then, constitutes a logical basis for asserting the value-specification by absence, because what specifies the value of the parameter as [positive] in (b’) can be claimed to be the absence of the [neg] morpheme *-ma*. Moreover, the apparent value-specification by absence is not unique to Uzbek, as can be seen in the following Bukharan Tajik counterpart of the Uzbek dialogue above:

[36] a. *tūppi bozi⁶⁷me-kun-i-mi?*

ball play impf-do-2sg-q

‘Will you play ball?’

b. *bozi na-me-kun-am.*

play neg-impf-do-1sg

‘I won’t.’

b’. *bozi me-kun-am.*

play impf-do-1sg

‘I will.’

A list of the constants and parameter is provided below:

	Value-specification
Action:	Constant [Playing ball]
Tense/Aspect:	Constant [Present imperfect]
Action performed:	Parameter > [Negative] in (b) / [Positive] in (b’)
etc.	Constants

What are the value-specifiers for the parameter ‘action performed’ in (b) and (b’)? In (b), it is obviously the [neg] morpheme *na-*. One might, then, expect there to be a [positive] morpheme as opposed to the [neg] morpheme *na-* in (b’), but there is not one. Thus, in

⁶⁷ *Bozī* in literary Tajik.

both of the Uzbek and Bukharan Tajik examples, what specifies the value as [positive] appears to be the absence of the [neg] morpheme.

Note that the pair of [neg] and [positive] is one of such pairs that ‘exhaustively divide some conceptual domain into two mutually exclusively compartments, so that what does not fall into one of the compartments must necessarily fall into the other’ (Cruse 1986:198-9, see also Zimmer 1964:22). Taking the above examples into consideration, we may assume the following: Value-specification by absence takes place in situations where the specified value has at least one possible alternative which is mutually exclusive with the specified value and has morphological representation.

One of the most frequently encountered examples of these kinds of situations, then, is arguably the one with the alternatives/values [positive] and [neg]. This is the case with the Uzbek and Bukharan Tajik examples above. Since [positive] and [neg] are dichotomously excluded from and incompatible with each other, specification of a value as either of them inevitably nullifies the other, which incompatibility is the prerequisite for absence to be the value-specifier for one of the values.

Thus, absence is a strong candidate for being a value-specifier. In addition, interestingly, absence of a morpheme appears to require the appearance of the same non-variables which the morpheme requires as its accompaniments. For example, the variable *-ma* in the Uzbek (b) requires the accompanying non-variables *o’yna-*, *-y*, and *-man* and so does the absence of *-ma* in (b’). This may imply that rule 1-2 is applicable regardless of whether a value-specifier is a morpheme or absence of a morpheme.

However, although absence as a value-specifier is an interesting subject to pursue, we confine our discussion to ‘present’ value-specifiers, i.e. variables, in the present thesis. This is because a discussion of absence as a value-specifier necessarily involves discussion of absence as an encoder of information, which is too big a topic to discuss in a section or even a chapter.

2.4.4.1.7. Summary for the ‘Ellipsis’ Subsection

In this subsection, we have analysed sentences which are considered to be ‘elliptic’ by many grammarians and linguists. Our analysis, for which we used the notions of parameters, constants, and variables, revealed that there is no need to assume ‘inaudible’

or ‘elided’ constituents to account for such sentences – the sentences consist of variables and non-variables which accompany them, the occurrence of which is governed by two simple rules: rule 1-1 and 1-2.

2.4.4.2. Potential (Partial) Applicability of Rule 1 in Fusional Morphology

In the preceding subsections, we have discussed the application/operation of rule 1 in agglutinative languages. However, as was explained in the introductory chapter, agglutination and fusion are two hypothetical ends of the axis of ‘segmentability of words into morphemes’, onto which languages are mapped. Consequently, even a highly agglutinative language like Turkish utilises fusional morphology. This relative nature of agglutination in languages necessitates us to take into account fusional morphology to a certain extent in setting up principles in our system of IS analysis although it is proposed exclusively for agglutinative languages. In this subsection, we examine the possible applicability of rule 1 in analysing examples with fusional morphology. However, before we discuss this issue, let us briefly revisit the English version of the sentences discussed before and review it in the light of our findings in this section:

- [37] a. Did you meet Ali?
b. I **didn’t**.

Having introduced and explained the terminology, we can now call the morpheme ‘n’t’ a variable. Its occurrence in (b) is obligatory, unlike some of the morphemes in the following allosentence of (b):

- b’. I **didn’t** meet him.

The obligatory appearance of the variable ‘n’t’, which coincides with the obligatory appearance of the [neg] morphemes in Turkish, Japanese, and Bukharan Tajik, supports the validity of rule 1-1.

On the other hand, ‘I’ and ‘did’ are both non-variables which are present to secure the grammatical correctness of the sentence. ‘N’t’ certainly cannot appear in isolation, which is the reason ‘did’ accompanies it. ‘I’ is there too because English, unlike Turkish, Uzbek, Japanese, or Bukharan Tajik, requires a subject in this sentence. This occurrence of ‘I’ in the answer establishes a contrast with the examples we observed in the last subsection in which subject pronouns do not have to appear. (However, the use of

sentences with no ‘I’ in some registers, e.g. informal messages/speech, in English might be used as a piece of evidence supporting the validity of rule 1-1.) This contrast, though in a remote way, supports the validity of rule 1-2, because the contrast indicates language-specificity in the occurrence of non-variables.

This analysis of the English example, however, does not involve fusional morphology. We would like to have a brief look at the possibility of using rule 1 for IS analysis in cases where fusional morphology co-occurs inseparably with IS analysis in the following paragraphs. First, observe the following English verb and its form in the past tense:

run ran

In a situation where the parameter is ‘tense’, for which the value [past] is to be specified, what in ‘ran’ should be identified as the variable? Obviously, /æ/ in ‘ran’ as opposed to /ʌ/ in ‘run’ cannot be isolated from the morpheme, though the change in the vowel is the most likely candidate for the specifier of the value [past].

Harris (1991:176-177) finds a solution to this problem in accepting the phonological change as an ‘alternant member’ of the past tense morpheme *-ed*. While this solution is probably feasible in its own right, we do not adopt it, because perhaps against what is expected, our system has little difficulty in identifying the variable. It follows from our definition of variables that they have to be morphemes (and not phonemes). Accordingly, the variable that specifies the value of [past] has to be not a part of or a phoneme in ‘ran’, but its entirety, i.e. ‘ran’. (Of course, we cannot identify the past ‘hit’ as a variable when the value to be specified is [past]. However, this is because ‘hit’, unlike ‘ran’, is not capable of specifying the value for the parameter ‘tense’ to [past].) If we did not define variables as they are, we might have had to identify, say, a distinctive feature of the vowel [æ], e.g. [+ front], as the variable.

The same principle of identifying exclusively morphemes as variables saves us the trouble of extracting value-specifiers from examples with a highly fusional character. For example, if the value which Arabic *kutub* ‘books’ specifies is [plural], any attempts to isolate what is responsible for value-specification from *kutub*, which is the broken plural of *kitāb* ‘book’, would face much difficulty. Our definition of variables allows us to simply call the entirety of *kutub* as the variable.

Thus, although our system of IS analysis is tailored for agglutinative languages, it has at least a potential to be useful in analysing IS in fusional languages. This means that our system can handle not only agglutinative morphology but also fusional morphology in the agglutinative languages which are the subjects of the present thesis.

2.4.4.3. Language Specificity in the Operation of Rule 1-2

Before we finish our discussion on ‘ellipsis’ and ‘null-subjects’, let us have a brief look at language specificity in grammatical requirements in the operation of rule 1-2.

Compare the sentences in the following chart where answers to ‘did you meet Ali?’ in the languages which we have looked at so far are listed:

<i>tanışmadım</i> (Turkish)	[meet-neg-past-1sg]
<i>awanakatta</i> (Japanese)	[meet-neg-past]
<i>naşudam</i> ((Bukharan) Tajik)	[neg-become.past-1sg]

The variables, i.e. [neg] morphemes, are present in all of these sentences as rule 1-1 predicts. However, non-variables which accompany them are different from one language to another. For example, the [neg] morpheme in the Turkish sentence, according to rule 1-2, requires the accompaniment of the [1sg] morpheme as does its (Bukharan) Tajik counterpart. On the other hand, in Japanese, which has no subject-verb agreement suffix, [1sg] does not appear.

Thus, non-variables whose occurrence is called for by rule 1-2 are language-specific. That is, a non-variable whose occurrence is required in one language may not have to be present in another, because the grammar of a language may call for morphemes which the grammar of another language does not. This is the case even among genetically closely related languages such as Turkish and Uzbek. Observe the following Turkish and Uzbek counterparts of the English ‘this is yours!’⁶⁸:

[38]	Turkish	<i>Bu senin!</i>	(<	<i>sen-in</i>)
		this your		you-gen
	Uzbek	<i>Bu seniki!</i>	(<	<i>sen-ning-ki</i>)
		this yours		you-gen-nmlz

⁶⁸ Not ‘yours is this (one)!’ or ‘this is yours!’ as an answer to ‘which one is mine?’, in which case the Turkish equivalent can be *bu seninki!* or *seninki bu!*

*Bu *sening!* (< *sen-ning*)
 this your you-gen

The Turkish *senin*, etymologically speaking, consists of *sen* ‘you’ and the genitive suffix *-nin*. On the other hand, the Uzbek *seniki* consists of *sen* ‘you’, the genitive suffix *-ning*, and a nominalising suffix *-ki*. The Turkish and Uzbek equivalents of ‘this is mine!’ might be translated morpheme by morpheme, though strange as an English utterance, roughly as ‘This is your!’ and ‘This is yours!’, respectively. In this Uzbek sentence, the suffix *-ki*, which does not have to appear in the Turkish counterpart, is indispensable. This is to say that, if we assume a case where the [you] morpheme is a variable in these sentences, rule 1-2 requires different non-variables to accompany the variable in Turkish and Uzbek.

2.4.4.3.1. Orthographic Words

In most of the Turkish, Uzbek, and Tajik examples we have seen so far, every string of a variable and its accompanying non-variables is represented as an orthographic word. That is, a variable and its accompaniments always form a chunk immediately before and after which orthographic breaks are inserted. For example, the string of the variable *-ma* and its accompaniments *tanış-*, *-di*, and *-m* is written as one orthographical word in (27b):

ben (orthographic break) *tanışmadım* (orthographic break)
 I meet-neg-past-1sg

and is not divided by (an) orthographic break(s), as indicated by the unacceptability of:

**ben* (orthographic break) *tanış* (orthographic break) *ma* (orthographic break) *dım*
 (orthographic break)

However, do variables and their accompaniments always equal orthographic words? There is a tendency in the orthographies of Turkish and Uzbek for strings of variables and their accompaniments to coincide with units divided by orthographic spaces. For example, a negative answer to the question ‘did you meet Ali?’ is, in Uzbek, the following, which is almost identical to its Turkish counterpart:

(orthographic break) *tanış-ma-di-m* (orthographic break)
 meet-neg-past-1sg

‘I didn’t’

The above sequence of morphemes, like the Turkish *tanişmadım*, consists of a variable and its accompaniments and is also immediately preceded and followed by orthographic breaks. We might, then, be tempted to think that there is a correlation between orthographies and the units that the operation of rule 1 yields. However, while it is probably right to say that morpho-informational chunks are reflected in the orthography to some extent, the reflection is not thoroughgoing. In some orthographies, the units may be bounded by orthographic breaks, but this is not necessarily the case; orthography has essentially little to do with the units formed by the operation of rule 1.

The lack of constant correspondence between orthographic ‘words’ and the units formed by the operation of rule 1 is most aptly exemplified by the Tajik counterpart of the above examples: *našudam*. *Našudam* is a unit formed by the operation of rule 1 and is written as a single orthographic word. However, *našudam* has not always been written as a single orthographic word.

According to an outline of the orthographic rules for Tajik written in Latin alphabet, ‘the particle *na* in verbs is written separately’⁶⁹, although there was disagreement among the members of the orthography committee about whether *na* should be orthographically prefixed (see Komissiyai imloi alifboi lotinii navi tojikī 1929:2). The rule was seldom observed in the journal in which the outline was published. (Indeed, the rule was neglected even in the body text of the outline itself.) It however was observed in a number of publications thereafter for several years⁷⁰, until another change was made to the orthography and the particle started to be prefixed to verbs⁷¹. For example, the unit *našudam* has been written either as one orthographic word *našudam* or as two words *na* (break) *šudam*, as a result of the unstable orthographic rules. This fact demonstrates the non-essential nature of the correspondence between orthography and the strings of morphemes that rule 1 yields.

⁶⁹ Like ‘not’ in English.

⁷⁰ See, e.g., Tospulotuf et al. (1932). However, judging from Asimova’s (1982:65) illustration of the application of the orthography in practice, the confusion regarding the orthography of *na-* persisted at least until 1931.

⁷¹ See, for example, Ajnī (1939a, 1939b).

This nature is even more conspicuous in Japanese orthography which does not employ writing with orthographic breaks (*wakachigaki*). The current orthography does employ (*tō*)*ten* or comma, but its use is not fully standardised despite its long history.⁷²

Moreover, it is not necessary to use (*tō*)*ten* in current Japanese orthography (see Sanseidō Henshūjo 1991:204)⁷³. This, then, is an example where orthography lacks direct correlation with rule 1.

Thus, orthography is a set of writing conventions that may or may not be linguistically motivated. There are few linguistically sound reasons why, say, the Turkish interrogative morpheme *mI* [Q], whose phonological behaviour is not different from that of many (other) suffixes in the language (and whose cognate in Uzbek *-mi* is a suffix), is written separately from the morpheme that precedes it.⁷⁴ There are also a number of morphemes that can be written either as an orthographic ‘word’ or as a suffix in Turkish, e.g. *ile/-(y)lE*⁷⁵, *ise/-(y)sE*, etc.⁷⁶

In sum, orthographic words may coincide with the morpheme strings that rule 1 yields, but this is not necessarily so. The fact that orthographic words do often coincide with the morpheme strings that rule 1 yields, either in principle or in practice, is interesting, but we will not discuss this issue further in this thesis.⁷⁷

⁷² The oldest form of (*tō*)*ten* is found in a *kanbun-kundoku* document hand-copied before 745 (Kaneko 1989:60). (See Lange (1970:142) for an explanation of *kanbun-kundoku*).

⁷³ See also Sproat (1992:12-13) for a discussion of the ambiguity that the lack of orthographic breaks brings about in Japanese.

⁷⁴ The Uzbek cognate of the morpheme *-mi* [Q] is written as a suffix in Uzbek orthography. So is the conjunction *-ki* [that], which is written as an independent word, *ki* [that], in Turkish orthography.

⁷⁵ See Bassarak (1988).

⁷⁶ See also Ülkütaşır (1973:122-3 cited in Lewis 1999:35) for the short-lived orthographic rule about suffixion of *için* ‘for’.

⁷⁷ Another issue of interest, but one that is also beyond the scope of this thesis, is that of whether or not morphemes should, in fact, be regarded as the smallest meaningful units in the mental lexicon of a native speaker. We are not entirely sure whether, say, *emir-ler* [order-pl] is always (i.e. in *both* sentence production and recognition) a sequence of two morphemes for Turkish speakers. There is evidence that suggests *emirler* ‘orders’ may be a single unit in native Turkish speakers’ visual recognition (See Gürel 1999). See also Taft and Forster (1975), Caramazza et al (1988), Emmorey (1988), Frauenfelder and Schreuder (1992), Colé et al (1997), Waksler (1999), and Segalowitz and Lane (2000) for discussions on lexical access and morphological processing.

2.4.4.3.2. Summary

We have posited rule 1-2 which complements rule 1-1 and mentioned the involvement of grammatical acceptability in rule 1. What we did not mention, however, is the nature of what we refer to with the expression ‘grammatically required’. For example, we have said before that ‘English grammar requires a subject’ in ‘I didn’t meet Ali’. Indeed, CS grammar requires ‘I’ to appear in ‘I didn’t meet Ali’, but in, say, informal writing, we find many acceptable and yet grammatically incorrect sentences in the sense that they do not comply with CS grammar, e.g.: ‘Seems okay to me’, ‘Home to dinner soon, then more work...’, ‘Have been busy with correcting and doing tax form...’, (all of which have been taken from e-mail messages) etc. Is the occurrence of subjects in these examples, then, ‘grammatically required’?

In this thesis, ‘grammatically required’ means ‘required to comply with CS grammar’. Accordingly, the English examples shown above are considered to be grammatically incorrect, despite their acceptability. Such sentences will be called pre-grammatical sentences and will be discussed at length in the subsection ‘pre-grammatical sentences’.

2.4.4.4. Echoing and Unattended Parameters

2.4.4.4.1. Echoing

In this subsection we are concerned with echoing. Here, echoing refers to the repetition of the utterance or part of the utterance of an interlocutor A by another interlocutor B. A good example of this is the notification of telephone numbers on the phone. It is customary for people to repeat the phone number s/he has just been told in confirmation that s/he has heard it correctly (see, e.g., Sakuma 1997:205). A Japanese dialogue example which may follow B’s utterance ‘(what is) the phone number?’ is shown below:

- [39] a. 284 80 44
284 80 44
‘(it’s) 284 80 44’
- b. 284 80 44 *ne*
284 80 44 *sfp*
‘284 80 44’

This may appear as an unnecessary repetition of non-variables. However, the phone number in (b) can be considered a variable, because, if B is absolutely sure about the correctness of the number he has memorised, there is no need for the confirmation of the number. In other words, if B accepts the number as being a constant without reservation (i.e. without confirmation), he would probably not echo the number. B's concern that he might have heard the number incorrectly makes the number a parameter, the value of which may be 284 80 44, 284 82 44, etc. Hence, the echoing in the above can be analysed as a case where the value of B's parameter happens to coincide with that of A's constant, thus:

	A's	×	B's	Value-specification:
Phone Number:	Constant		Parameter	284 80 44 > 284 80 44

The occurrence of echoing is not confined to the confirmation of phone numbers. This kind of echoing for confirmation of memory is common in taking orders at many restaurants. (It is even obligatory in family restaurant chains in Japan where waitresses/waiters are trained to repeat orders they have taken before they leave tables, reading aloud the orders they have entered into their handheld computers.) One echoes frequently in dialogues in situations where interlocutors cannot hear each other's utterances clearly. For example, the following may take place in a *taiyaki*⁷⁸ shop after A addresses the seller B over the busy counter:

- [40] a. *koshian muttsu!*
 strained bean jamsix
 'six (*taiyaki* with) strained bean jam'
- b. *koshian muttsu.*
 strained bean jamsix
 'six (*taiyaki* with) strained bean jam.'

The motive for this echoing is the confirmation of A's order, but the confirmation itself is not motivated by the possible flaw in memory, but distortion of sound around the busy counter. (See also Kumagai 1997:41)

Thus, although an echoed utterance may, on the face of it, appear to consist of non-variables, it in fact specifies values for parameters and hence consists of variables.

⁷⁸ A fish-shaped pancake stuffed with bean jam (*Shōgakukan Eiwa Jiten*).

2.4.4.4.2. Unattended Parameters

Aside from echoing, there is another complexity to which we did not attend in the simple analysis in the ‘ellipsis’ section, namely ‘unattended parameters’. Unattended parameters are parameters whose values remain unspecified after an occurrence/utterance of a sentence.

There are parameters whose values are unspecified, yet cause no deficiencies in informative communication. The parameters discussed in the preceding sections are not such parameters. For example, in Gencan’s previously cited example (repeated here for convenience) where the parameter is ‘performer of the action’:

Action:	Constant	sleeping	
Tense:	Constant	present	
Aspect:	Constant	progressive	
Performer of the action:	Parameter	Ayşe/Ahmet/Turgut/etc.	> Turgut
etc.	Constants		

failing to specify values for those parameters results in awkwardness/unnaturalness, as can be seen in the following:

- [41] a ‘Who is sleeping?’
 b” ## *Uyuyor.*
 sleep-pr.prog.3sg
 ‘(it/he/she) is sleeping’

In this example, B does not attend to the parameter to specify its value. This makes (b”), which consists only of non-variables, an awkward response to (a). However, not all parameters call for value-specification. There are parameters whose values remain unspecified in a dialogue and yet cause no deficiency to informative communication. Some examples of such unattended parameters are examined below. First, observe the following passage taken from a dialogue part of Muzaffer Buyrukçu’s diary (1998:112):

- «Ne içiyoruz Buyrukçu?» ‘What shall/do we drink, Buyrukçu?’
«Rakı,» dedim. ‘Rakı’ I said.

Let us extract the dialogue and add glosses to it:

- [42] a. *Ne iç-iyor-uz* *Buyrukçu?*

what drink-CONT-1PL Buyrukçu
'What do we drink, Buyrukçu?'

- b. *Rakı*
rakı
'Rakı.'

The only morpheme appearing in the answer is *rakı* (a Turkish liquor). Is this utterance sufficient in specifying the values of parameters which are there in this dialogue? It certainly specifies the value for the parameter 'object of the action (of drinking)'. However, this is not the only parameter in this dialogue.

B does not specify the value for the parameter 'identity of the prospective drinker'. In this sort of situation, at any rate in Turkish, the identity would be open to at least the following two possibilities: 'the answerer alone' and 'the answerer and the questioner'. (This is despite the appearance of *-uz* [1pl] in the question.) Accordingly, both of the following sentences are eligible to be (b)'s allosentences:

- b'. *biz rakı içiyor-uz*
we rakı drink-cont-1pl
'We drink rakı.'
- b''. *ben rakı içiyor-um*
I rakı drink-cont-1sg
'I drink rakı.'

The identity of the prospective drinker or drinkers is, then, a parameter the value for which is unspecified by (b). Though it is a parameter, neither Buyrukçu nor his friend makes any attempt to specify the value of it. Consequently, the parameter is left unattended by A and B. What becomes apparent from the observation of the above example is that the interlocutors are not interested in specifying the value for the parameter 'identity of the prospective drinker'.

This case shows that not all parameters' values have to be specified. This is because there are certain parameters' values which interlocutors think do not have to be specified. We call these sorts of either intentionally or unintentionally neglected parameters **unattended parameters**. Since values for such parameters are not specified, there is no variable specifying the value for the parameter.

Indifference of interlocutors to the value-specification for parameters is, however, hardly unusual or odd. Nor is it unique to Turkish. For example, in a previously cited Tajik dialogue, the same sort of indifference to value-specification as above is found:

- [43] a. *Metini tu kanī, hoy, pisaram?*
‘Where’s your pickaxe, my son?’
b. *Na-dor-am.*
neg-have-1sg
‘(I) don’t have (it/one).’

In this dialogue, the value for the parameter posited by A, namely ‘whereabouts of B’s pickaxe’ is unattended by B and its value is left unspecified.

Thus, there are a number of parameters’ values that go unspecified⁷⁹. There are, however, cases where interlocutors intentionally make their contribution less or more informative than is required. We will have a brief look at such cases in the next subsection.

2.4.4.4.2.1. Parameters Unattended/Attended by Intention

There can be different reasons for not attending certain parameters. In the cases we have observed in the preceding subsections, parameters are unattended more or less by interlocutors’ consensus, i.e. the parameters unattended by B are not considered by A to be necessary to be attended.

However, this is not always the case. There are cases where one does not specify values of parameters for which another interlocutor thinks values should be specified. Many of these cases involve intentional avoidance of value-specification by an interlocutor. There are also cases where one specifies values for more parameters than another thinks is appropriate. The former of these, i.e. intentional avoidance of value-specification, is often the case when an interlocutor wants to avoid making clear statements about certain issues.

An example of such an intentionally unattended parameter is observed in the following answer to a question in a questionnaire taken from a daily newspaper *Ulus* (1964.4.20,

⁷⁹ The motive for this selective value-specification of parameters may be ascribed, in some cases, to Grice’s (1975:45-46) maxims of quantity, namely ‘Do not make your contribution more informative than is required’.

cited in Demir and Akar (1994:105)). The question is ‘Is there any troublesome side to your relations with Turks which is caused by racial and religious differences?’⁸⁰. The answer belongs to an ethnic Greek tailor in Istanbul: *Münasebetlerimiz gayet iyidir*. ‘Our relationships are very good’. The negligence of the parameter set by the question may be intentional and ascribed to the increased anti-(ethnic-)Greek sentiment of the time⁸¹, which was the motive for the preparation of the questionnaire.

In yet other cases, the motive is to make one’s contribution less informative. The following passage taken from the English tabloid *the Sun* (van Dijk 1993:258 cited in Maynard 1997:86):

(Four Asians acquitted). They were among a mob of 50 Asians who smashed up an East London pub after a series of hammer attacks on other Asians. (1985.8.14)

Maynard says this sentence clearly informs the reader about the fact that a ‘mob’ of Asians smashed a pub, but does not inform who repeatedly attacked Asians with hammers. This, she claims, conceals the background that the Asian ‘mob’ attacked the pub to get even with the whites who had repeatedly attacked fellow Asians. The parameter ‘identity of hammer-attackers’ is, then, an example of an unattended parameter.

As is clear from this example, there are specific devices for not specifying the values of specific parameters. In the *Sun* article, the use of the noun ‘attack’ instead of the verb ‘attack’ is the device. Passivisation is another common device for not specifying the value of the parameter ‘subject’s identity’. In Bukharan Tajik (of young speakers) which does not utilise the *-a šudan* passivisation, the device for not value-specifying subjects is the use of the third person plural pronouns.

Let us briefly note the reverse of the above cases of unattended parameters. Some parameters are attended regardless of Grice’s maxim. Van Dijk (1993 cited in Maynard 1997:86) cites an article in the *Daily Telegraph* 1985.10.16 where the mayor of Bradford, Mr. Ajeeb is introduced as ‘former peasant farmer from Pakistan’. Maynard says his

⁸⁰ *İrk ve din ayrılığı nedeniyle Türklerle münasebetlerinizde aksayan taraf var mı?* in the original wording.

⁸¹ See Demir and Akar (1994) for a detailed description of the public opinion in Turkey which was being shaped in antagonism towards ethnic Greeks in Turkey and which paved the way for thousands of ethnic Greeks to leave Turkey.

personal history which has no relation to the content of the article is provided in order to threaten the mayor's authority or give a negative impression of him to the reader. This is, then, a case where more parameters than required are value-specified.

2.4.4.4.3. Summary

In this subsection, we have analysed examples of echoing and unattended parameters. The analyses revealed that there are variables which appear as if they are non-variables and that there are parameters unattended and parameters unnecessarily attended.

In the next subsection, we will look at pre-grammatical sentences – sentences which make good sense but are not 'correct' according to classical sentence grammar. We will start the analysis of pre-grammatical sentences with discussions on the notion of 'pre-grammaticality' as opposed to 'ungrammaticality'. The discussion involves the concept of 'grammatical acceptability', which concerns also the meaning of the expression 'grammatically required' which appears in rule 1-2.

2.4.4.5. Pre-grammatical Sentences

In the preceding subsection, we pointed out the possibility that many sentences which appear to be 'incomplete' can be accounted for with rule 1. However, despite their being 'incomplete', the examples which appeared in the preceding chapters were all grammatically acceptable.

Such sentences are considered incomplete, because native speakers' 'reconstruction' of their 'original' forms coincide in near unanimity. This is the basis of Palmer's claim that 'horses' as an answer to 'what are those animals in that field?' is to be reconstructed as 'they are horses'.

However, in this chapter, we look at some examples which would not attract a single unanimously accepted 'reconstruction' by native speakers. Some of the examples even violate CS grammar, which inapplicability of literary grammatical rules necessitates an alternative rule to account for the production of such examples. In the discussion that follows, we claim that the first half of rule 1, i.e. rule 1-1, is the alternative rule that takes the place of CS grammar in the production of such sentences. In other words, we will claim that exclusive operation of rule 1-1 may yield such sentences, which we will call

pre-grammatical sentences. The salient characteristic of such sentences is that they make good sense but are not ‘correct’ according to CS grammar.

We shall first briefly examine children’s language with some reference to studies on language acquisition. This may look like deviance from our topic ‘information structure’. However, as it will become clear in the course of the following discussion, it is hardly so. The following discussion on children’s language proves to be highly relevant to our discussion, insofar as it concerns the rules that seem to operate in children’s production of pre-grammatical utterances. It provides us with a unique insight into how pre-grammatical sentences in adult speech is produced.

However, before we proceed to the discussion of child language, we have to define pre-grammaticality, which is the subject of the following subsection.

2.4.4.5.1. Pre-Grammatical but not Un-grammatical

In this section, a few new words are added to our terminology. The words are hoped to be a clear representation of the notions referred to them, so that unnecessary terminological confusion is avoided. First, we would like to distinguish between **pre-grammatical** sentences and **ungrammatical** sentences. As was briefly explained before, pre-grammaticality is about sentences which make good sense and are acceptable but not ‘correct’ according to CS grammar.

‘Relative grammaticality’ is a term van Dijk (1981:78) uses in relation to sentences which make good sense but are not right if judged by rules set by ‘classical sentence grammar’. CS grammar very often exclusively concerns sentence-internal grammar and generally pays little attention to sentence-external, text/discourse-level linguistic actualities. As a result, there are acceptable sentences which are nevertheless judged to be ‘ungrammatical’ by CS grammar. Van Dijk uses the term ‘relative grammaticality’, which corresponds to our ‘pre-grammaticality’, to account for such sentences.

Some examples of pre-grammatical sentences taken from his book (ibid:77) are cited below:

- (5) A: Did you hit him?
B: No. He me.
- (6) A: With what has the postman been murdered?

B: John thinks with a knife.

(5B) and (6B) in the above examples are certain to be judged as ungrammatical sentences according to English CS grammar. However, equally certain is that they make good sense and are not even odd as answers to (A)s. (B)s are sentences which are characterised by their ‘pre-grammaticality’ in the terminology we introduce in this section – we refer to sentences that are not perceived to be odd and make good sense despite their inconsistency with CS grammar as pre-grammatical sentences. A simpler way of saying this is: pre-grammatical sentences are ungrammatical according to CS grammar, but are acceptable – such acceptable yet ‘ungrammatical’ sentences will be called pre-grammatical sentences.

The choice of the term ‘pre-grammatical’ rather than something like ‘quasi-grammatical’ is due to the fact that pre-grammatical sentences are salient in children’s language, particularly in its phase prior to children’s embarkation on adult grammar. Pre-grammatical sentences in children’s language will be discussed in details in the following section.

2.4.4.5.2. Child Language

Let us turn our attention to children’s language. In the first half of this subsection, we briefly review theories concerning children’s speech production, one of which provides a practical ground for our discussion.

Children start producing grammatically correct sentences in the course of their growth and maturation. There may be delays and errors, some of which are claimed to be possibly of genetic causes (see Chapman 2000:39-42 for some examples of such cases), but typical developing children eventually start speaking according to grammar. There are a number of processes involved in children’s language development (see Pinker 1994:262-296), but here we are specifically concerned with child speech at the one/two word utterances period.

The reason why we are interested particularly in this period is that this is the period where children’s speech has not embarked on adult grammar. However, there is great controversy regarding children’s utilisation of adult grammar. Some different standpoints

and theories are reviewed in the following subsection, following which, in light of the reviewed theories, our own standpoint will be determined.

2.4.4.5.2.1. Language Learning

There is immense controversy over what is called continuity/discontinuity in grammatical development. This controversy is essentially all about whether children under two years of age who are at the one-two word utterance stage produce utterances in accordance with an underdeveloped version of adult grammar. There are various theoretical positions on the acquisition of grammar such as behaviourist account⁸² and social/cognitive account⁸³. Among them, the following two have stimulated most debate and controversy in recent decades. They are also the most influential among the four and have attracted a large number of scholars from different disciplines:

Connectionism: ‘seemingly rule-governed behavior – such as speaking a language – can be explained without recourse to underlying rules’ (Hoff-Ginsberg 1996: 179)

Innate grammar position: ‘the acquisition of language is significantly supported by innate syntactic knowledge and language-specific learning procedures’ (ibid:175)

The first of the above two, namely connectionism (also known as parallel distributed processing), has been the most controversial of all the positions. To put it plainly, the connectionists’ position is that one does not necessarily need underlying rules to produce grammatically correct words (and possibly also sentences or utterances). It is better to cite Pinker’s (1997:547) reference to connectionism’s much cited and indeed controversy-provoking work, Rumelhart and McClelland (1986), to get an idea of what the position points to:

Connectionists point to a kind of artificial neural network, called a pattern-associator memory. In their classic computer model, David Rumelhart and James McClelland used a bank of input units for sounds in the verb stem, a bank of output units for sounds in the past-tense form, tunable connections between every input and output, and a learning

⁸² ‘[C]hildren imitate what they hear, and they are reinforced when they get it right and are corrected’ (Hoff-Ginsberg 1996:173)

⁸³ ‘[T]he starting point of language acquisition is provided by general cognition, as are the mechanisms of language development. The requisite experience for language acquisition is social interaction with other speakers’ (Hoff-Ginsberg 1996:174)

mechanism that recorded correlations among stem sounds and past-tense sounds (for example, -ing with -ang, and -alk with -alked). The model learned the past-tense forms of several hundred verbs, and generalized with fair accuracy to new verbs. Its errors resembled those made by children (such as come-comed), but it didn't use anything that looked like a rule, nor did it have distinct components for regular and irregular forms. So connectionists claim that symbolic rules, and any qualitative distinction between regular and irregular verbs, are fictions.

Thus, Rumelhart and McClelland conclude that their 'simple learning model shows, to a remarkable degree, the characteristics of young children learning the morphology of the past tense in English' (Rumelhart and McClelland 1986:266). Whether this single mechanism can explain the learning and production of English past tense and, by extension, indeed, the acquisition and production of language has been a focus of controversy in cognitive science (see, e.g., Plunkett 2000).⁸⁴

Connectionism, interesting as it is, has not yet seen many attempts of accounting for morphology more complex than past tense and plural formation (Plunkett and Juola 1999) at the time of writing of this thesis (see Christiansen and Chater 1999). Therefore, in the discussion that follows, we largely exclude connectionist accounts from our discussion. Instead, we shall discuss at length the other influential position, namely the innate grammar position, which is meant to be able to account for the entirety of morpho-syntax.

The position which is often considered to be the alternative (see Christiansen and Chater 1999:419) or sometimes opponent of connectionism is the second of the viewpoints listed above, namely the innate grammar position. The linguistic theory of Government and Binding theory is based on, or strongly associated with, this position. Syntacticians who practice this theory assume every child's innate knowledge of 'Universal Grammar', 'which consists of a set of principles that hold for every language and a set of options, or parameters, that have to be filled in by experience' (ibid:177) The notion of Universal Grammar naturally entails that, as O'Grady (1997:331) writes, 'the entire system of UG is available from the beginning of the acquisition process'. (See, however, O'Grady (1997: chapter 15) for different theories within the innate grammar position.)

Bickerton's (1990:111-112) claim cited below is a critical response to the idea of identifying an adult grammar in child speech. As Bickerton's criticism and the fact that there are different viewpoints on the acquisition of grammar show, the innate grammar position is, like the others, an assumption the validity of which is not undebatable⁸⁵. According to Bickerton (1990:114), 'the capacity to join different word-classes in fairly consistent patterns is just as much a characteristic of apes as it is of under-twos'. As is claimed in this comment, the capacity is perhaps not even uniquely human. Matsuzawa (1985) conducted an experiment where a chimpanzee named Ai, which had been taught to combine visual symbols as lexigrams, was required to name the number, colour, and object of sample items. The experiment is outlined below:

For example, when 5 red toothbrushes were shown in a display window as a simple item, it was necessary for Ai to press keys of '5', 'red' and 'toothbrush' in any order. Although no particular 'word order' was required, the chimpanzee favored two particular sequences almost exclusively among six possible alternatives: color/object/number and object/color/number. In both sequences, numerical naming was always located in the last position. (Matsuzawa 1996:198)

He claims the study demonstrates Ai's ability to spontaneously organise the 'word order' (Matsuzawa 1985:57). These facts, together with the case of Genie's language which will be explained below, appear to be in opposition to the attempt to identify the same syntactic rules as those which may be used by adults in children's two or three word utterances.

In opposition to such syntactic analysis of children's language, Bickerton (1990:111-112) claims the following:

The use of identical categories for child and adult grammars has been defended on the grounds that there is no principled basis for assuming that the child employs different types of grammar at different acquisition stages [...]. It has no force, however, against the proposal that children, from about age two onwards, do indeed use a single grammar, but

⁸⁴ A study by Marslen-Wilson and Tyler (1997), in which they demonstrated the neurological dissociation between the regular and irregular past tense production, is considered by some scholars to have posed a challenge to the view that the single mechanism produces both regular and irregular past tense forms.

⁸⁵ Syntacticians' justification for identifying innate grammar in speech of children under two years of age (18 month-2 years after birth is the period in which children start to put words together) is found in Atkinson (1992:173).

that younger children have no grammar at all, because what they are learning is not really language.

He presents various examples, including the language of Genie, a girl who had been kept from any exposure to language until she was thirteen, to claim that there is a 'language' which lacks distinguishing features (e.g. grammar) of language. Some fragments of Genie's speech cited in his book are: 'want milk', 'Mike paint', 'applesauce buy store', 'very sad, climb mountain' (ibid:116). Bickerton notes, 'only hearing the utterance in context could indicate whether *Mike paint* meant *Mike paints* or *Mike's paint*'. (In fact, without the help of context, 'Mike paint' could also be 'Mike painted' or 'Mike has painted' or 'Mike had painted' etc. since tense is not, or at least not always, indicated in her speech.)

Significantly, most of the difficulties which the comments cited above claim to exist in attempts to locate an adult grammar in children's speech appear to arise from the attempt to set up 'complete sentences' for all 'incomplete sentences' in children's speech. This assumption that children's sentences are incomplete realisations of sentences which have more constituents in them, is widespread. For example, Weisenberger (1976:275) says:

At the earliest stage of the child's language development his speech is limited by his inability to process more than a few words at a time in one utterance. From the situation in which an utterance occurs it is often clear that the child intends more than he is able to lexicalize. [...] how does the child choose from four or more possible words the few that he will lexicalize?

There is an assumption implicit in her view that children's utterances are 'elliptical'. The assumption, which seems to have certain popularity among some works published in 1970s, is that there are 'underlying or intended but unlexicalised words' behind 'one, two and three words that the child lexicalises in one utterance' (ibid:276). This standpoint has something in common with the postulation of inaudible elements. Hence the idea that certain 'categories of the syntactic deep structure are eliminated in the resulting surface structure [...]' (Bloom's (1970) concept of 'reduction transformation' examined in Miller (1978:454)).

A standpoint that is divergent from this is put forward by Greenfield (1978:450). She points out the principle that both children and adults follow in producing one-word utterances: 'express the single most informative element'. She does not refer to

transformational grammar or ellipsis, which diverts her analysis from syntactic analyses of children's utterances. As for two-word utterances, she says 'informativeness, as well as English syntax, play a role in the word order of two-word utterances' (Baker and Greenfield 1988:29).

The former of the two different views is based on the assumption that children share qualitatively identical grammar with adults. It assumes that an 'immature' version of adults' grammar is shared by young children. The latter, on the other hand, assumes a different system of production of utterances in young children from that in adults. Hoff-Ginsberg (1996:170), regarding this matter, writes:

Also remember that it is not at all clear that very young speakers have grammatical categories. Some have suggested that children's first word combinations are based either on structures learned word-by-word or on semantically based categories. If the latter is the case, then children start out with a very different kind of system than adults have, and somewhere in the course of development they undergo a qualitative change.

Weisenberger's study assumes continuity in grammatical development from early childhood to adulthood, while Greenfield's assumes discontinuity in it.

We can see now that the approach of these theories is in essence the same as the one we have found in many linguistic IS theories reviewed in the previous section. In the previous chapter, we have reviewed IS theories which assume inaudible elements in 'incomplete' sentences. On the other hand, in the theories reviewed above, two-word sentences are considered to be 'immature' or, indeed, 'incomplete' realisations of 'mature' and 'complete' sentences. These theories, then, share their essence with the theories which claim the existence of 'ellipsis' or 'inaudible elements' in the sorts of sentences discussed in the preceding sections. The explanation for 'elliptic' sentences and that for two- or three-word sentences are, thus, two products of the single assumption, namely 'sentences which have fewer components than they should have according to CS grammar miss some components'.

Compare this position with Greenfield's view on children's sentence production reviewed before. Greenfield's analysis of children's sentences, on which our analysis is based, is a bottom-up analysis where sentences are built up by combining constituents which are called for value-specification for parameters. On the other hand, Weisenberger's viewpoint is a top-down one according to which children's sentences are

incomplete realisations of sentences which would have been realised if produced by adults. The ‘top’ is hypothetical ‘complete’ sentences from which constituents are ‘elided’ to yield ‘incomplete’ sentences.

The fact is, however, that adults occasionally produce sentences which show resemblances to children’s one-/two-/three-word sentences. This may mean that both adults’ and children’s pre-grammatical sentences are produced in accordance with the same principle. What, then, can the principle be which is in operation in the production of both children’s and adults’ pre-grammatical sentences?

The principle that explained ‘elliptic’ sentences is rule 1. It seems to us to that the first half of rule 1, i.e. rule 1-1, is capable of providing a simple account of the production of pre-grammatical sentences. (After all, rule 1-1 is based on Greenfield’s observation of children’s one-/two-/three-word sentences.) Accordingly, we apply our method of analysis which we used for ‘elliptic’ sentences in the preceding section to such pre-grammatical sentences produced by children as well as adults.

In the remaining part of the present section, pre-grammatical sentences in child language and adult speech will be discussed. Our claim regarding pre-grammatical sentences is that rule 1 is responsible for production of pre-grammatical sentences regardless of the ages of the producers of them. The reason for the inclusion of child language in our discussion is the abundant and consistent appearance of pre-grammatical sentences and children’s irrelevance to rule 1-2, i.e. securing grammatical correctness. This allows us to observe more directly the operation of rule 1-1 and non-operation of rule 1-2 in pre-grammatical sentences in children’s language than we can in adult speech.

2.4.4.5.2.2. Children’s Pre-grammatical Sentences

In this section, pre-grammatical sentences in children’s language will be analysed to see whether the operation of rule 1-1 as well as non-operation of rule 1-2 can be identified in their production. Firstly, consider the following observation of Pietro, an Italian boy who was capable of producing both one-word and two-word utterances:

We have been playing in the park. Pietro has been throwing the ball to various people other than his mother. Now, he gets ready to throw once more, saying *mami*. Then he throws to his mother. (Greenfield 1982:3)

We can ascribe Pietro's choice of the only word to be verbalised to the word's status as the variable. If he said, say, 'ball' (in Italian), the utterance would have conveyed only a constant piece of information. In other words, Pietro specified the value for the parameter 'the receiver of the ball'. Thus, the 'variable only' principle which we have set up earlier in rule 1-1 seems to manifest itself here too. Secondly, let us examine utterances by children at two-/three-word stage. There have been arguments about whether children at two-word stage possess syntactic competence. The moderate opinion, which is tentatively adopted here is that there is both 'communicative competence' (Miller 1978:455) and syntactic competence. The example below is taken from Miller (1978:461):

In a second developmental phase the child produces utterances like *mama ummachen* meaning something like 'Mama, please turn over the (milk) can':

Meike (age: 1;10/MLU: 1.75)

[Meike is playing with a milk can]

kipp⁻ umkippen

'tip'

mama ummache' [gives M the can]

'mama turn over'

M: Nee. Mama will nichts umkippen.

'No. Mama will not turn over anything.'

Miller writes that the second utterance by Meike *mama ummache* means something like 'Mama, please turn over the (milk) can'. What, then, motivated the utterance of the two words? Our analysis of the two-word utterance would be as follows. *Mama* has Meike himself as an alternative agent to perform the action of turning over the can, while the action of turning over represented by the verb *ummachen* has an indefinite number of alternatives, i.e. other possible actions. In other words, the identity of the performer of the action and the action itself are parameters, whereas, say, the object of the action 'the (milk) can' is a constant as it is given from Meike to his mother at the time of utterance. Therefore, according to our previously proposed rule, the variables *mama* and *ummache* have to be present (while, for instance, 'the can' does not).

One possible account for the production of these sentences, which we adopt here, is that the exclusive operation of rule 1-1 yields such pre-grammatical sentences. Let us

examine this assumption with an example from one of the languages which are subjects of this thesis, Japanese:

At seventeen months her mother pointing at her father's hat asked her "Whose hat is this?"
She answered [to:tɕaɴ]. (Nakazima 1977:40)

Nakazima's analysis of this one-word utterance is as follows:

She knew it was her daddy's, but did not know the form of the possessive case [to:tɕaɴ no] (daddy's)

This may be so, but Nakazima had recorded [botɕɪ] 'hat' uttered by the girl when she was sixteen months old. If the girl uttered an arbitrarily selected word from words that are in association with the question, [botɕɪ] 'hat' could have well been uttered in place of [to:tɕaɴ] 'daddy' as, an 'elliptic' version of [to:tɕaɴ nobotɕɪ] 'daddy's hat'. It seems that the girl's exclusive selection of the latter can be ascribed to its being variable.

The object in the father's possession is constant: it is 'hat'. The relationship between the father and the hat, 'possession' whose morphological representation is the attachment of *no* [gen] to *x* is constant, too. Therefore rule 1-1 does not call for the appearance of *no* [gen] – it requires only the appearance of [to:tɕaɴ] 'daddy'. Thus, it is possible to account for pre-grammatical sentences produced by children with the 'variable only' rule. In other words, the exclusive operation of rule 1-1 can yield such pre-grammatical sentences.

2.4.4.5.2.3. Adults' Pre-grammatical Sentences

In this subsection, we will examine pre-grammatical sentences produced by adults to see whether the exclusive operation of rule 1-1 can account for their production.

In this section, most pre-grammatical sentences are taken from Turkish. There is a reason for this imbalance in the number of examples among the languages dealt with in this thesis. Turkish is abundant in pre-grammatical sentences whose deviation from the CS grammar is much more obvious than it is in, say, Japanese. In other words, the classical-sentence-grammatical unacceptability of Turkish pre-grammatical sentences is unmistakably apparent in Turkish, mainly because the production of Turkish pre-

grammatical sentences involves violation of subject-verb agreement morphology. This will be apparent in the following discussion.

Before we proceed to the examine Turkish pre-grammatical examples, we review an account from a linguistics textbook of the production of adults' pre-grammatical sentences. Perhaps not surprisingly, it is in line with the 'ellipsis' account which we have become familiar with in the preceding sections:

[...] I have put an appropriate preceding context in parenthesis:

(When is she coming?) I hope the eighth of May.

Paul thinks never.

John guessed soon.

All of these sentences are interpreted in the same way as corresponding expanded sentences. So we interpret *I hope the eighth of May* in this context (that is, following the question in parenthesis) the same way we interpret *I hope she is coming the eighth of May*. Sentences like these, which are missing parts, are called ELLIPTICAL. (Napoli 1996:300)

This is identical with the 'ellipsis' explanations which we reviewed in the "ellipsis' or 'inaudible elements'" section. It also shares with them the same problems of 1) being unable to explain why specifically the morphemes which appear in the answers have to appear and 2) arbitrariness in 'reconstructing' sentences with no 'missing' parts. In the "ellipsis' or 'inaudible elements'" section, we have abandoned the 'ellipsis' analysis of 'incomplete sentences' in preference to our own analysis which posits that such sentences are produced by rule 1-1 and rule 1-2.

However, there is a difference between the group of examples dealt with in that section and that which we will examine in this section. The former is grammatical, unlike the latter which is pre-grammatical.

The former is necessarily grammatical because there is rule 1-2, whose role is to secure grammatical correctness of sentences. However, remember that morphemes whose occurrence is called for by rule 1-2 are all non-variables which redundantly specify values for constants and hence are of secondary importance in terms of informative communication. On the other hand, variables called for by rule 1-1 are of primary importance.

These facts allow the assumption that some sentences may be acceptable with only rule 1-1 and without rule 1-2, though the sentences' correctness in terms of CS grammar may be lost without rule 1-2.

This means that exclusive operation of rule 1-1 may produce pre-grammatical sentences in adults' language as we claimed it does in children's language. Perhaps, then, adults occasionally produce pre-grammatical sentences relying on the same principle children rely on when they produce pre-grammatical sentences. In the following analysis of Turkish pre-grammatical sentences, we assume the validity of this hypothesis and see whether the assumption induces any problems in the process of the analysis.

Let us analyse the first adult pre-grammatical example where rule 1-1 is operating and rule 1-2 is not followed. Observe the following:

- [44] a. *Bir milyon kaç kişi?*
one million (lira) how many person
- b. *Bir milyon iki kişi!*
one million (lira) two person

The peculiarity of this dialogue is the oddity of the meanings of each utterance when interpreted in accordance with Turkish CS grammar. If interpreted according to Turkish CS grammar, the question can only be interpreted as 'how many people are one million?', while the answer can only be 'one million is two people!'. Observe the following rather lengthy description of the context in which the above dialogue took place:

The dolmuş 'passenger minivan'⁸⁶ driver to whom I and a friend had passed a one million lira banknote asked the question to us. My friend was sitting in the last row of seats in the dolmuş, from which the driver's seat was a few meters away. The banknote was forwarded to the driver by other passengers sitting between our seat and the driver's. The question was, therefore, yelled from the driver's seat in the beehive of traffic as well as the noise from the engine (the dolmuş had already left the dolmuş stop). The fare per person was three hundred and fifty thousand. Hence the one million could have been for either one or two persons, which is the reason

⁸⁶ Dolmuş used to be all passenger 'old American cars' which had been almost completely replaced by mini-vans by 1996. The translation 'passenger minivan' is used here because the example utterance is from 1999.

the question was asked. The question was answered by my friend. The question and answer means, in effect, something like ‘how many passengers is the one million (lira) for?’ and ‘it is for two passengers!’, respectively.

An informant asked to ‘reconstruct’ the dialogue presented various possible allosentences, some of which are shown below:

for a.: *(Bu) bir milyon (lira), kaç kişi-nin?*
 this one million lira how many person-gen
 ‘(lit) This one million is how many people’s?’
 Bu bir milyon-u kaç kişi ver-di?
 this one million-acc how many person give-past
 ‘How many people gave this one million?’
 Kaç kişi-den al-ıyor-um bu bir milyon-u?
 how many person-abl take-cont-1sg this one million-acc
 ‘(lit) From how many people I’m taking this one million?’

for b.: *İki kişi-den al-ıyor-sun.*
 two person-abl take-cont-2sg
 ‘You are taking (the one million) from two people.’
 İki kişi ver-di-k.
 two person give-past-1pl
 ‘(lit) Us two gave (the one million to you)’

A native speaker’s uncertainty regarding the sentences’ ‘reconstructed’ forms is suggestive, because it is in contradiction to the idea of ‘reconstruction’ itself, and hence also the postulation of ‘(in)complete’ sentences or ‘ellipsis’. As stated in the beginning of this chapter, the claim that ‘horses’ in Palmer’s example is to be reconstructed as ‘they are horses’ owes much of its validity to the fact that native speaker’s ‘reconstruction’ of its ‘original’ form coincide in near unanimity. However, the above Turkish example does not attract a single unanimously accepted ‘reconstruction’. This forces us to abandon the ‘incomplete sentence’/‘ellipsis’ approach to the Turkish example.

Note that even the semantic roles of the NPs *bir milyon* and *kişi* are not consistent in these ‘reconstructed’ sentences. It should be possible to account for the example dialogue within the framework of Government and Binding theory, claiming that the dialogue on the surface level can have various D-structures. However, its being possible does not

entail that it is the best way of analysing the example. As stated before, we do not assume deletion or empty categories in this thesis. Accordingly, the above example is analysed exclusively in terms of information structure.

Let us turn to the answer (b). The *bir milyon* part is a repetition of the same constituent in (a). This is echoing as explained in the section ‘echoing’. The dolmuş is running and there are a few meters of distance between the driver and the answerer. The echoing is therefore to confirm the denomination of the money the driver mentions. If B sat next to the driver and gave him the money, the answer could have been *iki kişi*, or just *iki* without *bir milyon*, because there would have been no need to verify their common constant.

İki kişi! then follows *bir milyon* to convey the information how many passengers’ fares should be taken by the driver from the one million lira. *Kişi* is also a echoed A’s non-variable (and B’s variable) which would not have to be uttered if B was next to the driver or inside the dolmuş is sufficiently quiet.

(It is also possible to ascribe *kişi*’s appearance to the utterers’ avoidance of the formation of the succession of *bir milyon iki* as an utterance, because the succession could be understood as the numeral ‘one million and two’. This avoidance of misunderstanding could constitute the third rule in rule 1, but we neglect this rule in this thesis for the following reasons: 1) the undesirable complication which the inclusion of a pragmatic rule in the present discussion may induce, and 2) analyses of few examples in this thesis involve this rule.)

One may think that the answer needs to specify also the identity of the payer of the one million, i.e. B. However, this is not necessary. The identity of the payer is not a parameter and is a constant the value of which is ‘(a) passenger(s) in the dolmuş’ for the driver, whose status as a constant is not challenged by the answerer. The identity of the payer is therefore a constant for both the driver and the answerer. Both sides are happy with the situation where the value of the constant is no more specific than ‘(a) passenger(s) in the dolmuş’. Accordingly, neither side attempts to re-specify the value. (Alternatively, ‘identity of the payer’ may be analysed to be an unattended parameter.)

On the other hand, the variable *iki* has to be present in an answer to (a), whatever the context is, because, unlike *bir milyon* and *kişi* which redundantly specify values for

constants, it is the variable that specifies the value for the parameter ‘number of passengers’.

Thus, the above can serve as an example of how an information structural rule, i.e. rule 1-1 can push aside (classical sentence) grammar and also, indeed, (classical sentence) syntax from language use in adult speech.

Let us look at some other Turkish examples in which rule 1-1 is in operation while rule 1-2 is not:

- [45] a. *İstoç var mı?*
İstoç exist Q
b. *Eskoop var!*
Eskoop exist

Context: In a bus operating between Eminönü and Başakşehir, the bus conductor yelled the above question to the passengers to ask whether there is anyone getting off at a bus stop in İstoç, a new industrial site. If there was not one, the bus would change its route and bypass İstoç. One passenger, mishearing İstoç for Eskoop, a stop between İstoç and the terminal, where he is getting off, yells back the sentence b. Immediately noticing that he misheard İstoç for Eskoop, he added the following c, letting the conductor know that he is not getting off at any of the bus stops in İstoç.

- c. *İstoç yok! İstoç yok!*
İstoç not exist İstoç not exist

All a, b, and c are, in fact, (CS grammar-wise) grammatically perfectly acceptable.

However, in the framework of CS grammar, they can be interpreted only as:

- a. Does İstoç exist?
b. Eskoop exists!
c. İstoç doesn’t exist! İstoç doesn’t exist!

Why would a bus conductor want to ask the passengers whether İstoç, which the conductor himself most certainly knows is existent, exists?

Of course, the bus conductor does not ask whether İstoç exists; The answerer, of course, does not inform the conductor about the existence of the industrial site. Both of them, particularly the conductor, whose bus passes through the site everyday, should know

very well that it exists – if he literally means what the utterance said according to CS grammar, he is most certainly out of his mind. Of course he was not asking the passengers whether İstoç exists – he asked whether there are any passengers getting off the bus in İstoç.

Likewise, the answerer did not inform the conductor about the existence of Eskoop, nor did he confirm the conductor’s statement about the non-existence of İstoç. He informed the conductor that he was getting off at Eskoop, mishearing İstoç for Eskoop. Naturally, the repetition of ‘İstoç doesn’t exist!’ which followed it had nothing to do with the existence of İstoç – he told the conductor that there was nobody who was getting off in İstoç.

Then, this question-answer pair, which is perfectly acceptable but incorrect in terms of CS grammar, is another example of pre-grammatical sentences. However, some questions remain. Why does the question comprise [İstoç], [exist], and [Q] but not, say, [get off]? Why does the answer consist of [Eskoop] and [exist], but not, say, [I]?

This is because both [get off] and [I] are pre-specified values for constants for the conductor, the answerer, and, indeed, everybody in the bus. On the other hand, the morphemes that are present in the answer are all (echoed and not echoed) variables:

	A’s	B’s	Value-specification:
Action:	Constant	Constant	Getting off from the bus
Location:	Constant	Parameter	İstoç (echo. In (b), for B, ‘Eskoop’)
Existence of performer:	Parameter	Parameter	> Positive (b) > Negative (c)
etc.	Constants	Constants	

The absence of any constant constituents will not lead to unacceptability despite the incorrect interpretation according to CS grammar, as is clear from the example – the exclusive operation of rule 1-1 does not cause unacceptability in (b). On the other hand, absence of any of the variables would induce not only grammatical inadequacy, which already exists in the utterances, but also information-structural inadequacy.

For example, if the variable for specifying the existence of a passenger getting off i.e. *var* ‘exist’ was absent in (b), the appearance of the remaining part of (b), *Eskoop* alone, though it is also a variable, would not make any sense in isolation in the dialogue. This

suggests the obligatory nature of the operation of rule 1-1 in any sentences. If rule 1-1 is bleached, the sentence loses its acceptability.

Thus, as in children's language, pre-grammatical sentences are produced in adult language too. Moreover, the pre-grammatical sentences are produced by the exclusive operation of rule 1-1, as is the case also in children's language.

However, there seems to be a difference between children's pre-grammatical sentences and adults' grammatical sentences. For example, adults usually do not say 'mummy blue dress', however readily acceptable the utterance is in the context where it is uttered. What, then, differentiates adults' pre-grammatical sentences from children's pre-grammatical sentences? We do not have a conclusive answer to this question, but we suggest some possible factors that may be involved in the differentiation in the subsection 'word etc.'.

2.4.4.5.2.4. Pre-grammatical Sentences with 'Dropped' Case Markers

So-called 'case marker drops' are a research focus within the field of traditional Japanese linguistics to which constant attention has been paid for over a century. The 'case marker drop' refers to the absence of case markers where their presence is predicted by CS grammar. One peculiarity of this 'drop' is that the absence, despite its nonconformity with the rules set by CS grammar, does not induce unacceptability. Sentences where this phenomenon is involved are, then, instances of sentences which are acceptable, yet not entirely grammatical, i.e. pre-grammatical sentences. In this subsection, we observe some examples of 'case marker drops' and see whether the analysis that explained the production of some pre-grammatical sentences in the previous subsections works also for this phenomenon.

There are a number of different accounts for the phenomenon of 'case marker drops'. In many studies, 'case marker drop' or 'elision' or 'omission' is the term typically used to refer to a lack of overt case marking in Japanese (e.g. Fukuda 1993). Among various other names given to the lack of case marking, there are 'ellipsis' (Martin 1970, 1975:50,

Watanabe 1986), ‘latency and omission of *ni*’⁸⁷ (Kudō 1950), ‘case marker omission’⁸⁸ (Fujiwara 1992), etc.

A number of grammars, notably Matsushita’s *Nihon Zokugo Bunten* (1901) ‘Compendium of Vulgar Japanese’ and *Hyoujun Nihon Kōgohō* (1961) ‘Standard Colloquial Japanese Grammar’, refer to absent case markers in colloquial Japanese. However, the lack of case marking has been usually considered to be outside the norm of CS grammar. Other explanations for ‘case-marker drops’ typically assume the existence of zero-morphemes where, according to CS grammar, there should be case markers. (see, for example, Hasegawa 1993, Tanba 1993, Maruyama 1995)

One observation shared by these studies is that many sentences without case markers make good sense despite the lack of case markers. Such sentences, then, are pre-grammatical sentences. Accordingly, we shall see whether the absence of case markers can be accounted for with the same principles that accounted for the production of pre-grammatical sentences in preceding subsections, namely the exclusive operation of rule 1-1 (and non-operation of rule 1-2). Observe the following examples of Japanese ‘case-marker drops’:

[46] *ningen to shite yakusoku mamore* (Itami 1992)

human as promise keep

‘as a man, keep the promise.’

sagamiko itta toki... (Noda and Ozu 1956)

lake Sagami went when

‘when (we) went to the lake Sagami...’

Almost all previous studies on sequences of a nominal and a verb such as these analyse the sequences to be results of ‘case marker drops’. According to them, in the examples

⁸⁷ *Ni no senzai to shōryaku* in Japanese. There are numerous other terms devised for the phenomenon, such as: *zero kigō hyōji* ‘zero-mark-representation’ (Mizutani 1996), ‘postposition drops’ (Hosaka et al 1992), *mujoshi meishiku* ‘NP without particles’ (Hosaka et al 1992), *joshiochi* ‘case marker drop’ (Yamamoto 1992), *mujoshi* ‘no particle’ (Hasegawa 1993), *o no shōryaku* ‘omission of *o*’ (Minashima 1993), *ippankaku* ‘general case’ (Matsushita 1961), *mujoshikaku* ‘no-case-marker case’ (Tanba 1993, Maruyama 1995), *joshi no shōryaku* ‘omission of case markers’ (Kitayama 1953), *namae-kaku* ‘name-case’ (Suzuki 1972), and *mukeika* ‘tangibility>intangibility conversion’.

⁸⁸ *Joshi shōryaku*.

above, the accusative case marker *o* and the dative case marker *ni* are ‘missing’ or have been ‘dropped off’. Martin (1975:50) notes in *A Reference Grammar of Japanese*, thus:

The surface versions of sentences – what we hear and see – often contain an optional omission of a marker, the result of ellipsis. The object particle *o* is very frequently dropped [...]

According to this analysis, the example sentences above have undergone the process shown below:

yakusoku o mamor- → omission of *o* → *yakusoku mamor-*
sagamiko ni ik- → omission of *ni* → *sagamiko ik-*

However, the ‘ellipsis’ explanation does not fully account for utterances where, on the face of it, case markers appear to have ‘dropped off’. For instance, there is no way one can identify the ‘dropped’ case markers with certainty in the latter of the examples above. This is, then, the same problem the ‘ellipsis’ approach had in the preceding subsections: the inevitable arbitrariness of ‘reconstruction’ of ‘complete’ sentences which the approach necessitates. Some may claim that the ‘dropped’ case marker is the allative case marker *e*, while others maintain it is the dative case marker *ni*. This remains in dispute because there seems to be few semantic or grammatical or syntactical conditions that affect the choice between *ni* and *e* for marking destinations.

The allative *e* and dative *ni* were functionally differentiated from each other in historical Japanese. In *Manyōshū* compiled in the eighth century, the former encoded movement from the present location to the distance, whereas the latter encoded the destination (Ōno 1978:21). However, in modern Japanese, *ni* and *e* are largely interchangeable as shown below:

[47] *sagamiko e itta toki...*
lake Sagami all went time
‘when (we) went to the lake Sagami...’
sagamiko ni itta toki...
lake Sagami dat went time
‘when (we) went to the lake Sagami...’

What this suggests is that in modern Japanese, both *e* and *ni* are used in marking destination⁸⁹, regardless of the direction of the movement to the destination or whether the destination is a terminus or not. This is to say that the ‘reconstruction’ approach necessarily fails here and so does ‘ellipsis’ as an explanation for ‘case marker drops’. In the remaining part of this subsection, we will see whether rule 1-1 accounts for production of pre-grammatical sentences involving ‘case marker drops’, as it did for production of pre-grammatical sentence in preceding subsections. The absence of case markers may be explained as a result of exclusive operation of rule 1-1, if they are non-variables.

In order to explain the absence of a case marker between *sagamiko* and *itta* in accordance with rule 1, the value which *e* or *ni* specifies must be pre-specified. However, in this example, there is no question in which pre-specification can take place. Moreover, the sentence makes sense even if it is uttered out of the blue. This suggests the existence of a constant the value for which is pre-specified without any value-specification by linguistic expression. It is, then, a constant that usually does not need to be specified, such as the ones we categorised under the label of ‘et cetera’ when we discussed Clark’s ‘building block’ example. (The constants in the ‘et cetera’ included, for example, ‘the language of communication’ and ‘B is hearing-impaired’ whose values were pre-specified as [English] and [Negative], respectively.) That is, there has to be a constant the value for which is pre-specified as [to]. It also has to be a constant shared by users of Japanese, such as this:

	Value
Action - object relation:	Constant to
(when action is ‘going (i.e. <i>iku</i> ⁹⁰)’ and object is location)	

Accordingly, we think it is possible to postulate a chart of constants such as the following:

Action:	Constant Going
Object:	Constant Lake Sagami
Action - object relation:	Constant to
(when action is [going (signified with the verb <i>iku</i>)] and object is location)	

⁸⁹ Or in marking NPs with the semantic role of goal, in the terminology of Case Grammar.

⁹⁰ *itta* ‘went’ is the past tense form of *iku* ‘go’

etc.

Constants

The third of the constants above, then, has to be valid in the entirety of Japanese. This in fact proves to be the case and there is ample evidence for its validity. Observe the following examples:

[48] *tōkōen ik-anai?* (Noda and Ozu 1956)

Tōkōen go-neg

‘shall we go to Tōkōen?’

daimaru it-te ne (Inoue 1984)

Daimaru go-ger sfp

‘(I) went to Daimaru (department store), then ...’

Note that neither of *tōkōen* and *daimaru* is marked with *ni* or *e*. Such absence of case markers is not unique to the Japanese of Tokyo and apparently widespread among dialects (see, e.g. Arai 1984, 1990, Kai 1991, Matsumoto 1982, Nakama 1984, Nohara 1979, Uemura 1972, Teru 1982) as well as in historical Japanese, which implies the prevalence of the constant in the Japanese as a whole.

This constant may not be very unique to Japanese. A similar structure to the example above can be found also in a variety of languages. For instance, Arzumanov and Sanginov (1988:226) presents the following as an example of an ‘omitted’ preposition *ba* ‘to’ in colloquial Tajik:

[49] *vay bozor raft*
s/he market go.past.3sg
‘s/he went to (the) market.’

In literary Tajik, *ba* ‘to’ has to appear in front of *bozor*. The acceptability of the example in colloquial Tajik, then, implies the existence of a constant that resembles the Japanese one: ‘action is *x* object when action is [going] and object is location’ whose value is pre-specified to be [to].

In Bukharan Tajik, the adposition is realised as a postposition/suffix, but the same constant seems to exist also in the dialect. A Bukharan informant even corrected our sentence *vay Amerika-ba raft* ‘he left to America’, claiming it should be *vay Amerika raft* in Bukharan Tajik. A phenomenon similar to this is found also in Persian. Mahootian (1997:165-6) presents the following examples, claiming that the use of the preposition *be* ‘to’, the Persian counterpart of Tajik *ba*, is optional:

bæcce-ha be mædrese ræft-ænd
 child-PL to school went-3P

‘The children went to school’

bæcce-ha mædrese ræft-ænd
 child-PL school went-3P

‘The children went to school’

This coincidence of the preferred or optional absence of the ‘to’ morpheme in the presence of both location and the [go] verb in Japanese and a few Iranian languages could suggest certain cross-linguistic nature of the constant.

It is worth noting that, the above discussed language-specific/inherent constant is probably one of the most simplest of its kind. The constant involves essentially only that the verb is [go] and the object is a location. In other words, they are the only conditions required for the constant’s value to be pre-specified without needing linguistic expression. The ‘constant’ status of other such constants may be dependent on more complex and numerous conditions. For example, in Bukharan Tajik, the absence of the [to] morpheme is realised with more complexity when the verb is [(tele)phone]. It seems to involve at least animacy of the object, as is understood from the following examples:

[50] *Šohida-ba telefon kard-am*

Shohida-to telephone did-1sg

‘I phoned Shohida’

Šohida telefon kard-am

In the above example, the occurrence of *ba* is obligatory, as it is in the following where the object is Shohida’s mother (*ona*):

[51] *Oni⁹¹ Šohida-ba telefon kard-am*

mother.iz Shohida-to telephone did-1sg

‘I called Shohida’s mother’

Oni Šohida telefon kard-am

However, in the following in which the object is *xona* ‘house’, the appearance of *ba* is not obligatory:

[52] *Xoni⁹² Šohida-ba telefon kard-am*

⁹¹ *ona* + izafet

⁹² *xona* + izafet

house.iz Shohida-to telephone did-1sg
'I called Shohida's home'
Xoni Šohida telefon kard-am

Thus, constants whose values need not be specified by linguistic expression may call for complex prerequisites to be constants.

In this subsection, pre-grammatical sentences with what appears to be 'ellipsis' of case markers was discussed in connection with IS. The discussion showed that production of a variety of pre-grammatical sentences can be ascribed to the operation of rule 1-1 and non-operation of rule 1-2.

2.4.4.5.3. Summary of 'Pre-grammatical Sentences'

We have examined pre-grammatical sentences the production of which has been explained in terms of 'ellipsis' in previously proposed theories. In the first few subsections, we have introduced Greenfield's view that children's production of pre-grammatical sentences is governed by information structure, following which we have pointed out the resemblances between children's pre-grammatical sentences and adults' pre-grammatical sentences.

Some shortcomings of the 'ellipsis' analysis, which is typically adopted to account for adults' pre-grammatical sentences, were pointed out. We have shown that we do not have to rely on the 'ellipsis' analysis to account for the production of pre-grammatical sentences and that rule 1-1 can explain how they are produced.

To sum up, our system of information structural analysis has proved to be useful not only in analysing grammatical sentences but also pre-grammatical sentences. Perhaps more importantly, the system accounts for the phenomena as diverse as pro-drop, ellipsis, and pre-grammaticality in a unified way, i.e. with just two information structural rules, namely rule 1-1 and rule 1-2.

2.4.5. Summary of 'Analysis Using Variables and Non-variables'

In the first half of this section, we analysed a phenomenon commonly referred to as 'ellipsis', following which echoing, unattended parameters, and pre-grammatical sentences were analysed. The production of all the examples involving these phenomena

has been explained by two simple principles, namely rule 1-1 and rule 1-2. It is notable that phenomena as diverse as these, for each of which separate syntactic, semantic, and pragmatic explanations have been proposed, are explained by a couple of very plain principles. This indiscriminate operation of rule 1 is the advantage which our system has over many previous theories.

2.5. Summary of the Chapter

In this chapter, we have reviewed the notion of ‘old/new information’, with which our notions of parameters and constants were contrasted. We then devised ‘morpho-informational’ units of variables and non-variables on the basis of these notions. These units proved to be of high usefulness as analytic devices which can account for a number of seemingly mutually dissimilar linguistic expressions.

Perhaps more important is the high simplicity and consistency the units bring into the analysis of such linguistic expressions. For example, previous analyses of linguistic expressions involving ‘ellipsis’, subject-less and pre-grammatical sentences typically necessitate separate syntactic, semantic, and pragmatic analytic devices such as ‘pro’, ‘yes-no contrast’, etc. On the other hand, our system of analysis accounts for all of such expressions with two simple principles based on the notions of variables and non-variables. This is an advantage our system of analysis has over a number of existing analyses.

Having demonstrated the usefulness of our system, we now turn to another type of linguistic expression, namely prominence. As we will see in the next chapter, prominence, like the various linguistic expressions which have been discussed in this chapter, has attracted a variety of analyses. Although the analyses invariably point to a certain relationship between prominence and information structure, they differ considerably in the ways they relate them. We will attempt to give an account for the prominence-IS relationship using variables and non-variables.

3. PROMINENCE

3.1. Introduction

The preceding chapter looked at the relationship between information structure and morphemes. In the present chapter, we will discuss the relation between information structure and prominence.

It should be noted in advance that the discussion in this chapter regarding the relationship between IS and prominence will not be comprehensive. We will largely confine our discussion to cases where prominence (the meaning of this term will be explained later) of a part of a sentence relative to the other parts in the same sentence is clearly perceptible by native speakers. That is, we will discuss in detail only sentences where identification of prominent parts is uncontroversial and attracts unanimous agreement among informants⁹³.

This leaves out of our discussion cases where informants' judgments as to where prominence is located or should be placed vary. It turns out that such cases involve prominence placement in sentences with multiple variables as well as some sentences where variables are bound morphemes (function words or predicators). Accordingly, we will not discuss extensively examples in which there are multiple variables or where bound morphemes are variables. However, we will not dismiss them completely, and we attempt to provide, although briefly, possible explanations regarding the uncertainty in prominence placement in such sentences. The discussion here, therefore, can be considered to be a preliminary study which, it is hoped, will suggest a direction for future research.

We will attempt, as we did in the preceding chapter, to connect morphemes with information statuses. A natural extension of this is an attempt to connect morphemes with prominence which we will then associate with information statuses. That is, this chapter explores a possible association between variables and prominence.

Before we proceed further, we would like to repeat a statement from the introductory chapter: no language universals are proposed in this thesis. We therefore have no interest in postulating universals regarding prominence placement in sentences.

3.2. Information Structure and Prominence

The relation between information structure or information statuses and prominence is not a novel topic in the literature of linguistics in English. The popular claim in the literature is that perceptual prominence⁹⁴ or auditory salience in a sentence somehow corresponds to the ‘focus’ or ‘new information’ in the sentence (e.g. Cruttenden 1997:73-87, Selkirk 1995, see also the section ‘Old/New information’ in the last chapter).

A similar sort of claim to this is found also in the linguistics of the languages which are the subjects of this thesis. For example, in their Tajik grammar, Rustamov and Ghafforov (1985:57-78) claim⁹⁵ that the ‘semantic centre’ (*markazi ma’’noī*)⁹⁶ in speech is distinguished by *zadai ta’’kid* ‘stress of emphasis’⁹⁷ (ibid:58, see also ibid:78). They also count separating new information (*xabari nav*) from old information (*xabari kūhna*)⁹⁸ as one of the functions of *zadai ta’’kid*. This description strongly suggests that they have in mind some sort of association between information statuses and prominence. They also claim that melody (*navo*) is used, jointly with other elements⁹⁹ of intonation, for distinguishing known (*ma’’lum*) information from unknown (*noma’’lum*) information¹⁰⁰,

⁹³ Except in the last section where we look at the lack of unanimity among native speakers’ judgements.

⁹⁴ Prominence in a sentence is variously called ‘sentence stress’ (Clark and Yallop 1995:342), ‘(H*) pitch accent’ (Selkirk 1995), or nucleus (in an ‘intonation-phrase’) in the British tradition of intonation description.

⁹⁵ They say that every single sentence has its semantic centre, that intonation is the basic means of specifying semantic centres in speech, and that intonation fulfils this duty with the ‘stress of emphasis’.

⁹⁶ However, they do not explicate the referent of the term, which makes the meaning of the term as vague as that of ‘focus’ or ‘new information’ in many works in English.

⁹⁷ There are other sorts of stress, namely ‘stress of syntagma’ (*zadai sintagma*) and ‘stress of phrase’ (*zadai fraza*). *Zadai sintagma*’s basic function is phonetic formation of *sintagma*, a kind of intonation-group/phrase. For the difference between *zadai sintagma* and *zadai fraza*, see Rustamov and Ghafforov (1985:63-64). They also mention *zadai émfatiki* ‘emphatic stress’ (ibid:59), but no detailed explanation accompanies the term.

⁹⁸ As well as to separate important (*muhim*) information from unimportant (*ghayrimuhim*) information.

⁹⁹ E.g. *tanfis* ‘interval or break’, *tamdid* ‘lengthning’

¹⁰⁰ Also for the expression of ‘logical relations’ (*munosibathoi mantiqi*) in a sentence and the formation of the sentence’s ‘sementico-intonational center’ (*markazi ma’’noiyyu intonatsionī*).

presenting the following set of examples where ‘distinguished’ parts are bold faced (ibid:60, the gloss and translation are mine):

<i>modar</i>	<i>az</i>	<i>duxtar</i>	<i>pursid</i>
<i>modar</i>	<i>az</i>	<i>duxtar</i>	<i>pursid</i>
<i>modar</i>	<i>az</i>	<i>duxtar</i>	<i>pursid</i>
mother	from	daughter	asked.3sg
‘mother asked her daughter’			

Obviously, some relationship between prominence and information statuses is assumed in Rustamov and Ghafforov’s grammar. An earlier Tajik grammar (Fitrat 1930:92) also writes about *zadaji mantiqī* ‘logical stress’ which indicates which word is important in a sentence.

This assumption is shared by Uzbek grammarians who posit *mantiqiy urg‘u* or *logik urg‘u* ‘logical stress’, which appears to correspond to *zadai ta’kid* of Tajik. *Mantiqiy* (or *logik*) *urg‘u*, is explained as the ‘stress which distinguishes a part of a sentence, strengthening/emphasising its meaning’ (Hojiev 1985:52), and works ‘for separating/distinguishing important elements’ (G‘ulomov 1947:28, cited in Oripov and Obidova 1994:50)¹⁰¹.

Thus, the idea of connecting information statuses and some phonetic, intonational, or perceptual/auditory prominence/salience is present in the grammars of two of the languages dealt with in this thesis. The same idea is not difficult to find also in the literature of linguistics in Turkish, Japanese, and English (e.g. Ergenç 1995:25, Kindaichi 1951:580, Mikami 1963:106, Grice 1978:121). As Ladd (1996:160) writes, ‘[i]t is now generally accepted that sentence accentuation [our prominence] reflects – in some way – the intended *focus* of an utterance’.

We also consider it reasonable to assume an association between prominence and particular parts of sentences, whether they are called ‘foci’ or ‘new information’ or ‘semantic centres’. However, there is a great deal of ambiguity in these explanations. The ambiguity derives from the existence of some questions that are not attended to in the explanations, such as: What is ‘focus’ or ‘new information’ or ‘important element’ or ‘semantic centre’? Does it have morphological representation? What is prominence?

¹⁰¹ According to Hojiev (1985:52), *logik urg‘u* is also called *ajratuvchi urg‘u* ‘separator stress’ and *ma’no urg‘usi* ‘meaning stress’.

Does it have phonological or prosodic representation or correlates? (What is the basis for considering, say, the multi-morphemic/lexical *az duxtar* [from daughter] in the example above to be a single ‘semantic centre’?).

The first half of these questions concerns how ‘new information’ is defined and how it is connected with morphemes. This has been dealt with in the first sections of the preceding chapter where we linked information statuses to morphemes, and therefore will not be redundantly discussed here. However, the latter half of the questions concerns the identity of prominence, which we have not discussed before. This will be the first topic to be attended to in the remaining part of this chapter. After the discussion on this topic, we will move on to associate prominence with morphemes.

3.3. Data

All examples in this chapter are, unless specified, taken from our field notes. The data of prominence placement in the examples cited in this chapter were, for the most part, obtained from our informants by asking eliciting questions. Informants were, after providing answers, asked whether the prominence patterns they provided were the only possible patterns. We also often had lengthy discussions with many of my informants about prominence in their languages and how prominence placement alters in different contexts. The data are also supported by our listening to many natural conversations, living in Bukharan Tajik and Turkish households. As for Japanese examples, we are the main source of data.

3.4. Terminology

Before we proceed to the discussion, what we refer to by the terms ‘word’, ‘accent’, and ‘stress’ needs to be made clear.

The term ‘word’ will be used, for a while, with the broadest or, rather, loosest possible definition that covers lexical entries in dictionaries and some ‘clitic groups’ consisting of clitics and their ‘hosts’ to parts of utterances which are variously called ‘intonation-groups’ or ‘intonation-phrases’ and ‘accentual phrases’. This lack of rigour in definition derives in part from the terminological confusion observable in comments on prominence in Japanese, Turkish, Uzbek, Tajik, and English works.

The term ‘stress’ will be used to mean prominence within words, however such prominence is achieved. On the other hand, the term ‘accent’ will be limited to prominence where, within words, pitch is involved. It is therefore equivalent to ‘pitch accent’. This distinction is based on Cruttenden’s (1997:13) distinction and following him, ‘stress’ is used in ‘the more general, less specified, way’. As will be explained in detail in the following subsection, ‘prominence’ will be used primarily to refer to prominence of a sentence. Accordingly, when stress or accent, which is itself prominent, coincides with (sentence) prominence, we may express the combined prominence as prominence/stress or prominence/accent.

In this thesis, the terms ‘stress’ and ‘accent’ are used to mean ‘lexical stress’ and ‘lexical accent’, respectively. On the other hand, prominence will be hereafter used in the sense of the ‘stress’ of a sentence¹⁰². That is, the point of prominence or salience of a word will be called ‘stress’ while that of a sentence will be called ‘prominence’, however the ‘prominence/salience’ is achieved.

Regarding what we call stress here, Fox (2000:115) writes: ‘There has been—and continues to be—disagreement about the phonetic nature of the phenomenon, its phonological role, and the appropriate mode of its description’. Accent is almost equally difficult to identify with phonetic measurements, as is indicated by the existence of phenomena such as the much debated *ososagari* ‘delayed drop’ where the pitch is higher at the unaccented mora following the mora which is phonologically and perceptually accented (see Sugito 1969, 1982:107-125). Identification of the stress of the sentence, i.e. prominence on the basis of phonetic data is at least as troublesome as that of stress and accent. Taking into account the difficulty, we will resort to perception to define prominence, but some explanation of what phonetic, phonological, and prosodic features could be relevant to prominence will be provided first.

The term ‘prosody’ will be used here to refer to linguistic information which involves segmentation of sentences¹⁰³. Prosody yields, most importantly for the following

¹⁰² Prominence in this sense is therefore approximate to ‘nucleus’ in the British tradition of intonation transcription.

¹⁰³ It is probably more reasonable to say ‘segmentation of speech’, rather than ‘segmentation of sentences’, considering that the sentence, an abstract piece of composite in speech, is based on intuition or psychological reality. However, since we have decided to call stretches of speech appearing in the examples sentences, we opt to use the term ‘sentence’ here too.

discussion, units often termed as intonational phrase and accentual phrase. The segmentation, in turn, may be based on phonetic features or their movement or change in time.

3.5. Prominence

The existence of **perceptual prominence of certain parts of a sentence relative to the other parts in the same sentence** has been a topic of discussion for some time, mainly in the literature on intonation. It is in this perceptual sense that we use the term prominence.

Prominence in sentences (or intonational phrases) is often called simply ‘prominence’ (Pierrehumbert and Beckman 1988:Chapter 4) or ‘sentence accent/stress’ (Clark and Yallop 1995:342), ‘intonation peak’ (Kornfilt 1997:505-9) or ‘nucleus’ (Gussenhoven 1983). However, they are defined on different grounds such as prosody, phonetics, and perception.

Prominence as it is used here is of perceptual nature and admittedly lacks rigour of definition in terms of phonetics. However, we think our reliance on perception as the basis for identification of prominence is justifiable for 1) the various sorts of unquantifiability and difficulty involved in production/perception of prominence in sentences and also for 2) our aim to use the notion of prominence in all the languages we take examples from in this chapter. (In addition to this, 3) we are interested in prominence perceived by people, whatever phonetico-prosodic nature it may have.)

In fact, some promising signs of the difficulties mentioned in 1) being tackled and partly solved are emerging in the work of Mary Beckman and her students. The prospect of prominence being defined in phonetico-prosodic terms in a convincing manner seems high in their work, and we will refer to their publications in the following subsection. However, for this thesis, we opt to use our perception-based definition of prominence, because most of the languages we cite examples from are largely outside the scope of their research. This prevents us from exploiting their achievement in analysing prosody (and hence also in prominence production/perception with prosodic cues) in which

language specificity is conspicuous¹⁰⁴. This is, however, not to say that we do not accept any phonetico-prosodic phenomena as being indicative of prominence. Wherever phonetics, phonology, or prosody is indisputably related to perceptual prominence, we will take advantage of the relation in our analysis.

In auditory terms, we see prominence as a product of certain phonetic and prosodic effects imposed on utterances. Brazil's (1997:19) explanation of English intonation, in which syllables are assigned prominence, can serve as a fairly straightforward exemplification of prominence in this sense:

‘Word accents’ are an automatic consequence of the assembly of particular lexical items:

- i. I shall 'try to 'see 'Tom and 'Mary

Prominent syllables are next assigned in one of the possible ways [...]:

- ii. i shall TRY to 'see 'tom and MARY
- iii. i shall 'try to SEE 'tom and 'mary

etc. [...]

Clearly, prominence is considered to be produced by the (super-)imposition of certain effects on a sentence, in which there are lexical items with their inherent stresses or accents, hence the presence of both the ‘word-accent’ (our stress) symbol ['] and prominence-indicating capitalisation in the sentences shown in (ii). In other words, Brazil's approach recognises the operation of at least two different phonetic-controlling components, namely ‘word accentuation/stress assignment’ and ‘prominence assignment’¹⁰⁵. This is also how we recognise prominence assignment in this chapter.

3.6. Overview of Possible Phonetico-Prosodic Nature of Prominence

Our prominence is based on perception, but, as stated in the immediately preceding section, we will take into consideration phonetic and prosodic cues which prominence may have. Accordingly, in this section, we review some comments on prominence that

¹⁰⁴ See, for instance, the difference observed between the two grammatically and syntactically close languages, Korean and Japanese in Vendetti et al (1996:309-310).

¹⁰⁵ Some works assume the existence of some other components, such as Fujisaki and Hirose's (1984) ‘phrase control mechanism’ (see also Ladd 1996:24-30).

are particularly relevant to the languages we are taking examples from in this thesis. The difficulties and problems involved in the identification of the phonetico-prosodic nature of prominence will also become apparent in the course of the overview.

Let us start with a review of comments on prominence in Turkish. Kornfilt (1997:505) writes the following regarding the example below where the ‘intonation peak’ (which seems to correspond to our prominence) is associated with what she calls the ‘preverbal constituent’ *ıstakoz* ‘lobster’ and assigned to a syllable in it, thus: ‘[...] the intonation peak will be located on whichever syllable carries primary word stress for that preverbal constituent’:

Hasán bugün	ıstakó ^z	ye	-di
Hasan today	lobster	eat	-Past

“Hasan ate (a) lobster(s) today”

This explanation is thus another exemplification of the prominence-imposition approach. Note that the prominence is associated first with the word *ıstakoz* ‘lobster’ and then assigned to the syllable *koz*.

An observation of this analysis brings about two crucial questions regarding prominence in Turkish: 1) what is indicative of prominence? 2) what is the unit with which prominence is associated?; Prominence is clearly assigned to the syllable *koz*, but what is the ‘constituent’ with which prominence is associated? If it is *ıstakoz*, on what basis is it chosen as a prominence-associated ‘constituent’? In other words, on what basis are sentences parsed into units which Kornfilt calls ‘constituents’? As we will see, these are, in fact, two interlocked questions.

3.7. Phonetic Nature of Prominence

Let us attend to the question 1 first. In Kornfilt’s analysis, how prominence (or intonation peak) of the syllable *koz* is achieved is not entirely clear. The impression her explanation gives is that the phonetic property of the syllable is somehow affected by being assigned prominence and made perceptually more prominent than it would be when it is not assigned prominence. However, it is unclear as to what property of the syllable is affected.

In fact, how prominence is brought about is not fully understood for most languages including the ones we deal with in this thesis. Some phonetic and prosodic features have been proposed as defining features of prominence in these languages, but the propositions are not necessarily compatible with each other. This variance among suggested phonetic and prosodic features of prominence is reflected in the lack of correspondence among the variety of comments on prominence in these languages. Let us first review such comments on prominence in Turkish and Uzbek.

In some works, prominence is discussed with little reference to its phonetic or prosodic properties. For example, in (Erguvanlı 1984:34) and (Göksel and Özsoy 2000:3) in which prominence is called ‘sentence-stress’ and ‘stress’, respectively, phonetic or prosodic properties which prominence may have are paid little or no attention.

There are also works in which some cues for identifying prominence are proposed, though the proposed phonetic or prosodic cues usually contain a certain amount of abstraction. For example, Ergenç (1995:25) associates prominence with *baskın ton* ‘lit. oppressive/powerful tone’, the meaning of which is obscure, and a pause preceding or following the syllable with prominence. Banguoğlu (1959:120) says *cümle vurgusu* ‘sentence-accent’ forms an ‘intensity peak’ (*yeğinlik doruğu*) higher than the stresses of ‘words’ (*kelimeler*) and ‘phrases’ (*öbekler*), but does not provide an explanation as to what it means for an ‘intensity peak’ to be ‘higher’ than ‘stresses’. On the other hand, Kornfilt talks about ‘intonation peak’. If we assume she means a distinctive usage of pitch (Kohler 1977:126)¹⁰⁶ by the term intonation, then she is probably associating prominence with pitch. We then have all of intensity, pause, and pitch as the proposed possible candidates for the properties of prominence. This wide variance among possible explanations of the phonetic nature of prominence naturally discourages us to identify prominence in Turkish using phonetic measurements.

Comments on the phonetic nature of prominence which do not necessarily agree with one another are found in Uzbek research too. Nurmanov (1990:14) talks about the ‘intonational centre’ (*intonacionnyj centr*), while G‘ulomov (1947:28 cited in Oripov and

¹⁰⁶ [...] *die distinktive Verwendung der Tonhöhe auf Wort- und Silbenebene* [...] (Kohler 1977:126). See also Edwards (1993:25-27). Some authors include other measurements such as ‘the fundamental frequency/pitch, intensity/loudness, and temporal structuring of utterances’ in the concept of intonation (Altman et al 1989:2).

Obidova 1994:50) associates prominence (i.e. *mantiqiy urg‘u*: see the section ‘Information structure and prominence’) basically with tone/melody (*ton/ohang*), and Oripov and Obidova point out the involvement of pause (*pauza*), saying ‘in general, a short pause occurs after *mantiqiy urg‘u*’ (ibid:49)¹⁰⁷.

Thus, there is disagreement among scholars as to what are correlates of prominence in Turkish as well as in Uzbek. However, none of the comments cited above are not supported by empirical evidence. One might, then, expect that an instrumental study on prominence will provide unarguable evidence of a certain phonetic feature or a set of phonetic features being indicative of prominence. Contrary to this expectation that instrumental studies will provide a clear picture as to how prominence is produced, they instead reveal the enormous difficulty which exists in defining prominence in phonetic terms.

There are a few instrumental studies on prominence in Turkish. Konrot (1991), in one of his instrumental studies, shows that change in fundamental frequency (and hence pitch contour) is more consistent with the presence of stressed words (*vurgulu sözcükler*) than acoustic amplitude (and hence intensity) is. Analysis of pitch (movement), then, seems to be a promising direction for the search of the phonetic nature of prominence. Note also that any of the terms ‘pitch’, ‘tone’, and ‘intonation’ appear in most of the comments cited above – this may suggest pitch or pitch movement’s being an indispensable element in prominence production/perception.

However, as Konrot says (ibid:31), locating prominence is certainly not simply the matter of locating the highest pitch in a sentence. One only has to be aware of downdrift/downstep (Kawakami 1977:103-104 cited in Vance 1987:78-79, Clark and Yallop 1995:337), a gradual declination (Cohen and ‘t Hart 1967:184 cited in Ladd 1996:16) of pitch in the course of an utterance to realise this. Some obvious examples of the lack of correspondence between the highest pitch in a sentence and intended prominence production/prominence can be observed in the tables of pitch contour in Selen (1973:40-2).

¹⁰⁷ They also say that the ‘word [*so‘z*] on which *mantiqiy urg‘u* falls is pronounced with such means as raising [*/making loud*] and lowering of voice/sound’ (Oripov and Obidova 1994:50), but no explication is provided in their work regarding this comment.

Thus, the ‘highest pitch = prominence’ assumption is not consistent with reality. Therefore, it has to be particular patterns in which pitch shifts that are characteristic of prominence. What sorts of pitch movement, then, are indicative of prominence in Turkish or Uzbek? Konrot (1991:31) makes an interesting comment on this matter. Having confirmed that ‘it is difficult to say that stressed [corresponds to our ‘prominent’] words have higher fundamental frequency in comparison with unstressed words’, he says a prominent word has a pitch that rises towards the end of the word, unless the word is at the end of a sentence, in which case pitch falls following the rise to mark the end of the sentence. Unfortunately, this observation is difficult to generalise, because the ‘words’ in his examples have stresses on their last syllables and the rise of pitch can be explained by this word-stressing (as Konrot himself claims, pitch is the most efficacious of the phonetic correlates of Turkish word-stress¹⁰⁸). That is, pitch rising towards the ends of ‘words’ will occur whether or not prominence is associated to them.

The discussion above, then, shows that an instrumental study does not necessarily provide a clear-cut solution to the problem of identifying or defining prominence in an objective way in Turkish. Attempts to identify prominence in Japanese by its phonetic nature can involve even greater difficulty than those in Turkish. In Japanese, an archetypal non-stress language, pitch has exclusive importance in word-accentuation (Beckman 1986). As for prominence, a number of scholars appear to consider that pitch movement is also the most efficacious of phonetic/prosodic features of the production/perception of prominence in Japanese, (Ōishi 1959, Kori 1989, Sugito 1980), although there may be minor involvement of duration and intensity (Miura 1992). However, the mismatch of the highest pitch in a sentence and prominence is the case in Japanese as it is in Turkish.

Japanese word-accentuation is, as the term itself suggests, pitch-determined. The fact that both accentuation and prominence exploit the same phonetic attribute means that both word-accent and prominence are mapped onto the same measurement, i.e. pitch movement, to form a single pitch contour. It may then be expected that prominence in a sentence manifests itself with distinctive pitch movement without distorting accentuation. It may be also expected that identifying prominence by examining pitch

¹⁰⁸ He also counts the width of the gap in amplitude between a stressed syllable and the following syllable as indicative of stress (Konrot 1991:26).

movement would not be a straight-forward process, because one has to isolate shifts in pitch contour which prominence may induce from those which word-accent cause.

What all this suggests is that, as in Turkish, it is probably feasible to think that pitch movement is the most reliable indication of the presence/location of prominence in Japanese. Accordingly, the question to be asked regarding identifying prominence using objective measurements in Japanese is also the same as the one asked for Turkish: What kind of pitch movement is indicative of prominence?

Let us review some analyses of prominence which attempt to identify/determine prominence in terms of pitch movement. We review the analyses primarily to show the elusiveness of perceptual prominence that exists despite attempts to identify it in terms of phonetics (or pitch) or prosody. The review is also for justifying our use of perception as the criterion for defining prominence.

3.8. Review of Previous Works

Most of the works introduced in the preceding section referred to phonetics to explain the nature of prominence. They explicitly or implicitly posit a unit to which prominence is associated — prominence is typically associated with a unit (following which prominence is assigned to particular syllables in the unit), so that word-accentuation or word-stress which takes place within the unit is not distorted by prominence (see, for example, Kawakami 1957, Kindaichi 1967:95-98, Abe 1998:370). The effect of the assignment of prominence on the stressed syllable is then explained in a variety of ways, e.g., rise in pitch for Japanese and high pitch, length etc. for Tajik. However, as we have seen in the preceding subsection, pitch movement is usually considered to be the most efficacious of all the phonetic attributes of prominence.

Kornfilt's analysis cited before may serve as an example of this view that there are prominence-associated units in which there are prominence-assigned syllables. In her example, *istakoz* 'lobster' is the unit with which prominence is associated, in which unit the syllable with prominence, i.e. *koz*, occurs.

However, in most works that adopt this approach to prominence, the unit with which prominence is associated is far from well defined. In Kornfilt's example *Hasan bugün istakoz yedi* cited before, the unit with which prominence is associated is the word

ıstakoz ‘lobster’. The prominence is then assigned to ‘whichever syllable carries primary word stress for that [...] constituent’, which is in this case *koz*. At first glance, this explanation appears entirely reasonable. However, when we re-read this passage, we find some unclarity in the explanation: why is specifically *ıstakoz*, rather than *koz* or *bugün* *ıstakoz* or *o* or some other unit, associated with prominence?

A simple answer to this question would be ‘because *ıstakoz* is a ‘word’’. If we assume that a noun constitutes a word, this answer adequately explains how prominence-associated units are determined in Kornfilt’s example. However, what a ‘word’ is in Turkish is a debatable issue. The very articulation that makes this explanation a feasible one, namely the equation of a noun and ‘word’, makes it difficult for us to explain why, for example, *sınava* in the following example from Ergenç (1995:26) constitutes a prominence-associated unit (slashes and underlines signify pauses and prominence, respectively):

Ahmet/ yarın önemli bir sınava girecek.

Ahmet yarın/ önemli bir sınava girecek.

Ahmet yarın/ önemli bir sınava girecek.

Ahmet yarın önemli bir /sınava girecek.

The gloss of this example follows:

[53] *Ahmet yarın önem-li bir sınav-a gir-ecek.*

Ahmet tomorrow importance-with one exam-dat enter-fut.3g

In this example, the units with which prominence is associated are not ‘words’. The unit *sınav-a*, which is associated with prominence¹⁰⁹ is a combination of a noun and a case marker. Why is it not only *sınav* ‘exam’ or only *-a* [dat], that is associated with prominence?

Interestingly, this parsing¹¹⁰ pattern is observed widely in other languages which are the subjects of this thesis. For example, in the following chart taken from Oripov and Obidova (1994:50), among the bold faced units (which they loosely term, as Ergenç does with his Turkish examples, *so ‘z* ‘word’) with which prominence is associated, only

¹⁰⁹ ‘the highest point of the sentence’s melody’ in Ergenç’s terminology (*Tümce nin ezgisi, odaklama yapılan bilgi öbeğinde en yüksek noktaya ulaşır.* (Ergenç 1995:25)).

¹¹⁰ In this chapter, the term ‘parsing’ does not signify ‘formally analysing the syntactic structure of a sentence’ but simply means ‘dividing sentences into units’ with no syntactic connotation.

bahor ‘spring’ and *ko‘k* ‘blue’ are single morphemes. All the other units consist of multiple morphemes, for example: *quyosh-i-ning* [sun-iz-gen], *nur-lar-i* [light-pl-3sg], *dengiz-ga* [sea-dat]. Their comment accompanying the chart is: ‘reading any one of the words (*so‘zlar*) in this sentence with [lit. giving] the *mantiqiy urg‘u* [prominence in our terminology; see the section ‘Information structure and prominence’] is possible’¹¹¹, but what exactly is the ‘word’?:

Bahor quyoshining nurlari ko‘k dengizga sig‘may toshadi.

Bahor **quyoshining** nurlari ko‘k dengizga sig‘may toshadi.

Bahor quyoshining **nurlari** ko‘k dengizga sig‘may toshadi.

Bahor quyoshining nurlari **ko‘k** dengizga sig‘may toshadi.

Bahor quyoshining nurlari ko‘k **dengizga** sig‘may toshadi.

Bahor quyoshining nurlari ko‘k dengizga **sig‘may** toshadi.

Bahor quyoshining nurlari ko‘k dengizga sig‘may **toshadi**.

The gloss for the above is provided below:

[54] *Bahor quyosh-i-ning nur-lar-i ko‘k dengiz-ga sig‘-ma-y*
 spring sun-iz-gen light-pl-3sg blue sea-dat fit in-neg-ger
tosha-di.
 overflow-past

‘The light of the sun of spring overflowed, not being contained in the blue sea.’

Surprisingly, the kinds of parsing resembling these ones are widespread also in the literature of Japanese prosody. For example, Akamatsu (1997:236-237) says when the noun *kōchō* ‘headmaster’ in the following sentence is made to stand out, the whole of the noun plus the nominative case marker *kōchō-ga*¹¹² [headmaster-nom] is ‘pushed up’ in pitch:

*kōchō-ga kinō sendai-e ik-i-mashita*¹¹³
 headmaster-nom yesterday Sendai-all go-infinitive-past indicative
 ‘the headmaster went to Sendai yesterday’

Why is prominence associated with the unit *kōchō-ga* and not *kōchō* or *kōchō-ga kinō*?

What is the motive for this way of parsing of sentences?

¹¹¹ *Bu gapdagi so‘zlarning har biriga mantiqiy urg‘u berib o‘qish mumkin.*

¹¹² [ko- [o-çço-o-ŋa]] in Akamatsu’s phonetic transcription.

¹¹³ Akamatsu’s [ko [o çço o ŋa] k’ i [no o [se] n da i e i [k’ i ma] çj ta].

Among traditional Japanese grammarians as well as a number of engineers working on computational utterance parsing (e.g. Takeda and Ichikawa 1991, Yamamoto et al 1992), this parsing of sentences into units consisting of ‘a noun plus any following postposition’ (Beckman 1996:41), despite the apparent lack of empirical motivation for it¹¹⁴, has popularity as well as currency. The units this parsing yields are called *bunsetsu* which Martin (1975:156) sees as a ‘basic phonological phrase, i.e. accent phrase’ while Beckman (1996:41) sees as ‘a syntactic grouping that corresponds roughly to a minimal NP’. Like Akamatsu, Kindaichi (1951:578) also associates prominence with this unit. Thus, the parsed units that are claimed to be associated with prominence may or may not be single morphemes in the Turkish, Uzbek, and Japanese studies reviewed above. Having reviewed some popular parsing patterns in Turkish, Uzbek, and Japanese which show curious resemblance to each other (such as the production of the units in the ‘noun plus postposition’ structure¹¹⁵), the questions arise as to what the basis of the parsing is. We will attend to this question in the next subsection.

3.9. Basis of Parsing

Is the kind of parsing explained in the preceding subsection based on prosody or perception or syntax? One would expect it to be based on a prosodic (or phonetic or phonological) basis since this way of parsing yields units with which prominence is associated. Therefore, it is also expected that there are objective criteria according to which the parsing is executed, and also that the criteria are based on quantifiability. However, Kornfilt, Ergenç, Oripov and Obidova, and Akamatsu do not provide an explanation of what criteria they use in this parsing.

In some of the examples cited above, the parsing looks syntactic, while in some others it looks phonological. For example, some of the parsed units seem to correspond to some extent to domains of the operation of certain phonological rules such as stress assignment, vowel harmony, and accentuation (see, e.g. Clarke and Hamamura 1981:8).

¹¹⁴ See Tsukishima (1964:103) for an explanation of *bunsetsu* in the traditional Japanese grammar.

¹¹⁵ Prominence-associated units in Tajik grammars also shows resemblance to those explained above. See the set of examples from Rustamov and Ghafforov (1985:60) cited in the section ‘Information structure and prominence’ where the preposition *az* ‘from’ is grouped together with the following noun *duxtar* ‘daughter’.

Some others resemble the unit Nespov and Vogel call ‘clitic groups’ which consist of clitics and their hosts.

One thing that the parsed units have in common is that they are units that have only one stress or accent¹¹⁶ whose location is generally considered by the parsers to be pre-determined in them. For instance, in the above Japanese example, the units with which Akamatsu associate prominence seem to coincide with the domains of ‘lexical’ accentuation. Explanation of the domains of ‘lexical’ accentuation in Japanese follows.

I will not go into the details of the accentuation system of Japanese, but instead adopt McCawley’s (1977) account (see also Haraguchi (1977) for a largely similar account within Autosegmental Phonology). McCawley proposes the following¹¹⁷ (the square brackets in (ii) were added by me):

- i. Make everything high pitched.
- ii. Make everything after a ['] low pitched.
- iii. Make the first syllable low pitched if the second is high pitched.

The symbol ‘’ in (ii) represents the fall in pitch, which is lexically pre-determined. Some examples of accent patterns the rules yield are taken from McCawley (1977:261):

ma'kura ga	ma kura ga	[pillow nom]
koko'ro ga	ko ko ro ga	[heart nom]
atama' ga	a a tama ga	[head nom]
sakana ga	sa sa kana ga	[fish nom]

¹¹⁶ This is a fall in pitch in Japanese signified in Japanese ToBI with H*+L (see Vendetti 1997). See McCawley (1977) for an explanation. Haraguchi (1977) provides a largely similar account to McCawley’s within Autosegmental Phonology. Complications outside the scope of the general rules explained in these works (e.g. pre-accenting morphemes such as *-na* [neg], *-sase* [caus], etc.) shift accent to the preceding mora in certain words (Shibata et al 1980, Vance 1987). To further complicate matters, there is also such a phonetic phenomenon as the much debated *ososagari* ‘delayed drop’ where the pitch is higher at the unaccented mora following the mora which is phonologically and perceptually accented (see Sugito 1969, 1982:107-125).

¹¹⁷ The rules may be paraphrased and find somewhat different representation in some other works. For example, another description of Japanese accentuation may posit the following as the accentuation rules in Japanese: 1) The pitch of the first mora and that of the second always differ, i.e. if the first is high, the second must be low, and vice versa. 2) Once the pitch falls, it will not rise in the following morae.

According to the terminology popular in the western literature of Japanese phonology, the first three in which there are falls in pitch (shown with the symbol “”) are accented, while the fourth where no fall in pitch occurs is unaccented. Thus, in Japanese, whether and where the pitch falls bears exclusive importance in accentuation.¹¹⁸

(Adopting another popular convention in the western literature of Japanese phonology, I will mark the fall of pitch with the symbol for acute accent, ‘’’. Accordingly, the first four of the above examples will be shown as *mákura ga*, *kokóro ga*, *atamá ga*, and *sakana ga*, respectively, in this chapter. Morae with the acute accent symbol will be called accented morae.)

As is clear from the third of the examples above, the pitch fall can take place after a noun and before a postposition. This means that one has to observe the pitch movement of the set of a noun and the following suffix or postposition to analyse accentuation.¹¹⁹

Otherwise, for example, the accentual difference between the two nouns shown below could not be fully accounted:

<i>haná ga</i>	ha ^{na} ga	[flower nom]
<i>hana ga</i>	ha ^{na} ga	[nose nom]

Because of this accentuation system, if one assumes that prominence does not distort accentuation, one is probably inclined to adopt the parsing exemplified above. This is probably why Akamatsu’s prominence-associated units seem to coincide with the domains of accentuation, which is ‘with at most a single rise and at most a single fall in pitch’ (Vance 1987:102).

The same logic seems to be present in the Turkish and Uzbek parsing explained before. In the Turkish example, for instance, the units such as *sınav-a* [exam-dat] or *gir-ecek* [enter-fut] are commonly considered to be domains of stress assignment rules in each of which there is a single stressed syllable. Let us observe an example taken from Demircan

¹¹⁸ Note, however, that phonetic reality does not always coincide with the clear phonological pitch distinction of high and low. There is such a phonetic phenomenon as the much debated *ososagari* ‘delayed drop’ where the pitch is higher at the unaccented mora following the mora which is phonologically and perceptually accented (see Sugito 1969, 1982:107-125, Hasegawa and Hata 1995).

¹¹⁹ The genitive case marker *no* behaves differently from the other postpositions in accentuation. See Arisaka (1941:356) for details.

(1975:333, the symbol indicating stress is changed from ‘˘’ in Demircan to ‘˙’. The glosses and translations are mine):

<i>sés</i>	[sound]	‘sound’
<i>ses-síz</i>	[sound-without]	‘silent’
<i>ses-siz-lík</i>	[sound-without-nmlz]	‘silence’
<i>ses-siz-liğ-in</i>	[sound-without-nmlz-gen/2sg]	‘silence’s’ or ‘your silence’
<i>ses-siz-liğ-in-í</i>	[sound-without-nmlz-3sg/2sg ¹²⁰ -acc]	‘your/his/her/its silence (acc.)’

As Comrie (1997:896) writes, ‘in regularly stressed, non-compound words, main stress falls on the last syllable’¹²¹. The units shown above, which Comrie would call ‘words’, are all domains of the application of a stress assignment rule and each of them has a single (primary) stress. This is the case with units in Ergenç’s example, e.g. *sınav-á*, *gırecék*.

Thus, Ergenç’s parsing of the previously cited Turkish sentence can be understood as being in accordance with units in which stress assignment rules¹²² are in operation, just like the parsing by Akamatsu of the Japanese sentence is done according to domains of the operation of accentuation rules. The parsing of the Uzbek example can also be accounted by the fact that the units the parsing yields are domains of stress rules (see examples in Oripov and Obidova (1994:48) that resemble Poser’s (32)). This is also the case with Tajik (see Rustamov and Ghafforov 1985:52-56).

Now we have the most likely answer to the first of the questions posed at the end of the preceding section. We can safely assume that Kornfilt, Akamatsu, Oripov and Obidova, and Ergenç’s parsing is motivated by phonology, or more specifically, accentuation and stress rules. The most likely answer to the question regarding the basis of the popular parsing explained in the previous subsection is now obtained. However, the second question remains unanswered. Is associating prominence to the units the parsing yields justifiable? This is discussed in the next subsection.

¹²⁰ When the [3sg] suffix *-I* is followed by another suffix, *n* appears between them, making the resulting succession of *In* indistinguishable from the [2sg] suffix *-In*.

¹²¹ There are a large number of non-regularly stressed words as we will see later.

¹²² For a brief introduction to Turkish word-stress, see Demircan (1975, 1976, 1977), Potapova (1978, 1982), or Tekin (1991). For a general explanation of non-final word stress, see Sezer (1983).

3.10. Unit with Which Prominence is Associated

The style of parsing where boundaries of units largely coincide with those of domains of operation of accentuation or stress assignment is thus preferred by a number of scholars. The parsing is adopted by a number of Japanese, Turkish, Uzbek, and Tajik linguists alike. Why the popularity of this particular way of parsing? The popularity derives probably from the fact that this parsing prevents stress and accent from being distorted by super-imposition of prominence onto them.

If such parsing is not adopted, prominence may be associated directly with the postposition *ga* in the Japanese example *haná ga* [flower nom] above. If one assumes that assignment of prominence involves raising of pitch, association of prominence with *ga* could raise the pitch of *ga* which should be lower than the preceding mora *ná*, thus blurring the accentual distinction between *haná ga* [flower nom] and *hana ga* [nose nom]. This may result in obscuring the meaning which the sequence of the three morae represents and appears to be considered not desirable by scholars who adopt the parsing based on the domain of the application of accentuation or stressing rules.

Thus, the equation of the prominence-associated unit with the domain of the application of accentuation or stressing rules can be understood as a way of reconciling prominence and accentuation/stressing – prominence is associated with a domain of accentuation/stressing rules so that it is assigned to a particular syllable/mora in the domain in such a way that the assignment does not distort the ‘pre-determined’ accent/stress. However, is associating prominence to domains of the operation of accentuation/stressing rules justifiable? The answer appears to be negative. Because there exist cases where perceptual prominence is not on syllables which are predicted to receive prominence by accentuation/stressing rules. Let us observe such cases in Japanese.

Ōishi (1959) discusses cases where prominence is assigned to morae not in accordance with lexically determined accentuation (which, of course, is the basis of the previously explained popular way of parsing). Among the examples he cites, there many kinds of examples of prominence distorting accentuation.

Let us observe some examples¹²³. Recall that, as is explained in McCawley's explanation of Japanese accentuation cited before, the Japanese accentuation rules requires all morae after the accented mora to have the low pitch. The pitch movement which the accentuation rules predict are shown in the right column. In the left column, Ōishi's examples are shown (see also examples in Vardul 1989:245):

ké npoo	WA	[constitutional law top]	ké npoo wa
ná nika	NI	[something dat]	ná nika ni

Clearly, in these examples, prominence is associated not with domains of accentuation rules and actually distorts accentuation by being associated with (and also assigned to) the postpositions. There is also empirical evidence for the rising of pitch which postpositions with prominence induce. Muranaka and Hara's (1994:397) instrumental study of prominence records rising of pitch at the postpositional topic marker *wa* with prominence despite the low pitch assigned to it by accentuation rules.

Thus, Ōishi and Muranaka and Hara's observations suggest that the answer to the second of the questions asked at the end of a preceding section is likely to be negative – associating prominence with domains of accentuation is, at least in Japanese, not an adequate way in which to analyse prominence. (There is further evidence for the inadequacy of this approach – pitch contours of sentences often do not support the parsing of sentences into domains of accentuation. See Vardul's (1989) discussion of what he calls 'phrasemes'.)

The implicit assumption behind the approach to prominence reviewed above is that prominence manifests itself using the same phonetic qualities which stress/accent exploits. Indeed, as we have seen before, the likelihood of prominence exploiting pitch (movement) which is also used for accentuation and stressing, seems irrefutable. However, this does not necessarily eliminate the possibility of prominence exploiting phonetic or prosodic attributes which it does not share with stress or accent. Prominence can and does exploit such attributes.

¹²³ In Ōishi, rise and fall in pitch by accentuation are indicated respectively with the symbols [and] (which will be indicated as the acute accent mark over the accented morae here), while those induced by prominence are indicated with the same symbols with a line above them (these will be shown with capital letters in this chapter).

For example, if we do not take for granted the popular way of parsing explained before, there is at least one prosodic attribute which prominence can exploit and accent/stress cannot, namely parsing itself. Parsing, i.e. how sentences are divided into units, itself may contribute to production/perception of prominence.

Beckman and her colleagues'¹²⁴ approach to prominence is in sharp contrast with that in the works explained above. In their works, sentences are divided at the points where what they call 'boundary tones' are present. The units divided by the 'boundary tones' are then called 'intonational phrases' or 'accentual phrases' (see Vendetti et al 1996 and Vendetti 1997). There are a number of novelties in their approach, one of which is its exclusive use of pitch movement as the criterion of primary importance for parsing. Unlike the approach that equates prominence-associated units with domains of accentuation/stress rules, in their approach, there are no pre-determined units that exist regardless of the presence or absence of prominence. Rather, prominence also takes an active part in determining the boundaries of units. As a consequence, prominence alters prosodic structures of sentences.

For example, Vendetti et al (1996:308-309) present an example from Chonnan Korean where a unit starts with a postposition, dividing the sequence of noun and postposition which would be considered to form a unit according to the previously introduced approaches. In other words, a boundary tone divides a domain of operation of accentuation rules. They also say that this kind of parsing 'can easily occur in Japanese, too'.

Vendetti et al do not say that the postposition is associated with or assigned prominence, but say the parsing puts 'the postposition first in the phrase [unit], where the peak of the phrase's demarcative accent will fall on it'. This appear to mean that, taking one of Ōishi's examples as an example, $\overset{na}{nika} \overset{NI}$, which is a domain of operation of accentuation rules and hence would be considered to form the unit $\{nanika\ ni\}$ according to the previously explained approach, may have a unit boundary running through it, thus: $\{nanika\} \{ni \dots\}$.

The analysis by Vendetti et al, then, suggests the inadequacy of the association of prominence with *bunsetsu* or the domains of the operation of accentuation rules at least

¹²⁴ See the web sites of the Ohio University Linguistics Department (<http://www.ling.ohio-state.edu/phonetics>).

in Korean and Japanese. Vendetti et al thus suggest, more clearly than Ōishi does, the inaptness of the kind of parsing employed in the theories introduced in 3.8. Moreover, their work suggests the need for prosodic analysis of sentences, i.e. how sentences are parsed, for identifying prominence.

Is identification of prominence, though, possible only with analyses of pitch and parsing? Analyses of pitch and parsing are tasks with enough complexity to make identification of prominence difficult, but there is a possibility that we need to look at more diverse attributes for prominence-identification.

Takeda and Ichikawa (1991) classify prominence into various categories (ibid:387), following which they identify phonetic attributes that bring *bunsetsu* or specific morae to prominence (ibid:389-390). Their study shows that there are considerable individual differences/preferences in speakers' choice of phonetic attributes for producing prominence. It is also worth mentioning that length, which is not discussed in the works introduced so far, is claimed by them to be the salient characteristic of (consciously produced) prominence.

Having reviewed different approaches to prominence, one thing becomes clear: identification or production of prominence is a confusingly complex matter, which defies attempts to account for it in simple phonetic or prosodic terms.

3.11. Summary

Identification of prominence on a quantifiable basis is thus a complicated matter. The prosodic analysis of prominence by Vendetti et al appears to offer better prospects for prominence identification than many other analyses do. However, not enough research in their framework is done to enable one to make any definite statement regarding the relationship between parsing and prominence in Japanese. As for prominence in Turkish, Uzbek, and (Bukharan) Tajik, such studies are almost non-existent.

This lack of enough research on this topic makes it unrealistic to set up a definition of prominence based on quantifiable measures. This means that, if we employ a phonetico-prosodic definition of prominence, inclusion of ambiguity and unquantifiability in the process of identifying prominence would be inevitable. Moreover, the ambiguity appears to be greater than that which a perception-based definition brings about.

Therefore, as we stated before, we resort to perception in defining prominence and employ the definition of prominence: **perceptual prominence of (a) certain part(s) of a sentence relative to the other parts in the same sentence**. What the term ‘part’ in this definition represents is the topic of the next section.

As has been stated before, the review in this section was, in part, for justifying our use of perception as the criterion for defining prominence. Having done this, in the next section, we will attempt to associate morphemes, or more specifically, variables, with prominence.

Our attempt in the following section, then, has an implication for our overall argument of this thesis – if prominence, which is linked to information structure/status (see the section ‘Information structure and prominence’), is associated with variables, the ‘prominence-variable’ association would serve as supporting evidence for our theory where we attempt to link information statuses with morphemes.

3.12. Prominence – Morpheme

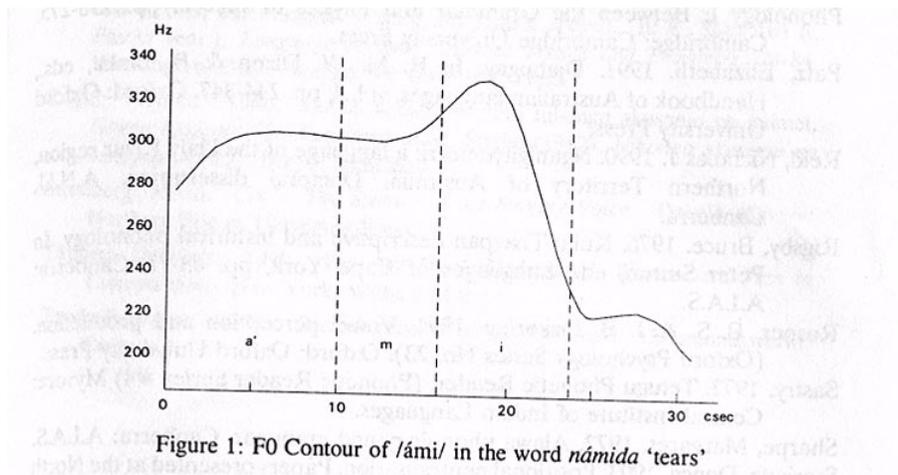
We have confirmed in the preceding section that there are problems in associating prominence with units determined according to rules of accentuation or stress assignment. What, then, do we associate prominence with? The subject of this thesis is to connect morphology and information structure – naturally, the units we shall attempt to associate with prominence are morphemes. Since we have linked morphemes to information statuses in the last chapter, our attempt to link prominence with morphemes is at the same time an attempt to link prominence, morphemes, and information statuses to one another.

In this section, we attempt to associate prominence with a particular kind of morphemes, namely variables. As we shall see the association very often (but not always) induces also assignment of prominence to the variables. In such cases where variables are assigned prominence, we will find it appropriate to indicate the prominence by capitalising the variables (not particular syllables in them), as in the following example copied from Ladd (1996:192):

Eski müdür bir KİTAP yazdı.

‘The former director wrote a book.’ (lit. former director one BOOK wrote)

This way of indicating prominence also has practical value. It allows us to compare prominence in different languages on the same (perceptual) basis, abstracting differences in prominence production in the languages. By adopting this way of transcribing prominence, we also gain the advantage of avoiding unnecessary complications such as transcription of *ososagari* (delayed drop) in Japanese. (When *ososagari* occurs, for example, the fall in pitch in *námida* ‘tears’ can occur in the second syllable *mi* although in perception it occurs in the first syllable *ná*. This figure is taken from Hasegawa and Hata (1995:143).)



3.13. Prominence – Variables

We attempt in this section to link prominence with variables. This ‘link’, however, does not necessarily assign prominence to variables – it only associates prominence with variables.

The claim we make in this section will be that variables are associated with prominence, although they may not be assigned prominence. In other words, variables are prominence-associated units which may or may not bear prominence. We will see examples where variables are prominence attractors but are not necessarily prominence-bearers in the following subsections.

3.13.1. Variable = Free Morpheme

Firstly, observe the following simple Turkish question-answer pair:

- [55] a *ne öl-dü?*
 what die-past
 ‘what died?’
- b *köpek öl-dü.*
 dog die-past
 ‘a dog died’

The constants and parameter for B can be tabulated as follows:

Action:	Constant	Dying
Tense:	Constant	Past
Action performed:	Constant	Positive
Performer of action:	Parameter	> Dog
etc.	Constants	

Obviously, the variable in (b) which specifies the value for the parameter ‘performer of action’ as [dog] is *köpek* ‘dog’. We indicate variables with bold faced letters, accordingly, the information structure of (b) can be shown thus:

- b ***köpek*** *öl-dü.*

In the above dialogue, the prominence of (b) has to be assigned to *köpek*, thus:

- b ***KÖPEK*** *öl-dü.*

In this example, we can see a coincidence of variable and prominence. This coincidence is found also in Uzbek, Bukharan Tajik, and Japanese. In the following translations of the above Turkish example, prominence is assigned necessarily to the words for ‘dog’:

- | | | | | | |
|----------------|----------------|-----------------|---|---------------------|--------------------|
| [56] Uzbek | <i>nima</i> | <i>o‘l-di?</i> | — | <i>IT</i> | <i>o‘l-di.</i> |
| | what | die-past.3sg | | dog | die-past.3sg |
| Bukharan Tajik | <i>či</i> | <i>murt?</i> | — | <i>KUČUK</i> | <i>murt.</i> |
| | what | died.3sg | | dog | died.3sg |
| Japanese | <i>nani ga</i> | <i>shin-da?</i> | — | <i>INU</i> | <i>ga shin-da.</i> |
| | what nom | die-past | | dog nom | died-past |
| | | ‘what died?’ | | | ‘(a) dog died’ |

Thus, a correspondence between variables and prominence is easily found in these languages. Even the most common exchange of greetings displays the correspondence, as can be observed in the Bukharan Tajik example below:

- [57] a *čitū š(u)mo?*
 how you¹²⁵
 ‘How are you?’
- b *man NAGHZ.* (or only *naghz* without *man*)
 I fine
 ‘I’m fine.’

Can we, then, conclude that prominence is consistently assigned to, as well as associated with, variables? There are some complications that prevents us from claiming the consistent correspondence between prominence and variables, despite the examples above. One such complication is the existence of so-called ‘unstressable’ morphemes: morphemes which are considered not to bear stress and hence also prominence. What happens if one of such morphemes becomes a variable? We will turn to this problem in the following subsection.

3.13.2. Variable = Bound Morpheme

First, recall one of the Bukharan Tajik examples we have become familiar with in the preceding chapter:

- [58] a. *Ali kati šinos* *šud-i-mi?*
 Ali with acquaintance became-2sg-q
 ‘Did you meet Ali?’
- b. ***na-šud-am.***
 neg-became-1sg
 ‘(I) didn’t.’

In (b) in this example, unlike in the previously cited examples, the variable is a function/bound morpheme. Would the variable bear prominence the same way *köpek* and *naghz* do in the examples cited in the last subsection? The variable indicated with bold faced letters in (b) is *na-* [neg] and is indeed prominent:

NA-šud-am.

It seems that the variable-prominence correspondence which we observed in the previous subsection exists also in (b) in the example above. Thus, as far as the examples presented

¹²⁵ Polite 2sg. or 2pl.

(In (b''), the syllable with prominence is the second of the first morpheme: *taNIŞ-ma-di-m*.) If the prominence-variable correspondence was consistent in every circumstance, the assignment of prominence to *taniş-* in (b'') would not have taken place. This suggests operation of a rule other than the prominence-variable correspondence in (b''). The rule is probably a Turkish stress assignment rule.

The variable *-ma* [neg] happens to be one of the group of morphemes which are said to be 'pre-stressing' in Turkish. (See the index in Lees (1961:16-20) for a list of prestressing morphemes.) That is, it shifts stress to the immediately preceding syllable (except in the aorist¹²⁷), which is in this case *niş* in *taniş-*.

The assignment of prominence to *niş* in (b'') then, can be ascribed to the interference of Turkish phonology, i.e. the stressing rule to prominence-variable correspondence. Our interpretation of the assignment of prominence to a non-variable in the above example is the following: Prominence is not associated with *taniş*, but with the variable *-ma*.

However, since *-ma* happens to be a morpheme which does not take prominence/stress, the prominence is, in accordance with the phonological rule of 'pre-stressing', sent to the preceding syllable, which happens to be in another morpheme.

This interpretation entails that, prominence assigned to a syllable in the non-variable *taniş-* has nothing to do with the information structure of the sentence. That is, the syllable preceding the pre-stressing variable happens to be in a non-variable. An examination of the Japanese counterpart of the above Bukharan Tajik and Turkish question-answer pairs confirms the plausibility of this entailment and hence also of our interpretation:

- [60] a. *Ali ni at-ta?*
 Ali dat meet-past
 'Did (you) meet Ali?'
 b. *aw-a.na-katta*
 meet-neg-past
 'I didn't'

¹²⁷ In the negative aorist, *-mE* [neg] itself bears stress. However, in the aorist too, there is a case where the syllable preceding *-mE* [neg] bears stress, as a result of another stress-affecting rule. This is when the stress-attracting impossibilitative *-E* precedes *-mE* [neg]. The following example is taken from Lees (1961:34): *silméz/silémez*.

In (b), the negative morpheme is, like in the Bukharan Tajik and Turkish counterparts of this example, the variable. The variable *-na* is a ‘pre-accenting’ morpheme which, like the Turkish *-ma*, shifts accent to the mora preceding it (except when the verb is ‘unaccented’, i.e. has no lexically determined pitch fall in it. See Vance (1987:86,91) and McCawley (1977:268) for details). Therefore, as in the Turkish example, the variable, despite its being a variable, does not bear prominence/accent in (b):

aw-á.na-katta

aw-a.ná-katta

The morpheme *-na* has the epenthetic vowel *a*¹²⁸ which appears between the negative *-na* and verb stems ending with a consonant. It is to this epenthetic vowel that prominence is assigned. Thus, a comparison of the Turkish example with its Japanese counterpart shows that the assignment of prominence to *taniş-* has no relationship with the morpheme’s information status.

In sum, the discussion in this section shows that prominence-association and prominence-assignment are two separate processes and that the simple view of assuming the ‘prominence-variable’ correspondence fails to explain prominence occurring outside variables.

3.13.3. Summary

The analysis of prominence in this section is different from a number of other analyses in that it associates prominence not to supra-morphemic units, but to morphemes which specify values for parameters, i.e. variables.

Our analysis assumes the application of two constraints in prominence-assignment, namely phonological and information-structural constraints. The former consists of language-specific phonological rules (i.e. stressing/accenuation rules) while the latter, prominence-variable association, operates consistently among the languages dealt with in this thesis. Examples in this section have been adequately explained with this theory of double-application of phonological and information-structural constraints. There are,

¹²⁸ There is another analysis that take *-ana* rather than *-(a)na* as the ‘underlying’ form of the negative morpheme. (see Vance 1987:191, Bloch 1970:19). We follow Kiyose (1995) in regarding the *a* preceding *na* as a buffer vowel.

however, exceptional cases for which this theory cannot fully account. They are the subjects of the next section.

3.14. Cases Where Information-Structural Constraints Rule Over Phonological Constraints

In the discussion in the preceding section, we mentioned the operation of two constraints in prominence assignment, namely phonological constraint and information-structural constraint. What we did not mention is that the application of these constraints is not consistent in all circumstances.

If every sentence complies with the two constraints, it would be difficult to explain the cases cited before by Vendetti et al and Ōishi, because prominence in their examples does not comply with the accentuation rules. Ōishi's examples copied below are apt examples of such cases:

$\begin{matrix} \text{ké} & \text{WA} \\ \text{npoo} & \end{matrix}$	[constitutional law top]	$\begin{matrix} \text{ké} \\ \text{npoo wa} \end{matrix}$
$\begin{matrix} \text{ná} & \text{NI} \\ \text{nika} & \end{matrix}$	[something dat]	$\begin{matrix} \text{ná} \\ \text{nika ni} \end{matrix}$
$\begin{matrix} (\text{ná} & \text{DE} \\ \text{na} & \text{ban no}) & \text{má} \\ & & \text{e} \end{matrix}$	[front loc]	$\begin{matrix} (\text{ná} & \text{DE} \\ \text{na} & \text{ban no}) & \text{má} \\ & & \text{e de} \end{matrix}$

The accentuation rules predict the accentuation patterns in the right column, which patterns are distorted by the assignment of prominence to the postpositions in the actual utterances shown in the left column. These are, then, examples where prominence-variable correspondence is in operation while accentuation rules are not.

This type of prominence assignment violating the phonological constraint is by no means unique to Japanese. Observe the example from Turkish:

[61] *kapát-ma*
 close-neg
 ‘don’t close (it)’

The negative suffix *-ma* is a pre-stressing morpheme, whose lack of stress is predicted by the phonology of Turkish. Despite this, an informant maintained that *-ma* [neg] can bear prominence, in such a situation as follows: A two-year-old child keeps on shutting a door despite an adult’s repeated order to the child not to shut the door, *kapatma!* ‘do not close

(it)!’. Running out of patience, the adult yells: *‘kapatMA’ dedim sana!*¹²⁹ ‘I said “don’t close it” to you!’.

What, then, induces such negligence of phonological constraints by the speakers?

Various reasons can be assumed, but we think one of the reasons is that the information-structural constraint overrides phonological constraints.

We have seen in the chapter ‘Pre-Grammatical Sentences’ cases where rule 1 breaches grammatical rules – if an information structural principle can operate at the expense of grammatical correctness, another information structural rule, i.e. prominence-variable correspondence, may violate phonological constraints as well. In pre-grammatical sentences, variables have priority over non-variables in the probability of occurrence — we assume variables’ priority over non-variables exists also in prominence assignment. That is, variables can be assigned prominence at the expense of phonological correctness, violating accentuation/stressing rules. Naturally, this would take the form of subordination of phonological rules (i.e. accent/stress rules) to the information-structural rule (variable-prominence correlation).

In the next subsection, we shall examine examples where the variable-prominence correlation takes place despite the existence of phonological rules which prevent it.

3.15. Variable-prominence correlation overriding phonological rules

Observe the following example taken from Lambrecht (1994:215):

And then, when we’d finished talking about pigs, we started talking TO the pigs.

Lambrecht maintains that:

[...] the predicator *to* [...] cannot supply an element of information whose addition to a presupposition would result in an assertion.

However, according to our system which explains information structure in terms of the interlocutors’ constants and parameters it is fully eligible to be a variable. In order to

¹²⁹ His pronunciation of *kapatMA* here was clearly distinct from that for *kapatma* ‘closing’, the verb *kapat-* with the nominalising *-ma*. (See Demircan 1976:199):

kapat-má

close-nmlz

‘closing’

confirm this, let us modify the example above to devise examples in Japanese and Bukharan Tajik. A Turkish example will not be devised for a reason explained later. Firstly, imagine a situation where a Japanese speaker, whom we call A here, has a set of constants such as the following and believes that B talks about pigs:

	A's	Values
Action:	Constant	Talking
Action performed:	Constant	Positive
Performer of action:	Constant	B
Object of action:	Constant	Monkey
Object's relation with action:	Constant	about
Tense:	Constant	Non-Past
etc.	Constants	

However, in a rather unlikely case where the talker B in fact talks to a monkey, say, as part of his research on animals' language comprehension, B may attempt to correct A's belief by changing his/her constants B in the following way:

	A's	B's	Values
Action:	Constant	Constant	Talking
Performer of action:	Constant	Constant	B
Execution of action:	Constant	Constant	Positive
Object of action:	Constant	Constant	Monkey
Action - object relation::	Constant	Parameter	about > to
Tense:	Constant	Constant	Non-Past
etc.	Constants	Constants	

B's utterance towards A, which may be in response to A's utterance such as 'so you talk about monkeys', would then have a Japanese equivalent of English 'to' as a variable (bold-faced), thus:

[62] *saru* ***ni*** *hanasu* (*no*)
 monkey dat talk (sfp)
 'I talk **TO** monkey(s)'

Thus, the Japanese equivalent of 'to' is fully eligible for being a variable. Is prominence, then, assigned to the variable *ni* [dat] in this sentence?

We consciously selected a noun which has a fall of pitch in it, i.e. a noun which is ‘accented’, because the fall brings *ni* to the low pitch — if any rise in pitch is involved in prominence production, prominence assigned to a low-pitched *ni* is more likely to distort the stress pattern than that assigned to a high-pitched *ni*. This could make prominence more easily recognisable and is the reason why the accented *sáru* ‘monkey’ is preferred to the Japanese translation of ‘pig’, *buta*, which is unaccented. Observe the accentuation of the domain of the operation of accentuation rules *saru ni* as it is predicted by Japanese phonology:

sáru ni ^{sá}_{ru} ni [monkey dat]

If the prominence-variable correlation has priority over the phonological rule, the prominence, assuming production of prominence involves rise in pitch, would distort this stress pattern, which is in fact the case:

^{sá} NI
ru

This is exactly the kind of distortion of a pitch pattern observed in Ōishi’s examples cited before. In this example, as in Ōishi’s, prominence alters a phonologically predicted stress pattern, that is, in the determination of pitch of *ni*, the pitch-raising effect of the prominence-variable correlation has priority over the pitch-lowering accentuation rule. Thus, there exist cases where phonological (accentuation) rules are overruled by the prominence-variable correlation.

However, as we have seen in the analysis above, this kind of prominence assignment involves the contradiction of the prominence-variable correlation and phonological rules. It is therefore possible that some speakers opt to give priority to phonological rules, the violation of which occasionally results in ambiguity (recall the *hana ga* [nose nom] and *haná ga* [flower nom] pair distinguished from each other by accentuation), rather than to the prominence-variable correlation.

This suggests that priority may be given by choice or intention to the prominence-variable correlation over accentuation/stressing rules. In other words, there may be speaker-specificity in prominence-assignment which contradicts with accentuation/stressing rules. This speaker-specificity was observed among six of Bukharan Tajik speaking informants with whom we discussed this type of prominence-assignment. We will turn to its discussion now. (Bukharan Tajik phonology, unlike the

Turkish one, does not assign stress to the last syllable in a domain of stressing rules, which is why a Bukharan Tajik example, rather than a Turkish example, is used here — in Turkish, the variable coincides with the stressed syllable, thus: *domuz-á konuş-acağ-ım* [pig-dat talk-fut-1sg] and hence is unsuitable for our discussion in this section.)

Observe the following Bukharan Tajik sentence:

[63] *čũčqa-ba gap zadan-i man*
 pig-to talk-fut.ptpl I
 ‘I will talk TO pig(s)’

The postposition/suffix *-ba* is usually unstressed in Bukharan Tajik (which is also the case with its standard Tajik equivalent¹³⁰, preposition *ba*¹³¹) as can be seen in the following domains of the operation of stressing rules:

xoná-ba [house-to] (Rastorgueva 1964:115)
čũčqá-ba [pig-to]
bozór-ba [market-to]

Regrettably, there is not an extensive study of the phonology of Bukharan Tajik. We are therefore reluctant to call the apparently consistent lack of stress on *-ba* an established phonological ‘rule’. We should, therefore, be content with showing that at least *-ba* in *čũčqá-ba* is not stressed and hence not prominent.

However, when the same set of constants and parameter as the ones for the Japanese example above (except that the animal is, as in Lambrecht’s example, ‘pig’ rather than ‘monkey’ here and that the tense is not ‘non-past’ but ‘future’ which Japanese does not have) was assumed¹³², two informants uttered immediately the following without thinking:

čũčqa-BA gap zadani man

The prominence assigned to the variable *-ba* ‘to’ clearly does not comply with the stress pattern of *čũčqá-ba*. The example above, then, can be regarded as an example of the

¹³⁰ The literary Tajik preposition *ba* is considered to be one of the ‘words and morphemes without stress’ (*kalima va morfemahoi bezada*). See Rustamov and Ghafforov (1985:56).

¹³¹ The literary Tajik preposition *ba* is a suffix/postposition in many dialects (see Rastorgueva 1964:115).

¹³² Their utterances were often responses to such an utterance as *čũčqa haq-aš-ba gap zadan-i (mi) š(u)mo* [pig truth/respect-3sg-to talk-fut.ptpl (Q) you (sg.plt)] ‘You will talk about pigs(?)’ where they assign to themselves such imaginary roles as ‘scientists doing a research on pigs’ language recognition’.

prominence-variable correlation overruling the stress-assignment. However, the prominence assignment to the variable in this case was not unanimous among the informants. The other four among the six informants, after premeditation, provided either or both of the immediately above one and the one in which the stress pattern is retained, i.e. in which *-ba* is not prominent.

Thus, by some informants, the variable *-ba* is assigned prominence, despite the stress in *čũčqá*. However, the other informants were not comfortable with *čũčqa-BA*, which may suggest the existence of individual differences in the degree of acceptability of prominence assigned against stress patterns.

3.16. Summary

Having observed the Japanese and Bukharan Tajik examples, the following has become clear: 1) The prominence-variable correlation often assigns prominence to variables even when accentuation/stressing rules predict their non-prominence. 2) There probably are conditions (such as personal preferences) which take part in the process of determining whether such accent/stress-violating prominence takes place.

3.17. Summary of the Chapter

In this chapter, we have attempted to explain prominence placement in terms of the operation of two often contradicting principles, namely the prominence-variable correlation and phonological (accentuation/stressing) rules.

It has been argued that prominence is associated with variables and that variables are also assigned prominence unless there are phonological rules that prevent the variables to bear prominence/stress/accent. This means that operation of phonological (accentuation/stressing) rules usually has priority over the prominence-variable correlation.

However, there are also cases where variables are assigned prominence despite the existence of phonological rules preventing them from being prominent/stressed/accented. Our analysis is that they are instances in which the prominence-variable correlation overrules the phonological rules.

Thus, the discussion in this chapter revealed the association between prominence and variables which association can be often obscured by the operation of phonological rules that block direct assignment of prominence to variables.

The implication which this revelation has to our theory is clear. The prominence-variable correlation calls for a link between information (structure) and morphemes, which link manifests itself as variables and non-variables, as its prerequisite. Therefore, the existence of the correlation is in support of our theory where we attempt to associate pieces of information with morphemes.

4. SUMMARY AND CONCLUSION

In this thesis, we have tried to devise a simple analytic system for incomplete sentences in languages which utilise distinctively agglutinative components in their morphology.

In the introductory chapter, we explained why we were interested specifically in agglutinative languages. Agglutinative languages were of particular interest for us because they are languages in which words are readily segmentable into morphemes. The morphological segmentability that characterises agglutination allowed us to link pieces of information, which are usually associated with syntactic units, directly to morphemes. Based on the linkage between pieces of information and morphemes, we constructed a morpheme-based system of IS analysis in the second chapter.

In this second chapter, we proposed a very simple set of information-structural principles, which we named rule 1-1 and rule 1-2. We showed that the simple rules can account for a number of linguistic phenomena associated with incomplete sentences, which would otherwise attract complicated syntactic accounts or often ambiguous semantic/pragmatic explanations.

The third chapter, ‘Prominence’, discussed the relationship between information structure and perceptive prominence in sentences. We found a clear correlation between variables and the placement of prominence. However, we also found that the correlation could often be abstracted by word-level stress/accent placement.

All these chapters are indispensable for our discussion, but the most important of them is the second; that is the chapter in which we devised and explained in detail our system of analysis in its entirety. The system exhibits several unique characteristics such as simplicity and observability.

Simplicity was given definite priority over sophistication in the process of devising the system. However, we believe that the simplicity of our system was not achieved at the expense of sophistication. Rather, the priority given to simplicity contributed greatly to the clarity of the system by forcing us to introduce observability of the operation of the principles (i.e. rules 1-1 and 1-2) that the system comprises. Our pursuit of simplicity also eliminated much unwanted complication, which is often a characteristic of some

elaborate linguistic theories. Hopefully, it also made our system accessible to those whose area of speciality is not linguistics.

However, it must be acknowledged that some problems were identified and mentioned in the preceding chapters which need to be resolved for the system to be fully functional. The most fundamental of the problems is arguably in identification of morphemes or determination of morpheme boundaries by native speakers.¹³³

It is also important to note that the applicability of our system is restricted to agglutinative languages. This restriction may appear to be a self-imposed one. However, as we have explained in the introductory chapter, it is in fact a restriction the data that we used in this thesis call for. The data on which our system is based are taken from only four languages, namely Turkish, Uzbek, Bukharan Tajik, and Japanese. Clearly, these languages do not exhaust all types of languages. They represent only a tiny fraction of languages and hence can not be representative of the languages of the world. On the other hand, the four languages have one morphological characteristic in common, namely relatively high level of agglutination, and it is on agglutination that our system of IS analysis is based. Accordingly, while we consider it erroneous to claim universality or wide cross-language applicability of principles elicited from a set of data taken from such a limited number of languages, we do think the applicability of rule 1 is extendable to agglutinative languages in general.

The most significant characteristic of our system is its unconventional approach to information structure. It identifies a piece of information in every morpheme, which can either be replaced by another piece of information or remain unchanged. This approach does not comply with the popular approach to the study of IS which usually sees the proposition as the basic unit of information. Our approach, however, has at least one obvious advantage over the conventional one – it brings a high degree of observability into analysis of incomplete sentences by linking pieces of information with the ‘tangible’ units of morphemes.

¹³³ For example, is *-ama* in the following example copied from Oflazer (1994:137) one morpheme or a combination of two morphemes *-a* and *-ma* as Underhill’s (1976:402) grammar holds it? (Etymologically speaking, *-ama* is the contracted *-a*, *al-*, and *-ma* (Shōgaito 1989:948).): *Osman-lı-laş-tır-ama-yabil-ecek-ler-imiz-den-miş-siniz-cesine* ‘(behaving) as if you were of those whom we might consider not converting into an Ottoman.’

In conclusion, in the present thesis, we have tailored a system of morpheme-based IS analysis using data from several languages with agglutinative morphology. The system, which is composed essentially of two very simple principles, has proved to have certain explanatory power for incomplete sentences in agglutinative languages.

APPENDIX 1: Glossary

Accent: Prominence where, within words, pitch is involved; Pitch accent; Lexical accent.

Allosentences: Sentences which, despite their morphological and syntactic differences, specify the same value for a given parameter.

CS Grammar: Classical Sentence Grammar.

Information Structure (IS): Distribution of variables and non-variables in sentences.

IS: See Information Structure

Non-variables: Morphemes other than variables.

Parameters and Constants: See 2.4.2.

Pre-grammatical Sentences: Sentences which are acceptable but not ‘correct’ according to classical sentence grammar.

(Sentential) Prominence: Used in this thesis primarily to refer to perceptive prominence of a sentence.

Stress: Prominence within words, however such prominence is achieved; Lexical stress.

Unattended Parameters: Parameters whose values are unspecified.

Variables: Morphemes which specify values for parameters.

APPENDIX 2: Transliteration

Japanese

Japanese examples are transliterated in the modified Hepburn romanisation.

Uzbek

Most of the Uzbek publications cited in this thesis are in Cyrillic script. We transliterate them into the new Latin alphabet which consists only of ASCII symbols. For an overview of the history of the orthography of Uzbek, reference is made to Allworth (1964:Chapter XVII).

The Russian genitive plural suffix *-ev* in the names of the editors of *O‘zbek tilining imlo lug‘ati* ‘Orthographic dictionary of the Uzbek language’ are spelt *-yev*, though the same suffix is spelt *-ev* in the name surname *Tojiev* on page five. We did not insert *y* in transliterating Cyrillic *e* to Latin *e*.

Tajik

For Tajik, we devise a system of transliteration which is based on the Cyrillic transliteration system of the International Organisation for Standardisation.

For an overview of the history of the orthography of Tajik, see Perry (1997).

In reference to Tajik papers/books published in Latin script, we retain their original spelling, except the small letter *B* which is replaced with *b*.

Vowels are inserted in accordance with the orthography of modern standard Tajik (Kalontarov 1974, Maniëzov and Mirzoev 1991) to the titles of Tajik articles written in Arabic script.

Russian

Russian words and names are transliterated according to the ISO system with two exceptions: the Cyrillic *х* and *э* are transliterated as *x* (instead of *h/ch* of the ISO system) and *é* (as in the BSI system), respectively.

Transliteration Used in This Thesis

	Tajik	Uzbek	Russian
Аа	a	a	a
Бб	b	b	b
Вв	v	v	v
Гг	g	g	g
Дд	d	d	d
Ее	e	e	e
Ёё	yo	yo	jo
Жж	ž	j (ž)	ž
Зз	z	z	z
Ии	i	i	i
Йй	y	y	j
Кк	k	k	k
Лл	l	l	l
Мм	m	m	m
Нн	n	n	n
Оо	o	o	o
Пп	p	p	p
Рр	r	r	r
Сс	s	s	s
Тт	t	t	t
Уу	u	u	u
Фф	f	f	f
Хх	x	x	x
Цц	(ts)	(ts)	c
Чч	č	ch	č
Шш	š	sh	š
Щщ	(šč)	(shch)	šč
Ъъ	”	’ (”)	”
Ыы			y
Ьь	,	,	,
Ээ	é	é	é
Юю	yu	yu	ju
Яя	ya	ya	ja
Ғғ	gh	g’	
Йй	ī		
Ққ	q	q	
Ўў	ū		
Хх	h		
Чч	j		
Ўў		o’	

APPENDIX 3: A Short Bibliography of Languages in Bukhara

- Babaev, K. R. 1991. Kommunikativnoe povedenie evreev Buxary. Sovetskaja Etnografija, 5, Sept-Oct. 86-94.
- Grigor'ev, V. V. 1861. Obozrenie dialektičeskix otličij upotrebitel'nogo u buxarskix tadžikov persidskogo jazyka. Kazan'.
- Gulkarov, Josef. 1998. Etimologičeskij slovar' buxarsko-evrejskogo jazyka s perevodom na russkij, angliiskij i ivrit = Etimologycal [sic] dictionary of Bukharian - Russian - English - Hebrew = Lughati zaboni yahudiëni Buxoroī - Rusī - Angliši - Ibrī = Milon Bukhari - Rusi - Angli - 'Ivri. Tel-Aviv: Berit yots'e Bukharah.
- Hojiev, S. 1966. Ba"ze xususiyathoi morfologii zaboni adabii tojik dar ibtidoi asri XX (dar asosi «Tūhafi ahli Buxoro»-i Mirzo Siroji Hakim). IN 1966. Tezishoi ma"rūzaho dar konferentsiyai ilmii olimoni javon. Dušanbe: Akademiyai fanho. 83-85.
- Hojiev, S. 1968. Ba"ze xususiyathoi xosi iste"moli fe"l dar «Tūhafi ahli Buxoro». IN Tohirova, Q. / Tursunov, M. / Zikriëev. eds. 1968. Maqolaho oid ba ba"ze mas"alalohi zabonšinosii tojik. Leninobod: Vazorati maorifi xalqi RSS Tojikiston. 51-60.
- Hojiev, Saidmurod. 1987. Zaboni adabii tojik dar ibtidoi asri XX: Dar asosi materiali "Tuhafi ahli Buxoro"-i Mirzosiroji Hakim. Dušanbe: Doniš. [A linguistic analysis of "Tuhafi ahli Buxoro"¹³⁴ written by Mirzosiroji Hakim (1877-1914), a native of Bukhara¹³⁵.]
- Kerimova, A. A. 1959. Govor tadžikov buxaru. Moskva: Izdatel'stvo vostočnoj literatury.
- Kerimova, A. A. 1980. O ritmike tadžikskoj razgovornoj reči: Na osnove govora tadžikov Buxary. Iranskoe jazykoznanie, Ežegodnik 1980. 224-227.
- Mahmudov, M. / Berdiev, B. eds. 1989. Lughati muxtasari lahjahoi Buxoro. Dušanbe: Doniš.

¹³⁴ Hakim (1913).

¹³⁵ See Hojiev (1987:11-14) for a concise biography of Hakim.

- Mardonova, A. 1970. Pešoyand va pasoyandho dar ševai éroniēni Kogon. IN Ghafforov, R. / Jūraev, Gh. eds. 1970. Mas"alahoi ševašinosii tojik. Dušanbe: Doniš. 99-116.
- Melex, N. A. 1962. Gižduvanskij govor tadžikskogo jazyka (Tezisy doklada). IN Naučnaja konferencija po iranskoj filologii. Leningrad.
- Melex, N. A. 1964. Fonetičeskie osobennosti gidžuvanskogo govora tadžikskogo jazyka. IN Iranskaja filologija: Trudy naučnoj konferencii po iranskoj filologii (24-27 janv. 1962 : Leningrad, R.S.F.S.R.). Leningrad: Izd. Leningradskogo universiteta. 27-32.
- Melex, N. A. 1968. Gižduvanskij govor tadžikskogo jazyka (Avtoreferat. kand. diss.). Leningrad.
- Mirzaev, M. M. 1965. Buxarskaja grupa govorov uzbekskogo jazyka (Avtoreferat dokt. diss.). Taškent.
- Mirzaev, M. M. 1969. O'zbek tilining Buxoro grupa shevalari. Toshkent: O'zbekiston SSR «Fan» Nashriyoti.
- Paper, Herbert H. 1982. Judeo-Tajik: Textual Sources for Pre-1917 Tajik Vowel Phonology. *Folia Slavica*, 5, 1-3. 336-343.
- Rajāi, Ahmad Ali. 1964. Yāddāšti dar bāre-ye lahje-ye boxārai (Entešārāt-e Dānešgāh-e Mašhad, 8). Mašhad: Dānešgāh.
- Saidova, Q. / Muruvvatov, J. 1960. Boz yak asari nav dar ilmi ševašinosī: Taqriz ba kitobi A. A. Kerimova "Govor tadžikov buxaru". *Maorif va Madaniyat*, Dec. 13rd, 1960. [A review of Kerimova 1959.]

REFERENCES

- Abdoraĥmonov, G'ani. 1996. O'zbek tili grammatikasi: Sintaksis. Toshkent: O'qituvchi.
- Abe, Isamu. 1998. Intonation in Japanese. IN Hirst, Daniel / di Cristo, Albert eds. 1998. Intonation Systems: A Survey of Twenty Languages. Cambridge University Press. 360-375.
- Ajnī (Aynī), Sadriddin. 1939a. Margi sudxūr. Stalinobod: Naşrijoti Davlatiji Toçikiston Şū'baji Adabijoti Bacagona.
- Ajnī (Aynī), Sadriddin. 1939b. Tirozi çahon. Stalinobod: Naşrijoti Davlatiji Toçikiston Şū'baji Adabijoti Bacagona.
- Akamatsu, Tsutomu. 1997. Japanese Phonetics: Theory and Practice. München, Newcastle: Lincom Europa.
- Akiner, Shirin. 1989. Uzbekistan: Republic of Many Tongues. IN Kirkwood, Michael II ed. 1989. Language Planning in the Soviet Union. Houndmills; Basingstoke; Hampshire; London: McMillan Press. 100-122.
- Allworth, Edward. 1964. Uzbek Literary Politics. London; the Hague; Paris: Mouton.
- Altmann, Hans / Batliner, Anton / Oppenrieder, Wilhelm. 1989. Das Projekt 'Modus-Fokus-Intonation': Ausgangspunkt, Konzeption und Resultate im Überblick. IN Altmann, Hans / Batliner, Anton. / Oppenrieder, Wilhelm eds. 1989. Zur Intonation von Modus und Fokus im Deutschen. Tübingen: Max Niemeyer Verlag. 1-20.
- Arai, Kōichi, 1984. Tōhoku Hōgen no Bunpō. Kokubungaku Kaishaku to Kanshō, 49, 1. 38-50.
- Arai, Kōichi, 1990. Shudai Kōsei to Nihongo: Hyōjungo to Hōgen no Idō. Kokubungaku Kaishaku to Kanshō, 55, 1. 77-86.
- Arisaka, Hideyo. 1941. Akusento no Kata no Honshitsu ni Tsuite. Gengo Kenkyū, 7-8. IN Shibata, Takeshi / Kitamura, Hajime / Kindaichi, Haruhiko. 1980. Nihon no Gengogaku dai 2 kan: On'in. Tokyo: Taishūkan. 352-363.
- Arzumanov, S. / Sanginov, A. 1988. Zaboni tojikī. Dušanbe: Maorif.

- Asimova, B. S. 1982. Jazykovoe stroitel'stvo v Tadžikistane (1920-1940 gg.) Dušanbe: Doniș.
- Atkinson, Martin. 1992. Children's Syntax: An Introduction to Principles and Parameters Theory. Oxford, Cambridge: Blackwell.
- Aynī, Sadriddin. 1928. Dar atrofi zaboni forsī va tojikī. Rahbari doniș, 28, 4-5.
- Azizī. 1929. Ba zaboni darī durri suxan suftan mexohem. Rahbari doniș, 16, 4-6.
- A'la zoda, F. 1930. Boz ham dar atrofi mas'alai alifboi navi tojikī. Rahbari doniș, 30, 5-7.
- A'lo-zoda, F. 1930. Dar girdi zaboni adabii tojik. Rahbari doniș, vol. 28. 3-4.
- Baker, Nancy D. / Greenfield, Patricia M. 1988. The Development of New and Old Information in Young Children's Early Language. *Language Sciences*, 10 (1): 3-34.
- Banguoğlu, Tahsin. 1959. Türkçenin Grameri. Ankara: Türk Dil Kurumu.
- Bassarak, Armin. 1988. Morphologisierungstendenzen im Türkischen. *Linguistische Studien, Reihe A: Arbeitsberichte*, 188, 1-13.
- Beckman, Mary. 1986. Stress and Non-stress Accent. Dordrecht: Foris.
- Beckman, Mary E. 1996. The Parsing of Prosody. *Language and Cognitive Processes*, 11, 1/2. 17-67.
- Berke, Vardar. / Güz, N. / Öztokat, E. / Senemoğlu, O. / Sözer, E. 1988. Açıklamalı Dilbilim Terimleri Sözlüğü. İstanbul: ABD Kitabevi.
- Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.
- Birner, Betty J. 1994. Information Status and Word Order: An Analysis of English Inversion. *Language*, volume 70, number 2. 233-259.
- Bloch, Bernard. 1970. *Bernard Bloch on Japanese: Edited with an introduction and analytic index by Roy Andrew Miller*. New Haven; London: Yale University Press.
- Bloom, Lois. 1970. *Language Development: Form and Function in Emerging Grammars*. Cambridge, Massachusetts; London, England: The MIT Press.
- Boboeva, A. 1978. Hozirgi o'zbek adabiy tilida to'liqsiz gaplar. Toshkent: O'zbekiston SSR «Fan» Nashriyoti.

- Bobomurodov, Š. / Mūminov, A. 1983. Lughati muxtasari kalimasozii zaboni adabii tojik. Dušanbe: «Maorif».
- Boeschoten, Hendrik. 1998. Uzbek. IN Johanson, Lars / Csató, Éva Ágnes. eds. 1998. The Turkic Languages. London; New York: Routledge. 357-378.
- Boeschoten, Hendrik / Vandamme, Marc. 1998. Chaghatay. IN Johanson, Lars / Csató, Éva Ágnes. eds. 1998. The Turkic Languages. London; New York: Routledge. 166-178.
- Brazil, David. 1997. The Communicative Value of Intonation in English. Cambridge; New York; Melbourne: Cambridge University Press.
- Bresnan, Joan / Mchombo, Sam A. 1987. Topic, Pronoun and Agreement in Chichewa. Language, 63, 4. 741-782.
- Brown, Gillian / Yule, George. 1983. Discourse Analysis. Cambridge; New York; Melbourne: Cambridge University Press.
- Buyrukçu, Muzaffer. 1998. Dillerinde Dünya. Ankara: T.C. Kültür Bakanlığı.
- Buzurgzoda, Lutfullo. 1940. Fonetikaji zaboni adabiji toçik: Ba tavri muqojisa kardan bo xususijathoji fonetikiji dialektho. Stalinobod; Leningrad: Naşrijoti Davlatiji Toçikiston Şū'baji Ta'lim-pedagogī.
- Caramazza, Alfonso / Laudanna, Alessandro / Romani, Cristina. 1988. Lexical Access and Inflectional Morphology. Cognition, 28. 297-332.
- Chafe, Wallace L. 1970. Meaning and the Structure of Language. Chicago, London: The University of Chicago Press.
- Chafe, Wallace L. 1974. Language and Consciousness. Language, 50. 111-133.
- Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definitiveness, Subjects, Topics, and Point of View. IN Li & Thompson eds. 1976. 25-55
- Chapman, Robin S. 2000. Children's Language Learning: An Interactionist Perspective. Journal of Child Psychology and Psychiatry, 44, 1, 33-54.
- Cheng Shih-liang / A-pu-tu-je-ho-man. 1987. Wu-tzu-pieh-ko yu chien chih. Beijing: Min tsu chu pan she.

- Chodzko (Xodzko), Alexander. 1852. Grammaire persane ou principes de l'iranien moderne, accompagnées de fac-similés pour servir de modèle d'écriture et de style pour la correspondance diplomatique et familière (Paris). [2nd ed. 1883 as Grammaire de la langue persane, 2ème édition, augmentée de textes persans inédits et d'un glossaire (Paris: Maisonneuve & Cie)].
- Christiansen, Morten H. / Chater, Nick. 1999. Connectionist Natural Language Processing: The State of Art. *Cognitive Science*, 23, 4. 417-437.
- Clark, Herbert H. 1996. *Using Language*. Cambridge University Press.
- Clark, John / Yallop, Colin. 1995. *An Introduction to Phonetics and Phonology Second Edition*. Oxford UK, Cambridge USA: Blackwell.
- Clarke, Hugh D. B. / Hamamura, Motoko. 1981. *Colloquial Japanese*. New York: Routledge.
- Cohen, Antonie / 't Hart, Johan. 1967. On the Anatomy of Intonation. *Lingua*, 19. 177-192.
- Colé, Pascale / Segui, Juan / Taft, Marcus. 1997. Words and Morphemes as Units for Lexical Access. *Journal of Memory and Language*, 37. 312-330.
- Comrie, Bernard. 1982. On the Morphological Typology of Balto-Finnic: A Reassessment. *Etudes Finno-Ougriennes* 15. Budapest: Akadémiai Kiadó. 91-99.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology* 2nd edition. Oxford: Basil Blackwell.
- Comrie, Bernard. 1997. Turkish Phonology. IN Kaye, Alan S. / Daniels, Peter T. eds. 1997. *Phonologies of Asia and Africa (including the Caucasus)*. Eisenbrauns. 883-899.
- Cruse, D. A. 1986. *Lexical semantics*. Cambridge [Cambridgeshire]; New York: Cambridge University Press.
- Cruttenden, Alan. 1997. *Intonation* 2nd edition. Cambridge University Press.
- Çokuğurel, Işıl. 1996. *Türkçede Eksiltili Anlatım ve Eksiltili Cümle* (Unpublished MA thesis). Adana: Çukurova Üniversitesi.

- Daneš, František. 1966. A Three-level Approach to Syntax. IN Daneš, F. et al. eds. 1966. Travaux linguistiques de Prague vol. I: University of Alabama Press. 225-240.
- Demir, Hülya / Akar, Rıdvan. 1994. İstanbul'un Son Sürgünleri. İstanbul: İletişim Yayınları (1st-3rd press) Belge Yayınları (1999, 4th press).
- Demircan, Ömer. 1975. Türk Dilinde Vurgu: Sözcük vurgusu. Türk Dili, 284. 333-339.
- Demircan, Ömer. 1976. Türk Dilinde Ek Vurgusu. Türk Dili, 294. Ankara: Türk Dil Kurumu. 196-200.
- Demircan, Ömer. 1977. Bileşik Sözcük ve Bileşik Sözcüklerde Vurgu. Türk Dili Araştırmaları Yıllığı Belleten, 1977. Ankara: Türk Dil Kurumu Yayınları. 263-275.
- Deny, Jean. 1995. Türk Dili Gramerinin Temel Kuralları (Türkiye Türkçesi). Ankara: Türk Dil Kurumu. [Translated from the French 'Principes de Grammaire Turque ("Turk" de Turquie)' by Oytun Şahin.]
- Devlet, Nadir. 1990. Sovyetler Birliği'nde Türk ve Müslümanların Nüfusu 60 Milyona Yaklaştı. Türk Kültürü, 325. 9-16.
- Duszak, Anna. 1994. On Thematic Configurations in Texts: Orientation and goals. IN Čmejrková, Světa / Štícha, František eds. 1994. The Syntax of Sentence and Text: A Festschrift for František Daneš. Amsterdam; Philadelphia: John Benjamins. 105-117.
- Eckmann, János. 1988. Çağatayca El Kitabı. İstanbul: İstanbul Üniversitesi Edebiyat Fakültesi.
- Edwards, Jane Anne. 1993. Principles and Contrasting Systems of Discourse Transcription. IN Edwards, Jane Anne / Lampert, Martin D. eds. 1993. Talking Data: Transcription and Coding in Discourse Research. Hillsdale (New Jersey): Lawrence Erlbaum Associates. 3-32.
- Emmorey, Karen D. 1988. Do People Parse? IN Wallace, Karen ed. 1988. Morphology as a Computational Problem (UCLA Occasional Papers #7: Working Papers in Morphology). Department of Linguistics, University of California, Los Angeles. 107-121.

- Enç, Mürvet. 1986. Topic Switching and Pronominal Subjects in Turkish. IN Slobin, Dan I. / Zimmer, Karl eds. 1986. *Studies in Turkish Linguistics*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Ergenç, İclâl. 1995. *Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü: Bir Deneme*. Ankara: Simurg.
- Ergin, Muharrem. 1997. *Türk Dil Bilgisi [Turkish Linguistics]*. İstanbul: Bayrak Basım-Yayın-Tanıtım. [Originally published in 1958.]
- Erguvanlı, Eser Emine. 1984. *The Function of Word Order in Turkish Grammar*. Berkeley; Los Angeles; London: University of California Press.
- Éšniëzov, M. 1977. *Dialektologiyai tojik (dasturi ta"limi) Qismi yakum*. Dušanbe: Universiteti Davlatii Tojikiston ba Nomi V. I. Lenin.
- Faensen, J. 1983. *Sprachen in der UdSSR*. Osnabrück.
- Firbas, Jan. 1994. Substantiating Daneš's View of Givenness as a Graded Phenomenon. IN Čmejrková, Světa / Štícha, František eds. 1994. *The Syntax of Sentence and Text: A Festschrift for František Daneš*. Amsterdam; Philadelphia: John Benjamins. 119-129.
- Fitrat, [A]. 1930. *Qoidahoji zaboni toçik: Sarf va nahv*. Istolinobod; Toşkand: Naşri toçik.
- Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure: The phonology of Suprasegmentals*. Oxford University Press.
- Franks, Steven. 1995. *Parameters of Slavic Morphosyntax*. New York, Oxford: Oxford University Press.
- Frauenfelder, Uli H. / Schreuder, Robert. 1992. *Contrasting Psycholinguistic Models of Morphological Processing and Representation: The Role of Productivity*. IN Booij, Geert / van Marle, Jaap. eds. 1992. *Yearbook of Morphology 1991*. Kluwer Academic Publishers. 165-183.
- Fujii, Takeo / Ōshima, Arata / Hayashi, Keiko. 1998. *Kotoba no shikumi to sono kenkyū*. IN Fujii, Takeo / Ōshima, Arata eds. 1998. *Kotoba no Sekai: Eigo no Sekai*. Osaka: Ōsaka Kyōiku Tosho. 1-38.

- Fujisaki, Hiroya / Hirose, Keikichi. 1984. Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese. *Journal of Acoustical Society of Japan (E)*, 5(4). 233-242.
- Fujiwara, Masanori. 1992. Joshi Shōryaku no Goyōronteki Bunseki. IN Tajima, Ikudō, Tanba, Kazuya eds. *Nihongo Ronkyū 3*. 1992. *Gendai Nihongo no Kenkyū*. Osaka: Wasen Shoin.
- Fukuda, Minoru. 1993. Head Government and Case Marker Drop in Japanese. *Linguistic Inquiry*, 20, 1, Winter. 168-172.
- Gencan, Tahir Nejat. 1979. *Dilbilgisi (Gözden Geçirilmiş IV. Baskı)*. Ankara: Ankara Üniversitesi Basımevi.
- Go'rkiy (Go'rkij). 1955. P'esalar. Tashkent(?)
- Göksal, Aslı / Özsoy, A. Sumru. 2000. Is There a Focus Position in Turkish?. (manuscript provided by Özsoy).
- Greenberg, Joseph H. 1960. A Quantative Approach to the Morphological Typology of Language. *International Journal of American Linguistics*, 26. 178-194.
- Greenberg, Joseph. 1974. *Language Typology: A Historical and Analytic Overview*. The Hague; Paris: Mouton.
- Greenfield, Patricia Marks. 1978. Informativeness, Presupposition, and Semantic Choice in Single-word Utterances. IN Waterson, Natalie / Snow, Cathrine. eds. 1978. *The Development of Communication*. Chichester; New York; Brisbane; Toronto: John Wiley & Sons. 443-452.
- Greenfield, Patricia Marks. 1982. The Role of Perceived Variability in the Transition of Language. *Journal of Child Language*, 9. 1-12.
- Greenfield, Patricia M. / Zukow, Patricia Goldring. 1978. Why Do Children Say What They Say When They Say It?: An Experimental Approach to the Psychogenesis of Presupposition. IN Nelson, Keith E. ed. 1978. *Children's Language volume 1*. New York; London; Sydney; Toronto: Gardner Press, Inc. 287-336.
- Grice, H. Paul. 1975. Logic and Conversation. IN Cole, Peter / Morgan, Jerry L. eds. 1975. *Syntax and Semantics volume 3: Speech acts*. New York; San Francisco; London: Academic Press. 41-58.

- Grice, H. Paul. 1978. Further Notes on Logic and Conversation. IN Cole, Peter. ed. 1978. Syntax and Semantics Volume 9: Pragmatics. New York; San Francisco; London: Academic Press. 113-128.
- Gussenhoven, Carlos. 1983. Focus, Mode and the Nucleus. *Journal of Linguistics*, 19. 377-417.
- Gürel, Ayşe. 1999. Decomposition: To What Extent? The Case of Turkish. *Brain and Language*, 68. 218-224.
- Gʻulomov, A. 1947. Oʻzbek tilida urgʻu. Toshkent.
- Gʻulomov, A / Tixonov, A. N. / Qoʻngʻurov, R. 1977. Oʻzbek tili morfem lugʻati. Toshkent: Oʻqituvchi.
- Hakim, Mirzosiroyi (Mirzo Sirojiddin binni Xojī Mirzo Abdurrahim). 1913/1330 h. Tuhafi ahli Buxoro. Kogon: Matbaai sangii Kogon.
- Halimov, S. 1974. Sadriddin Aynī va baʼze masʼalahoī inkišofi zaboni adabii tojik. Dušanbe: Irfon.
- Halliday, M. A. K. 1966. Intonation system in English. IN McIntosh, A. / Halliday, M. A. K. eds. 1966. Patterns of Language. London: Longman.
- Halliday, M. A. K. 1967. Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, 3:199-244.
- Haraguchi, S. 1977. The Tone Patterns of Japanese: An Autosegmental Theory of Tonology. Tokyo: Kaitakusha.
- Harris, Zellig Sabbettai. 1991. A Theory of Language and Information : A mathematical approach. Oxford; New York: Oxford University Press.
- Hasegawa, Yoko / Hata, Kazue. 1995. The function of F0-peak delay in Japanese. IN Proceedings of the Twenty-First Annual meeting of the Berkeley Linguistics Society, Feb 17-20, 1995. Berkely, California: Berkeley Linguistics Society. 141-151.
- Hasegawa, Yuri. 1993. Hanashikotoba ni okeru 'Mujoshi' no Kinō. *Nihongo Kyōiku*, 80. 158-168.
- Hattori, Shirō. 1959. Nihongo no Keitō. Tokyo: Iwanami Shoten.

- Haviland, Susan E. and Herbert H. Clark. 1974. What's New? Acquiring New Information as a Process in Comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 13. 512-521.
- Hepçilingirler, Feyza. 1999. Türkçe "Off" (10. Basım). İstanbul: Remzi Kitabevi.
- Hoff-Ginsberg, Erika. 1996. *Language Development*. Pacific Grove, Albany, Belmont, Bonn, Boston, Cincinnati, Detroit, Johannesburg, London, Madrid, Melbourne, Mexico City, New York, Paris, Singapore, Tokyo, Toronto, Washington: Brooks/Cole Publishing Company.
- Hojiev, Azim. 1985. *Lingvistik terminlarning izohli lug'ati*. Toshkent: O'qituvchi.
- Hojiev, Saidmurod. 1987. *Zaboni adabii tojik dar ibtidoi asri XX: Dar asosi materialii "Tuhafi ahli Buxoro"-i Mirzosiroji Hakim*. Dušanbe: Doniș.
- Hosaka, Junko / Takezawa, Toshiyuki / Uratani, Noriyoshi. 1992. Analyzing Postposition Drops in Spoken Japanese. *The 1992 International Conference on Spoken Language Processing ICSLP 92 Proceedings, vol-2*. University of Alberta, Department of Linguistics. 1251-1254.
- Hosaka, Junko / Takezawa, Toshiyuki / Uratani, Noriyoshi. 1992. *Taiwa Deetabeesu o Tsukatta Mujoshimeishikaku no Bunseki*. IN *Jinkō Chinō Gakkai Kenkyūkai Shiryō SIG-SLUD-9203-1*. Tokyo: Jinkō Chinō Gakkai (Japanese Society for Artificial Intelligence). 1-8.
- Hudson, R.A. 1980. *Sociolinguistics*. Cambridge: Cambridge University Press.
- Ido, Shinji. 2001. *Buhara no Tajikugo*. *Gengo*, vol. 3, no. 8. 98-103.
- Inoue, Kazuo. ed. 1984. *Ozu Yasujirō Sakuhinshū IV*. Tokyo: Rippū Shobō.
- Isaev, Magomet Izmajlovič. 1978. *O jazykax narodov SSSR*. Moskva: Izdatel'stvo Nauka.
- Ismatī, Obid. 1929. *Yak nigoh ba matbuoti inqilobii tojikī*. *Rahbari doniș*, 9. 3-5.
- Itami, Jūzō. 1992. *Mimbō no Onna* (unpublished scenario). Tokyo: Itami Film Co.
- Jahangiri, Guissou. 1997. *The Premises for the Construction of a Tajik National Identity, 1920-1930*. IN *Djalili, Mohammad-Reza / Grare, Frederic / Akiner, Shirin eds. 1997. Tajikistan: The Trials of Independence*. New York: St. Martin's Press. 14-41.

- Kai, Masumi. 1991. 'Wa' wa Ika ni Shite Shōryaku Kanō to Naru ka. *Nihongo Nihon Bunka*, 17.00. 113-128.
- Kalontarov, Ya. I. 1974. *Lughati imloi zaboni adabii tojik*. Dušanbe: Irfon.
- Kamio, Akio / Takami, Ken'ichi. 1998. *Danwa to Jōhō Kōzō*. Tokyo: Kenkyūsha.
- Kaneko, Akira. 1989. *Mojishi*. IN Okimori, Takuya ed. 1989. *Nihongoshi*. Tokyo: Ōfū. 45-62.
- Kao, Erh-chiang. 1985. *Ta-chi-ko yu chien chih*. Beijing: Min tsu chu pan she.
- Katada, Fusa. 1995. *On the Nature of Agglutination*. IN *PAMAPLA 19: Papers from the nineteenth annual meeting of the Atlantic Provinces Linguistic Association, Charlottetown, Prince Edward Island, November 10-11, 1995 = ACALPA 19: actes de dix-neuvieme colloque annuel de l'Association de linguistique des provinces atlantiques, Charlottetown, Isle-de-Prince-Edouard, le 10-11 novembre 1995*. 115-127.
- Kawakami, Shin. 1957. *Tōkyōgo no takuritsu kyōchō no Onchō*. reprinted IN Kawakami, Shin. 1995. *Nihongo Akusento Ronshū*. Tokyo: Kyūko Shoin. 76-91.
- Kawakami, S. 1977. *Nihongo Onsei Gaisetsu*. Tokyo: Ōfūsha.
- Kerimova, A. A. 1959. *Govor tadžikov buxaru*. Moskva: Izdatel'stvo vostočnoj literatury.
- Kerimova, A. A. 1997. *Tadžikskij jazyk*. IN Rastorgueva, V. S. / Moškalo, V. V. / Édel'man, D. I. eds. 1997. *Jazyki mira: Iranske jazyki I. Jugo-zapadne iranske jazyki*. Moskva: Indrik. 96-121.
- Kindaichi, Haruhiko. 1951. *Kotoba no Senritsu*. *Kokugogaku*, 5. (Reprinted in Shibata, Takeshi / Kitamura, Hajime / Kindaichi, Haruhiko. 1980. *Nihon no Gengogaku dai 2 kan: On'in*. Tokyo: Taishūkan. 561-593.)
- Kindaichi, Haruhiko. 1967. *Nihongo On'in no Kenkyū*. Tokyo: Tōkyōdo Shuppan.
- Kiss, Katalin É. 1995. *Introduction*. IN Kiss, Katalin É. ed. 1995. *Discourse Configurational Languages*. New York; Oxford: Oxford University Press. 3-27.
- Kitayama, K. 1953. *Joshi no Shōryaku, Kakarimusubi Nado: Genji Monogatari no Gohō Kenkyū*. *Kokugo to Kokubungaku*, August. 56-60.

- Kiyose, Gisaburō Norikura. 1995. *Japanese Grammar: A New Approach*. Kyoto: Kyoto University Press.
- Kohler, Klaus. J. 1977. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.
- Komissiyai imloi alifboi lotinii navi tojikī. 1929. Qoidahoi imloi alifboi lotinii navi tojik. *Rahbari doniś*, 19, 1-3.
- Kononov, Andrej Nikolaevič. 1960. *Grammatika sovremennogo uzbekskogo literaturnogo jazyka*. Moskva; Leningrad: Izdatel'stvo akademii nauk SSSR.
- Konrot, Ahmet. 1991. Sesbilim Çalışmalarında Nesnelik. IN Dil Derneği ed. 1991. *Dilbilim ve Türkçe*. Ankara: Dil Derneği. 23-35.
- Kori, Shirō. 1989. Fōkasu Jitsugen ni Okeru Onsei no Tsuyosa, Jizoku Jikan, F0 no Yakuwari (Acoustic Manifestation of Focus in Tokyo Japanese: The Role of Intensity, Duration, and F0). *Onsei Gengo*, 3. 29-38.
- Korkmaz, Zeynep. 1960. Türkçede Ek Yığılması Olaylarının Meydana Gelişi Üzerine. IN Levend, Agah Sırrı ed. 1960. *Türk Dili Araştırmaları Yıllığı Belleten 1960*. Ankara: Türk Dil Kurumu. 173-180.
- Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.
- König, Ekkehard. 1991. *The Meaning of Focus Particles: A comparative perspective*. London and New York: Routledge.
- Kudō, Rikio. 1950. Jōdai ni Okeru Kakujoshi ni no Senzai to Shōryaku. *Kokugo Kokubun*, 46, 5. 309-321.
- Kumagai, Tomoko. 1997. Kurikaesu. IN Sakuma, M. et al eds. 1997. *Bushō, Danwa no Shikumi*. Tokyo: Ōfū. 38-48.
- Kuno, Susumu. 1978. *Danwa no Bunpō*. Tokyo: Taishūkan.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge, New York, Melbourne: Cambridge University Press.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge; New York; Melbourne: Cambridge University Press.

- Lange, Roland. A. 1970. *The Phonology of Eighth-Century Japanese*. Tokyo: Sophia University.
- Lazard, Gilbert. 1970. Persian and Tajik. IN Seboek, Thomas A. ed. 1970. *Current Issues in Linguistics*, volume 6: *Linguistics in South West Asia and North Africa*. The Hague; Paris: Mouton. 64-96.
- Lees, Robert B. 1961. *The Phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.
- Lewis, Geoffrey L. 1967. *Turkish Grammar*. Oxford: Oxford University Press.
- Lewis, Geoffrey L. 1999. *The Turkish Language Reform: A Catastrophic Success*. Oxford University Press.
- Liberman, Mark / Sag, Ivan. 1974. Prosodic Form and Discourse Function. IN La Galy, Fox and Bruck eds. 1974. *Papers from the Tenth Regional Meeting*. Chicago: Chicago Linguistic Society. 416-427.
- Mahmudov, M. / Berdiev, B. eds. 1989. *Lughati muxtasari lahjahoi Buxoro*. Dušanbe: Doniś.
- Mahmudov, N. / Nurmonov, A. 1995. *O‘zbek tilining nazariy grammatikasi*. Toshkent: O‘qituvchi.
- Mahootian, Shahrzad. 1997. *Persian*. London; New York: Routledge.
- Maniëzov, Abduqodir / Mirzoev, Abdusattor. 1991. *Lughati imlo*. Dušanbe: Maorif.
- Marslen-Wilson, William D. / Tyler, Lorraine K. 1997. Dissociating Types of Mental Computation. *Nature*, 387. 592-594.
- Martin, Samuel E. 1970. Junctural Cues to Ellipsis in Japanese. IN Jakobson, Roman / Kawamoto, Shigeo eds. 1970. *Studies in General and Oriental Linguistics: Presented to Shirô Hattori on the Occasion of his Sixtieth Birthday*. Tokyo: TEC Company Ltd. 429-445.
- Martin, Samuel. E. 1975. *A Reference Grammar of Japanese*. New Haven, London: Yale University. (Republished by Charles E. Tuttle Company of Rutland, Vermont & Tokyo in 1988.)

- Maruyama, Naoko. 1995. Hanashikotoba ni Okeru Mujoshikakuseibun no Kaku. *Keiryō Kokugogaku*, 19, 8. 365-380.
- Matsumoto, Hirotake. 1982. Ryūkyū Hōgen no Shukaku Hyōgen no Mondaiten. *Kokubungaku Kaishaku to Kanshō*, 47, 9. 178-185.
- Matsushita, Daizaburō. 1901. *Nihon Zokugo Bunten*. Tokyo: Seishidō.
- Matsushita, Daizaburō. 1961. *Hyōjun Nihon Kōgohō*. Tokyo: Hakuteisha.
- Matsuzawa, Tetsuo. 1985. Use of Numbers by a Chimpanzee. *Nature*, 315, 57-59.
- Matsuzawa, Tetsuo. 1996. Evolution of Intelligence: A View from Linguistic and Cognitive Functions of Chimpanzees. IN McGrew, W.C. / Marchant, L.F. / Nishida, T. eds. 1996. *Great Ape Societies*. Cambridge University Press. 196-209.
- Maynard, Senko K. 1997. *Danwa Bunseki no Kanōsei: Riron, Hōhō, Nihongo no Hyōgensei*. Tokyo: Kurosio Publishers.
- McCawley, James D. 1977. Accent in Japanese. IN *Studies in Stress and Accent: Symposium on Stress and Accent, 1976*, University of Southern California. Los Angeles: Department of Linguistics, University of Southern California. 261-302.
- Melex, N. A. 1968. *Gižduvanskij govor tadžikskogo jazyka (Avtoref. kand. diss.)*. Leningrad.
- Mikami, Akira. 1963. *Nihongo no Ronri*. Tokyo: Kurosio.
- Miller, Max. 1978. Pragmatic Constraints on the Linguistic Realization of ‘Semantic Intentions’ in Early Child Language. IN Waterson, Natalie / Snow, Cathrine. eds. 1978. *The Development of Communication*. Chichester, New York, Brisbane, Toronto: John Wiley & Sons. 453-467.
- Miller, Roy Andrew. 1970. Introduction. IN Miller, Roy Andrew ed. 1970. *Bernerd Bloch on Japanese*. New Haven / London: Yale University Press. ix-xli.
- Miller, Roy Andrew. 1971. *Japanese and the Other Altaic Languages*. Chicago; London: the University of Chicago Press.
- Minashima, Hiroshi. 1993. Nihongo no Kakujoshi ‘o’ no Shōryaku ni Tsuite: Yūseisei to teisei no kanyo no kanōsei. *Gengogaku Ronsō Tokubetsugō*. 58-71.

- Mirzaev, M. M. 1969. O‘zbek tilining Buxoro grappa shevalari. Toshkent: O‘zbekiston SSR «Fan» Nashriyoti.
- Miura, Ichiro. 1992. Perception of Linguistic Focus on Japanese. IN Tajima, Ikudō / Tanba, Kazuya eds. 1992. Nihongo Ronkyū 1. Gengogaku to Sono Shūhen. Osaka: Wasen Shoin.
- Mizutani, Shizuo. 1996. Gendaigo no Kaku: Shiron. Keiryō Kokugogaku, 20, 7. 283-303.
- Muranaka, Toshiko / Hara, Noriko. 1994. Features of Prominent Particles in Japanese Discourse: Frequency, Functions and Acoustic Features. ICSLP 94, Yokohama. 395-398.
- Nakama, Mitsunari. 1984. Miyako Hōgen no Bunpō. Kokubungaku Kaishaku to Kanshō, 49,1. 172-182.
- Nakazima, Sei. 1977. A Comparative Study of the Speech Development of Japanese and American Children (Part Eight): Structure and function of one-word-sentences. Onsei Kagaku Kenkyū, 11. 35-42.
- Napoli, Donna Jo. 1993. Syntax: Theory and Problems. New York, Oxford: Oxford University Press.
- Napoli, Donna Jo. 1996. Linguistics: An Introduction. New York, Oxford: Oxford University Press.
- Nawada, Tetsuo. 1989. Tajikugo. IN Kamei, Takashi ed. 1989. Gengogaku Daijiten. Tokyo: Sanseido. 613-614.
- Nemenova, R. L. 1955. Kratkij očerk grammatiki tadžikskogo jazyka. IN Kalontarov, Ja. I. ed. 1955. Kratkij tadžiksko-russkij slovar’. Moskva: Gosudarstvennoe Izdatel’stvo Inostrannyx i Nacional’nyx Slovarj. 525-616.
- Ne’matov, Hamid (Ne”matov, ·Xamid / Nigmatov, Xamid). 1999. Personal communication.
- Niëzmuhammadov, B / Niëzi, Š. / Buzurgzoda, L. 1955. Grammatikai zaboni tojikī Qismi I: Fonetika va morfologiya baroi maktabhoi haftsola va miëna. Stalinobod: Našriëti davlatii tojikiston.
- Noda, Kōgo / Ozu, Yasujirō. 1956. Sōshun. unpublished scenario.

- Nohara, Mitsuyoshi. 1979. Amami Ōshima Hōgen no Bunrei. Ryūkyū no Hōgen, 5. Tokyo: Hōsei Daigaku Okinawa Bunka Kenkyūjo.
- Nurmanov, A. 1990. Presuppozitsional'nyj aspekt prostogo predloženija uzbekskogo jazyka. Sovetskaja tjurkologija, 6, nojabr'-dekabr'. 13-16.
- Ochs, Elinor. 1979. Introduction: What Child Language Can Contribute to Pragmatics. IN Ochs, Elinor / Schieffelin, Bambi B. 1979. New York, San Francisco, London: Academic Press. 1-17.
- Oflazer, Kemal. 1994. Two-level Description of Turkish Morphology. Literary and Linguistic Computing, 9, 2. 137-148.
- O'Grady, William. 1997. Syntactic Development. Chicago and London: The University of Chicago.
- Ōishi, Hatsutarō. 1959. Purominensu ni Tsuite: Tōkyōgo no kansatsu ni motodzuku oboegaki. IN Kokuritsu Kokugo Kenkyūjo ed. 1959. Kotoba no Kenkyū. Tokyo: Shūei Shuppan. Reprinted IN Shibata, Takeshi / Kitamura, Hajime / Kindaichi, Haruhiko eds. 1980. Nihon no Gengogaku dai 2 kan: On'in. Tokyo: Taishūkan. 594-612.
- Ōno, Susumu. 1978. Nihongo no Bunpō o Kangaeru. Tokyo: Iwanami.
- Oripov, Qobuljon / Obidova, Mahmuda. 1994. Ifodali O'qish. Toshkent: O'qituvchi.
- O'rinboev, B. 1982. O'zbek so'zlashuv nutqi. Toshkent: O'zbekiston SSR «Fan» Nashriyoti.
- Özkan, Nevzat. 1996. Gagavuz Türkçesi Grameri. Ankara: Türk Dil Kurumu.
- Öztürk, Rıdvan. 1997. Uygur ve Özbek Türkçelerinde Fiil. Ankara: Türk Dil Kurumu Yayınları.
- Palmer, F.R. 1981. Semantics 2nd edition. Cambridge: Cambridge University Press.
- Perry, John R. 1979. Uzbek Influence in Tajik Syntax: The Converb Construction. IN Clyne, Paul R. / Hanks, William F. / Hofbauer, Carol L. eds. 1979. The Elements, a parasession on linguistic units and levels, April 20-21, 1979 (Chicago Linguistic Society Conference on Non-Slavic Languages of the USSR). University of Chicago. 448-461.

- Perry, John R. 1996. From Persian to Tajik to Persian: Culture, Politics, and Law Reshape a Central Asian Language. IN NSL vol.8: Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics. Chicago: Chicago Linguistic Society, The University of Chicago. 279-305.
- Perry, John R. 1997. Script and Scripture: The Three Alphabets of Tajik Persian, 1927-1997. *Journal of Central Asian Studies*, 2, 1. 1-18.
- Pierrehumbert, Janet / Beckman, Mary. 1988. *Japanese Tone Structure*. Cambridge, Massachusetts; London, England: The MIT Press.
- Pinker, Steven. 1994. *The Language Instinct*. New York: William Morrow and Company, Inc.
- Pinker, Steven. 1997. Words and Rules in the Human Brain. *Nature* 387 (6633), 5 June 1997. 547.
- Plunkett, Kim. 2000. The Best of Both Worlds?. *New Scientist*, 165, 2222. 46-47.
- Plunkett, Kim / Juola, Patrick. 1999. A Connectionist Model of English Past Tense and Plural Morphology. *Cognitive Science*, 23, 4, pp. 463-490.
- Polinsky, Maria. 1999. Review of Knud Lambrecht 'Information structure and sentence form'. *Language*, 73, 3. 567-582.
- Potapova, T. D. 1978. Ob udarenii v tureckix toponimax: Opyt eksperimental'nogo issledovanija. *Sovetskaja tjurkologija*, 1978, 5, sentjabr'-oktjabr'. 79-91.
- Potapova, T. D. 1982. Slovesnoe udarenie v tureckom jazyke: K postanovke problemy. *Sovetskaja tjurkologija*, 1982, 3, maj-ijun'. 56-62.
- Prince, Ellen F. 1992. The ZPG Letter: Subjects, Definiteness, and Information-status. IN Mann, William C. / Thompson, Sandra A. eds. 1992. *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. Amsterdam; Philadelphia: John Benjamins Publishing Company. 295-325.
- Rahbari doniř (monthly journal). 1927.8.-(?)1931/1932. Tashkent (; Stalinabad?). [According to Ismatī (1929), the journal title was altered from the second issue from *Doniř-biniř* to *Rahbari doniř*].

- Rahmatullayev, Shavkat / Hojiyev, Azim. 1995. O‘zbek tilining imlo lug‘ati. Toshkent: O‘qituvchi.
- Rajabov, Nazar. 1996. O‘zbek shevashunosligi. Toshkent: O‘qituvchi.
- Ramat, Paolo. 1987. Linguistic Typology. Berlin; New York; Amsterdam: Mouton de Gruyter.
- Rastorgueva, V. S. 1956. Očerki po tadžikskoj dialektologii Vypusk 3: Leninabadsko-kanibadamskaja grupa severnyx tadžikskix govorov. Moskva: Izdatel’stvo akademii nauk SSSR.
- Rastorgueva, V. S. 1964. Opyt sravnitel’nogo izučeniya tadžikskix govorov. Moskva: Izdatel’stvo «Nauka».
- Rastorgueva, V. S. 1992. A Short Sketch of Tajik Grammar. Translated by Herbert H. Paper. Research Institute for Inner Asian Studies, Indiana University.
- Rastorgueva, V. S. / Bakaev, C. X. / Isaev, M. I. / Kerimova, A. A. / Pirejko, L. A. 1970. Different Types of Bilingualism among Iranian Peoples of the USSR. International Congress of Anthropological and Ethnological Sciences (7th, 1964, Moskva). Moskva: «Nauka». 715-721.
- Romaine, Suzanne. 1988. Pidgin and Creole Languages. London and New York: Longman.
- Rumelhart, D. E. / McClelland, 1986. On Learning the Past Tense of English Verbs. IN McClelland, James L. / Rumelhart, David E. and the PDP Research Group. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 2: Psychological and Biological Models. Cambridge, Massachusetts; London, England: The MIT Press. pp. 216-271.
- Rustamov, Šarofiddin / Ghafforov, Razzoq eds. 1985. Grammatikai zaboni adabii hozirai tojik Jildi I: Fonetika va morfologiya. Dušanbe: Našriëti «Doniš».
- Rustamov, Šarofiddin / Ghafforov, Razzoq eds. 1986. Grammatikai zaboni adabii hozirai tojik Jildi II: Ibora va sintaksisi jumlahoi sodda. Dušanbe: Našriëti «Doniš».
- Rutherford, E. 1988. Sarum. New York: Ivy Books.
- Sakuma, Mayumi. 1997. Bunshō, Danwa no Matomari. IN Sakuma, M. et al eds. 1997. Bunshō, Danwa no Shikumi. Tokyo: Ōfū. 181-207.

- Sanseidō Henshūjo ed. 1991. Atarashii Kokugo Hyōki Handobukku. Tokyo: Sanseidō.
- Schlegel, August Wilhelm von. 1818. Observations sur la langue et la littérature provinciales. Paris.
- Schwarz, Henry G. 1984. Tajik IN The Minorities of Northern China: A Survey (Studies on East Asia Vol. 17). Bellingham: Western Washington University Center for East Asian Studies. 217-227.
- Segalowitz, Sidney J. / Lane, Korri C. 2000. Lexical Access of Function versus Content Words. *Brain and Language*, 75. 376-389.
- Selen, Nevin. 1973. Entonasyon Analizleri. Ankara: Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Yayınları.
- Selkirk, E. O. 1984. Phonology and Syntax: The relation between sound and structure. Cambridge, Mass.: MIT Press.
- Selkirk, Elizabeth. 1995. Sentence Prosody: Intonation, Stress, and Phrasing. IN Goldsmith, John A. ed. 1995. *The Handbook of Phonological Theory*. Cambridge, Mass: Blackwell. 550-569.
- Sezer, Engin. 1983. On Non-final Stress in Turkish. *Journal of Turkish Studies*, 5. 61-69.
- Shibata, Takeshi / Kitamura, Hajime / Kindaichi, Haruhiko. 1980. *Nihon no Gengogaku dai 2 kan: On'in*. Tokyo: Taishūkan.
- Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge; New York; Port Chester; Melbourne; Sydney: Cambridge University Press.
- Shibatani, Masayoshi / Bynon, Theodora. 1995. Approaches to Language Typology. IN Shibatani, Masayoshi / Bynon, Theodora eds. 1995. *Approaches to Language Typology*. Oxford: Clarendon Press. 1-48.
- Shōgaito, Masahiro. 1989. Churuku Shogo. IN Kamei, Takashi ed. 1989. *Gengogaku Daijiten Vol. 2*. Tokyo: Senseidō. 937-950.
- Shukurov, Shamsudin. 1976. O‘zbek tilida fe‘l zamonlari taraqqiyoti. Toshkent: Fan.
- Sjoberg, André F. 1962. *Uzbek Structural Grammar*. Bloomington: Indiana University / The Hague: Mouton & Co.

- Soper, John. 1996. *Loan Syntax in Turkic and Iranian* (Revised and edited by Andras J. E. Bodroglegeti). Bloomington, Indiana: Eurolingua.
- Sproat, Richard William. 1992. *Morphology and Computation*. Cambridge, Massachusetts, London, England: The MIT Press.
- Sugito, Miyoko. 1969. Dōtai Sokutei ni yoru Nihongo Akusento no Kaimei. *Gengo Kenkyū*, 55: 14-39
- Sugito, Miyoko. 1980. Akusento, Intoneeshon no Hikaku. IN *Nichieigo Hikaku Kōza 1: Onsei to Keitai*. Tokyo: Taishūkan.
- Sugito, Miyoko. 1982. *Nihongo Akusento no Kenkyū*. Tokyo: Sanseidō.
- Suzuki, Shigeyuki. 1972. *Nohongo Bunpō-Keitairon*. Tokyo: Mugi Shobō.
- Swift, Lloyd B. 1962. *Some Aspects of Stress and Pitch in Turkish Syntactic Patterns*. Uralic and Altaic Series, vol. 13. Bloomington: Indiana University.
- Swift, Lloyd B. 1963. *A Reference Grammar of Modern Turkish*. Bloomington: Indiana University.
- Szabo, Laszlo. 1994. Agglutination: The Examples of Hungarian. IN *Papers from the Eighteenth Annual Meeting Atlantic Provinces Linguistic Association*. 116-122.
- Szatrowski, Polly, 1993. Nihongo no Danwa no Kōzō Bunseki: Kanyū no Sutoratejii no Kōsatsu. Tokyo: Kurosio.
- Taft, Marcus / Forster, Kenneth I. 1975. Lexical Storage and Retrieval of Prefixed Words. *Journal of Verbal Learning and Verbal Behavior*, 14. 638-647.
- Takami, Ken-ichi. 1995. Kinōteki Kōbunron ni yoru Nichieigo Hikaku. Tokyo: Kurosio.
- Takeda, Shōichi / Ichikawa, Kaoru. 1991. Nihongobun Onsei ni Okeru Puominensu no Inritsuteki Tokuchō no Kaiseki. *Nihon Onkyō Gakkaishi*, 47, 6. 386-396.
- Tanba, Tetsuya. 1993. Mujoshikaku no Kinō: Shudai to Kaku to Gojun. *Kokugo Kokubun*, 58, 10. 38-57.
- Tekin, Talât. 1991. *Türkçede Kelime Vurgusu Kuralları*. *Dilbilim Araştırmaları*, 1991. Hitit Yayınevi. 1-4.
- Teru, Hiromoto. 1982. Kikaijima no Hōgen. *Kokubungaku Kaishaku to Kanshō*, 47, 9. 74-86.

- Tomlin, Russel S. / Forrest, Linda / Pu, Ming Ming / Kim, Myung Hee. 1997. Discourse Semantics. IN van Dijk, Teun A. ed. 1997. Discourse as Structure and Process. London; Thousand Oaks; New Delhi: Sage Publications. 63-111.
- Tospulotuf, M. / Gitelmaxer, B. / Klimcitskij, S. 1932. Zaboni toçikī baroji avrupoijon. Stalinobod: Naşri Toçik.
- Türel, Orhan. 1968. Torukogo Bunpō·Kaiwa [Turkish Grammar-Conversation]. Tokyo: Maruzen.
- Timurtaş, Faruk K. 1977. Eski Türkiye Türkçesi: XV yüzyıl Gramer-Metin-Sözlük. İstanbul: İstanbul Üniversitesi Edebiyat Fakültesi Basımevi.
- Tsukishima, Hiroshi. 1964. Kokugogaku. Tokyo: Tōkyō Daigaku Shuppankai.
- Türk Dil Kurumu. 1942. Şeyhî Divanı. Ankara: Türk Dil Kurumu.
- Turska gramatika (in Bulgarian).
- Uemura, Yukio. 1972. Ryūkyū Hōgen Nyūmon. Gengo Seikatsu, 251. 20-37.
- Underhill, Robert. 1976. Turkish Grammar. Cambridge; London: The MIT Press.
- Unger, J. Marshall. 1990. Japanese and What Other Altaic Languages?. IN Bald, Philip ed. Linguistic Change and Reconstruction Methodology. Berlin; New York: Mouton de Gruyter. 547-561.
- Ülkütaşır, M. Şakir. 1973. Aratürk ve Harf Devrimi. Ankara: TDK.
- Vahhobov, T. 1980. Oid ba ba"ze mas"alahoi soxtmoni zabon dar solhoi 20-30. IN Ghafforov, R. et al eds. 1980. Zabonşinosii tojik. Dušanbe: Naşriëti «Doniş». 237-247.
- Vance, Timothy J. 1987. An Introduction to Japanese Phonology. New York: State University of New York Press.
- van Dijk, Teun A. 1981. Studies in the Pragmatics of Discourse. The Hague; Paris; New York: Mouton Publishers.
- van Dijk, Teun A. 1993. Elite Discourse and Racism. London: Sage Publications, Inc.
- van Dijk, Teun A. ed. 1997. Discourse as Structure and Process (Discourse Studies: A multidisciplinary introduction volume 1). London; Thousand Oaks; New Delhi: Sage Publications.

- Vardul, I. F. 1989. Accent and Word Boundary in Japanese. IN Permanent International Altaistic Conference (28th, 1985, Venice, Italy). 1989. Proceedings of the XXVIII Permanent International Altaistic Conference: Venice, 8-14 July 1985. Wiesbaden: Distributed by O. Harrassowitz. 241-248.
- Vendetti, Jennifer J. 1997. Japanese ToBI Labelling Guidelines. Ohio State University Working Papers in Linguistics, 50. 127-162.
- Vendetti, Jennifer J. / Jun, Sun-Ah / Beckman, Mary. 1996. Prosodic Cues to Syntactic and Other Linguistic Structures in Japanese, Korean, and English. IN Morgan, James L. / Demuth, Kathrine eds. 1996. Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers. 287-311.
- Venneman, T. 1975. Topic, Sentence Accent, and Ellipsis: A proposal for their formal treatment. IN Keenan, E. L. ed. 1975. Formal Semantics of Natural Language. Cambridge: Cambridge University Press. 313-328.
- Vygotsky, L.S. 1962. Thought and Language. Cambridge, Mass.: M.I.T. Press.
- Wade, Terence Leslie Brian. 1992. A Comprehensive Russian Grammar. Oxford; Cambridge: Blackwell Publishers.
- Waksler, Rachelle. 1999. Cross-linguistic Evidence for Morphological Representation in the Mental Lexicon. *Brain and Language*, 68. 68-74.
- Watanabe, Yasuko. 1986. Two Kinds of Ellipsis in Japanese Discourse: A Quantitative Text Study. *Studies in Language*, 10, 2, 337-351.
- Weisenburger, Janet L. 1976. A Choice of Words: two-year-old speech from a situational point of view. *Journal of Child Language*, 3, 275-281.
- Windfuhr, Gernot L. 1979. Persian Grammar: History and State of its Study. The Hague; Paris; New York: Mouton Publishers.
- Wittgenstein, Ludwig. 1958. Philosophical Investigations. Trans. G. E. M. Anscombe. 3rd edn. New York: Macmillan.
- Yamamoto, Mikio / Kobayashi, Satoshi / Nakagawa, Seiichi. 1992. Onseitaiwabun ni okeru Joshiochi · Tōchi no Bunseki to Kaiseki Shuhō. *Jōhōshorigakkai Ronbunshi*, 33, 11. 1322-1330.

- Yaman, Ertuğul / Mahmud, Nizamidden. 1998. Özbek Türkçesi - Türkiye Türkçesi ve Türkiye Türkçesi - Özbek Türkçesi Karşılıklar Kılavuzu. Ankara: Türk Dil Kurumu.
- Yaşaroğlu, Erdil. 1998. Komik Kitap: Komikaze 4. İstanbul: LeMan Basın Yayın Ltd. Şti.
- Zehnī, T. 1987. Az ta”rixī leksikai zabonī tojikī. Dušanbe: Naşriēti «Doniř».
- Zeynalov, Ferhat. 1993. Türk Lehçelerinin Karşılaştırmalı Dilbilgisi (Translated from the Azerbaijani by Yusuf Gedikli). İstanbul: Cem Yayınevi.
- Zimmer, Karl. 1964. Affixal Negation in English and Other Languages: An Investigation of Restricted Productivity. Word Monograph No. 5. New York.
- Zhongguo da bai ke quan shu zong bian ji wei yuan hui ed. 1993. Ren wen bai ke quan shu da xi: Min zu bai ke quan shu. Beijing: Zhongguo da bai ke quan shu chu ban she.