

Christopher, Nadezda (2018) Aspects of Information Structure in Kazakh : the Dynamic Syntax Approach. PhD thesis. SOAS University of London. <http://eprints.soas.ac.uk/32245>

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this thesis, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g. AUTHOR (year of submission) "Full thesis title", name of the School or Department, PhD Thesis, pagination.

Aspects of Information Structure in Kazakh – the Dynamic Syntax Approach

Nadezda Christopher

Thesis submitted for the degree of PhD

2018

Department of Linguistics
SOAS, University of London

Abstract

The Kazakh language is an under-researched Turkic language spoken in the Central Asian state of Kazakhstan and some neighbouring countries. While the grammar of this language is fairly well described, its information structural characteristics have not been examined in detail in the literature to date. This thesis aims to start filling this research gap by providing detailed descriptions of: the relation between information structure and word order; topic markers; and a pragmatically significant particle.

Original, contextualised language examples are used to reject previous limited and rigid understanding of the relation between information structure and word order in Kazakh. It is shown that the information structural configurations of a Kazakh sentence are far more diverse than had been assumed. This conclusion is not only a revelation in its own right, but can also serve as a foundation for further research on the information structure of Kazakh, and other under-researched Turkic languages.

This thesis also provides the first detailed descriptions of the three Kazakh topic markers. Numerous examples of their uses are presented in order to demonstrate the differences in these markers' distribution and meaning. Several grammaticalisation processes related to these topic markers are revealed; it is proposed that these processes are currently at different stages of progress.

Pragmatically significant particle *ǵoj* is examined in detail for the first time: its distribution and meaning are illustrated with contextualised examples from various sources. It is posited that there are two syntactically diverse variants of this item which do, however, share the same existential semantics.

The theoretical framework of Dynamic Syntax is employed throughout the thesis to underpin the first formal analyses of the phenomena under discussion.

Acknowledgements

I am grateful to the Wolfson Foundation for their generous financial support, without which this project would not have been possible. I am also grateful for grants from SOAS, University of London, and from Indiana University, Bloomington, which contributed to the costs of presenting my research at conferences.

I am deeply grateful to my primary Supervisor – Irina Nikolaeva – for generously sharing her linguistic and academic expertise with me, for her support, encouragement and belief in me, even at times when I did not believe in myself. My heartfelt thanks also go to my supervisory committee – Lutz Marten and Chris Lucas – who ignited the passion for Dynamic Syntax in my undergraduate self 6 years ago, and have been helping me maintain it ever since through many inspiring and engaging conversations. A big thank you goes to Monik Charette for always having time to talk about my thesis and my life – it meant and helped a lot.

I am thankful to Rahima Abduvalieva for her kind assistance in organising my research and life in Astana, and, more specifically, for putting me in touch with Duman Aitmagambetov at the Eurasian National University who helped immensely with my language learning and research; and with the wonderful Suzhanov family who treated me like one of their own. I am grateful to them, and to all my Kazakh friends and language consultants – *köp rakhmet sizderge!*

While I have no relatives in the UK, I have been lucky to meet some wonderful people who have become like family to me: Julia Malyshenko, Jasneet Flora, Natalie Dalla Volta, Francisca Everduim, and Kathryn Riley (in order of appearance). Thank you for all your love and support, and for accepting me as I am. I am grateful to Anna Lobova (St. Petersburg), and Galina Beresten' (Krasnoyarsk) for always being just one message or phone call away, and for loving and supporting me across the (thousands of) miles.

I dedicate this thesis to my parents – Ljubov' and Aleksey Cherjominy (Любовь и Алексей Черёмины) who have always believed me and in me, and loved me unconditionally. *Мамуля и Папуля – это посвящается вам!*

Table of Contents

DECLARATION	2
ABSTRACT	3
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS	5
LIST OF FIGURES AND TABLES	9
ABBREVIATIONS	10
1. INTRODUCTION	11
1.1 THE KAZAKH LANGUAGE.....	11
1.1.1. Socio-linguistic Information	12
1.1.2. General Features.....	14
1.1.3. Nouns and Pronouns.....	15
1.1.4. Postpositions.....	19
1.1.5. Verbal Agreement	21
1.1.6. Converbs and Present Tense	23
1.1.7. Participles and Relative Clauses.....	26
1.1.8. Verbal Tense and Mood.....	28
1.1.9. Conditionals and Subjunctives	30
1.1.10. Interrogative Clauses	32
1.1.11. Existential Clauses.....	34
1.2. OBJECTIVES OF THE THESIS.....	36
1.3. METHODOLOGY	39
1.4. STRUCTURE OF THE THESIS.....	41
2. THEORETICAL BACKGROUND	44
2.1. INFORMATION-STRUCTURAL CATEGORIES	45
2.1.1. Topic.....	46
2.1.2. Focus	57
2.1.3. Contrast.....	63
2.1.4. Specificity and Definiteness	67

2.1.5. The Role of Information Structure in the Grammar	70
2.2. DYNAMIC SYNTAX	72
2.2.1. The Principles of the Framework	73
2.2.2. The Dynamic Syntax Tools	75
2.2.3. Tree Growth Process.....	78
2.2.4. Computational Rules in the Tree Construction Process	83
2.2.5. Adjunction Rules and Unfixed Nodes	90
2.2.6 Lexical Information and Lexical Entries	98
2.2.7. Modelling SOV in Dynamic Syntax	100
2.2.8. Intonation Phrase Boundary.....	109
2.2.9. Dynamic Syntax Approach to Information Structure	118
2.3. CONCLUSIONS	122
3. WORD ORDER AND INFORMATION STRUCTURE.....	123
3.1. INTRODUCTION.....	123
3.2. NEUTRAL WORD ORDER IN KAZAKH – SOV	127
3.2.1. The Notion of Basic Word Order.....	127
3.2.2. Minimal Context Test.....	130
3.2.3. Unmarked Arguments Test.....	133
3.2.4. Typological Observations for SOV Languages.....	135
3.2.5. Information-structural Properties of SOV	137
3.3. SENTENTIAL POSITIONS AND IS.....	140
3.3.1. The Immediately Preverbal Position.....	141
3.3.2. IPV Position to Focus Field in Turkish	142
3.3.3. Contrastive Focus Field and New Information Focus Position in Kazakh	144
3.3.4. Sentence-initial position	149
3.3.5. Post-verbal Position in Kazakh	152
3.3.6. Conclusions	157
3.4. INFORMATION-STRUCTURAL PROPERTIES OF SCRAMBLED WORD ORDERS IN KAZAKH.....	158
3.4.1. OSV.....	158
3.4.2. Verb-medial Word Orders – SVO and OVS	165
3.4.3. Verb-initial Word Orders – VSO and VOS	169
3.4.4. Conclusions	173
3.5. KAZAKH WORD ORDER IN DYNAMIC SYNTAX	174
3.5.1. Modelling Kazakh in DS.....	174
3.5.2. SOV in DS.....	182

3.5.3. OSV in DS.....	188
3.5.4. Verb-medial and Verb-initial Word orders	192
3.5.5. Conclusions	196
3.6. CONCLUSIONS	197
4. DIFFERENTIAL OBJECT MARKING IN KAZAKH	199
4.1. INTRODUCTION.....	199
4.2. DIFFERENTIAL OBJECT MARKING IN KAZAKH	204
4.2.1. Previous Analyses of Differential Object Marking in Turkish	213
4.2.2. An Alternative Approach to Differential Case Marking	228
4.3. KAZAKH DOM - A DYNAMIC SYNTAX ANALYSIS.....	232
4.4. CONCLUSIONS	240
5. TOPIC MARKERS.....	242
5.1. INTRODUCTION.....	242
5.2. TOPIC MARKER <i>Še</i>	243
5.2.1. Kazakh Particle <i>Še</i> – Previous Research	244
5.2.2. Turkish Topic Marker <i>Ise</i>	250
5.2.3. Hypothesis on the Diachrony of <i>Še</i>	256
5.2.4. Identifying Two Types of <i>Še</i>	257
5.2.5. Meaning and Distribution of the Contrastive Topic Marker <i>Še</i>	262
5.2.6. Meaning and Distribution of the Contrastive Question Particle <i>Še</i>	265
5.2.7. Conclusions	268
5.3. TOPIC MARKER <i>BOLSA</i>	269
5.3.1. Topic Marker <i>Bolsa</i> in Related Languages	269
5.3.2. Kazakh <i>Bolsa</i> – Identifying the Topic Marker and its Distribution	271
5.3.3. Meaning of the Topic Marker <i>Bolsa</i>	279
5.3.4. Conclusions	280
5.4. TOPIC MARKER <i>DEGEN</i>	281
5.4.1. <i>Verba Dicendi</i> -related Topic Markers	281
5.4.2. Previous Research on <i>Degen</i>	282
5.4.2. Non-topic-marking Uses of <i>Degen</i>	292
5.4.3. Topic Marker <i>Degen</i> – Distribution and Meaning	300
5.4.4. Conclusions	305
5.5. DYNAMIC SYNTAX APPROACH TO KAZAKH TOPIC MARKERS	306
5.5.1. DS Analysis of Topic Markers.....	306

5.5.2. Conditionals as Topic Markers	312
5.5.3. Conclusions	317
5.6. CONCLUSIONS	318
6. PARTICLE ĞOJ	320
6.1. INTRODUCTION	320
6.2. PREVIOUS RESEARCH	321
6.3. ĞOJ IN IMPERATIVE CLAUSES	325
6.4. SYNTACTIC DISTRIBUTION OF ĞOJ	331
6.4.1. Post-predicative Ğoj	331
6.4.2. Post-nominal Ğoj	333
6.5. PRAGMATIC CONTRIBUTION OF ĞOJ	342
6.5.1 Pragmatic Contribution of Post-predicative Ğoj	342
6.5.2. Pragmatic Contribution of the Post-nominal Ğoj	350
6.6. SEMANTICS OF ĞOJ	352
6.6.1. Russian <i>Že</i>	353
6.6.2. Sanzhi Dargwa = <i>q'al</i>	359
6.6.3. Tundra Yukaghir Particle <i>Mə(r)=</i>	365
6.6.4. Ğoj as an Existential Operator	367
6.7. DYNAMIC SYNTAX ANALYSIS OF ĞOJ	373
6.7.1. The Dynamics of Kazakh Relative Clauses	374
6.7.2. The DS Analysis of Ğoj-clefts	382
6.7.3. The DS Approach to the Post-verbal Particle Ğoj	390
6.7.4. Unified Lexical Entry for Ğoj and Conclusions	395
6.8. CONCLUSIONS	397
7. MAIN RESULTS, CONTRIBUTION, AND FURTHER RESEARCH QUESTIONS	399
7.1. SUMMARY OF THE MAIN RESULTS	399
7.2 CONTRIBUTION TO THE FIELD	402
7.3. QUESTIONS FOR FURTHER RESEARCH	405
BIBLIOGRAPHY	408

List of Figures and Tables

Figure 1: Sentence-initial NP with an IP boundary	117
Figure 2: Sentence-initial NP – no IP boundary	117
Figure 3: Sentence-initial NP with an IP boundary	384
Figure 4: Sentence-initial NP – no IP boundary	384
Table 1: Transliteration of the Kazakh Alphabet	13
Table 2: Kazakh Case System	17
Table 3: Possessive Suffixes	18
Table 4: Declension of Personal Pronouns	19
Table 5: Declension of Demonstrative Pronouns <i>bul</i> ‘this’ and <i>sol</i> ‘that’	19
Table 6: The Pronominal Agreement Paradigm	22
Table 7: The Possessive Agreement Paradigm	22
Table 8: Declension of Interrogative Pronouns <i>kim</i> ‘who’ and <i>ne</i> ‘what’	33
Table 9: Lambrecht’s IS Sentence Types.....	59
Table 10: Cross-classification of definiteness and specificity.....	68
Table 11: Types in Dynamic Syntax	78
Table 12: Information structure of an SOV sentence (to be revised).....	140
Table 13: Information structure of an SOV sentence (revised)	157
Table 14: Information structure of an OSV sentence (to be revised)	162
Table 15: Information structure of an OSV sentence (revised)	164
Table 16: Information structure of an SVO sentence (to be revised)	166
Table 17: Information structure of an OVS sentence (to be revised)	167
Table 18: Information structure of an SVO sentence (revised)	169
Table 19: Information structure of an OVS sentence (revised)	169
Table 20: Information structure of verb-initial sentences	173
Table 21: Clause positional distribution of <i>še</i>	260
Table 22: Distribution of <i>še</i> with grammatical categories	261
Table 23: Conjugation of <i>bol-</i> in the conditional mood	271
Table 24: Conjugation of the verb <i>de-</i> in the <i>-ǴAn</i> past tense	293

Abbreviations

1	first person	IMP	imperative
2	second person	INF	infinitive
3	third person	INJ	interjection
ABL	ablative	INST	instrumental
ACC	accusative	IPFV	imperfective
ADD	additive	LOC	locative
ADJ	adjective	M	masculine
ADV	adverb	MOD	modal particle
ADVZ	adverbialiser	N	neuter
ATTR	attributive	NEG	negative
AUX	auxiliary	NMLZ	nominaliser
CAUS	causative	NOM	nominative
CF	counterfactual	OBJ	object
CL	clitic	PART	particle
COLL	collective	PASS	passive
COM	comitative	PFV	perfective
COMP	complementiser	PL	plural
COND	conditional	POL	polite
CONJ	conjunction	POSS	possessive
COP	copula	PRET	preterite
CQ	contrastive question	PROG	progressive
CVB	converb	PRS	present
DAT	dative	PRT	partitive
DEICT	deictic	PST	past
ERG	ergative	PTCP	participle
EVID	evidential	Q	question
EXCL	exclamatory	SBJV	subjunctive
EXST	existential	SG	singular
F	feminine	SUBJ	subject
FM	focus marker	TM	topic marker
FUT	future	TOP	topic
GEN	genitive	VN	verbal noun

1. Introduction

This thesis examines information structure of the Kazakh language with particular focus on word order, topic markers *še*, *bolsa* and *degen*, and particle *ğoj*, none of which have been adequately, or in some cases at all, described or examined in the literature. Despite being one of the widely spoken Turkic languages of Central Asia, Kazakh is extremely under-researched, and this thesis fills only a fraction of this large gap in linguistic literature.

As far as the information-structural characteristics of Kazakh are concerned, this thesis moves away from the simplistic view of Kazakh information structure in which topics are rigidly assigned to the sentence-initial and focuses – to the immediately preverbal positions. It is proposed that a more fine-grained approach is required if we are to produce a full description of the relation between word order and information structure in Kazakh.

This chapter provides some introductory information on the Kazakh language and on this thesis and proceeds as follows: Section 1.1 contains some socio-linguistic information and introduces some aspects of Kazakh grammar with attention given specifically to categories directly relevant to this thesis; Section 1.2 presents the main objectives of this thesis; Section 1.3 discusses the methods whereby the data used in this thesis were collected; in Section 1.4, synopses of all the succeeding chapters are given.

1.1 The Kazakh Language

This section presents some introductory information on the Kazakh language including aspects of its socio-linguistic and grammatical characteristics. The former are provided in 1.1.1. I limit the latter to the aspects of grammar that are specifically relevant for rest of this thesis namely: 1.1.2 considers the general features characteristic of Kazakh and some related languages; 1.1.3 provides some information on nouns and pronouns; 1.1.4. covers postpositions; in 1.1.5 verbal agreement paradigms are presented, while 1.1.6 looks at some of the converbial and present tense

forms. In 1.1.7 some of the participial verb forms are introduced, together with relative clauses formation, and 1.1.8 covers aspects of tense and mood indicators. Conditional and subjunctive forms are discussed in 1.1.10, and existential clauses in 1.1.11.

1.1.1. Socio-linguistic Information

Kazakh is the official state and national language of the Republic of Kazakhstan, and is spoken by around 13 million people, most of whom (around 10 million) reside in Kazakhstan (Smailov 2011), while the rest form Kazakh ethnic minorities in China (around 1.25 million), Uzbekistan (around 1 million), with smaller Kazakh-speaking communities also found in Russia, Turkmenistan, Mongolia, and Kyrgyzstan (Simons and Fenning, 2018).

Kazakh has been the official state language of the Republic of Kazakhstan since the declaration of its independence from the USSR in December 1991. Prior to its independence the Kazakh Soviet Socialist Republic followed the USSR language policy, which, while claiming respectful treatment of all national languages, was unofficially, but explicitly aimed at enforcing Russian as the lingua franca of the USSR from around the late 1920s (Isaev 1970).

The dominance of the Russian language caused long-lasting damage to the prestige and status of the Kazakh language, which, in turn, led to a steep decline in its use. The first language-related laws of the independent Republic of Kazakhstan reflected the weak position of Kazakh, as Russian had to be recognised as the second official language that is used on equal grounds with the Kazakh language.

The writing system of Kazakh is relatively new – it was developed in the early twentieth century by order of the Soviet Russian administration; prior to this Kazakh was an oral language due to the nomadic and seminomadic way of life of the cattle-breeding tribes who spoke it (Demirci 2006). The first alphabet was based on the modified Arabic script until the late 1920s, when all the Turkic languages of the USSR were switched to the Roman alphabet. About a decade later, the modified version of the Cyrillic alphabet was developed for Kazakh.

This modified Cyrillic alphabet is still used in Kazakhstan, although in 2012 plans were announced to change the Kazakh writing system once again to a Roman-based alphabet. Two versions of the new alphabet were released in 2017, with another released in February 2018. It appears that this February 2018 version will be the final, official version of the new alphabet. In this thesis I use a transliteration system loosely based on the Turkish alphabet – it is presented in Table 1 below.

Table 1. Transliteration of the Kazakh Alphabet

Kazakh Alphabet	Transliteration	Kazakh Alphabet	Transliteration	Kazakh Alphabet	Transliteration
а	a	н	n	ъ	''
ә	ä	ң	ñ	ь	'
б	b	о	o	э	ε
в	v	ө	ö	ю	ju
г	g	п	p	я	ja
ғ	ğ	р	r		
д	d	с	s		
е	e	т	t		
ё	jo	у, ұ	u		
ж	ž	ү	ü		
з	z	ф	f		
и, і	i	х, һ	h		
й	j	ц	ts		
к	k	ч	č		
қ	q	ш	š		
л	l	щ	šč		
м	m	ы	ï		

1.1.2. General Features

Together with Karakalpak, Kyrgyz and Nogai, Kazakh belongs to the Aralo-Caspian sub-group of Western Turkic languages. It is reported to be especially closely related to Karakalpak, which some sources claim to be a dialect of Kazakh (Kirchner 1998). Three main dialects are recognised in Kazakh – North-eastern Kazakh, Southern Kazakh, and Western Kazakh – however, the differences between them are minor and are mainly pertinent to lexicon and phonology. Most of the data collected for this thesis have been verified by native speakers from all three dialectal regions (see Section 1.3 for details), and no contradictory grammaticality judgements have been observed.

The information in the grammatical description presented in this and following sections predominantly comes from Kazakh grammars by Kažbulatova (2009)¹ and Muhamedowa (2016). The examples provided here are either from these sources or from my own fieldwork; where a source is not identified for an example, that example is from my fieldwork (see also Section 1.3).

Like in other Turkic languages, vowel harmony is a key feature in the phonology of the Kazakh language. In the process of vowel harmony, vowels in the affixes become assimilated to the vowels in the stems according to their [\pm front] and [\pm back] features.² Progressive and regressive consonant assimilation processes are also present in Kazakh. These phonological processes result in great variability of suffix allomorphs. Following the Turkological linguistic literature tradition, I use capital letters to show those parts of a suffix that are subject to alternation due to phonological processes: *-Men*, *-MA*, *-DI*, *-LAr*, etc.

Again, like other Turkic languages, Kazakh is predominantly agglutinative in its morphology, meaning that words are formed via affixation. Some analytic features

¹ This grammar is written in Russian. All translations from Russian sources used in this thesis are mine.

² Reader is referred to Kajdarov (1997), Aralbaev (1988), Balakaev et al. (1962), and Kirchner (1998) for further information on vowel harmony in Kazakh.

are also observed, however, especially in the verbal realm, where many forms (e.g. progressive or evidential forms) are constructed with the help of auxiliaries.

Grammatical relations between the constituents within phrases of all sizes are expressed by means of case suffixation and postpositions. The canonical constituent order within a clause is SOV (discussed in Chapter 3), however, this word order is not fixed, and all six possible permutations of it can occur in certain (sometimes very narrow) contexts.

As mentioned previously, Kazakh is under-researched, however, it is relatively well-described with descriptive tomes on its syntax, morphology, and phonology having been published in Kazakh, Russian, Turkish and English. In the following sections, I cover some aspects of Kazakh grammar with a view to providing a background for the chapters that follow.

1.1.3. Nouns and Pronouns

Muhamedowa (2016: 219) states that Kazakh nouns “distinguish the categories of number, person and case by means of grammatical suffixes”, and that the “distinction between definite and indefinite referents is encoded by means of case-ending variation and indefinite pronouns”. I have not been able to observe any grammatical suffixes marking person on nouns, nor does the author provide any examples of these. As far as the definite/indefinite distinction is concerned, this is addressed in detail in Section 3.5 of Chapter 3 which is dedicated to differential object marking.

The plural form of nouns is formed with the suffix *-LAR*, which can be realised as one of the following allomorphs: *-lar*, *-ler*, *-dar*, *-der*, *-tar*, *-ter*, as shown in (1).

- (1) a. *kitap* – *kitap-tar*
 book book-PL
- b. *gül* – *gül-ler*
 flower flower-PL

- c. adam – adam-dar
man man-PL

Plural suffixes are not used when a noun follows a numeral or a quantifier, as shown in examples (b) and (c) in (2) below, as compared with the example in (a).

- (2) a. Oquşı-lar kel-di.
student-PL come-PST(3)
'Students came.'
- b. On oquşı kel-di.
ten student come-PST(3)
'Ten students came.'
- c. Köp oquşı kel-di.
many student come-PST(3)
'Many students came.'

Substantivised word forms can also be marked for number, as shown in (3), where a substantivised adjective is pluralised.

- (3) Aqildi-lar ket-ti.
clever-PL leave-PST(3)
'Clever people/the clever ones left.'

Where several nouns occur in sequence, the plural suffix can be attached only to the last noun.

- (4) Üstel-de gazet, žurnal, kitap-tar žatır.
table-LOC newspaper magazine book-PL lay
'Newspapers, magazines, and books are (laying) on the table.'

Traditional grammars state that the case system in Kazakh comprises seven cases: nominative, genitive, accusative, dative, locative, ablative, and instrumental. Since the so-called nominative case has no overt case markers, I do not include the nominative case into the case system in the same way as Göksel and Kerslake’s (2005) do not include it into their description of the Turkish case system. Instead of talking about nouns in the nominative case, I refer to them as unmarked or bare nouns. Unmarked noun phrases³ and their syntactic behaviour are discussed in detail in Section 3.5 of Chapter 3, and in Chapter 4. Table 2 below presents Kazakh case markers.

Table 2. Kazakh Case System

Case	Underlying Suffix Form	Surface Forms
Genitive	<i>-NIñ</i>	<i>-niñ, -niñ, -diñ, -diñ, -tiñ, -tiñ</i>
Accusative	<i>-NI, -n</i>	<i>-ni, -ni, -di, -di, -ti, -ti, -n</i>
Dative-Directional	<i>-GA, -nA, -A</i>	<i>-ğa, -ge, -qa, -ke, -na, -ne, -a, -e</i>
Locative	<i>-DA, -ndA</i>	<i>-da, -de, -ta, -te, -nda, -nde</i>
Ablative	<i>-DAn</i>	<i>-dan, -den, -tan, -ten, -nan, -nen</i>
Instrumental	<i>-Men</i>	<i>-men, -ben, -pen</i>

In the example in (5) all case-markers are used to identify the relationship between the noun phrases to which they are attached, and other sentence constituents.

- (5) Maral Dana-men Astana-da sauda ortaliğ-i-na
 Maral Dana-INST Astana-LOC trade centre-POSS-DAT
 bar-ïp, Akunin-niñ žana kitab-ï-n
 go-CVB Akunin-GEN new book-POSS-ACC
 kitap düken-i-nen satıp al-dï.
 book shop-POSS-ABL buy-PST(3)

³ The term ‘noun phrase’ is used as a purely descriptive term throughout this thesis. Phrases with a noun as the lexical head can be referred to as noun phrases; the same applies to verb phrases.

‘Maral and Dana went to a shopping centre in Astana, and bought Akunin’s new book in (from) a/the book shop.’

Note the use of the possessive marker on the head noun in the compounds *sauda ortaliğ-i* ‘shopping centre’ and *kitap düken-i* ‘book shop’ – this is the usual strategy for linking two nouns within a noun phrase or a compound. Although Muhamedowa (2016) states that nouns can be put together as a compound without a ‘linking element’, this is mostly avoided in the language, and the strong preference is to include some sort of linking element. This is discussed further in Section 5.4 of Chapter 5, where topic marker *degen* is shown to be developing into a linking element in class phrases where the head is unmarked.

The possessive suffixes are presented in Table 3 below.

Table 3. Possessive Suffixes

Person	Underlying Suffix Form	Surface Forms
1SG	-(<i>I</i>) <i>m</i>	- <i>m</i> , - <i>im</i> , - <i>im</i>
2SG informal	-(<i>I</i>) <i>ñ</i>	- <i>ñ</i> , - <i>iñ</i> , - <i>iñ</i>
2SG formal	-(<i>I</i>) <i>ñIz</i>	- <i>ñiz</i> , - <i>ñiz</i> , - <i>iñiz</i> , - <i>iñiz</i>
3SG	-(<i>s</i>) <i>I</i>	- <i>i</i> , - <i>i</i> , - <i>si</i> , - <i>si</i>
1PL	-(<i>I</i>) <i>mIz</i>	- <i>miz</i> , - <i>miz</i> , - <i>imiz</i> , - <i>imiz</i>
2PL informal	- <i>LA</i> <i>rIñ</i>	- <i>larıñ</i> , - <i>leriñ</i>
2PL formal	- <i>LA</i> <i>rIñiz</i>	- <i>larıñiz</i> , - <i>leriñiz</i>
3PL	- <i>LA</i> <i>rI</i>	- <i>ları</i> , - <i>leri</i>

Personal pronouns are as follows: *men* – 1SG, *sen* – 2SG informal, *siz* – 2SG formal, *ol* – 3SG, *biz* – 1PL, *sender* – 2PL informal, *sizder* – 2PL formal, *olar* – 3PL. The second person formal and informal pronouns differ in that the former are typically used in informal situations, and to address friends, relatives, children, or people who are clearly younger than the speaker; the latter are used in formal situations, and to address colleagues, seniors (in age and rank), strangers or people you meet for the first time. The full declension paradigm for personal pronouns is shown in Table 4.

As can be deduced from there being only one form for the third person singular pronoun, Kazakh does not mark gender, and one and the same pronoun is used for the English ‘she’, ‘he’, and ‘it’.

Table 4. Declension of Personal Pronouns

	1SG	2SG	2SG.FRM	3SG	1PL	2PL	2PL.FRM	3PL
Unmarked	<i>men</i>	<i>sen</i>	<i>siz</i>	<i>ol</i>	<i>biz</i>	<i>sender</i>	<i>sizder</i>	<i>olar</i>
Genitive	<i>meniñ</i>	<i>seniñ</i>	<i>sizdiñ</i>	<i>oniñ</i>	<i>bizdiñ</i>	<i>senderdiñ</i>	<i>sizderdiñ</i>	<i>olardiñ</i>
Dative	<i>mağan</i>	<i>sağan</i>	<i>sizge</i>	<i>oğan</i>	<i>bizge</i>	<i>senderge</i>	<i>sizderge</i>	<i>olarğa</i>
Accusative	<i>meni</i>	<i>seni</i>	<i>sizdi</i>	<i>onı</i>	<i>bizdi</i>	<i>senderdi</i>	<i>sizderdi</i>	<i>olardı</i>
Locative	<i>mende</i>	<i>sende</i>	<i>sizde</i>	<i>onda</i>	<i>bizde</i>	<i>senderde</i>	<i>sizderde</i>	<i>olarda</i>
Ablative	<i>menden</i>	<i>senden</i>	<i>sizden</i>	<i>ondan</i>	<i>bizden</i>	<i>senderden</i>	<i>sizderden</i>	<i>olardan</i>
Instrumental	<i>menimen</i>	<i>senimen</i>	<i>sizben</i>	<i>onimen</i>	<i>bizben</i>	<i>sendermen</i>	<i>sizdermen</i>	<i>olarmen</i>

There are several demonstrative pronouns in Kazakh, many of which are synonymous: *bul*, *mina*, and *osi*, for example, mean ‘this’, and *sol* and *ana* – ‘that’. Table 5 gives the declension for *bul* ‘this’ and *sol* ‘that’.

Table 5. Declension of Demonstrative Pronouns *bul* ‘this’ and *sol* ‘that’

	<i>bul</i>	<i>sol</i>
Genitive	<i>buniñ</i>	<i>soniñ</i>
Dative	<i>buğan</i>	<i>soğan</i>
Accusative	<i>bunı</i>	<i>sonı</i>
Locative	<i>bunda</i>	<i>sonda</i>
Ablative	<i>bundan</i>	<i>sondan</i>
Instrumental	<i>bunimen</i>	<i>sonimen</i>

1.1.4. Postpositions

Postpositions are the only type of adposition available in Kazakh which has no prepositions, as would typologically be expected of a head-final SOV language (see

also Section 3.2 of Chapter 3). Like prepositions, postpositions are a predicative category, meaning that they can impose type restrictions on their complements/arguments, assign thematic roles/cases, and have a semantic content which may be underspecified (Saint-Dizier 2006).

Postpositions follow their arguments/complements, as shown in the examples (6)-(10) below.

- (6) kitap bojinša
book according.to
'according to the book'
- (7) bilim üšin
knowledge for
'for the knowledge'
- (8) kezdesu-ğa dejin
meeting-DAT before
'before the meeting'
- (9) žumis-tan kejin
work-ABL after
'after work'
- (10) dos-pen birge
friend-INST together
'with a/the friend'

As these examples show, postpositions can be preceded by unmarked nouns, as in (6) and (7), and nouns in different cases: dative-directional in (8), ablative in (9), and instrumental in (10).

Postpositions are used to express spatial, temporal, cause-and-effect, instrumental, directional, aboutness, and other types of relation between the postpositional phrase (postposition + complement) and other sentential constituents.

1.1.5. Verbal Agreement

The infinitive form of Kazakh verbs usually ends in *-u* or *-ju*, as in *kelu* ‘to come’ or *qoju* ‘to place, to stand, to stop’, while the stem coincides with the second person singular informal imperative form as in *kel!* ‘come’ or *qoj!* ‘stop!’. Most verb forms use the verb stem as their derivation base.

Subject agreement obligatorily occurs in all finite clauses in Kazakh, and is expressed via suffixation on the predicate, as demonstrated by (11), where the predicate agrees with the first person plural subject expressed by the pronoun *biz* ‘we’; (12) is ungrammatical due to omission of the agreement suffix.

(11) Biz Kazakstan-da tur-a-miz.
 1PL Kazakhstan-LOC live-PRS/FUT-1PL
 ‘We live/will live in Kazakhtan.’

(12) *Biz Kazakstan-da tur-a.
 1PL Kazakhstan-LOC live-PRS/FUT
 Intended: ‘We live/will live in Kazakhstan.’

Two main agreement paradigms are identified in Kazakh: the *z-* or the pronominal paradigm, and the *k-* or the possessive paradigm; these are presented in Tables 6 and 7 respectively. The labels ‘*z-*’ and ‘*k-*’ paradigm originate from the difference in the forms of the first person plural suffix.

Table 6. The Pronominal Agreement Paradigm

Person	Singular	Plural
1	<i>-MIn</i>	<i>-MIz</i>
2	<i>-sIñ</i>	<i>-sIñdAr</i>
2.FRM	<i>-sIz</i>	<i>-sIzdAr</i>
3	<i>-Ø/-DI</i>	<i>-Ø/-DI</i>

Table 7. The Possessive Agreement Paradigm

Person	Singular	Plural
1	<i>-m</i>	<i>-K</i>
2	<i>-ñ</i>	<i>-ñdAr</i>
2.FRM	<i>-nIz</i>	<i>-ñIzdAr</i>
3	<i>-Ø</i>	<i>-Ø</i>

The possessive agreement paradigm is used with the *-DI* past forms, as well as the conditional and optative moods; in all other instances the pronominal agreement paradigm is applied. Non-finite predicates, such as converbial or participial verb forms do not indicate agreement with the subject.

Following Bisang (2007), I only consider verb forms that can head a matrix clause to be finite, which means that only verb forms carrying agreement suffixes can be considered to be finite in Kazakh. Kazakh non-finite verb forms, such as converbs and participles discussed below, cannot head independent clauses. However, unlike non-finites in Indo-European languages, they can carry some tense and aspect features. This is not unusual typologically, as observed by Nikolaeva (2007) who notes that there is no universality in the category of finiteness since neither tense nor agreement is a decisive feature of it cross-linguistically.

1.1.6. Converbs and Present Tense

There are several converbial suffixes in Kazakh, which are added to the verb stem to form converbs: *-(I)p*, *-A*, *-MAy*, *-ĜAll*, *-ĜAsIn*, *-MAyIšA*. I focus on the first two converbial forms which are of particular importance and interest to this thesis, since they can participate not only in the formation of subordinate clauses, but also in the construction of compound verb forms. In Russian linguistic tradition they are referred to as *deprichastije*, and in Kazakh linguistics they are labelled *kösımše*.

-A converbs are formed via the addition of an appropriate variant of the suffix to the verb stem. Surface forms *-a* and *-e* appear after consonants and harmonise with the final vowel of the verb stem; the suffix is realised as *-j* after vowels. This converb is roughly equivalent to the English gerund form, and would typically be approximated into English as such, as shown in (13).

- (13) Olar bir-biri-ne külimde-j qara-dı.
3PL each.other-DAT smile-CVB look-PST(3)
'They looked at each other smiling.'

This converbial form can be used in various compound verb forms, where the lexical verb appears as an *-A* converb, and a light or auxiliary verb contributes tense, aspect, modality, and possibly some other additional meanings, as shown in (14) and (15).

- (14) Tereze-ni žab-a sal!
window-ACC close-CVB AUX.IMP(2SG)
'Close the window!'

(Muhamedowa 2016: 119)

- (15) Ol buriš-qa otir-a qal-dı.
3SG corner-DAT sit-CVB AUX-PST(3)
'He suddenly sat in the corner.'

(Muhamedowa 2016: 110)

The English approximation of the example (15) includes the additional meaning of ‘suddenness’ contributed by the auxiliary, however, it is rather difficult to do the same for the sentence in (14). According to Muhamedowa (2016), when the verb *sal-* ‘lay down’ is used as an auxiliary or a light verb, as in (14), it indicates that the action described by the converbial form of the lexical verb can be or was carried out without any efforts, almost incidentally. Thus, in the example in (14), the use of the imperative form of this verb might indicate that the action of closing the window would not take the addressee of this command much effort to perform – this subtle additional meaning is lost in translation.

In Chapter 6, I return to -A converbs and their participation in complex verb forms in imperative sentences. I propose an alternative analysis for an imperative construction that has been erroneously – in my view – analysed as containing the particle *ğoj* in imperatives.

-A converb provides the basis for the present/future tense form, which is constructed via the addition of an appropriate suffix from the pronominal agreement paradigm (Kažbulatova 2009, Muhamedowa 2016). The interpretation of this verb form as indicating either a present (usually habitual) or a future action depends on the context, as the strings in (16), (17), and (18) exemplify.

(16) Biz künde žaz-a-miz.
 1PL daily write-PRS-1PL
 ‘We write every day.’

(17) Biz erteñ žaz-a-miz.
 1PL tomorrow write-FUT-1PL
 ‘We shall write tomorrow.’ / ‘We are writing tomorrow.’

(18) Ol biz-ge kel-e-di.
 3SG 1PL-DAT come-PRS/FUT-3
 ‘He usually comes to us.’ / ‘He will come to us.’

Note, that the converbial vowel *A* is no longer glossed as CVB, but as PRS or FUT where the tense is clear from the context as in (16) and (17), and as PRS/FUT where the tense is not contextually signalled or restricted, as in (18).

Converbial *-(I)p* form is constructed from the verb stem by the addition of *-ip* or *-ip* to the stems ending in a consonant, and *-p* to stems ending in a vowel. An *-(I)p* converb can function as a predicate of a subordinate clause, as shown in (19) and (20). The subject of the subordinate clause can be the same as that of the main clause as in (20), or different as in (19). The action expressed by the *-(I)p* converb precedes the action of the main verb.

- (19) Köktem kel-ip, gül-der gülde-di.
 spring come-CVB flower-PL blossom-PST(3)
 ‘The spring came, and the flowers blossomed.’

- (20) Quan-ïp, biz-ge qara-dï-ñ.
 become.happy-CVB 1PL-DAT look-PST-2SG
 ‘You became happy, and looked at us.’

When used in compound predicates, *-(I)p* converbs can either denote an action preceding that expressed by the second verb in the compound, as in (21), or combine with the second verb to create a new meaning, as in (22), where the verb *şiq-* ‘to come out’ adds the aspectual meaning of completeness.

- (21) Men erte ojan-ïp tur-dï-m.
 1SG early wake.up-CVB get.up-PST-1SG
 ‘I woke up early and got up.’ / ‘Having woken up early I got up.’

- (22) Biz mina kitap-tï oqï-p şiq-tï-k
 1PL this book-ACC read-CVB exit-PST-1PL

‘We have read this book (to the end).’ / ‘We finished reading this book.’

-(I)p converbs are combined with auxiliaries *žat-* ‘lie’, *otir-* ‘sit’, *tur-* ‘stand’, or *žur-* ‘walk’ to form the so-called observed present (Kažbulatova 2009), the present tense form, which could be most closely approximated into English via the present continuous tense. It can be used to describe an action happening at the moment of speaking, or at the present period of time in general, as (23) and (24) show.

(23) Ol universitet-te oq-ıp žatır.
3SG university-LOC study-CVB AUX(3)
‘He goes to university.’

(24) Sen hat žaz-ıp otir-siñ.
2SG letter write-CVB AUX-2SG
‘You are writing a letter.’

There are slight semantic differences between the four auxiliaries that can be used to construct this tense form, which can contribute to the overall meaning of the verb form, and also impose some restrictions on the lexical verbs they can combine with. These semantics are not entirely transparent, and a detailed in-depth study would be of great interest.

Now that we have considered the converbial forms and their roles in the formation of some of the tenses, I move to the participial forms and some of the past tense forms available in Kazakh.

1.1.7. Participles and Relative Clauses

Three participial forms – present, past, and future – are identified in the grammars, and are referred to as *prichastie* in Russian, or *esimše* in Kazakh. Non-finite clauses headed by participles can: a) occur in argument position; b) be complements of postpositions; and c) serve as relative clauses (modifying noun phrases).

(27) Sen ol žaz-ğan kitap-ti oqi-di-ñ be?
 2SG 3SG write-PTCP book-ACC read-PST-2SG Q
 ‘Did you read the book he wrote?’

(28) Sen bar-atin is sapar-ğa biz de
 2SG go-PTCP work trip-DAT 1PL too
 bar-a-miz.
 go-PRS-1PL
 ‘We are also going on the business trip on which you are going.’

Relative clauses headed by participles are of relevance to several sections in Chapter 6, where a previously unseen cleft construction is presented and discussed.

1.1.8. Verbal Tense and Mood

Pronominal agreement suffixes are added to the past participle verb form to construct one of the past tense forms, namely, the form that Kažbulatova (2009) refers to as ‘(reliable) distant past tense’ (*dostovernoe davnoproshedšee vremja*). Many Kazakh grammars refer to this tense as *buringi ötken šaq* ‘distant past’ (lit.: ‘long ago passed time’), however, this label is not entirely accurate as this tense form can be used to describe events that occurred recently. I refer to this tense as the *-ĜAn* past.

This tense is used to refer to past events of which the speaker is certain even if she did not witness them first-hand. It is not, therefore, a direct evidential form, since the speaker might not have witnessed it, however, this form contains some sort of epistemic authority that the speaker holds over the proposition. Additionally, there is typically no reference to a specific time at which the event took place, but rather a general declaration of the event having occurred at some point in the past.

(29) Sen osi film-di kör-gen-siñ.
 2SG this film-ACC see-PTCP-2SG

‘You watched this film (at some point).’ / ‘You’ve seen this film before.’

Straughn (2011) discusses this and other Kazakh verb forms within the domain of evidentiality, and in particular via the application of the notion of confirmativity (Aronson 1967, Friedman 1978). While he makes some interesting observations about evidentiality in Kazakh, I believe much further research is required, as far as evidential and epistemic meanings of various tense forms are concerned.

Another past tense form that frequently appears in the examples throughout this thesis is labelled by Kažbulatova (2009) as ‘evidenced past tense’ (*ochevidno-prošedšee vremja*). Muhamedowa (2016) refers to it as *-DI* past, and Straughn (2011) as simple *-DI* past. Kazakh grammars refer to this past tense form as *žedel ötken šaq* ‘recent past tense’, as opposed to *burinǵı ötken šaq* ‘distant past’ mentioned above, but, again, I refrain from using this label due to its inaccuracy. This verb form is constructed by the addition of the suffix *-DI* to the verb stem plus the appropriated agreement suffix from the pronominal suffixation paradigm.

The *-DI* simple past form is preferred for describing past events that occurred at a specific time in the past, habitual or recurring events in the past, and events of which the speaker was a direct witness. In the example in (30), this tense form is used to describe an event which happened within a specified temporal frame in the past, and in which the speaker was a participant and, therefore, a direct witness.

- (30) Biz keše teatr-ǵa bar-dı-q.
1PL yesterday theatre-DAT go-PST-1PL
‘We went to the theatre yesterday.’

The difference between the *-ǴAn* and the *-DI* past tenses in respect to their relation with a specific time in the past is especially clearly observed in clauses with predicate negation, as illustrated by the examples in (31) and (32). Negation is achieved with the help of the suffix *-MA*, which has surface forms *-ma*, *-me*, *-ba*, *-be*, *-pa*, *-pe*.

The negated *-ĜAn* past tense form yields the interpretation that the speaker has never been to the theatre, while the negated *-DI* form indicates that the speaker did not attend the theatre on a specific occasion.

(31) Men teatr-ġa bar-ma-ġan-mĭn.
 1SG theatre-DAT go-NEG-PTCP-1SG
 ‘I have never been to the theatre.’

(32) Men teatr-ġa bar-ma-di-m.
 1SG theatre-DAT go-NEG-PST-1SG
 ‘I did not go to the theatre.’

1.1.9. Conditionals and Subjunctives

The conditional form of the verb is constructed by adding the suffix *-sA* to the verb stem followed by a possessive agreement suffix. Verbs in conditional form appear in protases of conditional sentences which can be introduced by *eger* ‘if’, however, unlike in languages like English and Russian, for instance, *eger* ‘if’ is completely optional, and almost always omitted in spoken Kazakh.

The consequent, or the main clause of a conditional sentence, can contain a verb in the imperative mood as in (33), in the *-A* present/future tense form usually expressing a future action as in (34), as well as other verb forms which might be used to express a hypothetical consequence of a hypothetical condition, as in (35).

Note the use of the past participial form of the lexical verb plus the conditional form of the auxiliary verb *bol-* ‘be’ which acts as an auxiliary verb in this case, for expressing a hypothetical condition; the combination of a present participle *-AtIn* and the auxiliary *e-* (defunct diachronic form of the verb ‘to be’) in the simple past tense form are used to express the hypothetical consequent.

(33) (Eger) sen Anglija-ġa bar-sa-ñ, maġan
 if 2SG England-DAT go-COND-2SG 1SG.DAT

kiškentaj bir sījliq äkel!
 small a gift bring.IMP
 ‘If you go to England, bring me a small present!’

(34) Biz Bolat-ti kör-se-k, onı
 1PL Bolat-ACC see-COND-1PL 3SG.ACC
 toj-ğa šaqır-a-miz.
 wedding-DAT invite-FUT-1PL
 ‘If we see Bolat, we’ll invite him to the wedding.’

(35) Men hat žaz-ğan bol-sa-m, ol žauap
 1SG letter write-PTCP AUX-COND-1SG 3SG answer
 qajtar-atın e-di.
 return-PTCP AUX-PST(3)
 ‘If I wrote (him) a letter, he would reply.’

I return to conditionals in more detail in several sections of Chapter 5; in one of the sections I propose a hypothesis explaining the use of the conditional form of the verb *bol-* as a topic marker. Let us now briefly consider the Kazakh subjunctive mood.

As Predolac (2017) observes, there is much confusion as to which suffixes are to be considered the markers of the subjunctive mood in Turkish. Kazakh subjunctive mood has not been well researched to date, therefore, I use this term with the caveat that much further research is required in this area. I follow Ippolito (2013) in referring to the verb forms presented in (36) and (37) below as subjunctive conditionals. These verb forms are constructed with the help of suffix *-GAndA* added to the verb stem.

(36) Men hat žaz-ğanda, ol žauap qajtar-atın
 1SG letter write-SBJV 3SG answer return-PTCP
 e-di.
 AUX-PST(3)
 ‘Had I written to her, she would have replied.’

- (37) Sen muğalim bol-ma-ğanda, dāriger bol-atın
 2SG teacher be-NEG-SBJV doctor be-PTCP
 e-di-ñ.
 AUX-PST-2SG
 ‘Had you not become a teacher, you would have become a doctor.’

Subjunctive verb forms are mentioned again in Section 5.2 of Chapter 5, where they are shown to be used together with particle *še* to form a distinct type of interrogative sentences.

1.1.10. Interrogative Clauses

Polar questions are formed with the help of the question particle *MA*, which has the following surface forms: *ma*, *me*, *ba*, *be*, *pa*, *pe*. The particle follows the predicate, and no word order alternations are required, as the examples in (38)-(40) demonstrate.

- (38) Sen mīna kitap-ti oqı-di-ñ ba?
 2SG this book-ACC read-PST-2SG Q
 ‘Did you read this book?’

- (39) Biz teatr-ğa bar-a-miz ba?
 1PL theatre-DAT go-FUT-1PL Q
 ‘Are we going to the theatre?’

- (40) Ol dāriger me?
 3SG doctor Q
 ‘Is she a doctor?’

In Section 5.2 of Chapter 5 I examine another Kazakh question particle – question particle *še*, which is in complementary distribution with *MA*. I label *še* a contrastive

question particle, and describe its distribution and meaning in the aforementioned section. I posit that this contrastive question particle originated from contrastive topic marker *še*, which is also described and examined. An example containing question particle *še* provided in (41) below will suffice for this introductory section.

- (41) Biz teatr-ğa bar-a-miz, sen še?
 1PL theatre-DAT go-FUT-1PL 2SG Q
 ‘We are going to the theatre, and what about you?’

It can be seen from the English approximations of the sentences in (39) and (41), that question particles *MA* and *še* are used in different ways and contribute diverse meanings to the interrogative sentences in which they appear.

Wh-questions are formed with the help of interrogative pronouns: *kim* ‘who’, *ne* ‘what’, *qalaj* ‘how’, *qašan* ‘when’, *qajda* ‘where’, *qanša* ‘how much/many’, *nege* ‘why’, and some others. Only the first two pronouns from this list decline, as shown in Table 6 below.

Table 8. Declension of Interrogative Pronouns *kim* ‘who’ and *ne* ‘what’

	<i>kim</i>	<i>ne</i>
Genitive	<i>kimniñ</i>	<i>neniñ</i>
Dative	<i>kimge</i>	<i>nege</i>
Accusative	<i>kimdi</i>	<i>neni</i>
Locative	<i>kimde</i>	<i>nede</i>
Ablative	<i>kimden</i>	<i>neden</i>
Instrumental	<i>kimmen</i>	<i>nemen</i>

Some examples of wh-questions are presented in (42)-(44) below.

- (42) Keške teatr-ğa bizben kim bar-a-dī?
 in.the.evening theatre-DAT 3PL.INST who go-FUT-3
 ‘Who is going to the theatre with us tonight?’

- (43) Sen qašan žumīs iste-j-siñ?
 2SG when work do-FUT-2SG
 ‘When will you work?’
- (44) Ol qajda qara-p otir?
 3SG where look-CVB AUX(3)
 ‘Where is he looking? / What is he looking at?’

As these examples demonstrate, interrogative pronouns are usually placed immediately before the verb, which is also considered the focus position in Kazakh. This is explored in detail in Chapter 3.

1.1.11. Existential Clauses

Kazakh existential sentences are constructed with the help of the existential element *bar* ‘exist’ / ‘there is’, or its negative counterpart *žoq* ‘does not exist’ / ‘there is not’. Existential sentences are used to indicate existence, availability, or presence of something or someone somewhere, or possession of something by someone. Two most commonly used types of existential constructions can be identified: the genitive-possessive and the locative constructions.

The genitive-possessive constructions are exclusively employed to indicate possession: the possessor is marked with the genitive case, and the possessed takes a possessive suffix, as shown in (45), (46), and (47).

- (45) Meniñ mīsīğ-īm bar.
 1SG.GEN cat-POSS.1SG EXST
 ‘I have a cat.’
 Lit.: ‘My cat exists.’

- (46) Eki bala-m bar.

two child-POSS.1SG EXST

‘I have two children.’

Lit.: ‘My two children exist.’

- (47) Onıñ bala-sı žoq.
3SG.GEN child-POSS.3SG NEG.EXST
‘He does not have children.’

As can be seen from (46), if the constituent in the genitive case indicating the possessor is contextually obvious, it can be omitted.

In the locative existential construction the sentence initial noun in the locative case indicates location in the broadest sense possible, as the examples below demonstrate. It is followed by an unmarked noun indicating the entity that is present at the location identified by the preceding location adverbial. The English ‘there is...’ construction typically provides the best interpretation of Kazakh locative existential constructions into English, as can be seen in (48) and (49), however, as the example in (50) shows, a locative existential sentence can also denote a possession relationship.

- (48) Osı qala-da köp mejramhana bar /žoq
this city-LOC many restaurants EXST NEG.EXST
‘There are / aren’t many restaurants in this city.’

- (49) Bir apta-da žeti kün bar.
a week-LOC seven days EXST
‘There are seven days in a week.’

- (50) Mende üš žüz kitap bar.
1SG.DAT three hundred book EXST
‘I have three hundred books.’

Locative existential sentences with *bar* are often used to introduce a new referent into discourse, with the intention to supply some new information about it in subsequent utterances, as demonstrated in (51).

- (51) Astana-da Bayterek de-gen monument bar.
 Astana-LOC Bayterek say-PTCP monument EXST
 Ol elorda-niñ eñ körnekti ğımarat-tar-ı-niñ
 3SG capital-GEN most prominent building-PL-POSS-GEN
 bir-i.
 one-POSS
 ‘There is a monument called Bayterek in Astana. It is one of the most prominent buildings of the capital.’

The first sentence in (51) declares the existence of Bayterek in Astana; the sentence following it has Bayterek as its topic referent, and provides an update or comment about it. I return to these locative existential constructions in Chapter 5, where the addition of the particle *ğoj* to this kind of sentences is discussed.

1.2. Objectives of the Thesis

This thesis pursues two principal objectives. The first objective is to provide the first detailed, synchronic, data-rich description of the information-structural characteristics of the Kazakh language, which includes accounts of word order variations and information-structural/pragmatic markers. While Kazakh is the state language of one of the largest countries in the world, it is significantly under-studied (along with many other Turkic languages of Central Asia), especially in comparison to the related Turkish language, for example. The description of the information-structural characteristics of Kazakh presented in this thesis provides an important contribution to the body of descriptive linguistic literature on this language.

Additionally, this description can be of interest from a comparative typological perspective in Turkological, and wider cross-linguistic contexts. Being the first work

wholly dedicated to describing Kazakh information structure and related phenomena, this thesis aims to contribute to the slowly expanding body of research conducted on the lesser researched Turkic languages.

It is due to the lack of research on Kazakh and related Turkic language of Central Asia, that this thesis heavily relies on previous research of information structure and information-structural phenomena in the related Turkish language. Parallels are drawn between Turkish and Kazakh throughout most of this thesis, and in some chapters previous research on a particular phenomenon in Turkish is even taken as basis for my research of the same phenomenon in Kazakh.

Every chapter of this thesis contains new, previously unseen and not discussed data, which are fully glossed and appear within at least minimal contexts. Thanks to these new data, as well as the descriptions and analyses thereof, this thesis proposes novel ways of approaching some of the phenomena, and highlights some erstwhile undescribed constructions and processes that occur in the Kazakh language. In each chapter, the descriptive part is followed by the analytical part, where the latter relates to the second objective of this thesis.

The second objective of this thesis is initiating formal modelling of some of the information-structural phenomena observed in Kazakh from the perspective of the syntactic theoretical framework of Dynamic Syntax (Kempson et al. 2001, Cann et al. 2005). To the best of my knowledge, this is the first attempt of a formal syntactic analysis of Kazakh data in general, and certainly is the first attempt of a Dynamic Syntax analysis in particular.

The Dynamic Syntax framework is adopted for its parsing-oriented approach which does away with syntactic hierarchical positions, movement, or feature assignment in favour of focusing on the dynamics of the parsing process represented by semantic trees. This allows Dynamic Syntax formalisms to be applied to verb-final languages – of which Kazakh is one – as easily as to verb-medial languages, without resorting to postulating some type of movement.

In Dynamic Syntax, all strings are parsed on a left-to-right, word-by-word basis relative to their context, and such information-structural concepts as topic and focus are derived through the process of semantic tree constructions rather than being

thought to be encoded in the syntax directly, or to belong to some other layer of information.

This approach significantly simplifies the task of describing the relation between word order and information structure, since topic, focus, and other information-structural concepts are ultimately seen as pragmatic effects of the parsing process, and can be used as descriptive notions without the need for any further speculation as to their syntactic status or in-situ position. This is shown in the Dynamic Syntax analysis of neutral and scrambled word orders in Kazakh (Chapter 3), where no recourse is made to particular information-structural syntactic positions, features, or movement.

Applying Dynamic Syntax formalisms to differential object marking (Chapter 4) provides a fresh approach to the phenomenon, which had previously only been examined within static hierarchical frameworks, or in reference to the difference in the ‘meaning’ of marked vs unmarked direct objects. Specifically, it is shown that the difference between marked and unmarked direct objects lies in the way they are parsed – onto fixed and unfixed nodes of a semantic tree respectively. While these parsing strategies are standardly available in the framework, their application towards differential object marking demonstrates the interaction of the ways in which a constituent is parsed with the pragmatic effects associated with the absence and presence of a case marker on a direct object.

When applied to Kazakh topic markers (Chapter 5), Dynamic Syntax analysis demonstrates that all three topic markers under consideration perform the same function of ensuring that the topical element is parsed onto a Linked structure thus providing the immediate context against which the rest of the utterance is parsed. Since two of the topic marking forms have originated from conditional verb forms, the Dynamic Syntax formalisms are used to explore the reason behind this development.

Dynamic Syntax formalisms are also used in the analysis of the Kazakh pragmatic marker *ǵoj* (Chapter 6), which is shown to function as a copula and a particle. The Dynamic Syntax approach to these items allows for an elegant formalisation of their diverse syntactic distribution patterns and common semantics within one lexical entry.

1.3. Methodology

The data presented in this thesis consists of elicited data, corpus data, and naturalistic data from observed, film, or cartoon dialogues. This section provides details on these data and the methods whereby they have been collected.

My elicitation work started in late 2014 in London, where I was very lucky to have met Bolat Nassygazy – at the time an MA student (in Education Management and Leadership) at University of Warwick. Bolat was born in 1988 in the city of Ural'sk (Kazakh name Oral, north-western Kazakhstan), in a Kazakh-speaking family. He grew up and attended a Kazakh school there as well. He subsequently moved to Almaty to read for a degree in Kazakh language at the Abai Kazakh National Pedagogical University. While Bolat is bilingual in Kazakh and Russian, Kazakh is his mother tongue, and he uses it daily in his personal and professional life.

Bolat returned to Kazakhstan after completing his studies in the UK and now works in Astana, in the education sector. His assistance with the language data for this thesis continued throughout, and he even created a messenger group of 8 native speakers of Kazakh with whom I could consult to cross check Kazakh data. I am grateful to the participants of that messenger group as well.

Most of the elicited data used in this thesis were collected during a nine-month long field trip to Astana, where I lived from September 2015 until May 2016. The data were mainly collected during elicitation sessions with several native speakers of Kazakh who come from different regions of Kazakhstan, but now reside in Astana. All elicitation sessions were conducted in Russian, since all the language consultants are to a higher or lesser degree proficient in it.

The elicitation sessions mainly took place at the Eurasian National University (ENU) in Astana, and I am grateful to Duman Aitmagambetov, the Head of the Rector's Office, for his kind help with arranging weekly sessions for me at the Department of Practical Kazakh headed by Quandiq Kul'manov.

I worked with several Kazakh language consultants at that department: Šolpan Isakulova, Lāzzat Bajbolat, Satıbalđı Berdibaev, and Aldan Musabekova. I most

closely worked with Šolpan, Lāzzat, and Satı̄baldı̄, all of whom are native speakers of Kazakh, Russian being their secondary language. They use Kazakh as their main language in their personal, professional and day-to-day life, occasionally switching to Russian if necessary (where their interlocutor does not speak Kazakh fluently).

Šolpan was born in 1981 in the Zhambyl region of Kazakhstan, where she attended a Kazakh school, and later Taraz State University where she read for a degree in Kazakh language and literature. She has been teaching at ENU since 2003, and is currently a docent at the Department of Practical Kazakh.

Lāzzat was born in 1966 the Almaty region of Kazakhstan. From 1999 to 2003 she studied at the South Kazakhstan Pedagogical University in Shymkent, and graduated with a diploma of a teacher of Kazakh language and literature. In 2005 Lāzzat obtained a Master's degree in Kazakh Philology from Akhmet Yassawi University in Turkestan, after which she worked as a teaching fellow (Kazakh language) at the South Kazakhstan Pedagogical University and Shymkent Socio-Pedagogical University. She has been a senior teaching fellow (various Kazakh language courses) at ENU since 2012.

Satı̄baldı̄ was born in 1967 in the Almaty region of Kazakhstan. He received a Master's degree from Al-Farabi Kazakh National University in Almaty, and has been teaching Kazakh at different levels since 1992. He has been a senior teaching fellow at ENU since 2014.

I obtained consent from all language consultants prior to recording the elicitation sessions using Zoom H1n voice recorder. The sessions usually took place in an office in ENU. The prosodic data used in Chapters 3 and 6 were recorded then as well.

Following recommendations outlined in Senft (2009) and Chelliah and De Reuse (2011), a variety of methods were used in the elicitation sessions: direct translation tasks (from Russian into Kazakh and vice versa), discourse-completion tasks, and semi-structured interviews. Some of the translation tasks were based on those from the Questionnaire on Information Structure (Skopeteas et al. 2006), however, they had to be adapted and translated into Russian first. The examples have

been cross-checked with several language consultants, regardless of the method by which they were originally collected.

The thesis also contains many natural, non-elicited examples, some of which were collected by me while in Astana. Throughout the thesis, all language examples that were not elicited or collected by me are referenced.

As mentioned earlier, some of the data used in this thesis come from the Almaty Corpus of Kazakh Language (ACKL), available online at: <http://web-corpora.net/KazakhCorpus>. This open access corpus comprises more than 40 million word tokens, and offers a convenient and straightforward search tool. The corpus contains written materials of diverse genres from publicistic to literary and scientific, which provides for a rich source of linguistic data. The data from this corpus prominently feature in Chapter 5, where the results of several corpus searches are presented and analysed. Linguistic examples from the ACKL are referenced as such throughout the thesis.

Many language examples come from various works on Kazakh, such as Kazakh grammars and PhD theses, which are not numerous, especially in the English language; these are referenced as usual, with the surname of the author and year of publication. The examples referenced as KT are taken from the website kaz-tili.kz, which has been developed for Russian speaking learners of Kazakh by Valjaeva (2018).

Films, TV series, and cartoons in Kazakh provided another useful source of naturalistic language data. For the examples taken from a film or a cartoon, the following abbreviations are used: *Miñ Bir Tün* ‘One Thousand and One Nights’ – MBT; ‘Monsters University’ – MU; ‘Alice Through the Looking Glass’ – ATLG. All examples from published works are referenced as usual, with the author’s name and the year of publication.

1.4. Structure of the Thesis

This thesis is divided into seven chapters: the first two chapters lay the foundation necessary for the descriptions and the analyses presented in the subsequent four core

chapters; the final, seventh chapter summarises and concludes the thesis. Each core chapter first describes an information structural phenomenon under consideration, and then presents a DS analysis of it.

Chapter 2 focuses on the information-structural notions relevant for this thesis, and on Lambrecht's approach to these notions, which is followed for the descriptive parts of this thesis. It also introduces the theoretical framework of Dynamic Syntax, its tools and means of representations, as well as its approach to the central information-structural notions of topic and focus.

Chapter 3 examines the relation between word order and information structure in Kazakh. In this chapter I demonstrate that the information-structural characteristics of a Kazakh sentence go beyond the traditional rigid assumptions of a sentence-initial topic and an immediately preverbal focus. It is shown that in order to arrive at a more accurate description of the information-structural configurations available in a Kazakh sentence, the notions of topic and focus must be expanded to include at least their contrastive counterparts. It is also shown that the notion of focus field rather than focus position ought to be employed in descriptions of Kazakh information structure. The Dynamic Syntax approach to information-structural characteristic of diverse word order permutations in Kazakh is also presented in this chapter.

Chapter 4 deals with the phenomenon of differential object marking, which is linked with the matters of information-structural status and sentential position of a direct object. A detailed description of this phenomenon is provided along with previous analyses of its Turkish equivalent, since no formal analytical work had been done on Kazakh differential object marking. This chapter also comprises a Dynamic Syntax analysis of Kazakh differential object marking, which can be applied to this phenomenon in related languages. The proposed analysis allows to explicate and formally represent the phenomenon without the recourse to syntactic hierarchical structures or movement.

In Chapter 5 I take a detailed look at Kazakh topic markers, which have not received much attention in linguistic literature to date. These are first described in detail, before a general Dynamic Syntax analysis of their function in a sentence is presented. This is followed by a more focused Dynamic Syntax consideration of topic

markers which originated from conditional verb forms, and a link between the former and the latter is explained through Dynamic Syntax formalisms.

Chapter 6 is dedicated to *ǰoj* – usually referred to as a particle – many uses of which are associated with the information-structural notions of givenness or contrast. The first detailed account of this item’s distribution is presented, and it is demonstrated that two syntactically diverse types of *ǰoj* exist in Kazakh, both of which, however, contribute the same pragmatic effect of givenness or shared information. Both the syntactic differences and the pragmatic similarities are accounted for in the Dynamic Syntax analysis of *ǰoj*, which results in a single lexical entry for both types of *ǰoj*.

Chapter 7 summarises and concludes this thesis, and poses several questions for further research.

2. Theoretical Background

“What is thought clothed in? In meanings, or in forms?”

Abdul Qaahir al-Jurjaani, 11th century grammarian

The term ‘Information Structure’ (IS) generally refers to the way in which an utterance is structured in order to optimise the transfer of information it carries; it is also referred to as information packaging (Chafe 1976). Studies of IS are directly connected with such areas of linguistics as semantics, pragmatics, syntax, morphology, and prosody; as well as with general cognitive facilities and world perception and knowledge (Zimmermann and Féry 2009).

The interest in IS can be dated back to at least medieval Arabic scholars. Contemporary IS studies commenced around the late 1920s, when the Prague School linguists introduced the *theme-rheme* dichotomy (also referred to as old-new or topic-focus divisions). The field of IS research has particularly expanded over the past four to five decades (Féry and Ishihara 2016).

The primary aim of some of the IS studies produced during this time has been to discover the meaning of IS and its components, while other studies have been aimed at identifying the place of IS and IS notions within grammar. Thus, it can be said that IS studies have been concerned with two main groups of questions: 1) questions concerning the content of IS and the meaning of such information-structural notions as focus, topic, and contrast, amongst others; and 2) questions on the role of IS in the architecture of grammar.

The first part of this chapter – Section 2.1 and its sub-sections – explores the information-structural and pragmatic notions relevant for the rest of this thesis. It would not be controversial to say that the notions of topic and focus could be seen as the central notions of the information-structural studies, and these are considered in the first two sub-sections of this section. These are followed by sub-sections dedicated to the notions of contrast, and specificity.

Section 2.2 begins with a brief overview of the ‘traditional’ approaches to the second set of IS-related questions, which are then rejected in favour of the Dynamic

Syntax treatment of information-structural phenomena as effects naturally occurring during the contextualised parsing process. The theoretical syntactic framework of Dynamic Syntax (DS), its principles, tools, as well as its approach to the place of information structure in the grammar are then introduced in detail. Section 2.3 provides a summary of the preceding sections and concludes this chapter.

2.1. Information-structural Categories

Since the literature attempting to answer the two main groups of questions identified in the introduction above is too vast to cover in a section of a thesis, I only identify and define the information-structural and pragmatic notions that are directly relevant to this thesis. Occasional references are made to some alternative approaches in order to provide some general background information, however, this is kept to a minimum.

I generally follow Lambrecht's (1994) approach to IS, which is viewed as:

that component of sentence grammar in which propositions as conceptual representations of states of affairs are paired the lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts (1994: 5).

In other words, the study of IS, according to Lambrecht (1994) is concerned with the ways in which the pragmatic structure of a proposition is expressed formally by means of grammatical markers, prosody, form and order of constituents within a sentence, form of the complex grammatical constructions, as well as choices between related lexical items.

Lambrecht's attribution of IS to the realm of sentence grammar simply means that when we examine IS, we are not concerned with the structure and organisation of discourse, but rather with the internal organisation of sentences within discourse/context. That is to say, that while we certainly must take into account the discourse/context in which a sentence is produced – since one and the same sentence

might display different information-structural properties in different contexts – we are not concerned with the structure of that discourse or context itself.

One of the most important points in Lambrecht’s approach to IS and the notions of topic and focus is their relational nature which allows for compatibility of Lambrecht’s approach to the notions of topic and focus with the formal syntactic framework of DS as shown further in this chapter. Lambrecht does not consider topic and focus to be independent primitives, but rather pragmatic notions which are construed from and within the context of the interaction of an utterance with the larger discourse picture and with the speaker’s assumptions of the hearer’s knowledge.

This shared knowledge between the speaker and the hearer is often referred to as Common Ground (CG) – a notion proposed by Stalnaker (1974, 1978, 2002). Each utterance produced by interlocutors updates the content of the CG, which means that CG is continuously modified throughout the communication process.

Let us now consider the notions of topic, focus, contrast, and specificity in subsections 2.1.1, 2.1.2, 2.1.3, and 2.1.4 respectively. These notions cannot be considered independently from other information-structural and pragmatic concepts – these are introduced and briefly explicated as and when required throughout the following subsections. In 2.1.5 some proposals regarding the role of IS within the architecture of grammar are presented.

2.1.1. Topic

Topic is one of the central notions in Lambrecht’s representation of IS and its main categories. Unlike Reinhart (1982), who rejects ‘aboutness’ as a sufficient definition of sentence topic, Lambrecht proposes that:

a referent is interpreted as the topic of a proposition IF IN A GIVEN DISCOURSE the proposition is construed as being ABOUT this referent, i.e. as expressing information

which is RELEVANT to and which increases the addressee's KNOWLEDGE OF this referent (1994: 127).⁴

In other words, a topic is a pragmatically construed sentence relation, such that the rest of the sentence is about its referent.

Lambrecht clearly delimits the notions of topic referent and topic expression, underlining that the term 'topic' must be understood as identifying the entity/referent about which information is being conveyed in a proposition. A topic expression then, is a linguistic expression which designates a topical entity/referent in a sentence, however, it is not obligatorily present in every sentence. That is to say, a sentence may be constructed about a topical referent, but have no overt topic expression.

The distinction is exemplified by the answers to the same question in (1) and (2), where one and the same topical referent – *Alexey* – is expressed by a pronominal topical expression in the first sentence, but is not overtly present in the second sentence.

- (1) Q: What did Alexey do next?
A: [He]_{Topic Expression} finished his dinner and left.
- (2) Q: What did Alexey do next?
A: [∅]_{Topic Expression} Finished his dinner and left.

It is clear, however, that both sentences are constructed about one and the same topical referent, and both these sentences, under Lambrecht's analysis would be of Topic-Comment structure.

Contrary to this, Vallduví (1992), and Vallduví and Engdahl (1996), for example, refer to topic expressions as links, and base their four-way classification of possible IS configurations on the overt presence of the link in a sentence. Thus, an example

⁴ Accents here and in the following quotations from Lambrecht (1994) are by Lambrecht.

like that in (2) would be considered an All-Focus, rather than a Topic-Comment sentence type, since no overt topic expression or link is present there.

Additionally, Nikolaeva (2001) observes that, a sentence like that in (3) would also be considered an All-Focus sentence under Vallduví's (1992) analysis. However, under Lambrecht's model, which distinguishes not only topic referents from topic expressions, but also primary and secondary topics, this sentence would be analysed as the "topic-secondary topic-focus" structure. There is no overt topical expression for the primary topic – identified as a 'zero' pronoun in the example below, while the secondary topic is expressed by a clitic (Nikolaeva 2001: 14). A sentence with a primary and a secondary topics conveys information about the relation between the two topical referents.

- (3) Q: How does the boss feel about broccoli?
A: *pro* [l'odia]
it.hates
'He hates it.'

(Vallduví 1992: 74)

Although the importance of the notion of secondary topic cannot be denied, I do not examine secondary topichood in this thesis. This is left for further research into topichood, as well as relations between primary and secondary topics in Kazakh and other Turkic languages of Central Asia.

Another point to be made about the notion of topic referent is to do with its givenness status: a topic referent must be 'given'. Féry and Ishihara define a 'given' referent as a referent that is "active in the consciousness of the interlocutors" (2016: 6). Rochemont (2016) further observes that givenness has many forms and functions, but in all its uses it is to do with the informational or cognitive status of a referent as being present in the discourse model/CG in some way.

Givenness can be divided into text-givenness and context-givenness, where the former is to do with the referent having been mentioned in the discourse, and the latter with the referent being contextually salient. Thus, a referent might be considered

‘given’ by virtue of simply being present in the interlocutors’ field of vision, which makes this referent available for immediately becoming a topic referent.

In Lambrecht’s model, givenness is not a categorical feature (i.e. given vs. not given), but rather a scale expressing the degree of a referent’s discourse salience. In this thesis I predominantly keep to the two extremes of that scale as I present and discuss the most canonical cases of information-structural statuses of different constituents in simple Kazakh sentences. The question of how various degrees of givenness of a referent are reflected in the syntactic form, prosody, or clausal position of the constituent expressing that referent in Kazakh and other under-researched languages is left for further research.

To sum up, Lambrecht considers a topic to be a pragmatically construed sentence relation identifying a text-/context-given discourse referent about which the rest of the sentence is construed. This topic referent may, but does not have to be overtly present in a sentence as a topic expression. For consistency and clarity, in this thesis, I only deal with sentences with overt topic expressions conveyed by nominal or pronominal elements.

Another important distinction among topic types must be made in relation to continuity of the topic referent. Where the topic referent does not change from one sentence to the next, as in (4), for example, the topic is considered to be a continuous aboutness topic.

(4) Q: What did [Alexey]_i eat?

A: [He]_i ate the soup.

In the example above, the topic referent in the question is expressed by the proper name *Alexey*; the same topic referent is expressed by the pronoun *he* in the answer. In Lambrecht’s view, if such continuous topic referents have an overt topic expression in a sentence, then this expression would canonically be an unaccented pronominal element, like that shown in the answer in (4).

The example in (5) presents a situation where the topic referent changes from one sentence to another: in the first clause of the answer the topic referent is the same as that in the question, but a different topic referent – *Ljubov'* – is found at the beginning of the second clause. I refer to such non-continuous topics as switch or contrastive topics.

(5) Q: What did Alexey eat?

A: He ate the soup, and *Ljubov'* ate the pasta.

Contrastive topics are discussed in more detail in sub-section 2.1.3 dedicated to the notion of contrast.

Lambrecht also underlines that: a) not all sentences have a topic; and b) topicality cannot always be easily assigned to one particular constituent, and that there are degrees to which elements in a proposition qualify as topics – that is to say that different elements can be more or less topic-worthy. According to Lambrecht, this is precisely what accounts for the absence of distinct formal marking of the topic relation in many languages.

Topics can be marked in a number of different ways across languages or even in one and the same language. Some of the devices used for topic marking include special sentential position, pronominalisation, agreement morphology, and dedicated topic markers; the latter are of particular importance for this thesis.

Topic markers are special morphemes that exist in a number of the world's languages, and are used to identify a topic expression in a sentence. Topic markers are optional, and, as reported by Gundel (1988), they are not the only means of coding the topic-comment structure in any of the 30 languages from 13 different language families examined by her. Lambrecht (1994: 119) observes that topic markers do not always accurately reflect relative degrees of topicality of the referents, however, this does not preclude them from playing an important role in grammars of many languages.

Dalrymple and Nikolaeva (2011) demonstrate close association of topic marking with certain grammatical functions in some languages, as, for example, is observed in differential subject marking in Somali (Semitic), where only topical subjects take on the nominative form, while non-topical or focused subjects must appear in the absolutive form.

Topic markers can mark a variety of syntactically and semantically diverse elements, as the examples from Korean in (6)-(10) below demonstrate (these examples are from Gundel (1988), however, the gloss has been revised by me). In the first set of examples given in (6), the SOV sentence in (a) does not contain a topic marker; this sentence can be a topic-comment sentence, an all-focus sentence, or a sentence with a narrow focus on the direct object – in the same way an SOV sentence with neutral placing of the main sentential stress (on the immediately preverbal syllable) can be in Kazakh (Chapter 3, Section 3.2).

The sentence in (b) is also an SOV sentence and is identical to that in (a), apart from the absence of the subject-marking morpheme *-ka* on the subject *John*, and the presence of the topic marker *-nin* immediately after it. The all-focus reading is no longer available for a sentence like this, and it can only be parsed as a topic-comment sentence, where the subject is also the topic of the sentence. The sentence in (c) is an OSV sentence, in which the sentence-initial direct object carries a topic-marker; again, only the topic-comment reading is available for this sentence.

- (6) a. John-ka i salam-lil mannaassta.
 John-NOM this man-ACC met
 ‘John met this man.’
- b. John-nin i salam-lil mannaassta.
 John-TOP this man-ACC met
 ‘(As for) John, (he) met this man.’
- c. Salam-nin John-ka mannaassta.
 man-TOP John-NOM met

‘(As for) this man, John met (him).’

(Gundel 1988: 217)

In sentences (b) and (c) in (6) the topic marker is attached to an argument of the main verb of the sentence, but this does not have to be the case, as the example in (7) shows. In the sentence in (7), the topic marker is attached to a gerund-type element, which is not an argument of the predicate.

- (7) Suyeng-nin Waikiki-ka chota.
 swimming-TOP Waikiki-NOM good
 ‘(As for) swimming, Waikiki is good.’

(Gundel 1988: 217)

In the examples in (8)-(10) the topic marker marks time and place expressions. Radetzky (2002) claims that this is not accidental, and proposes that topic markers originally developed from locative and temporal markers via the following grammaticalisation path: locative/temporal marker > contrastive marker > topic marker.

- (8) Ecey-nin John-ka naksicil-lil kaassta.
 yesterday-TOP John-NOM fishing-ACC went
 ‘Yesterday, John went fishing.’

- (9) San-ey-nin namu-ka manhta.
 mountain-at-TOP tree-SUBJ many.
 ‘At the mountain, there are many trees.’

- (10) San-ey kal ttay-ey-nin moca-lil ssi-la.
 mountain-to go time-at-TOP hat-ACC wear-IMP
 ‘When you go to the mountain, wear a hat.’

(Gundel 1988: 217)

The examples in (7)-(10), as well as the Mandarin Chinese example in (11) below, confirm Li and Thompson's (1976) conclusion that topics, unlike subjects, do not have a selectional relation with any verb in a sentence, that is, they do not have to be an argument of a predicate. This is in line with Lambrecht's (1994) understanding of the notion of topic presented above, where a referent is considered topical if the proposition is construed to be about it within a given context. In other words, topicality is of pragmatic rather than syntactic nature, and topicality is defined on referents rather than on linguistic expressions.

We can see that in the examples in (7), (8) and (9) both the topic marker and subject case-marker are present and the topical element is not the same as the subject. These sentences are traditionally referred to as "double subject" constructions (Chao 1968; Li and Thompson (1976); Lin 1981, amongst others), which do, however, follow the Topic-Comment structure. Such constructions are pertinent of topic-prominent languages such as Japanese, Korean, and Mandarin Chinese amongst others. Chappell (1996) provides the following example of a canonical double subject construction from Mandarin Chinese:

- (11) xiàng bízǐ cháng
 elephant nose long
 'Elephants have long trunks.' / 'As for elephants, their trunks
 are long.'

(Chappell 1996: 467)

Returning to the general characteristics of topic markers, Gundel (1988) also notes that topic markers tend to be optional – unlike case marking particles in such languages like Japanese and Korean –, that is to say, a sentence can be complete, grammatical and pragmatically felicitous without the use of a topic marker, whereas case-marking of the subject, direct and indirect objects, and adjuncts is obligatory.

It is often observed that topic markers coincide with forms that mark relative clauses, or protases of conditional sentences (Haiman 1978, Seiler 1985, Gundel 1988,

Radetzky 2002), as demonstrated by the examples below from Hua (Yagaria, Papua New Guinea) and Imonda (Waris, Papua New Guinea). The examples from Hua in (12) and (13) show the same morpheme used for topic marking and for marking the protasis of the conditional sentence respectively.

- (12) Dgai-mo-ve baigu-e
 I-C.P-TOP will.stay-1SG
 ‘(As for) me, I will stay.’

- (13) E-si-ve baigu-e
 come-3SG.FUT-INT will.stay-1SG
 ‘If he comes, I will stay.’ (lit. ‘Will he come? I will stay.’)
 (Gundel 1988: 219 citing Haiman 1978)

According to Gundel (1988), marking of the protases and relative clauses with the same marker as the topics is to be expected as they share the property of being given in relation to the rest of the sentence. In the example in (14) from Imonda the topic marker *-fa* marks a noun phrase as a topic.

- (14) Ōgòt-fa, ah-ia ale-f
 enemy-fa Q-LOC stay-PRS
 ‘Where is the enemy?’

(Seiler 1985: 206)

In the examples in (15) and (16) the same morpheme is used to mark the protases of the conditional sentences (it is realised as *-ba* if attached to bilabial nasal or to a verb, as in (15)).

- (15) po feha-f-ba ka ale-f
 water fall-PRS-fa I stay-PRS
 ‘If it rains, I will stay.’

(Seiler 1985: 204)

- (16) ude ale-ta-ba ed-fa ne-m
dog stay-IRR-fa it-fa 2-GOAL
ue-ne-t
CL-eat-CF

‘If the dog had been here, then it would have eaten you.’

(Seiler 1985: 205)

In (17), the same topic marker is used to mark a relative clause.

- (17) ōm ka nagla-na-ba ednèi uōgō ka
yesterday I see-PST-fa that.one drum I
f-ia-i
CL-get-FUT

‘I want to get the drum I saw yesterday.’

(Seiler 1985: 66)

Considering the examples like those given above, Haiman (1978: 567) goes as far as stating that “conditionals are topics (= givens, presuppositions) of their sentences”. However, it must be noted that this is not the case cross-linguistically as Caron (2006) demonstrates with the examples from several Chadic languages in which the protases are marked with focus, and not topic markers. This is shown in the examples in (18) and (19) from a Chadic language of Northern Nigeria – Polci, where the identifying copula *kən* marks a focused constituent and the protasis respectively.

- (18) wún gi kən yu ɲen a ga:
girl DEICT COP pour milk in calabash
gi.
DEICT

‘The GIRL poured milk into the calabash.’

- (19) Gǎrbà kən ndzəŋ ɬo: wú də kə
 Garba COP cut meat ACC INJ 2SG.SUBJ
 fǔ:-m.
 tell-1SG
 ‘If Garba slaughters an animal (lit: GARBA slaughters an animal), tell
 me.’

(Caron 2006: 78)

Chapter 5 of this thesis is dedicated to topic markers in Kazakh; the connection of some of the Kazakh topic markers with conditionals is described, and then explicated via the formalisms of DS.

To sum up, in this sub-section the notions of topic, topic referent, topic expression, givenness, Common Ground, and topic marking have been discussed and exemplified. I follow Lambrecht’s approach to the notion of topic being a referent about which the rest of the proposition is constructed. Lambrecht underlines that a topic expression must not be equated with a topic referent, since the absence of the former in a sentence does not automatically mean that the sentence is not ‘about’ a particular topic referent.

Four types of topic have been identified: primary, secondary, aboutness, and contrastive. The first two topic types are to do not only with their referents’ relation to the discourse, but also with their relation to each other within a particular sentence. This is not examined in this thesis, and is left for further research.

The latter two topic types are to do with the relation of the topic referent in the sentence under discussion with that in the preceding context. A continuous aboutness topical expression identifies a topic referent that remains unchanged from the preceding context into the current utterance, while a contrastive topic diverges from the preceding one. The notions of continuous and contrastive topics are widely used in Chapters 3 and 5 of this thesis, and the notion of a contrastive topic is discussed in more detail in 2.1.3.

2.1.2. Focus

Before considering Lambrecht's definition of focus, two other notions must be explicated: presupposition and assertion. Lambrecht lists both these notions alongside the notions of topic and focus as some of the main information-structural concepts. This is what he writes about presupposition and assertion:

Let us refer to the 'old information' contained in, or evoked by, a sentence as the PRAGMATIC PRESUPPOSITION (or simply PRESUPPOSITION), and let us refer to the 'new information' expressed or conveyed by the sentence as the PRAGMATIC ASSERTION (or simply the ASSERTION) (1994: 52).

There are two important points to be made in regards to the statement above. Firstly, we must note Lambrecht's addition of the label 'pragmatic' to presupposition and assertion, lest these are confused with their semantic counterparts. This especially applies to the notion of presupposition; much literature has been dedicated to the distinction between pragmatic and semantic presupposition, which are closely related, but certainly are not one and the same thing. For further details on this the reader is referred, for example, to Soames 1989, Horn 1997, Simons 2006.

It will suffice to say here that semantic presupposition is derived directly from the meanings of the words included in a proposition, and the way they are syntactically combined with each other. A classic example of semantic presupposition is the sentence *The King of France is bald*, which presupposes the existence of a king of France. Pragmatic presupposition, on the other hand, is the sum of 'knowledge' evoked in a sentence, and assumed by the speaker to be either already available in the hearer's mind at the time of utterance, or can be taken for granted by the hearer.

The second important point to note is that the concept of presupposition should not be equated to the limited understanding of the concept of 'old information'. As Lambrecht himself clarifies, 'old information' is not necessarily the information that has been explicitly shared between the interlocutors, but "the sum of 'knowledge' evoked in a sentence which a speaker assumes already available in the hearer's mind at the time of utterance" (1994: 50). As Abbott (2000) observes, Lambrecht expands

his understanding of presupposition even further by stating that the set of pragmatically presupposed propositions also includes those that the speaker expects the hearer to be “ready to take for granted at the time the sentence is uttered” (1994: 52). Topic is a part of the presupposition, but is not identical to it.

Assertion seems to be a less controversial notion, and is usually simply linked to ‘new information’, or the information that is newly added to the hearer’s ‘knowledge’ by the utterance in question. However, Lambrecht highlights the importance of not completely equating assertion with ‘new information’, since ‘information’ is to do with the act of a speaker increasing a hearer’s knowledge of the world by adding a new proposition to it, while ‘assertion’ is the added proposition itself. Thus, pragmatic assertion is “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” (1994: 52).

Now that the notions of presupposition and assertion have been clarified, we can move to Lambrecht’s definition of focus. Focus for Lambrecht is:

[...] the element of information whereby the presupposition and the assertion DIFFER from each other. The focus is that portion of a proposition which cannot be taken for granted at the time of speech. It is the UNPREDICTABLE or pragmatically NON-RECOVERABLE element in an utterance. The focus is what makes an utterance into an assertion” (1994: 207).

Thus, focus is a part of assertion without being its equivalent. While topic is optional, focus is an obligatory component of a sentence.

Cross-linguistically a variety of cues can be used to signal the focused part of an utterance. In English, for example, a complex system of phonological cues is utilised to encode a variety of focus-related interpretative effects (see Steedman (2000) for a comprehensive descriptive overview). According to Wedgwood (2003), and contrary to many semantics-based approaches to focus, all these cues are aimed at assisting the addressee’s recognition of the asserted part of the utterances from the pragmatically presupposed parts. This general understanding of the notion of focus is adopted for the descriptive chapters of this thesis.

Lambrecht identifies three IS sentence types based on the focus articulation: predicate focus, argument focus, and sentence focus. These are presented and described in Table 9 below, based on Lambrecht (1994: 226-235). Focal accent is indicated by capitalisation in the examples column.

Table 9. Lambrecht’s IS Sentence Types

Sentence Type	Description	Examples
Predicate Focus	- The predicate is the focus and the subject (plus any other topical elements) is the presupposition. - Also referred to as the unmarked topic-comment structure.	A: What happened to your car? B (English): It broke DOWN. B (French): Elle est en PANNE.
Argument Focus	- Focus identifies the missing argument in a presupposed open proposition. - Also referred to as identificational sentence type.	A: I heard your bike broke down? B (English): My CAR broke down. B (French): C’est ma VOITURE qui est en panne.
Sentence Focus	- Focus extends over both the subject and the predicate (minus any topical non-subject elements). - Also referred to as the event-reporting or presentational sentence type.	A: What happened? B (English): My car broke down. B (French): J’ai ma voiture qui est en panne.

What Lambrecht identifies as Sentence Focus can also be referred to as broad, all-sentence focus, while Lambrecht’s Argument Focus can be labelled narrow focus. In Argument Focus sentences, the non-focused argument (if present) is usually topical, while the predicate can be labelled ‘presupposed’ or ‘backgrounded’, since verbs are not referential and therefore cannot be topical.

Lambrecht’s Predicate Focus – or the unmarked Topic-Comment structure – is also sometimes considered an instance of broad rather than narrow focus, since an utterance with this information-structural configuration could include a topical subject and focused predicate plus direct object. The descriptive terms of broad and narrow focus are adopted and used interchangeably with Lambrecht’s descriptors throughout this thesis.

In contrast to Lambrecht’s understanding of predicate focus, this term is also used in the literature to refer to “all instances of focus on lexical verbal predicates, such as V and VP, and on multi-functional elements in the extended verbal projection, like tense, aspect and mood (TAM-focus)” (Zimmerman 2016: 314). As can be

deduced from this definition of predicate focus, it excludes focus on nominal terms of adjuncts, as well as verum focus (which is discussed below). The examples of predicate focus are presented in (20) and (21).

(20) [What did Peter do? / Peter kicked the cat.]

(No,) Peter PETted the cat.

(21) [Peter is petting the cat.]

(No,) Peter HAS petted the cat (but he's done now).

(Zimmermann 2016: 315)

In (20), the focus indicated by prosody is on the verb *petted*, which is in contrastive relationship with the verb *kicked* in the immediately preceding utterance, while in (21) the auxiliary *has* is prosodically marked to indicate focus on the aspect of the verb, which contrasts with that of the preceding proposition.

Zimmerman (2016) argues against identifying the concept of predicate focus as a separate information-structural category due to the lack of unique features that would grant such categorisation. Using varied cross-linguistic data Zimmerman (2016) convincingly demonstrates that the marking devices (especially the prosodic ones) are consistently identical for signalling term and predicate focus. She additionally notes that both term and predicate focus can be reliably elicited with the help of the same empirical diagnostic methods, such as answers to wh-questions, corrections and association with exclusive focus particles. Zimmerman's view is followed in this thesis, and the term 'predicate focus' is only used in Lambrecht's sense presented above. Where one of the TAM categories is in focus, it is simply observed that a certain tense, aspect, or mood is in focus.

The final focus type that must be discussed in this section is that of verum focus. Höhle (1988, 1992) is credited for coining the term 'verum focus', however, the phenomenon itself had been previously noted and described by Halliday (1967), and Watters (1979), who used the term 'polarity focus'. In Höhle's understanding of the phenomenon, verum focus is used by the speaker to affirm the truth of her thought.

In other words, verum focus is the focus on the truth value of a proposition. In German, for example, this is indicated by a pitch accent in the left periphery of a clause, as shown in the example in (22) below.

- (22) Karl HAT den Hund gefüttert.
Carl has the dog fed
'Carl DID feed the dog.'

(Lohstein 2016: 290)

According to Höhle, the accentuation of the finite auxiliary in sentences like that in (22) is due to the assignment of the semantic element Verum to it. Gutzmann (2012) classifies this approach to verum focus as 'Focus Accent Thesis' – where Verum is considered to be a silent operator, which can become focused. Verum focus, then, occurs when this covert element is in focus, thus making verum focus a sub-type of focus. This approach is also supported by Büring (2006), and Zimmerman and Hole (2008), amongst others. The other way to look at this, is to view Verum as a conversational operator, which is present if verum focus is realised, and absent otherwise. This is the so-called 'Lexical Operator Thesis', and is supported, for example, by Romero and Han (2004), and Gutzmann and Castroviejo Miró (2011).

For the purposes of this thesis, I follow Stommel (2011) in her identification of verum focus as contrastive focus, and consider verum focus to be contrastive focus on the polarity of an utterance, especially in the context of a semantically identical proposition of opposite polarity. The example of verum focus in (22) provides a suitable illustration for this description, since this particular utterance would only be felicitous in the context of a semantically equal proposition of opposite polarity, e.g.: *Carl did not feed the dog*. The verum focus effect arises from this replacement of one polarity for the opposite one. This is in line with the understanding of contrastivity and contrastive focus presented in 2.1.3 below.

Before moving to that, however, such focus-related concepts as focal accent and focus projection need to be considered. Focal accent is the prosodic prominence of a word or minimal segment, usually associated with that word or segment being

focal or in focus. However, Lambrecht observes that the accented expression does not necessarily constitute the whole of the focus expression. This is what is referred to as the phenomenon of focus projection, where the same sentence with the same focal accent can express a variety of focus sizes.

The notion of focus projection was first described for English by Halliday (1967), and introduced into generative linguistics by Chomsky (1971), followed by work on this topic by Selkirk 1995, Rochemont 1998, Godjevac 2000, Büring 2006, Féry and Samek-Lodovici 2006, Arregi 2016, amongst many others. The usual representation of what is referred to as focus projection is shown on the sentence in (23) where the same constituent – *giraffe* – is accented regardless of the size of the focus.

(23) [She [bought a [GIRAFFE]_F]_F]_F.

Thus, the sentence in (23) can felicitously serve as the answer to questions inducing sentence, predicate, or argument focus: *What happened?*, *What did she do?*, *What did she buy?*. This example demonstrates that focus can project in a bottom-up manner from the most embedded constituent, which in a simple English SVO sentence is the direct object, onto larger constituents which dominate the element carrying the main sentential stress, until it reaches parts of sentences containing pragmatically presupposed material; if such material is not present, the entire clause is focused (all-focus sentences). It must be noted that the basic example of focus projection given in (23) does not cover some of the more complex cases of focus projection, but it is sufficient for the cases considered in this thesis.

For the purposes of this thesis I employ the notion of focus projection as a descriptive term without reference to some hierarchical structure assumed in most syntactic theories. It is shown in Chapter 3 that the concept of focus projection is also relevant for the Kazakh language, although it looks slightly different from the example in (23) due to the difference in the position of the direct object.

To sum up, in this sub-section, Lambrecht's definition of focus has been introduced, along with the three focus-based information-structural sentence types.

Other focus types, such as broad, narrow, and verum focus have been presented, and discussed in comparison to Lambrecht's terminology. The notion of contrastive/corrective focus has been introduced, but it is discussed in more detail in sub-section 2.1.3 dedicated to contrast and contrastivity. The notions of focal accent and focus projection have also been covered in this sub-section, and it was demonstrated that the constituent receiving the main focal accent is not necessarily the only constituent in focus.

2.1.3. Contrast

In this sub-section I consider the notion of contrast, which, according to Repp (2010), has been assumed to be an independent information-structural notion (see Molnár 2006) which often co-occurs with other IS categories, such as topic and focus, thus producing their subtypes – contrastive topic and contrastive focus respectively. Generally speaking, contrastivity indicates some sort of digression from or contradiction to some of the information present in the CG.

Lambrecht (1994) does not view contrastivity as a category of grammar, but suggests instead that the impression of contrastiveness arises from certain inferences that we draw based on given conversational contexts. Lambrecht follows Bolinger (1961) in his gradient approach to contrastiveness, expressed in the following excerpt:

In a broad sense, every semantic peak is contrastive. Clearly in *Let's have a picnic*, coming as a suggestion out of the blue, there is no specific contrast with *dinner party*, but there is a contrast between picnicking and anything else the group might do. As the alternatives are narrowed down, we get closer to what we think of as a contrastive accent (1961:87).

Lambrecht sees the advantage of this approach in that it allows for clear and less clear cases of contrastiveness. He states, however, that although “some pretheoretical notion of contrastiveness may apply to both” (1994: 291) contrastive focus and topic, the function of a contrastive focus is different from that of a contrastive topic. Let us first consider the notion of contrastive focus, before moving on to contrastive topic.

In Lambrecht's view, the clearest instances of contrastive focus are those where a stated or predicted focal referent is explicitly contradicted or replaced by another referent. In English, this is usually indicated by a contrastive accent, as shown in the example below.

(24) Pat said they called HER.

(Lambrecht 1994: 286)

Even in the absence of preceding context, contrastive accenting on the pronoun *she* indicates that some other referent is being replaced by the referent of this pronoun. In other words, as Gussenhoven (2007) puts it, a focus can be understood to be contrastive when it is on the constituent that is “in direct rejection of an alternative, either spoken by the speaker himself (‘Not A, but B’), or by the hearer” (2007: 90). Gussenhoven (2007) further suggests that the term contrastive focus can be used interchangeably with the that of corrective focus; this suggestion is followed in this thesis. The non-corrective/contrastive type of focus is referred as presentational or new-information focus (NIF).

Corrective/contrastive focus type is not to be confused with narrow focus, however, a corrective focus is most likely to be narrow (whereas a narrow focus can be NIF or corrective). Some languages have been reported to make a formal distinction between the expression of corrective/contrastive and presentational focus types. Navajo, for instance, has been claimed to have a neutral negative form *doo...da*, as shown in (25), and another negative form – *hanii* – which is used with contrastively/correctively focused terms, as shown in (26) and (27).

(25) Jáan doo chidí yiyíłchø-da
 John NEG car 3-PST-wreck-NEG
 ‘John didn’t wreck the car.’

(26) Jáan hanii chidí yiyíłchø
 John NEG car 3-PST-wreck
 ‘JOHN didn’t wreck the car (someone else did).’

- (27) Jáan chidí hani yiyíłchø
 John car NEG 3-PST-wreck
 ‘John didn’t wreck the CAR (he wrecked something else).’
 (Gussenhoven 2007: 91)

Chapters 3 and 6 show that contrastive focusing in Kazakh can be achieved either via contrastive accenting, or via a *ğoj*-cleft construction.

Moving to the notion of contrastive topic, let us consider the example below used by Lambrecht (1994) to illustrate this phenomenon.

- (28) I saw Mary and John yesterday. SHE says hello, but HE’s still angry at you.
 (Lambrecht 1994: 291)

According to Lambrecht, the accented pronouns in the second sentence encode two topic referents that are contrasted with one another. Lambrecht states that “the function of such contrastive topics is entirely different from that of contrastive foci” (1994: 291), with the main difference being in the fact that, unlike a contrastive/corrective focus, a contrastive topic is not associated with the idea of correction or contradiction.

Lambrecht observes that the two kinds of contrastive elements formally differ in Japanese, for example.

- (29) [Roommates Hanako and Mary discussing household chores:]
- | | | | | | |
|----|------------|----------|----------|-------|----------|
| H: | Mary-san, | anata-wa | osoji | shite | kudasai, |
| | Mary-VOC | you-TOP | cleaning | do | please |
| | watashi-wa | oryori | shimasu | kara. | |
| | 1SG-TOP | cooking | do | CONJ | |
- ‘Mary, you do the cleaning, I’ll do the cooking.’

M: Ie, watashi-ga oryori shimasu kara;
 No 1SG -NOM cooking do CONJ
 anata-wa hokanokoto shite kudasai.
 you-TOP other thing do please

‘No, I’ll do the cooking; YOU do something else.’

(Lambrecht 1994: 292)

According to Lambrecht’s analysis, the marker *-wa* indicates contrastive topics, while the exhaustive marker *-ga* signals contrastive focus. Note that regardless of the presence or absence of special topic-marking morphemes in a particular language, contrastive topics are expected to appear sentence-initially, since they indicate a change of topic referent in regards to which the rest of the utterance is to be constructed.

Taking Lambrecht’s observations into consideration, the notions of contrastive and shift topics are used interchangeably and are taken simply to indicate a topic which is divergent from the preceding aboutness topic; that is to say, a clear change of the topic is observed either in parallel structures within one utterance, as in the example in (30), or in question-answer discourses, as in (31). In (30), *Nina* is the contrastive/shift topic, since it differs from the preceding topic *John*; while in the response in (31), *my sister* diverges from the topic present in the question – *your siblings* – thus becoming a contrastive/shift topic.

(30) John went to the cinema, and Nina went shopping.

(31) [Where do your siblings live?]

My sister lives in Surrey.

Some examples of contrastive/shift topics in Kazakh are provided in Chapter 3, while Chapter 5 is dedicated to topic markers in Kazakh, some of which are shown to be contrastive topic markers, which means that they indicate that the topic referent of the utterance in which they appear has shifted from that of the preceding utterance.

Summing up, in this sub-section the notions of contrastive focus and topic have been considered. It has been shown that while both these notions have a contrastive nature, the way it interacts with their pragmatic behaviour within context results in the contrastive focus differing from the contrastive topic. Contrastive focus can also be referred to as corrective focus, and is manifested where a previously established focal referent is rejected and replaced by a new referent. A contrastively focused expression can be indicated by a contrastive accent or morpho-syntactically via special morphemes or syntactic constructions.

Contrastive topic is also labelled switch topic, and is observed when the topic referent of a clause differs from the topic referent of the preceding clause; it is usually overtly signalled by contrastive accentuation of the topical expression or special morphemes.

2.1.4. Specificity and Definiteness

While specificity and definiteness might not be considered information-structural notions per se, they certainly are pragmatic, discourse-anchored notions. Lambrecht considers these notions within his discussion of characteristics of topicality, and states that definiteness is a grammatical category (1994: 79), while specificity is a semantic notion which is linked with identifiability, and has no direct correlate in the definite/indefinite contrast (1994: 80).

While this approach might be applicable for languages with definite articles, such as English, German, or French, for instance, it is not entirely suitable for languages like Kazakh, which lacks definite articles. Interestingly, Lambrecht considers some Turkish examples – similar to those in (36) and (37) – and uses the terms ‘identifiable’/‘unidentifiable’, as well as ‘specific’/‘non-specific’, and ‘definite’ and ‘indefinite’ in his discussion of these examples.

In order to avoid any inconsistency or confusion between what is or is not considered to be a grammatical category, I follow von Heusinger’s (2002) approach to the notions of definiteness and specificity. This is especially relevant for Chapter 4 dedicated to differential object marking in Kazakh.

Von Heusinger (2002: 247) provides the following informal characteristics of specificity:

- (i) the interpretation of a specific NP does not depend on the interpretation of the matrix predicate or semantic operators such as modal verbs.
- (ii) the referent of a specific NP is functionally linked to the speaker of the sentence or to another referential expression in the sentence such as the subject or object.
- (iii) the lexical item ‘a certain’ prominently marks a specific reading of an indefinite NP.

Furthermore, he rejects identifying specificity as a subcategory of indefinite noun phrases (NPs), and posits that it should be treated as an independent category which interacts with definiteness thus creating a cross-classification. This cross-classification is presented in Table 10 below (from von Heusinger 2002: 253).

Table 10. Cross-classification of definiteness and specificity

	discourse old	discourse new
referentially anchored to discourse referents	specific definite NPs	specific indefinite NPs
referentially bound by operators	non-specific definite NPs	non-specific indefinite NPs

The four NP types resulting from this cross-classification are illustrated by the examples in (32)-(35), where the appropriate NPs are italicised.

(32) Specific, definite NP
The body was found in the river yesterday.

(33) Non-specific definite NP
The tiger has stripes.

- (34) Specific indefinite NP
A body was found in the river yesterday.
- (35) Non-specific indefinite NP
A tiger has stripes.

As can be seen from these examples, in the English language, specificity is not expressed by a particular article, unlike definiteness, which is indicated by the article *the*. However, as noted by Lyons (1999: 59), articles that mark “specificity, or something close to specificity, rather than definiteness are fairly widespread”.

Von Heusinger uses Turkish as an example of a language that morphologically marks specificity on direct objects via the accusative case marker *-I*, as the examples in (36) and (37) demonstrate.

- (36) Ben bir kitap oku-du-m.
 1SG a book read-PST-1SG
 ‘I read a book.’
- (37) Ben bir kitab-ı oku-du-m.
 1SG a book-ACC read-PST-1SG
 ‘I read a certain book.’

(von Heusinger 2002: 255)

In both sentences the direct object *kitab* is preceded by the indefinite article *bir* ‘a/one’, but only in (37) is the direct object also marked with the accusative case marker, which yields the indefinite specific reading for that noun. This is referred to as differential object marking, and is discussed in more detail in Chapter 4.

In summary, this sub-section presented von Heusinger’s (2002) approach to teasing apart the concepts of specificity and definiteness. Unlike many prior approaches, which considered specificity a subcategory of indefiniteness, von Heusinger recognises it as a separate category, and proposes a cross-classification of

it and definiteness. Specificity of an NP in von Stechow's understanding indicates that this NP is referentially anchored to another discourse object, which might, but does not have to be, the producer of the proposition.

2.1.5. The Role of Information Structure in the Grammar

The second line of IS research, namely, the line concerned with the role of IS within the architecture of grammar, is understandably heavily dependent on the syntactic theoretical framework assumed by a particular researcher. To illustrate this point, I briefly discuss two examples of approaching this question, namely, the syntactocentric approach of the Chomskyan generative grammar, and the multi-stratal model of Lexical-Functional Grammar (LFG).

The syntactocentric approach posits that the basic IS units, such as topic and focus, belong in the formal language system, and that they are directly represented on the phrase structure trees, either in the form of IS-related phrasal nodes (TopF or FocF, for instance), or in the form of IS features assigned to particular positions in trees. This approach was initially developed by Chomsky (1971) and Jackendoff (1972), and is followed by the Minimalist programme and other transformational generative syntactic theories (see e.g.: Rizzi 1997, Neeleman and van de Koot 2008, Abels and Muriungi 2008, Kučerová 2012, amongst many others).

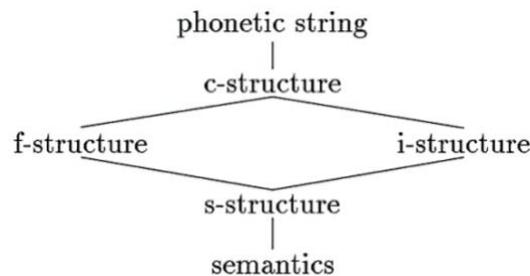
According to Schwabe and Winkler (2007), the main argument for proposing that focus (and other IS notions) is directly and unambiguously encoded in syntax comes from the so-called focus movement phenomena. For instance, a f(ocus)-feature is present within the syntactic clause structure and may cause the focused constituent to move into the specifier position of a functional projection in the left-peripheral domain of the clause, as É. Kiss (1998) reports for the identificational focus in Hungarian, for example. Being rooted within syntax, the f-feature receives its prominent phonological form (PF) in accordance with certain phonological rules, and is translated into a semantic representation at the logical form (LF).

LFG proposes a non-transformational multi-tiered syntactic model, which allows us to capture the IS-syntax interactions without positing specific clausal

positions for different IS notions. The c(onstituent)-structure and f(unctional)-structure occupy separate tiers in LFG and, prior to King's (1997) seminal paper, information or discourse function-structure was encoded within the f-structure via annotations on the c-structure.

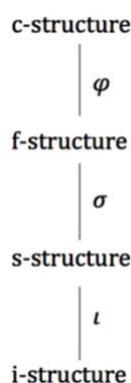
King (1997) noticed, however, that while it worked well for topic and focus structurally encoded on arguments, it did not work so well in cases where f-structure heads were focused or topicalized (focus on the verb within a verb phrase, for instance), which resulted in incorrect scoping – more material than intended appeared topicalized or focused as the f-structural and information-structural constituents did not always coincide. This prompted the proposal of a separate i(nformation)-structure projection, distinct from the f-structure projection, as shown in the schema in (38); s-structure represents semantic structure.

(38) King's levels of representation



LFG assumes that constituents of one level of structure are related to those in another level through a one-to-one or many-to-one relation. Butt and King (2000), Mycock (2006), Dalrymple and Nikolaeva (2011) amongst other LFG researchers follow King's (1997) proposal to treat information structure as a separate level of representation, independent from f-structure. However, Dalrymple and Nikolaeva's model (2011: 90) shown in (39) differs from King's model and other proposals within LFG.

(39) Dalrymple and Nikolaeva's levels of representation



The levels of linguistic structure – c-structure, f-structure, s-structure and i-structure – are related to each other through the functions φ , σ and ι . The relation between the s-structure and the i-structure through the function ι forms an important aspect of Dalrymple and Nikolaeva's (2011) proposal as it shows that all parts of a clause share the same information structure, that is, there is a “single information structure for the entire clause” (2011: 91).

LFG and other non-transformational multi-stratal syntactic frameworks provide a convenient representational and analytical tool for analysing IS and IS-related phenomena, and for establishing connections between IS and syntax. However, as Kempson et al. (2006) observe, the additional level of IS “lacks independent motivation and implies a non-minimalist architecture for the grammar” (p. 59). The same observation can be applied to syntactic movement triggered by IS postulated by generative grammarians, as there is no independent motivation for the highly theory-dependent proposals. The syntactic theoretical framework of DS followed in this thesis avoids that; an introduction to this framework and its approach to information structure are presented in the following section.

2.2. Dynamic Syntax

This section introduces the theoretical framework of DS (Kempson et al. 2001; Cann et al. 2005). This framework is used for the formal analyses of Kazakh word order-IS

relation, differential object marking, Kazakh topic markers, and the Kazakh particle *goj* described and discussed in Chapters 3, 4, 5, and 6 respectively.

Firstly, in 2.2.1, I provide a general introduction to the framework; in 2.2.2-2.2.6, DS formal tools, mechanisms, and rules are presented. In 2.2.7 the DS parsing process of a Korean SOV sentence is given, while 2.2.8 examines the role of an intonation phrase boundary in DS tree construction. Finally, 2.2.9 deals with the DS approach to the central notions of IS.

2.2.1. The Principles of the Framework

The formal syntactic framework of Dynamic Syntax (DS) is motivated by the intuition that language capacity is the ability to process language input, and that knowing a language means knowing how to parse it (Cann et al. 2005). DS is interested in the ability of the hearer – the parser – to build semantic representations from a linguistic sequence in the order in which parts of this sequence are encountered. As observed by Kiaer (2014: 120), DS “presents a ‘process-oriented’ instead of a ‘result-oriented’ view”, which means that DS is not interested only in the static result of the parsing process, but also in the steps of its dynamic, incremental structural development, underpinned by the notions of underspecification and update.

DS also acknowledges that this incremental development does not happen in isolation – context is taken into consideration in this framework, and the linguistic strings are interpreted and enriched by the context within which they are produced. This inclusion of contextual environment into the syntactic theory makes DS appropriate for investigating and formalising such highly context-dependent phenomena as constituent order alternations, contrast, focus and topic realisations. DS views on information structure and its components is explored in detail in 2.2.9.

The linear and incremental nature of the parsing process – as it unfolds in natural communication – is reflected in DS formalisms, which are based upon the dynamics of left-to-right, word-by-word parsing of language within context. This contextualised parsing is formally represented as a monotonic semantic tree growth process. According to Kempson et al. (2006), the central claim of DS is that “syntactic

properties of natural language reside exclusively in the progressive growth of such tree-structure representations strictly following the dynamics of left-right processing” (2006: 60). The semantic tree is the only level of representation employed by DS, in contrast with many other theoretical syntactic approaches, which use multiple levels of representation (e.g.: Lexical Functional Grammar or the Minimalist Program).

As the parsing process unfolds step-by-step, the semantic tree incrementally develops until the linguistic string is completed and parsed. This incremental growth process represents syntax in DS, and it is through this incremental growth that DS can account for many syntactic phenomena which are challenging for the bottom-up or top-down static approaches to syntax (see Kiaer 2014). Focus on the dynamic progress of tree development means that all the intermediary steps of that development are as important – and, in a way, more so – as the final tree.

This is one of the principal and crucial differences of the dynamic approach to syntax from the mainstream syntactic theory. The static representations of some hierarchical constituent structure of complete sentences are rejected, and the surface structures of linguistic strings are considered to consist of incrementally parsed ‘instructions’ which allow the parser to construct certain types of semantically structured propositional forms.

Parsing and not production is taken as the basic linguistic activity in DS for several reasons. Firstly, as observed by Cann et al. (2005), the task of parsing is more basic compared to production as far as language acquisition is concerned – perception infallibly precedes the production; there is strong evidence that at all stages of language acquisition children are able to understand structures that are more complex than those they can themselves produce, usually quite significantly so.

Secondly, Cann et al. (2005) note that that parsing and production largely rely on the same strategies – constructing representations of content from information in the context – with one constraint in that the structures constructed during production must match the intended context.

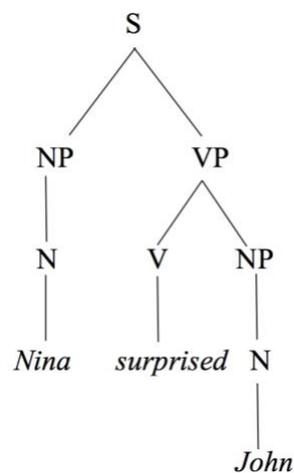
Thirdly, as Sperber and Wilson (1995) observe, human cognitive activities are inference based, or, as Cann et al. (2005: 21) put it: “Humans are information processing animals”. This applies to all types of informational input experienced and

processed by humans in a reflex-like way, not just the verbal input, and our capacity to assess and select information in linguistic communication is a manifestation of our general capacity to deal with information.

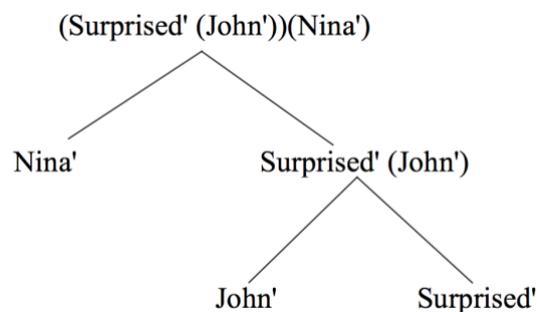
2.2.2. The Dynamic Syntax Tools

In this section I take a detailed look at the semantic trees mentioned in the preceding section. Let us begin with a comparison of a familiar ‘structure over words in a string’ type syntactic tree and a DS semantic tree, as shown in the examples (40) and (41) respectively.

(40) A hierarchical tree structure over a string of words



(41) Propositional tree structure



The tree in (40) illustrates the assumed properties of words in strings which determine the structure over those strings. For example, the label VP is used for the node that consists of a verb plus another word functioning as a noun and a noun phrase. As can be seen in (40), the words from the linguistic string *Nina surprised John* are presented to exist separately from the syntactic structure above them; the latter is constrained and governed by rules and principles that have no direct connection to the linguistic string.

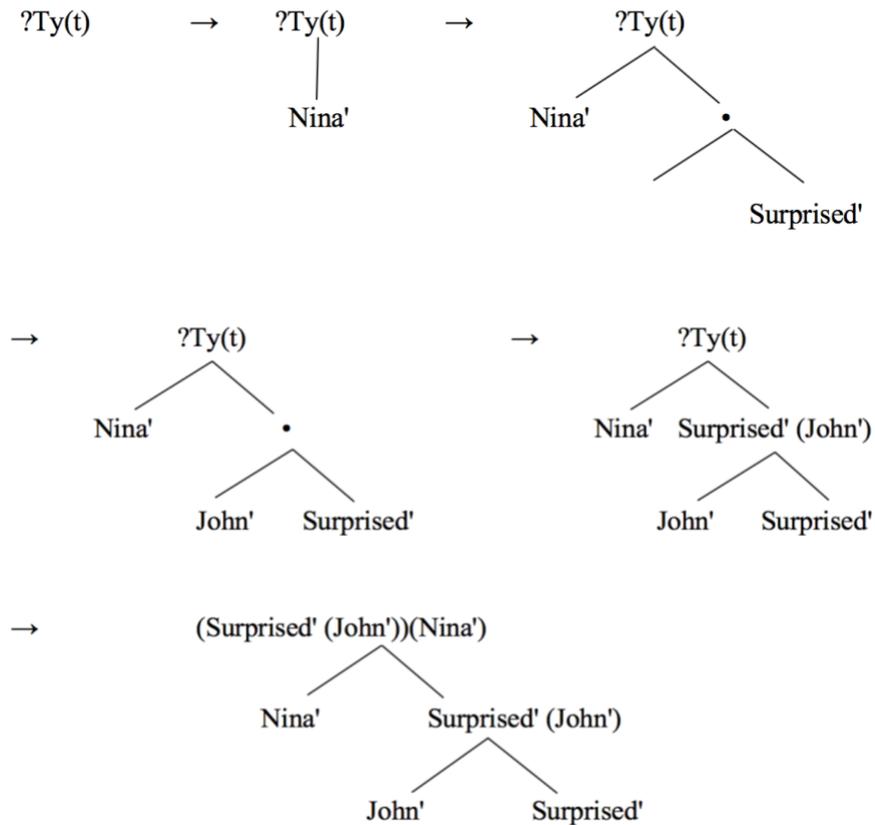
Unlike the tree in (40), the tree in (41) does not contain any information on the word order of the linguistic string. What it represents instead is the semantic structure of the propositional context expressed by the string *Nina surprised John*. The labels decorating the nodes of this tree are compositionally derived concepts expressed in some lambda calculus, like in some versions of categorical grammar (inter alia Morrill 1994, Carpenter 1998). The prime (') is used to indicate that the concept and not the word itself is used to decorate the nodes of the tree. It is assumed that from the English word *John* we construct the concept *John'*, which denotes a particular individual called *John*.

In the tree in (41) the linguistic string does not exist on a separate level from the structure determined by that string. Instead, it is a structured representation of the interpretation assigned to linguistic strings in the context in which they are produced.

As mentioned in the previous section, the steps by which the final tree is reached are as important as that final tree itself. Unlike in the Minimalist Program (Chomsky 1981, 1995), the process by which a semantic tree like that in (41) is built up is strictly time linear, and unfolds step-by-step, from the initial word to the final one.

This process is presented in (42) below; it begins with the overall goal to establish some propositional formula ($=?Ty(t)$) – according to Sperber and Wilson 1995 this goal is pertinent of any parsing effort.

(42) Parsing *Nina surprised John*



The semantic concepts decorating the nodes of the tree are called formulae. These are usually expressed as values of a predicate Fo , so for the semantic information contained in the string *Julia slept* we write (ignoring the tense) $Fo(\text{Sleep}'(\text{Julia}'))$. This is just one of the labels that can decorate a node.

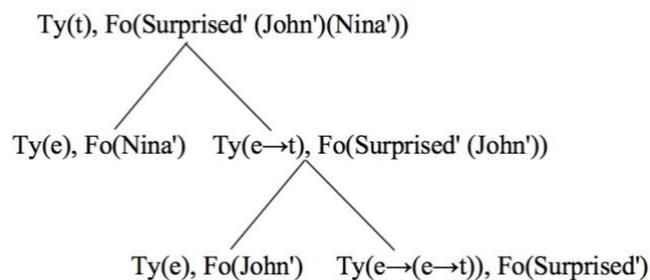
Another kind of label provides information on the type of the formula expression, its semantic category. An expression of propositional type t denotes a truth value, and an expression of type e denotes an entity; a functor type expression $e \rightarrow t$ denotes a one-place predicate: when it combines with a term (of type e) it yields a proposition (of type t) and denotes a set. The third and final basic type employed by DS is cn which expresses a nominal and is assigned to common nouns in which the formula consists of an ordered pair of a variable plus a propositional formula where that variable occurs free; the expression $cn \rightarrow e$ is assigned to quantifiers. The most common types used in DS are shown below in Table 11.

Table 11. Types in Dynamic Syntax

Type	Description
Ty(e)	Individual term
Ty(t)	Proposition
Ty(e→t)	One-place predicate
Ty(e→(e→t))	Two-place predicate
Ty(e→(e→(e→t)))	Three-place predicate
Ty(t→(e→t))	Proposition Predicate
Ty(cn)	Nominal
Ty(cn→e)	Quantifier

In (43) we can see the final tree from (42), which has been decorated with the formulae and types annotations presented above.

(43) Representation of content *Nina surprised John*



The tree in (43) illustrates how information from the functor nodes – on the right – combines with information from the argument nodes – on the left – thus resulting in the complex formula at the mother nodes. Let us now consider the stages of tree growth in more detail.

2.2.3. Tree Growth Process

The tree growth process in DS represents the parsing process and takes the central place in this framework. The parsing process is driven by the need to fulfil the requirements – that is, to provide updates for underspecified information. A requirement is shown by a question mark in front of the label that needs to be instantiated. A requirement signifies that by the end of the parsing process the node at which it appears must be updated with an instance of the required type; all requirements must be satisfied for the tree to be complete.

The universal requirement to establish propositional content is represented by the requirement $?Ty(t)$, as in the example in (42), for instance. This requirement forms a part of the minimal initial tree of any derivation, and is introduced as part of the rule known as the Axiom shown in (44) below.

(44) The Axiom

$$T_n(0), ?Ty(t)$$

The Axiom provides the initial minimal ‘tree’ consisting of a single node which is underspecified for the propositional content, but contains the goal of deriving a formula of type t . This initial node is referred to as the root node and is marked with the tree node address $T_n(0)$. Once a formula of type t has been constructed at the root node, we can say that the requirement has been satisfied and the question mark can be removed yielding $Ty(t)$ decoration, as at the root node of the tree in (43), for example. The same applies to requirements for other formula types – the question mark indicating an unfulfilled requirement is removed as soon as the requirement is satisfied.

There are numerous ways in which requirements can be satisfied, depending especially on the type of formula that is required. A requirement for a type t formula, for instance, might not be able to be satisfied in one step (i.e. by parsing just one word), so the goal of fulfilling such a requirement would need to be broken down into smaller goals the satisfaction of which would consequently lead to the satisfaction of the

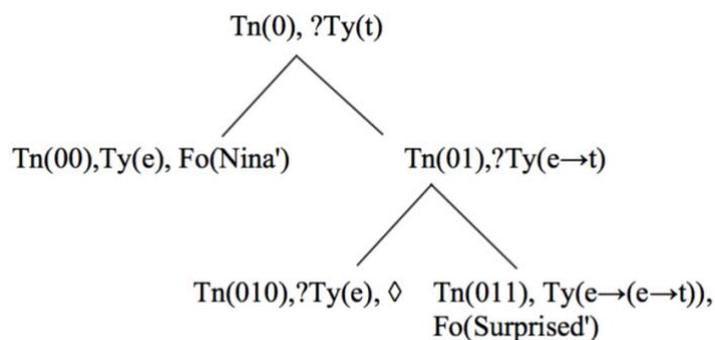
overall goal. This is to say that a requirement may remain unfulfilled up until the very end of the parsing process, as long as it is fulfilled before it is complete.

In order to indicate the node at which the derivation process is currently taking place, a so-called pointer is used – \diamond . When the pointer is at a particular node, it indicates the present point of tree development in accordance with the requirements held at that node. Thus the minimal or initial tree, shown in (44), should also include the pointer in its representation, as shown in (45) below.

(45) $T_n(0), ?Ty(t), \diamond$

With this information in hand, let us return to the tree from (43) and take a look at a snapshot of it before completion, namely at the stage where only the first two words *Nina* and *surprised* have been uttered, as shown in (46).

(46) Parsing *Nina surprised*...



Observe, that at this stage of tree development there are three unfulfilled or outstanding requirements: $?Ty(t)$, $?Ty(e \rightarrow t)$, and $?Ty(e)$. The positioning of the pointer indicates that the current node under development is the internal argument node of the predicate. This has been determined by the algorithm contained in the lexical entry (discussed in detail in Section 2.2.6 below) of the verb *surprised* when this verb was parsed. All the tree nodes are identified by a numerical index following the convention that the left daughter node of a node n is labelled with the index $n0$ and

the right daughter is labelled with the index $n1$. By another convention, in a DS tree arguments are always left daughters, and functors are right daughters of a mother node.

Taking a closer look at the tree in (46) we can see that the pointer is currently at the node $Tn(010)$, which contains a requirement for a formula of type e , or, in other words, a requirement for a term. As we already know, the next word in our linguistic string here is *John*, which can be successfully introduced into this tree at the current node, as it, being a type e expression, satisfies the current requirement. On the contrary, an expression like *saw* ($Ty(e \rightarrow)$) could not be successfully parsed at this node – even though it would satisfy a requirement present elsewhere within this tree – as the pointer is currently not at the node with that particular requirement. This means that a word like *saw* could not provide an update required at the current stage of the parsing process. The parsing process would not go ahead and would be aborted if the verb *saw* appeared after the verb *surprised* (**Nina surprised saw*), as this would result in ungrammaticality. Thus the position of the pointer plays a significant role in the notion of grammaticality in DS along with the need for all requirements to be satisfied by the end of the parse.

As briefly mentioned above, terms like ‘mother’ and ‘daughter’ node can be used to describe relationships between the nodes, however, DS also has more formal terms for this from the language of the Logic of Finite Trees (LOFT) (Blackburn and Meyer-Viol 1994; Kempson et al. 2001). The following two basic LOFT modal operators are used: one which corresponds to the daughter relation, $\langle \downarrow \rangle$, down; and one corresponding to the mother relation, $\langle \uparrow \rangle$, up. The angled brackets are used to indicate an existential statement, thus the notation like $\langle \uparrow \rangle$ can be interpreted as ‘there is a mother node’.

A numerical subscript can be added to these operator symbols if it is necessary to specify that an argument and a functor branch, thus, $\langle \downarrow_0 \rangle$ refers to an argument daughter, and $\langle \downarrow_1 \rangle$ refers to a functor daughter. The modality operators can be combined recursively, and used to describe the properties of one node from the position of the other. For example, for one of the nodes from the tree in (46), a statement like the one presented in (47) below can be made.

(47) $\langle \uparrow_0 \rangle \langle \uparrow_1 \rangle \langle \downarrow_0 \rangle, \text{Ty}(e), \text{Fo}(\text{Nina})'$

This statement is made from the node $T_n(010)$, and refers to the node $T_n(00)$. It can be read as follows: if you go up an argument daughter node, then up a functor daughter node, and then down an argument daughter node, you will arrive at the node with the decoration $\text{Ty}(e), \text{Fo}(\text{Nina})'$; or, in other words, at my mother's mother's argument daughter, $\text{Ty}(e), \text{Fo}(\text{Nina})'$ holds.

These modal operators are widely used in DS as parts of algorithms in lexical entries, as they facilitate the local linguistic context checks, annotation of nodes other than the current one, and construction of new nodes in the tree. In combination with the notion of requirement, modal operators can also be used to add constraints on further development of the tree, or a filter on the output tree. Thus, an annotation like $\langle \downarrow_0 \rangle \text{Fo}(\alpha)$ provides a requirement that $\text{Fo}(\alpha)$ must decorate the node's argument daughter at some point in the current tree construction.

It is important to note that terminal and non-terminal tree nodes are distinguished in DS. A terminal node is indicated by the *falsum* symbol: \perp ; it is used in conjunction with the 'down' modal operator in square brackets: $[\downarrow]\perp$. The square brackets here signify a universal statement, and the whole statement reads: 'for every node dominated by the current one the *falsum* holds', meaning that the current node is a terminal node and cannot have any daughters. The bottom restriction is assumed to belong in the lexical entries of all full content words which decorate terminal nodes, therefore, it is not necessarily shown in the trees unless absolutely crucial for the analysis.

Modal operators $\langle \uparrow \rangle$ and $\langle \downarrow \rangle$ can also be combined with the Kleene* (Kleene star, Kleene operator or Kleene closure), which essentially identifies a 'zero or more' set. As Cann et al. (2005) observe, the use of the Kleene* operator in DS is similar to the concept of 'functional uncertainty' in LFG (Kaplan and Zaenen 1989), although the final analyses are disparate. In combination with the modal operators, the Kleene star denotes an underspecified tree relation. For example, the expression $\langle \downarrow^* \rangle$ means 'if you go zero or more steps down there is a node...'; and the expression $\langle \downarrow_0^* \rangle$ means 'if you go zero or more steps down argument nodes there is a node...'. This node can

be any node under the current node, or, indeed, the current node itself since the number of steps to be taken can also be zero.

The combinations of the up and down operators and the Kleene* are also widely used in DS to express that some decoration holds at either some node above or below without specifying the precise location for this node. The annotation $\downarrow^* \text{Fo}(\alpha)$, for example, means that $\text{Fo}(\alpha)$ holds at a node which is zero or more steps down from the current one, and it can be used when the exact location in the tree is yet unknown.

2.2.4. Computational Rules in the Tree Construction Process

There are three types of actions that propel the process of tree building in Dynamic Syntax: a) computational actions; b) lexical actions; c) pragmatic actions. In this section we consider the computational (also known as transition or construction) rules that provide the computational actions as the basic mechanism of semantic tree construction. These rules follow the format shown in (48) below, where the top line reading presents the input tree description, and output tree description is given in the bottom line.

(48) Computational Rule Schema

Input Tree Description

Output Tree Description

A rule presented in this format can be interpreted as: from the given input tree, the described output tree can be derived. The input tree description contains a goal (requirement), which holds at some node, and which is broken down into sub-goals (other type requirements) in the output tree description. This multiplication of requirements that must hold at subsequent (daughter) nodes propels the tree growth process.

The computational rules constitute a close set, which is considered to be universally available, however, Bouzouita's (2008) and Gibson's (2012) claims that

the rules of Introduction and Prediction (presented in detail below) are not available for Medieval Spanish and Rangi respectively, contradict the cross-linguistic availability of these rules.

Recall that in DS the initial goal of any parsing process is the requirement to establish a proposition of type t ($=?Ty(t)$). The rule of Introduction is an inference from this initial goal, and breaks it down into two sub-goals, the fulfilment of which will ultimately fulfil the initial goal too. As shown in the Introduction rule presented in (49), and in the linear description of the tree growth process inducted by this rule shown in (50), the rule of Introduction unpacks the requirement for a proposition ($?Ty(t)$) into requirements for daughters of types e and $e \rightarrow t$. Note that if the expressions of the latter two types are combined through functional application, the resulting expression will be of the type in the initial requirement.

Thus, we can see that the larger goal gets divided into sub-goals, the achievement of which subsequently results in the achievement of that initial larger goal. Here, these rules are presented not in their formal appearance, but in the form that shows tree growth directly, using the tree-decorating expressions introduced in the previous sections; the formal definitions of the rules can be found in Cann et al. (2005).

(49) Introduction (Rule)

$$\frac{\{\dots \{Tn(n), ?Ty(t) \dots \diamond\} \dots\}}{\{\dots \{Tn(n), ?Ty(t), ?\langle \downarrow_0 \rangle Ty(e), ?\langle \downarrow_1 \rangle ?Ty(e \rightarrow t) \dots, \diamond\} \dots\}}$$

(50) Application of Introduction

$$Tn(n), ?Ty(t), \diamond \rightarrow Tn(n), ?Ty(t), ?\langle \downarrow_0 \rangle Ty(e), ?\langle \downarrow_1 \rangle ?Ty(e \rightarrow t), \diamond$$

As can be seen from (49) and (50), the output tree in the rule of Introduction only differs from the input tree in the amount and types of requirements, but still only consists of one node. The construction of the daughter nodes (the requirements for

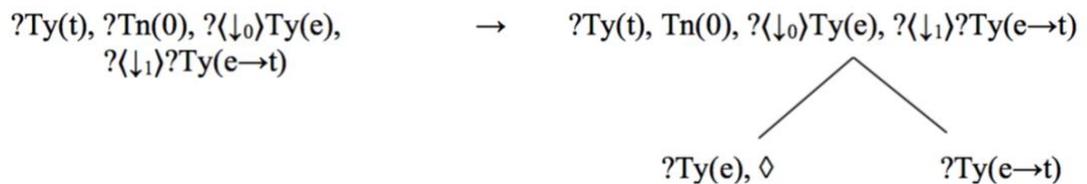
which are added during Introduction) is achieved by the rule of Prediction, presented in (51) below.

(51) Prediction (Rule)

$$\frac{\{\dots \{Tn(0), ?\langle \downarrow_0 \rangle Ty(e), ?\langle \downarrow_1 \rangle ?Ty(e \rightarrow t) \dots, \diamond\}\}}{\{\dots \{Tn(0), ?\langle \downarrow_0 \rangle Ty(e), ?\langle \downarrow_1 \rangle ?Ty(e \rightarrow t)\}, \{\langle \uparrow_0 \rangle Tn(0), ?Ty(e), \diamond\}, \{\langle \uparrow_1 \rangle Tn(0), ?Ty(e \rightarrow t)\}\}}$$

The tree growth steps resulting from the rule of Prediction are shown in (52) below.

(52) Application of Prediction



In (52) we see the subject and predicate nodes being constructed, and the pointer moving to the subject node, where the requirement for a type *e* expression holds.

As can be seen from the Introduction and Prediction rules presented above, these rules apply specifically in situations where no nodes apart from those introduced by the Axiom exist in a tree. It can also be seen, however, that these rules begin a tree building process suitable for a language with a strict SVO word order (like English), which ensures that the subject is always parsed before the verb or the object. Bouzouita (2008) and Gibson (2012) oppose the claim that the rules of Introduction and Prediction are universally applicable.

Bouzouita (2008) bases her rejection of the cross-linguistic nature of Introduction and Prediction on the fact that in Spanish the lexical characteristics of the verb (predicate) alone can unfold the construction of the subject-predicate template; this observation is indeed applicable to the other subject pro-drop languages.

Gibson (2012) demonstrates that Rangi, also being a subject pro-drop language, constructs the propositional structure from an inflected verb form – whenever one such is encountered in the parsing process. According to Gibson (2012), instead of following the Introduction and Prediction rules outlined above, in Rangi, the overt subjects are modelled onto a parallel tree which is connected to the main tree via a LINK relation.

In Chapter 3, Section 3.5 I present the DS model of Kazakh clause structure, which demonstrates that the Introduction and Prediction rules presented above are also not applicable in the parsing process of this pro-drop SOV language with rich case morphology. I now turn to the computational rules that, unlike the rules of Introduction and Prediction, are not concerned with the unfolding of tree structure during parsing, but rather deal with the values of the tree nodes. These rules need to provide a means of removing requirements; a means of moving the pointer from the completed nodes; and a means of collecting and compiling the information from the terminal nodes of the tree to satisfy the requirements held at the higher nodes.

The first rule from this group – Thinning – deals with removing the requirements and simplifying the information held at the node. The formal presentation of this rule is given in (53) below. The variable ‘X’ here and in the rules shown below stands for any expression that can potentially occupy that position – in this case it stands for any type of requirement.

(53) Thinning (Rule)

$$\frac{\{\dots\{\dots, X, \dots, ?X, \dots, \diamond\}\dots\}}{\{\dots\{\dots, X, \dots, \diamond\}\dots\}}$$

The rule of Thinning takes the input tree node description, in which both the requirement for a value and the value itself are present, and eliminates the requirement. The effect from the application of this rule is that all the satisfied requirements are removed from the node, as shown in the example in (54). In (54), the tree from (46) (*Nina surprised...*) is shown at the stage where only the first word – *Nina* – has been

parsed, and the node $T_n(00)$ still carries both the requirement for an expression of type e and the expression of this type itself. We then see the same node after the rule of Thinning is applied – the satisfied type requirement is removed and only the expression that satisfied it remains.

(54) Application of Thinning



Unlike the other transitional rules, the rule of Thinning is not optional and, despite appearing trivial in application, plays an important part in the completion of a tree. Recall, that a tree cannot be completed successfully if it contains any outstanding requirements.

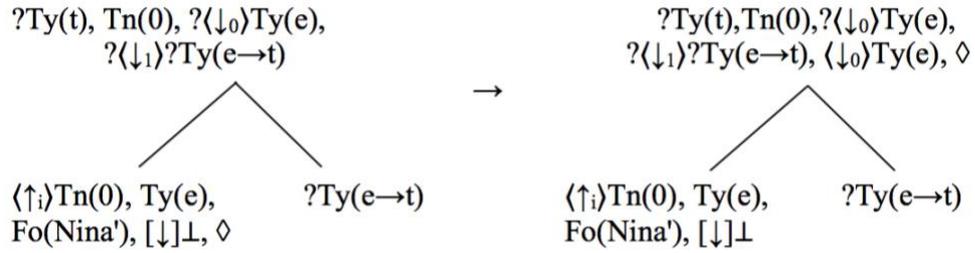
The rule of Completion is the rule that moves the pointer from a daughter node to a mother node, and, most importantly, annotates the mother node with the information about the daughter. This rule can be considered a pointer-moving rule that enables the parsing process. The formalism for this rule is presented in (55) below, and the application of the rule in the parsing process is shown in (56).

(55) Completion (Rule)

$$\frac{\{ \dots \{ \dots T_n(n) \dots \}, \{ \langle \uparrow_i \rangle T_n(n), \dots, Ty(X) \dots, \diamond \} \dots \}}{\{ \dots \{ T_n(n), \dots, \langle \downarrow_i \rangle Ty(X), \dots, \diamond \}, \{ \langle \uparrow_i \rangle T_n(n), \dots, Ty(X), \dots \} \dots \}}$$

$i \in \{0, 1, *\}$

(56) Application of Completion



There are two major differences between the tree before and after the application of the rule of Completion in (56): firstly, the root node in the second tree now carries the information about its argument daughter node; and, secondly, the pointer is moved from the node $Tn(00)$ up to the root node $Tn(0)$. At this point, the rule of Thinning will apply at the root node to remove the requirement for a type e expression at the argument daughter node ($=\langle\downarrow_0\rangle Ty(e)$), since it has been fulfilled and the appropriate decoration ($=\langle\downarrow_0\rangle Ty(e)$) has been added to the root node by the rule of Completion.

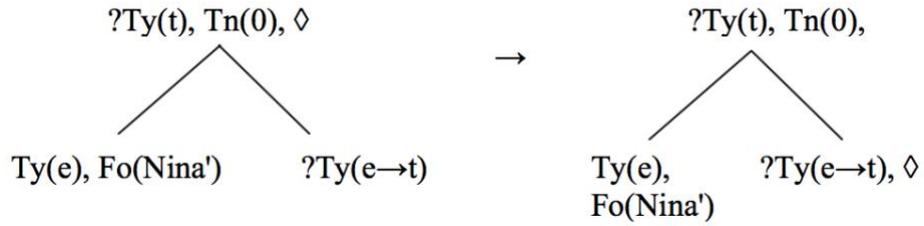
There is also a rule that moves the pointer down – the rule of Anticipation. This rule moves the pointer from a mother to a daughter node with an outstanding requirement; it is presented in (57).

(57) Anticipation (Rule)

$$\frac{\{\dots\{\dots Tn(n), \dots, \diamond\}, \{\langle\uparrow\rangle Tn(n), \dots, ?X \dots\} \dots\}}{\{\dots\{Tn(n), \dots\}, \{\langle\uparrow\rangle Tn(n), \dots, ?X, \dots, \diamond\} \dots\}}$$

In (58) we see the application of the rule of Anticipation to the tree growth process.

(58) Application of Anticipation



The pointer in the second tree in (58) has moved from the root node down to the functor daughter node with an outstanding requirement for a type $e \rightarrow t$ expression. Once this requirement is satisfied at this node (in this case by the parsing of the verb *danced*), the rules of Thinning, and Completion will apply, removing the satisfied requirement from the node $\text{Tn}(01)$, bringing the pointer back up to the root node, which will also receive an update as to the information held at the functor daughter node.

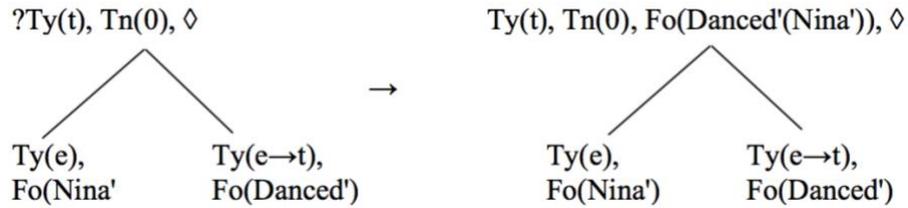
At that point, then, the requirements at both the argument and the functor daughter nodes are satisfied, but the parsing process cannot be considered completed until the last outstanding requirement ($=\text{?Ty}(t)$) is satisfied at the root node. This is where the rule of Elimination applies. Provided there are no outstanding requirements at the daughter nodes, the rule of Elimination takes the formulae from these nodes, performs functional application over them, and annotates the mother node with the resulting formula, which satisfies the outstanding requirement on that node. The formalism of the rule of Elimination is presented in (59), and its effect on the tree growth process is shown in (60).

(59) Elimination (Rule)

$$\frac{\{ \dots \{ \text{Tn}(n) \dots \text{?Ty}(X), \langle \downarrow_0 \rangle (\text{Fo}(\alpha), \text{Ty}(Y)), \langle \downarrow_1 \rangle \text{Fo}(\beta), \text{Ty}(Y \rightarrow X), \dots, \diamond \} \dots \}}{\{ \dots \{ \text{Tn}(n) \dots \text{?Ty}(X), \text{Fo}(\beta(\alpha)), \text{Ty}(X), \langle \downarrow_0 \rangle (\text{Fo}(\alpha), \text{Ty}(Y)), \langle \downarrow_1 \rangle (\text{Fo}(\beta), \text{Ty}(Y \rightarrow X), \dots, \diamond) \} \dots \}}$$

Condition: $\langle \downarrow_i \rangle \text{?}\phi$ does not hold, $i \in \{0,1\}$

(60) Application of Elimination



The rule of Elimination concludes this section concerned with the computational/transitional rules that are considered to be applicable in all tree construction processes. As mentioned earlier in this section, the rules of Introduction and Prediction might not be as universally applicable as claimed in Cann et al. (2005), however, all the other rules presented here are essential for the tree growth process, which cannot progress or indeed end without the application of these rules. As these rules are included in the tree growth process by default, they are assumed to apply whenever necessary, and are not overtly stated for each tree growth process. In the next section, the rules of Adjunction are considered.

2.2.5. Adjunction Rules and Unfixed Nodes

These rules rely on the notions of underspecification and unfixed nodes. An unfixed node is a node which is not yet fixed within the tree under construction, which means that it does not have its own Tn ‘address’, but is rather identified via its relation to its mother. The contribution of an unfixed node to the overall tree structure is established after some more material is parsed. Unfixed nodes are represented by a dashed line in the tree growth schemata.

An unfixed node with the requirement for an expression of type e ($=?Ty(e)$) is introduced into the tree by the *Adjunction rule; as before, the Kleene star operator used to characterise an underspecified tree relation. The rule of *Adjunction defines a transition from a partial tree consisting of only one node – the root node – decorated with the requirement for type t ($=?Ty(t)$), and a tree node address $Tn(a)$ to a tree with an new unfixed node which is dominated by the root node. The new unfixed node has some sort of daughter relation to the root node, represented by the modality $\langle \uparrow^* \rangle$, however, the exact relation is established during the parsing process. To ensure this

occurs, and the unfixed node obtains a fixed position, the requirement for a fixed tree node address ($=?\exists x.Tn(x)$) is added to the new unfixed node. The formalism for the rule of *Adjunction is presented in (61) below, and the application of the rule is in (62).

(61) *Adjunction (Rule)

$$\frac{\{\dots \{\{Tn(a), \dots, ?Ty(t), \diamond\} \dots\}}{\{\dots \{\{Tn(a), \dots, ?Ty(t)\}, \{\langle \uparrow^* \rangle Tn(a), ?\exists x.Tn(x), \dots, ?Ty(e), \diamond\}\} \dots\}}$$

(62) Application of *Adjunction

$$\begin{array}{ccc} ?Ty(t), & \diamond & \rightarrow & Tn(n), ?Ty(t) \\ & & & \vdots \\ & & & \langle \uparrow^* \rangle Tn(a), ?Ty(e), ?\exists x.Tn(x), \diamond \end{array}$$

As mentioned above, and as can be seen from (61) and (62), the *Adjunction rule is applied to a partial tree which only consists of a root node containing the requirement for a type t expression ($=?Ty(t)$). The rule is used for the parsing of fronted elements, such as topics, wh-question words, and other left-dislocated constituents (Kempson et al. 2001: 150-189, Cann et al. 2005: 153-154, Bouzouita 2008 – Chapter 5, Gibson 2012 – Chapter 6). My analysis of topicalised and other left-dislocated constituents in Kazakh also employs the *Adjunction rule.

The next rule – the rule of Merge – applies when a fixed node requiring a certain formula type appears in a tree with an unfixed node decorated with that formula type. In other words, an unfixed node built by the rule of *Adjunction merges with a fixed node which carries the requirement for the type of formula already held at the unfixed node, provided these nodes do not carry any conflicting specifications. The formalism in (63) represents the rule of Merge; DU stands for Declarative Unit.

(63) Merge (Rule)

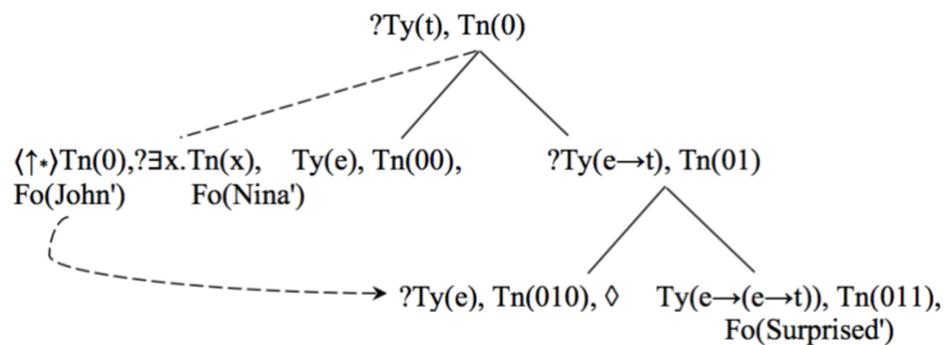
$$\frac{\{\dots\{\dots, DU, DU', \dots\}\dots\}}{\{\dots\{\dots, DU \cup DU', \dots\}\dots\}}$$

$$\diamond \in DU'$$

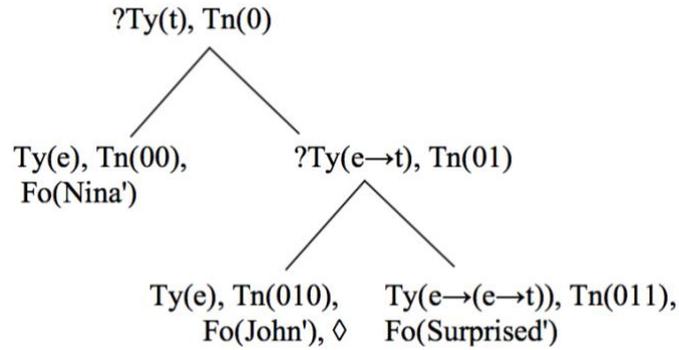
This general rule quite simply unifies two DUs, and the constraints are: (i) that the pointer is at the fixed node, and (ii) that the two DUs are able to unify – i.e. they have no conflicting specifications, as mentioned above.

In the tree growth schemata the merge process is indicated by a dashed arrow going from the unfixed node participating in the process to the fixed one, as shown in the schema in (64) for the string ‘*John, Nina surprised*’. This string contains a left-dislocated constituent – *John* – which is parsed onto an unfixed node via the rule of *Adjunction. Once the transitive verb *surprised* is parsed, the direct object node Tn(010) with the requirement for a formula of type *e* is constructed. As the unfixed node contains a formula of type *e*, and there are no conflicting specifications on the unfixed node and node Tn(010), the Merge process is triggered. In (65) we see the tree from (64) after the Merge process took place.

(64) Parsing *John, Nina surprised* with Merge



(65) Parsing *John, Nina surprised* – after Merge



Note that the structure of this tree is the same as the structure for the sentence without the left-dislocated element *Nina surprised John*, as shown in (43). This is an important point, since it demonstrates the difference of DS from most current theories of syntax, where separate layers of grammar, or specific locations for various functional projections are assumed. The structural and information-structural differences between two strings are not encoded in the DS representation, but are shown to derive from the differences in which the final semantic tree is established.

The next Adjunction rule to be considered is the Local *Adjunction rule. This is a restricted version of the *Adjunction rule and is used to build an unfixed node relative to a local node with the requirement for type t . The formalism of this rule is presented in (66), and the application is shown in (67).

(66) Local *Adjunction (Rule)

$$\frac{\{ \dots \{ \text{Tn}(a), \dots, \text{?Ty}(t), \diamond \} \dots \}}{\{ \dots \{ \text{Tn}(a), \text{?Ty}(t) \dots \} \dots \{ \langle \uparrow_0 \uparrow \uparrow \rangle \text{Tn}(a), \text{?Ty}(e), \text{?}\exists x. \text{Tn}(x), \diamond \} \}}$$

(67) Application of Local *Adjunction

$$\begin{array}{ccc}
\text{Tn(a), ?Ty(t)} & \rightarrow & \text{Tn(a), ?Ty(t)} \\
& & \vdots \\
& & \langle \uparrow_0 \uparrow_1^* \rangle \text{Tn(a), ?Ty(e), ?\exists x. \text{Tn}(x), \diamond}
\end{array}$$

At first glance the effect from the application of the Local *Adjunction rule might appear identical to that of the *Adjunction rule, however, there is an important difference between the two – the composite operator $\langle \uparrow_0 \uparrow_1^* \rangle \text{Tn}(a)$ present in the Local *Adjunction rule. The introduction of this operator ensures that, should an unfixed node built by the *Adjunction rule already be present in the tree, it would not collapse with the unfixed node built by the Local *Adjunction rule. The presence of a distinct tree node identifier in the form of $\langle \uparrow_0 \uparrow_1^* \rangle \text{Tn}(a)$ becomes a conflicting specification preventing the node with this decoration from merging with the unfixed node decorated with $\langle \uparrow^* \rangle \text{Tn}(a)$.

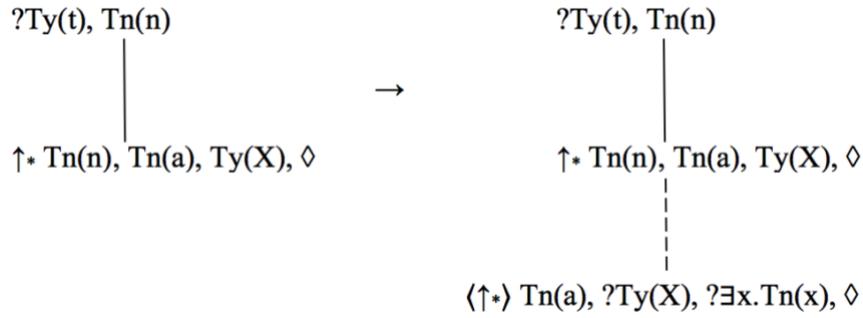
Cann et al. (2005) use the Local *Adjunction rule to capture and analyse Japanese scrambling phenomena, while Kiaer (2007) shows its employment in the parsing of terms in Korean, where each term is parsed onto an unfixed node via the Local *Adjunction rule, until the case morpheme fixes it. I take the same approach to analysing the parsing of nominals in Kazakh.

The Late *Adjunction rule differs from the other Adjunction rules presented thus far in that it builds an unfixed node with a requirement for the same formula type as the fixed node from which it is projected. The rule of Late *Adjunction is shown in (68) below, and its application is presented in (69).

(68) Late *Adjunction (Rule)

$$\frac{\{\text{Tn}(n), \dots, \{\uparrow^* \text{Tn}(n), \text{Tn}(a), \dots, \text{Ty}(X), \diamond\}, \dots\}}{\{\text{Tn}(n), \dots, \{\uparrow^* \text{Tn}(n), \text{Tn}(a), \dots, \text{Ty}(X)\}, \{\langle \uparrow^* \rangle \text{Tn}(a), ?\text{Ty}(X), ?\exists x. \text{Tn}(x), \diamond\}, \dots\}}$$

(69) Application of Late *Adjunction



Cann et al. (2005) employ this rule in their DS account of sentence extraposition phenomena in non-pro-drop languages such as English, where an overt ‘dummy’ subject is essential when a syntactic subject expressed by a clause appears post verbally, like in (70), for instance.

(70) *(It) is certain that she is right.

The last Adjunction rule to be considered in this section is the Link Adjunction rule. The notion of Link in DS signifies a connection between two parallel trees, which share some information. Most commonly, a Link relationship would connect a node with a $Ty(e)$ formula of one tree, and the root node of another – an illustration is provided in example (73) below. Modal operators $\langle L \rangle$ and $\langle L^{-1} \rangle$ participate in the Link Adjunction rule: the former ‘points’ to a tree linked to the current node, and the latter ‘points back’ to the node found at the other end of the link. The formal Link Adjunction rule is presented in (71).

(71) Link Adjunction (Rule)

$$\frac{\{ \dots \{ Tn(a), Fo(\alpha), Ty(e), \diamond \}, \dots \}}{\{ \dots \{ Tn(a), Fo(\alpha), Ty(e) \} \dots \} \{ \langle L^{-1} \rangle Tn(a), ?Ty(t), ?\langle \downarrow^* \rangle Fo(\alpha), \diamond \}}$$

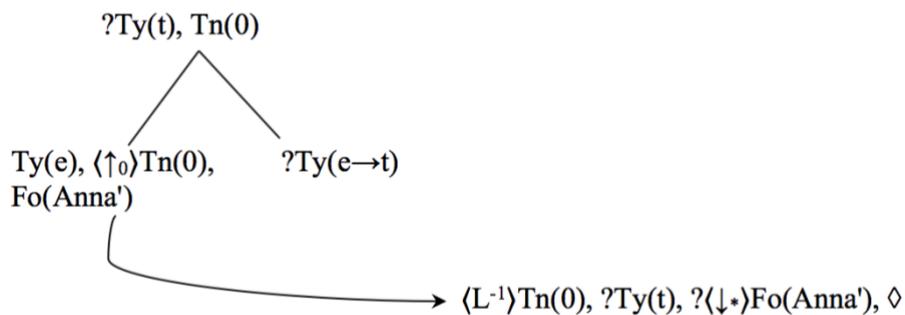
The input tree in this rule represents the head, while the output tree contains that same head and the Link node, which has two requirements - $?Ty(t)$, and $? \langle \downarrow^* \rangle Fo(\alpha)$. The first of these requirements is essentially a requirement for a new tree to be built, while

the second requirement stipulates that a copy of $Fo(\alpha)$ from the head must be found somewhere in the new tree, thus ensuring an anaphoric relation between the two tree structures. In other words, anaphora resolution is achieved by building a partial tree that creates context against which another tree develops.

Cann et al. (2005) employ Link structures for the DS analysis of English constructions with relative clauses, left-dislocated ‘hanging’ topic constructions, and adverbial clauses. Kiaer (2007) uses Link structures in her analysis of Korean topic markers, and Gibson (2012) proposes to parse Rangi NP expressions onto Link structures in the initial parse, and utilises them in her analysis of relative and some subordinate clauses, as well as cleft constructions in Rangi. The tree growth schema in (73) below shows the use of a Link structure in the parsing of the first constituent English sentence with a non-restrictive relative clause from (72).

(72) Anna, whom Galina loves, returned.

(73) Parsing *Anna...* with Link Adjunction



As mentioned earlier, Link structures are used in the Analysis of the so-called ‘hanging’ topics, like in the English example in (74).

(74) As for this film, I didn’t like it much.

The Topic Structure Introduction and the Topic Structure Requirement rules are in use here as well; the former builds a Linked structure from the root node introduced by

the Axiom, and the latter uses the $\langle D \rangle$ operator to impose a requirement for the copy of the term under consideration during the construction process without imposing any structural restrictions. The rules are presented in (75) and (76) respectively.

(75) Topic Structure Introduction (Rule)

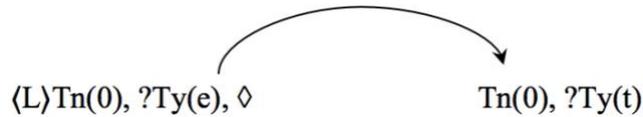
$$\frac{\{\{Tn(0), ?Ty(t), \diamond\}\}}{\{\{Tn(0), ?Ty(t)\}, \{\langle L \rangle Tn(0), ?Ty(e), \diamond\}\}}$$

(76) Topic Structure Requirement (Rule)

$$\frac{\{\{Tn(0), ?Ty(t)\}, \{\langle L \rangle Tn(0), Fo(\alpha), Ty(e), \diamond\}\}}{\{\{Tn(0), ?Ty(t)\}, ?\langle D \rangle Fo(\alpha), \diamond\}, \{\langle L \rangle Tn(0), Fo(\alpha), Ty(e)\}}$$

Tree growth processes induced by the rules from (75) and (76) are shown in (77) and (78) below.

(77) Application of Topic Structure Introduction



(78) Application of Topic Structure Requirement



As can be seen from (78), two parallel trees connected by a Link structure are being constructed. The partial tree on the left consists of a completed node with a satisfied requirement of type e , while the main tree on the right contains a requirement for a

copy of this term to be present somewhere in this tree (=?(D)Fo(Film')). Link structures are used extensively in the DS analyses of phenomena presented in Chapters 3, 5, and 6.

2.2.6 Lexical Information and Lexical Entries

Having considered the computational DS rules we now turn to the principal driving force of the parsing process – the lexicon. DS views the information contained in the lexicon as procedural. A lexical entry of a word can contain algorithms that go far beyond simple annotation of terminal nodes, and parsing of a word can result in the addition of information to a non-terminal node, the addition of further requirements, as well as the construction of a partial tree or even a full propositional structure. In this section we take a closer look at the structure of and the actions used in lexical entries.

The algorithm of actions contained in the lexical entry of a word or a morpheme is triggered when the element in question has been uttered and, subsequently, is being processed by a parser. In my representation of the parsing process in Kazakh I also demonstrate that such prosodic element as intonation phrase (IP) boundary also has a lexical entry – that is, it plays a role in the construction of the semantic tree in the parsing process. This general observation was arrived at independently from Kiaer (2007, 2014), who analyses the IP boundary as an indicator for a non-local structural growth in Korean; this is discussed in more detail in Chapter 3.

Unlike the computational/transitional rules considered in the preceding sections, the actions contained in the lexical entries are not optional – their application is obligatory once the processing of a lexical (or prosodic) item carrying them occurs. Additionally, computational/transitional rules do not have lexical triggers, however, appropriate input material must be present for a computational/transitional rule to apply.

Lexical entries follow the same general structure of a conditional statement, which consists of a Trigger, Actions, and an Elsewhere Statement; these correspond

to the IF, THEN, and ELSE parts of a lexical entry respectively. The format of lexical entries is shown in (79) below.

(79) Lexical Entry (Format)

IF	?Ty(X)	Trigger
THEN	...	Actions
ELSE	...	Elsewhere Statement

The IF statement presents the conditions that must be present at the point where the lexical item is being parsed for that parsing to successfully proceed. If the conditions in the IF statement are met, the lexical actions contained in the THEN statement can be performed. If the conditions in the IF statement are not met, however, the parsing process moves to the ELSE statement, which usually contains the instruction to terminate the parsing process – ‘Abort’.

The Actions are expressed via the predicates ‘make(...)’, which builds a new node; ‘go(...)’, which moves the pointer to the specified node; and ‘put(...)’, which annotates a node with the specified information. As mentioned above, the instruction to terminate the parsing process is presented in the form ‘Abort’. The order in which the instructions occur within the THEN statement is of crucial importance, as ‘put’ before ‘make’ means ‘add information to the current node, and then build another node’, while ‘make’ before ‘put’ means ‘build a new node and add information to it’. An example of a lexical entry for the English transitive verb *surprised* is provided in (80) below.

(80) *surprised* (lexical entry)

IF	?Ty(e→t)	Predicate Trigger
THEN	go((↑ ₁)?Ty(t));	Go to propositional node
	put(Tns(PAST));	Tense information
	go((↓ ₁)?Ty(e→t));	Go to predicate node
	make((↓ ₁));	Make functor node
	put(Fo('Surprised'), Ty(e→(e→t)), [↓]↓);	Annotation
	go((↑ ₁));	Go to mother node
	make((↓ ₀))	Make argument node
	go((↓ ₀))	Go to argument node
	put(?Ty(e));	Annotation
ELSE	Abort	Elsewhere Statement

Summing up, the ground principles of the theoretical framework of DS have been presented in this section so far, together with its essential formal tools. The computational/transitional rules were introduced, along with the Adjunction rules, responsible for the construction of unfixed nodes and Linked structures. The next section is dedicated to describing the DS approach to information structure, and, more specifically, to the notions of Topic and Focus. Although these notions are still employed throughout this thesis as descriptive terms, the underlying mechanics of these pragmatic effects are assumed to be as outlined in the following section.

2.2.7. Modelling SOV in Dynamic Syntax

As observed in Gibson (2012), there is a variation between languages in the balance between lexical and computational/transition rules involved in the establishment of propositional structure. To the best of my knowledge, no work has so far been done on modelling SOV Turkic languages in DS, however, proposals have been made for some other SOV languages, such as Japanese (Cann et al. 2006), and Korean (Kiaer 2007, 2014), for example. Like Kazakh and other Turkic languages, Japanese and Korean are pro-drop, head-final languages with rich case morphology. In this section I specifically focus on Kiaer's (2007, 2014) DS analysis for processing Korean SOV

and S-IO-DO-V sentences, and include a sample DS derivation of a Korean sentence. This will serve as the basis for modelling the derivation of a Kazakh utterance, presented in Section 3.5 of Chapter 3.

Canonical SOV languages present an issue for static syntactic theories in that the verb, which is meant to be ‘heading’ the whole structure, does not make an appearance until the very end of a sentence, thus leaving the other constituents somewhat ‘up in the air’, that is, without an established position within the hierarchical structure, and, essentially, without meaning. Dynamic Syntax approach is able to solve this issue by considering how each constituent is parsed incrementally, taking into consideration case-markers and even prosody.

As far as case is concerned, DS views it as fulfilling two distinct roles: 1) defining the filter on the output tree; and 2) initiating a structure-building action within the incremental tree-building process. The first function is expressed in the form of a requirement for a certain type of node to be found at a certain position from the node at which the case marker is processed. For instance, the nominative case would contribute the decoration $\langle \uparrow_0 \rangle \text{Ty}(t)$, which would add such a constraint that the tree under construction would only be well-formed if the current node is immediately dominated by a node with the decoration $\text{Ty}(t)$. In other words, the node with this decoration must be the subject node.

Typologically speaking, there are three strategies for parsing subject expressions available in DS. One of these strategies is to employ the rule of Link Adjunction to parse the subject expression – the subject noun phrase is parsed onto a tree which is constructed in parallel with the main tree; the Linked tree only contains the information from the supposed subject NP. This is one of the subject-parsing strategies proposed for Bantu languages; Gibson (2012) employs it in her analysis of Rangi.

Another strategy for parsing subject expressions, also adopted in the DS analyses for Bantu languages, involves the *Adjunction rule, and the subject is parsed onto an unfixed node with a requirement for an expression of type t to be found somewhere above the current node - $\langle \uparrow^* \rangle \text{Ty}(t)$. The node is fixed later in the derivation process, once the fixed structure is introduced. In Swahili, for instance, the

fixed structure is projected by the tense-aspect marker, and the subject expression is fixed as the logical subject of the sentence once that tense-aspect marker is parsed.

The third strategy for parsing subjects is based on the Local *Adjunction rule, which builds a locally unfixed node. This unfixed node receives a fixed tree address later in the parsing process through the use of case markers, for instance. This strategy has been applied in the DS analyses of languages with overt case markers, or, with what has been referred to as constructive case. Amongst these languages are Korean (Kiaer 2007, 2014), Standard Modern Greek (Chatzikiyriakidis 2010), and Romance languages, such as Latin (Kempson and Chatzikiyriakidis 2009). As Gibson (2012) points out, this strategy is not available for Bantu languages since they have no grammatical case which could perform as the constructive case and fix a locally unfixed tree node as the subject (or indeed any other term). This use of case in Korean is described in detail below.

Let us look at different cases in action on a sample Korean sentence given in (81) below.

- (81) Jina-ka sakwa-lul Mina-hanthey cwuesse.
Jina-NOM apple-ACC Mina-DAT gave
'Jina gave an apple to Mina.'

(Kiaer 2014: 154)

We can see that in the sentence in (81) every nominal phrase is distinctly marked with a case morpheme (or case particle, as Kiaer (2014) refers to them). Via the computational rule of Local *Adjunction (Section 2.2.5. Adjunction Rules and Unfixed Nodes), the first constituent of the sentence above is first parsed onto a locally unfixed node the position of which is identified only as $\langle \uparrow_0 \rangle \langle \uparrow_1^* \rangle T_n(a)$.

This is the only logically appropriate way for beginning a parse in a free word order language like Korean (and other languages with rich case morphology), as the linearly first constituent does not have to be the subject here – hence the omission of the rules of Introduction and Prediction (shown in Section 2.2.4. Computational Rules in the Tree Construction Process).

Note, that the underspecified relation expressed by $\langle \uparrow_0 \rangle \langle \uparrow_i^* \rangle T_n(a)$ can only potentially develop into three possible forms as constrained by valency of verbs in natural language. These three possible relations are those of a subject, direct object and indirect object expressed as $\langle \uparrow_0 \rangle T_n(a)$, $\langle \uparrow_0 \rangle \langle \uparrow_1 \rangle T_n(a)$, and $\langle \uparrow_0 \rangle \langle \uparrow_1 \rangle \langle \uparrow_2 \rangle T_n(a)$ respectively. This list could be expanded if adjuncts are considered to be optional arguments, however, the set of possible relations would still be restricted (Marten 2002).

Returning to the sentence from (81), let us consider how the parsing process unfolds for the first constituent *Jinaka*. The lexical entry for the nominative case marker is presented in (82) (Kiaer 2007: 120).

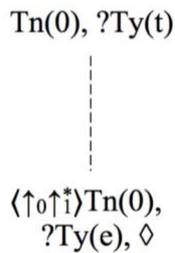
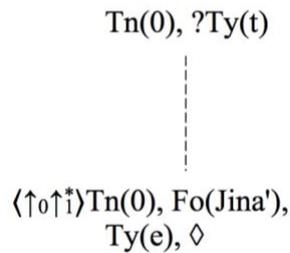
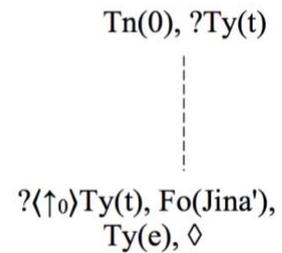
(82) Lexical entry for *-ka/-i* (nominative case)

IF	Fo(α), Ty(e)	
THEN	IF	$\langle \uparrow_0 \uparrow_i^* \rangle (T_n(a), ?Ty(t))$
	THEN	put($? \langle \uparrow_0 \rangle Ty(t)$)
	ELSE	Abort
ELSE	Abort	

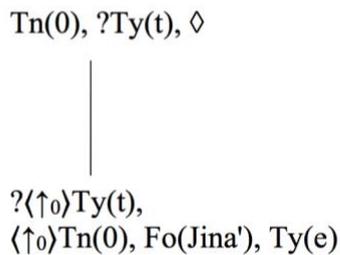
The lexical entry in (82) shows that the Local *Adjunction rule is applied as the general rule for introduction of argument nodes since the decoration $\langle \uparrow_0 \uparrow_i^* \rangle (T_n(a), ?Ty(t))$ is expected to already hold at the node by the time the parser gets to the case marker. The filter on the output function of the case marker is shown in the second THEN statement – the node carrying the nominative-marked entity must be dominated by the node with Ty(t) decoration, thus ensuring that the node at which the nominative case marker is parsed is the subject node. The application of this lexical entry is shown in (83) – closely modelled after Kiaer (2007: 121).

(83) Parsing *Jina-ka*

(a) Local *Adjunction

(b) Parsing *Jina*(c) Parsing *Jina-ka*

(d) Routinised Update



As the parsing process progresses, the tree node develops from an unfixed node with the requirement for the formula of type e , built by the Local *Adjunction rule in (a) in (83) above, to the node decorated with the semantic formula for *Jina* in (b), which satisfies the requirement for a type e entity, and then to the (still unfixed) node decorated with the formula for *Jina* and the requirement for the node $\text{Ty}(t)$ to be located up along an argument node from the current node, as shown in (c). Since this unfixed node is immediately dominated by the root node $\text{Tn}(0)$, a routinised update shown in (d) can follow. Kiaer (2007: 121) assumes that in Korean, the procedure of updating a nominal via the case marker is routinised and automatic, provided there are no prosodic signals to the contrary. The routinised update shown in (d) in (83) fixes the node as $\langle \uparrow_0 \rangle \text{Tn}(0)$ – in other words, this node becomes $\text{Tn}(00)$ – and moves the pointer to the top node.

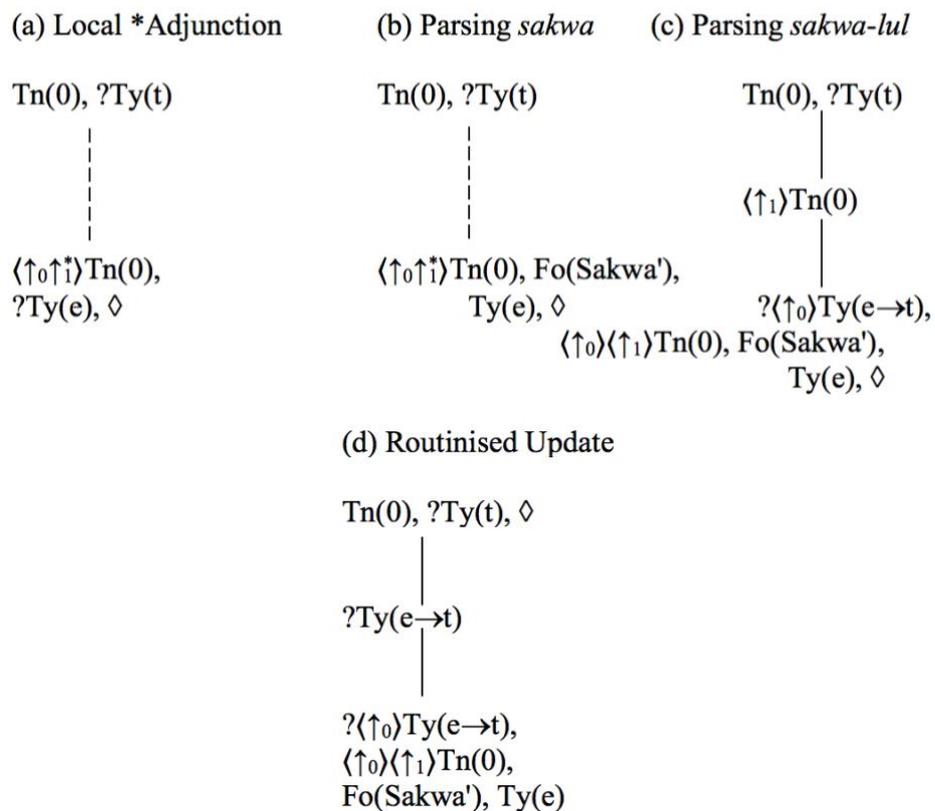
Let us move to the next constituent of (81) – *sakwa-lul* – and examine what Kiaer refers to as the “object-case marker” (2007: 121) *-lul/-ul*. The lexical entry for it is shown in (84) below (Kiaer 2007: 120).

(84) Lexical entry for *-lul/-ul* (accusative)

IF	Fo(α), Ty(e)
THEN	IF $\langle \uparrow_0 \uparrow_1 \rangle$ (Tn(a), ?Ty(t))
	THEN put(? $\langle \uparrow_0 \rangle$ Ty(e \rightarrow t))
	ELSE Abort
ELSE	Abort

As in the previous lexical entry for the nominative case marker, the Local *Adjunction rule is shown to be applied as the general computational action for parsing an argument node. The parsing of the accusatively-marked direct object from (81) is shown in (85) below.

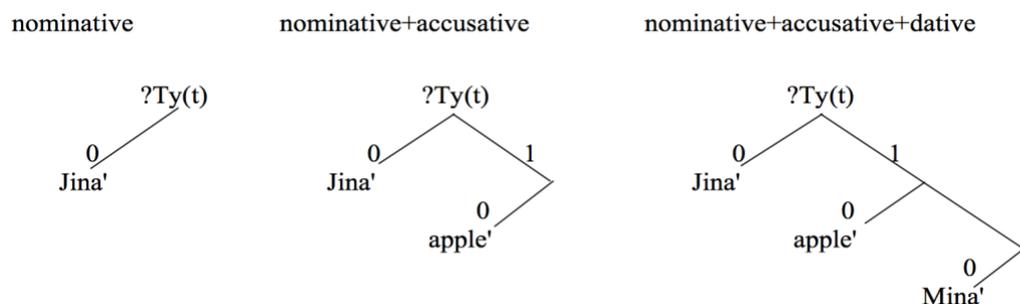
(85) Parsing *sakwa-lul*



It must be noted that the application of the Local *Adjunction rule is only possible if no other unfixed node introduced by the same rule is already present in the tree under construction; the same goes for the *Adjunction rule. If two locally unfixed nodes with the address $\langle \uparrow_0 \uparrow_i^* \rangle Tn(a)$ are present in the same tree, it would simply mean that they are one and the same node with two diverse decorations: $\langle \uparrow_0 \rangle Tn(a)$ for subject, and $\langle \uparrow_0 \rangle \langle \uparrow_1 \rangle Tn(a)$ for direct object. As we saw earlier in (83) and (85), in Korean, the unfixed node relations are resolved immediately and in accordance with the case filters, thus illuminating the issue of two unfixed nodes being present in the tree under construction.

The last argument from (81) – *Mina-hanthey* – is marked with the dative case marker. The nominal would once again be first parsed onto an unfixed node via the Local *Adjunction rule, and then the parsing of the case-marker would fix the node at the address $\langle \uparrow_0 \rangle \langle \uparrow_1 \rangle \langle \uparrow_1 \rangle Tn(0)$, thus identifying it as the indirect object in that clause. Kiaer (2007) proposes that this constructive use of case together with routinised updates leads to partial trees being constructed, as shown in (86) below.

(86) Partial tree-construction in Korean via case markers



Kiaer (2014) notes, however, that structures can be built without the case markers, which is commonly observed in spoken Korean, where subjects frequently appear without the nominative case marker. Kiaer further posits that it is not the nominative case marker, but the sentence initial positioning of a lexical item that projects anticipation for it to be the subject by definition. According to Kiaer, an overt nominative marker is used to add such expressive meanings as (contrastive) focus or

agent-hood. Interestingly, Kazakh does not mark its subjects at all and in a neutral/canonical sentence they appear sentence-initially; this is discussed in detail in Chapter 3.

Let us now move to the final element of the sentence in (81) – the verb *cwuesse* ‘gave’. As Cann et al. (2005) point out for Japanese, and Kiaer (2007) for Korean, in fully pro-drop languages, such as the two languages mentioned, all arguments are optional. It is therefore safe to assume for such pro-drop languages that a verb can project the complete predicate-argument structure, the argument values of which can be updated directly from the context. The predicate-argument structure building algorithms, then, are a part of lexical entries of Korean verbs. The verb from our example in (81) is a three-place predicate, which means that it projects a structure with three argument nodes decorated with metavariables *U*, *V*, and *W*. Metavariables are simply placeholders for some value with content, and are thus associated with a requirement for establishing such value – $\exists x.Fo(x)$ – added to the argument nodes. This requirement needs to be satisfied either within the tree-construction process or from the context. The lexical entry for the stem *cwu-* of *cwuesse* is presented in (87) below (from Kiaer 2007: 204); I omit any tense-aspect contributions made by this verb as these are not directly relevant to our discussion, and would unnecessarily complicate it.

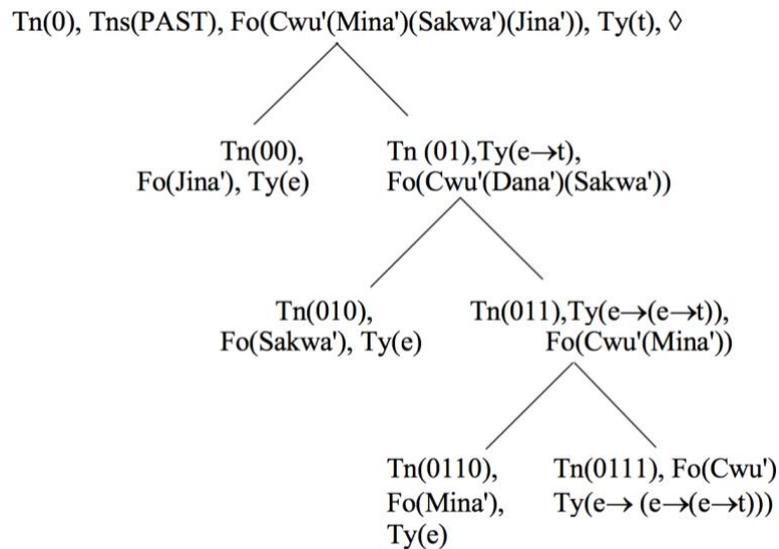
(87) Lexical entry for *cwu-* (simplified)

IF	?Ty(t)
THEN	make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(U), Ty(e), $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(?Ty(e→t)); make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(V), Ty(e), $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(?Ty(e→(e→t))); make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(W), Ty(e), $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(Fo(Cwu'), Ty(e→(e→(e→t))), \perp);
ELSE	Abort

As can be seen from this lexical entry, parsing of the verb *cwuesse*, or any other three-place predicate for that matter, results in construction of a tree with three argument

nodes, all of which are decorated with a place-holder metavariable, and a requirement for this metavariable to be replaced by a term in the *Formula* language. The final tree for the sample sentence is presented in (88) below.

(88) Parsing *Jinaka sakwalul Minahanthey cwuesse*.



It was shown earlier that constructive cases build partial tree-structures, in accordance to the schemata presented in the table in (86); thus, when the parser gets to processing the verb, some of the argument nodes constructed as part of the verb's lexical entry are already present in the tree. This does not pose an issue, and the nodes with the same address are essentially treated as one and the same node, meaning that the Merge rule does not need to be applied. The requirements for a term in the *Formula* language are instantly satisfied by the formulae already decorating the argument tree nodes.

In the case of a pro-drop construction or pronominal arguments (which also project metavariables onto the argument nodes), the metavariables are updated with the contextually appropriated formulae as part of the pragmatically driven process of Substitution (see Cann et al. 2005: 69 for a detailed explanation). Once Substitution occurs and the metavariables are replaced, no record is left of pronominal or pro-drop constructions, as the final tree is decorated with full terms in the *Formula* language.

Summing up, in this section, the parsing process for a sample Korean sentence was presented to demonstrate the DS approach to an SOV language, which is similar to Kazakh. It has been shown that while the verb projects the full predicate-argument structure, partial structure is built up via constructive case use prior to the parsing of the verb.

2.2.8. Intonation Phrase Boundary

Now that we have seen as sample parse of a Korean SOV sentence, we can introduce another concept that plays a significant role in the parsing process – the Intonation Phrase (IP) boundary. Although this prosodic device is not directly visible in writing in the same way as case or other overt morphological markers are, it plays an important constructive role. This has been demonstrated by Fodor (1998), according to whom readers not only create prosodic boundaries, but also take them into consideration and align syntactic boundaries with them, just as if they were a part of the string. This forms the basis of the Implicit Prosody Hypothesis (IPH):

In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction. (Fodor 2002: 2)

Fodor (2002) emphasises that the IPH is empirically supported by evidence from several constructions in several different languages. Fodor and colleagues were able to demonstrate the direct connection between the implicit prosody and the cross-linguistic differences in the preference of Relative Clause (RC) attachment. Jun's (2003) findings on the relation between the interpretation of RC attachment and prosodic phrasing in seven languages (English, Spanish, French, Greek, Farsi, Korean and Japanese) support the IPH. It is thus clear that overt or implicit prosody plays a crucial role in structure building in many languages.

This section first considers the notion of IP boundary and its possible effects as outlined in Kiaer (2007, 2014) for Korean. It is then demonstrated that IP boundaries play an important role in the parsing of Kazakh sentences too. An IP boundary is also included into the parsing process of a sample Kazakh sentence presented in Section 3.5 of Chapter 3.

Intonation phrase is one of the core notions in the framework of intonational phonology (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Ladd 1996), along with the notions of intermediate phrase and accentual phrase (see the same works for further details). An intonational phrase consists of one or more accentual phrases, and its boundary can be indicated in a number of ways: a) final lengthening; b) tone variation; or c) pause. As observed by Truckenbrodt (2015), pause is the least reliable indicator of an IP boundary. In German, for instance, Kohler et al. (2005) report that an actual pause was observed only in 37.3% of the 2470 IP boundaries investigated. Kiaer (2007) also states that the pause is an optional indicator of an IP boundary in Korean; instead, she relies on the final syllable lengthening as the indicator of an IP boundary. This also applies to Kazakh, as demonstrated further in this sub-section.

The important role IP boundaries play in resolving local ambiguity has been observed by many researchers on the basis of data from various languages (Price et al. 1991, Speer et al. 1996, Schafer 1997, Schafer and Jun 2002, Kim 2004, Kang and Speer 2005, among many others). IP boundaries and other prosodic cues (along with all syntactic and semantic information) have been demonstrated to be utilised by parsers from the early stage of processing, and to be facilitating the construction of syntactic structures, as illustrated in the simple examples in (89) below; the IP boundary is indicated by %.

- (89) a. She washed and fed the cat.
b. She washed% and fed the cat.

The difference between the (a) and (b) sentences in (89) is in the IP boundary after the verb *washed*, which is only present in the (b) sentence. The presence of this boundary

changes the meaning of the sentence from the cat having been washed and fed in the sentence in (a), to the female human agent washing herself and feeding the cat in the (b) sentence. The semantic structure of the sentence changes from the cat being the only beneficiary of the agent's actions indicated by the two predicates to the agent being the beneficiary of the first action, and the cat being the beneficiary of the second action. To put it even more precisely, the presence of the IP boundary indicates to the parser that no overt object is to be expected after the verb *washed*, and that in this instance it is to be processed as a reflexive verb.

Kiaer (2014: 153) identifies and characterises the constructive use of IP boundary tones as follows:

- a. Absence of IP boundary tones: 'continue' the structure building – structure building is not yet complete (cf. rising tones indicate completeness of local structure building but at the same time indicate that the overall structure building should continue).
- b. Presence of IP boundary tones: either 'delocalize (or disconnect)' structure building or 'finish-off' structure building.

In other words, in the absence of an IP boundary, structure building proceeds as normal ('continue'), whereas the presence of an IP boundary at the beginning/in the middle of a string indicates that what follows has to be constructed differently from the preceding structure ('delocalise/disconnect'). IP boundary tone appearing towards the end of a string signals the end of that tree construction process.

In DS this essentially means that, in the absence of an IP boundary somewhere before the end of the string, the tree building process continues in an uninterrupted manner as demonstrated in the preceding section. An IP boundary occurring at the beginning of a string indicates to the parser that whatever follows the boundary should not be built on to the same ongoing structure.

Kiaer (2007) uses numerous long distance dependency examples from Korean to demonstrate that prosody plays a crucial role in the syntactic structure building. Namely, an IP boundary occurring after the first constituent of a string helps the parser to detect a long distance dependency relation significantly before the verb is reached.

The Korean examples in (90) and (91) demonstrate this point; as previously, the symbol % indicates an IP boundary.

(90) Yumi-hanthey Mihe-ka key-lul cwuessta-ko
 Yumi-DAT Mihe-NOM key-ACC gave-COMP
 Kiho-ka saynggakhaysse.
 Kiho-NOM thought
 ‘Kiho thought than Mihe gave a key to Yumi.’

(91) Yumi-hanthey% Mihe-ka key-lul mantulessta-ko
 Yumi-DAT Mihe-NOM key-ACC made-COMP
 Kiho-ka malhaysse.
 Kiho-NOM said
 ‘Kiho said to Yumi that Mihe copied a key.’

(Kiaer 2007: 149)

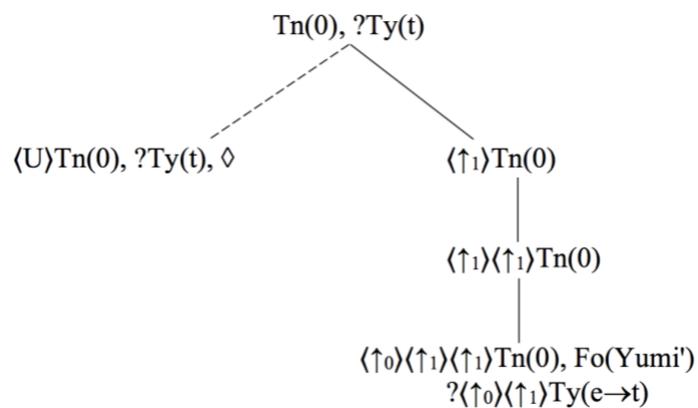
The example in (90) does not contain a long-distance dependency relation, while the sentence in (91) does, even though the visible sequences of words preceding the first embedded verb are identical. The only difference between these sequences is the audible IP boundary after the sentence initial dative noun phrase, which is absent in (90) and present in (91). Kiaer (2007, 2014) proposes the following lexical action for IP boundary (BREAK):

(92) Lexical Action of BREAK

IF	?Ty(t), Tn(a)
THEN	IF $\langle \downarrow^* \rangle Ty(x), ?\exists x Tn(x)$
	THEN 1
	ELSE make($\langle D \rangle$); go($\langle D \rangle$); put($\langle U \rangle Tn(a)$)
ELSE	Abort

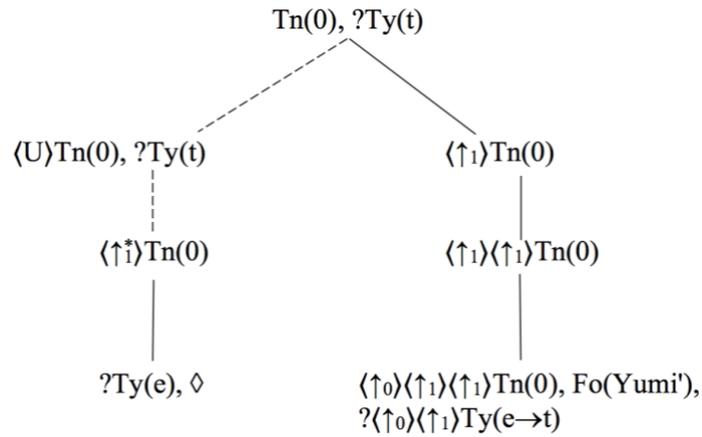
Kiaer uses ‘1’ to denote an instruction to stay at the current node without carrying out any actions; this can be replaced with a ‘stayput’ command as used, for example, by Lucas (2014). This lexical entry states that if the pointer is at the ?Ty(t) node, and there is an underspecified argument under it, then no further action is to be performed. However, if there is no such node, the ELSE statement instructs to make a weak dominance relation, go there and decorate it with ⟨U⟩Tn(a); this is demonstrated in (93) below.

(93) Parsing *Yumi-hanthey%*...



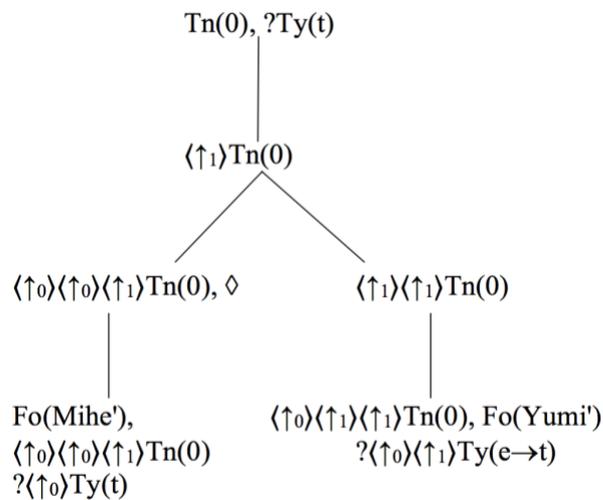
Note that the pointer is now at the newly constructed node, meaning that the parse of the next nominal constituent – *Mihe-ka* – will start here, with the Local *Adjunction rule, as shown in (94) below.

(94) Parsing *Yumi-hanthey Mihe-ka*...



When *Mihe-ka* is parsed, a local subject node is built, and all the appropriate decorations are added to it, as shown in (95). Note that since the node $\langle \uparrow_1 \rangle Tn(0)$ has already been introduced by the preceding dative noun phrase, and the new node must be dominated by a functor node ($\langle \uparrow_1^* \rangle Tn(0)$), the new node is automatically dominated by the existing $\langle \uparrow_1 \rangle Tn(0)$ node.

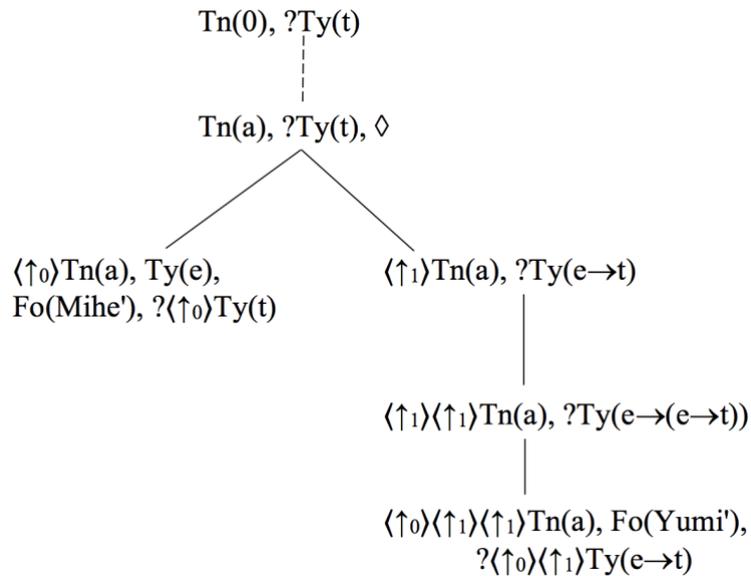
(95) Parsing *Yumi-hanthey Mihe-ka*...



The tree given in (96) below represents the beginning of the parse for the sentence in (90) without the IP boundary between *Yumi-hanthey* and *Mihe-ka*. Comparing (95) and (96), we can see that the presence of the IP boundary affects the tree-construction process. In other words, like Kiaer (2007, 2014) observes,

phonology feeds step-by-step structure building, which makes sense, since prosody is not added post-production, but is utilised by the producer to aid the parser's comprehension of the string being uttered.

(96) Parsing *Yumi-hanthey Mihe-ka*



Having conducted a number of off-line, on-line, and corpus studies, Kiaer (2007: 175) reports that parsers of Korean signal non-local resolution by prosodic cues such as IP boundaries, as shown in (91), for instance. IP boundaries are used by both producers and parsers of speech as auditory cues for non-local resolution of an expression. These observations are of importance not only for the DS analysis of long-distance dependencies, but also for the fundamental question of the relationship between syntax and phonology. On the basis of her experimental data Kiaer claims that “syntax does not feed phonology, but phonology feeds step-by-step structure building” (2014:152). That is to say, the incrementally processed phonological information (along with other information) determines the step-by-step construction of the syntactic structure.

Prosody plays an important part in the parsing of Kazakh strings. Unlike Korean, Kazakh does not allow long-distance dependency relations – all underspecified relations must be resolved as quickly and as locally as possible in this language; thus a sentence like that in (90) would be completely ungrammatical in Kazakh. As far as nominal elements are concerned, this preference for quick and local resolution is satisfied by the numerous case markers present in the Kazakh language (see Section 1.1 of Chapter 1). Recall, however, that there is also an unmarked noun form, which, should theoretically remain unfixed if overt case marking is all we go by.

I posit, however, that prosody must be taken into consideration, and that including an IP boundary into the parsing process is crucial for its comprehensive representation. Let us consider two sentences:

(97) Bolat kitap-ti Dana-ğa berdi.
Bolat book-ACC Dana-DAT gave
'Bolat gave Dana a/the book.'

(98) Bolat turali bil-me-j-min.
Bolat about know-NEG-PRS-1SG
'I don't know (about) Bolat.'

Unmarked proper name *Bolat* appears in both (97) and (98) but performs two different functions: in the former it is the subject of the sentence, and in the latter – the complement of postposition *turali* 'about'. The question is: does the parser know that the two 'Bolats' are different before the items following them are parsed? Even a cursory examination of these clauses' prosody yields an affirmative answer to this question.

The sentences from (97) and (98) were recorded by a native speaker of Kazakh and edited in PRAAT (Version 6.0.40, retrieved 11 May 2018 from <http://www.praat.org/>) – free speech analysis software created by Boersma and

Weenik (see Boersma 2001 for further information). Figures 1 and 2 represent sentences (97) and (98) respectively.

As you recall, for Korean, Kiaer (2007) considers final syllable lengthening as the indicator of an IP boundary. Figures 1 and 2 demonstrate the same for Kazakh: comparing the circled areas in Figures 1 and 2 we can see that the final vowel of *Bolat* is noticeably longer in the former than in the latter, which points to the presence and absence of an IP boundary respectively.

Figure 1. Sentence-initial NP with an IP boundary

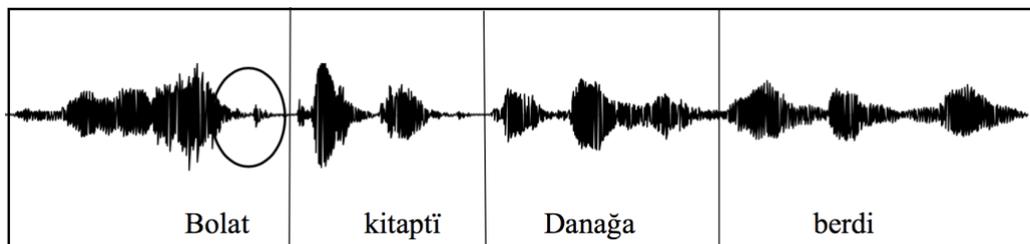
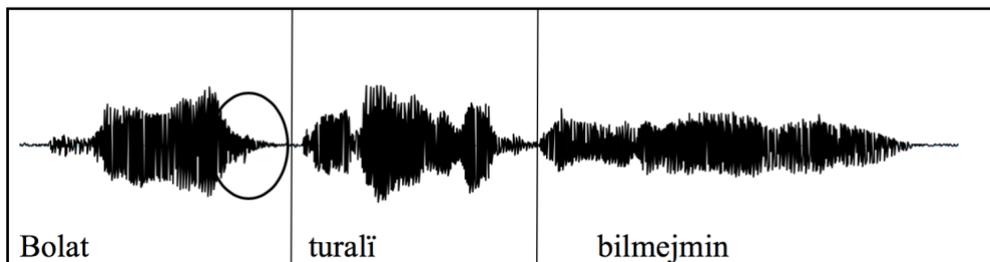


Figure 2. Sentence-initial NP – no IP boundary



The IP boundary in Figure 1 indicates that the parsing of the constituent is finished, and that it can be parsed onto a fixed subject node. No IP boundary in Figure 2 signals that this constituent is a part of a larger prosodic and syntactic phrase, and that it must be parsed and fixed in relation to the following constituent.

On the basis of these observations, I propose that in an SOV sentence, the non-case-marked nominal that functions as the subject, is fixed as such by an IP boundary. In a way, then, an IP boundary functions as a case-marker, and, more specifically as a

nominative case marker, as shown for Korean in the preceding section. I therefore suggest the following lexical entry for IP boundary in Kazakh.

(99) Lexical entry for IP boundary

IF	Fo(α), Ty(e)
THEN	IF $\langle \uparrow_0 \uparrow_1^* \rangle$ (Tn(a), ?Ty(t))
	THEN put(? $\langle \uparrow_0 \rangle$ Ty(t))
	ELSE Abort
ELSE	Abort

The effect the parsing of an IP boundary has on a tree construction process is shown in the sample parse of the sentence from (97) which is presented in Section 3.5 of Chapter 3.

As mentioned earlier, the absence of an IP boundary, as can be seen in Figure 2 for the sentence from (98), indicates that the unmarked nominal remains unfixed until the following constituent is parsed. This is also discussed further in Chapter 3, where it is posited that a nominal can remain unfixed only before predicate-type constituents.

Summing up, in this sub-section it has been observed that a parsing process cannot be based solely on written language, but must take prosody into consideration. The notion of IP boundary has been introduced, and its application in the DS tree-building process has been illustrated. The role of an IP boundary in a subject-initial Kazakh sentence has been discussed, and a lexical entry for an IP boundary has been proposed. Now that DS and its tools have been sufficiently introduced, let us consider the DS approach to IS and its principal notions – topic and focus.

2.2.9. Dynamic Syntax Approach to Information Structure

As explicated in Section 2.2.1, DS does not posit any additional layers of syntax – both the parsing process and its final outcome are schematically represented with the

help of semantic trees, whose growth and development illustrate different stages of processing. Thus, unlike many other syntactic theories, where information-structural characteristics are assumed to exist on a separate layer, or in the form of additional features – as outlined in Section 2.1.5 – DS views the information-structural notions of topic and focus as pragmatic effects arising in the process of contextualised parsing; no additional layer of syntax, direct encoding into syntax, or special syntactic feature assignment is posited.

The tree-construction process takes place in context, which consists of a non-empty sequence of labelled trees, even at the beginning of a parsing task. This ability to explicitly formally represent context gives DS a considerable advantage over other syntactic frameworks, where formalism is predominantly restricted to the static sentential level. In DS, such highly context-dependent concepts as Topic, Focus, and other information-structural phenomena, which are traditionally narrowly defined as pertaining to (the independent level of) Information Structure, belong to the wider group of context-dependent phenomena, such as reference assignment, or pragmatic strengthening, for example. These IS concepts are derived through the parsing process – represented as the tree construction process – rather than belonging to some other level of information, or being encoded directly in the syntax.

From the DS point of view, then, topic is a minimal context of sorts, against which some update is constructed. When a topic is explicitly present in a sentence, that is, when a sentence contains a topic expression, it serves as a point of departure for the parsing process. In other words, the speaker makes a choice to construct the minimal context relative to which the update carried by the rest of the utterance should take place. A topic might be anaphorically linked to the preceding context, or indicate a departure from it; in the former case, the topical element is continuous, while in the latter, the topic is contrastive.

As mentioned in Section 2.2.5 and demonstrated in the rules and schemata presented in the examples in (75)-(78), the concept of Linked structures is used to formally represent topics as partial trees constructed during a single utterance process with the function of providing an antecedent for subsequent identification by some

anaphoric device (Kempson et al. 2006: 69); this antecedent-anaphora relation essentially constitutes the topic effect.

Contrastive or shift topics are also parsed onto a Linked structure, but they themselves are not anaphorically linked to the preceding context, thus diverging from it and creating the contrastive effect. Topic markers, then, are analysed as containing the instruction to parse the element they follow onto a Linked structure – this is demonstrated in detail in Chapter 5, Section 5.5. The same effect can also be achieved by prosody or punctuation; recall the example from (64) with a left-dislocated topical object *John* – in spoken language the hearer would be able to identify it as a topic thanks to intonation, and the comma following it would assist with this in the written version of that string.

The notion of focus is viewed in DS as “the provision of some structure identified as providing the update to some constructed structure that yields a completed propositional formula” (Kempson et al. 2006: 64). In other words, the focus effect occurs when an update required for completion of the proposition is provided; the proposition in this case is taken as context.

This update can happen in two ways: the tree-construction process can present both the structure requiring an update and the update itself, which is formally represented via the function of *Adjunction with subsequent application of the Merge, which unifies the two nodes: the node requiring an update, and the node providing that update.

Alternatively, the context provides the structure, and the new string provides an update which is characteristic of elliptical focus – as is often found in answers to wh-questions, for instance. Let us briefly consider one such example, as shown in (100).

(100) Q: Who annoyed his mother?

A: John.

According to the DS view, parsing of the elliptical focus *John* in the second sentence in (100) does not happen as a standalone operation, but rather starts at that node of the preceding utterance which requires an update. This assumption of trans-boundary relation between context and update eliminates the need to posit the silent utterance of the full sentence (*John annoyed his mother*).

Crucially, this means that all the participants of a verbal interaction construct the same tree (or sequence of trees) when they parse what one of them produces, and this tree is then available in context for the other participants to add to or re-use. If a node requiring an update is present in this context tree, as it is in (100), after the question has been asked, one of the hearers becomes the speaker and provides the required update. This update produces the focus effect, which can only be parsed as such within context. As Wedgwood puts it: "...the interpretation of some expression as focus is of necessity something that emerges dynamically in context" (2003: 3). Positing a syntactic focal position, then, would mean incorporating into the syntax a feature which is largely based on contextually anchored interpretative effects.

Viewing focus as a pragmatic effect from an update being provided to the context is compatible with Lambrecht's (1994) approach to focus, as outlined in Section 2.1 of this chapter. The non-focused material present in an utterance is presupposed and serves to aid the construction of the immediate context for the update provided by the focused element(s). The presupposed material together with the update form an assertion, which links the presupposed propositions with the non-presupposed ones.

As previously, presupposition is taken to be a purely pragmatic notion in Stalnaker's (1978) understanding of it as a condition associated with a sentence, which a speaker would expect to hold in the CG between the interlocutors when that sentence is produced. However, as Wedgwood (2003) points out, an utterance may be considered to include a presupposition in a cognitive sense even when the preceding context does not contain salient material relating to this presupposition. This, in turn, indicates, that the context of an utterance consists of a set of assumptions which were constructed during the communication process, and that it is not simply some type of

‘common ground’ available prior to the production of an utterance. Again, this is comparable and compatible with Lambrecht’s view presented earlier.

To sum up, DS does not separate IS into a separate level of grammar, assign information-structural features to tree-nodes, nor posit topic- or focus-headed phrases as part of a syntactic tree structure. Instead, topic and focus are considered to be pragmatic effects naturally arising during the parsing process. This view is compatible with Lambrecht’s treatment of topic and focus as pragmatically construed relations.

2.3. Conclusions

In this chapter a lot of theoretical ground has been covered with a view to introducing central information-structural notions and the approaches to them followed in this thesis. More specifically, the notions of topic, focus, contrast, specificity and definiteness have been discussed (among several others), together with Lambrecht’s approach to the first three and von Heusinger’s understanding of the fourth.

Lambrecht’s understanding of information-structural notions of topic and focus have been adopted due to their theoretical neutrality. As the author himself declares, the goal of his seminal work is to define IS and its components in a way that can be applicable cross-linguistically and incorporated into any syntactic theoretical framework. Lambrecht achieves his goal, since his approach is sufficiently descriptive while remaining theory neutral.

The syntactic theoretical framework of Dynamic Syntax has also been introduced in this chapter along with its rules, tools and approach to IS. The parsing process of a sample Korean SOV sentences have been presented.

3. Word Order and Information Structure

This chapter is concerned with the information-structural characteristics of alternative word orders available in the Kazakh language, which is considered to be a free word order language; Section 3.1 provides an introduction into the notion of free word order and some related notions.

In Section 3.2 the basic or neutral word order in Kazakh is identified, and the description of its information-structural properties is given – it is shown that different types of focus and topic have different restrictions on possible positions within a sentence; various sentential positions and their typical information-structural characteristics are discussed in Section 3.3.

Section 3.4 considers scrambled, that is, non-SOV, word orders; it is demonstrated that sentences with scrambled word orders are restricted to certain contexts depending on their information-structural properties. Section 3.5 provides the DS analyses of diverse word orders and the information-structural configurations associated with them. Section 3.6. concludes this chapter.

3.1. Introduction

Kazakh, along with such languages as Turkish, Japanese, Irish, Czech, and Russian, amongst many others, is usually understood to be a ‘free’ word order language. Simply put, this means that all six word orders presented in (1) are possible for monotransitive Kazakh clauses with accusative case-marked direct objects⁵ in appropriate contexts.

- | | | | | |
|-----|----|---------------------------|----|-----|
| (1) | a. | S[ubject] O[bject] V[erb] | d. | OVS |
| | b. | OSV | e. | VSO |
| | c. | SVO | f. | VOS |

⁵ The unmarked direct objects and the phenomenon of differential object marking in Kazakh are discussed separately in Chapter 4.

Word order flexibility like that illustrated in (1) is typically observed in languages with relatively well-developed morphological case-marking systems (Erguvanlı 1984, Kučerová 2007, Neeleman and van de Koot 2008, Titov 2012, amongst many others), of which Kazakh is one, as described in Chapter 1, Section 1.1. These languages, unlike English, for example, do not typically use sentential positions to encode grammatical relations.

As Erguvanlı (1984: 5) observes for Turkish, word order in this language has no “primary” grammatical function, such as signalling grammatical relations (e.g. subject or direct object), or the syntactic form (e.g. question or embedded clause); however, there are exceptions to this observation, as discussed below; this also applies to Kazakh. The question then arises as to the purpose and meaning of these word order variations in free word order languages, as Payne (1992: 1) puts it: “When there are several possible order patterns in a language, what is the communicative function of one, rather than another order?”

Since Ross (1967), reordering of constituents within a clause has been referred to as scrambling, and has been given a lot of attention. The notion of scrambling presupposes that every free word order language has a basic or canonical word order, from which other, non-canonical or marked word orders are produced by movement; movement analyses for various languages with scrambling have been proposed by Hale (1983), Saito (1985), Kuroda (1988) Speas (1990), Diesing (1992), Takano (1998), and many others.

Scrambling, however, is not like other types of movement, and is considered to be optional, not feature-driven, and not to affect the semantics/truth-conditionality of a sentence. As noted by Neeleman and van de Koot (2016:383), optionality of scrambling in free word order languages is especially problematic for generative approaches to syntax, due to the common assumption that linguistic representations are subject to some kind of economy.

Öztürk (2005) provides the examples in (2) to illustrate scrambling in Turkish, the unmarked/basic word order for which is assumed to be SOV.

(2) Turkish Scrambling

- a. Ali ev-i bul-du.
Ali house-ACC find-PST
'Ali found the house.'
- b. Evi Ali buldu.
- c. Ali buldu evi.
- d. Evi buldu Ali.
- e. Buldu Ali evi.
- f. Buldu evi Ali.

Öztürk (2005: 166)

Sentence (a) in (2) shows the canonical or basic SOV word order for Turkish, while the sentences (b)-(f) present scrambled word orders: OSV, SVO, OVS, VSO, and VOS respectively. The terms – the subject and the direct object – can appear pre- and post-verbally.

In this chapter I demonstrate that SOV is the neutral or canonical word order in Kazakh, and that the primary purpose of constituent scrambling is to encode different information-structural configurations. This means that all the word order variants in (1) (and in (2) for Turkish) may be truth-conditionally identical, while receiving diverse pragmatic readings and being confined to certain (sometimes very narrow) contexts. It can thus be claimed that any deviation from the neutral SOV word order is not random, but is information-structurally/contextually/pragmatically motivated which shows that the so-called 'free' word order is not so free after all.

This might appear to put Kazakh into the category of discourse configurational languages (together with Turkish according to Erguvanlı (1984)), in which discourse-semantic, and not the theta role or case considerations, play the primary role in sentence articulation (É. Kiss 1995). É. Kiss (1995: 6) identifies two main properties on the basis of which a language might be categorised as discourse-configurational:

A: The (discourse-)semantic function of ‘topic’, serving to foreground a specific individual that something will be predicated about (not necessarily identical with the grammatical subject), is expressed through a particular structural relation (in other words, it is associated with a particular structural position).

B: The (discourse-)semantic function ‘focus’, expressing identification, is realized through a particular structural relation (that is, by movement into a particular structural position).

É. Kiss (1995) notes that not all discourse-configurational languages display both properties A and B. Kazakh appears to have tendencies for both properties A and B, but, as the data in this chapter show, these indeed are only tendencies rather than rules. Besides, due to my employing the theoretical framework of Dynamic Syntax in this thesis, I do not aim to establish a ‘structural position’ for any of the elements of the sentences I consider, unless that structural position refers to the linear order in which one or another sentential constituent is parsed.

Surányi (2016) points out the importance of distinguishing discourse-configurationality from discourse-prominence. While the former postulates a systematic and exclusive association of a particular phrase structure configuration with an information-structural category that falls under the notions of topic and focus, the latter, being a much less restrictive notion, simply assumes that a sentential word order of a language favours a grammatical description that attributes a key role to information-structural categories.

Surányi discusses Spanish, for which sentence-final focus position has been reported (Zubizaretta 1998), and posits that for this language, as well as for many other languages (including the verb-final ones such as Japanese or Dutch) that have been erroneously labelled as discourse-configurational, focus simply appears in a particular sentential position at the level of prosodic representation rather than that of phrase structure. In other words, a focused element appears in a position where it would satisfy the nuclear stress-focus correspondence requirement. This chapter shows that this observation can also be attributed to Kazakh, which is a discourse-prominent, rather than a discourse-configurational language.

3.2. Neutral Word Order in Kazakh – SOV

This section examines the basic, canonical or neutral word order in Kazakh. Firstly, the notion of a basic or canonical word order is briefly discussed, and the understanding of it followed in this thesis is clarified. The section then proceeds to present three diagnostics that have been used to identify SOV as the neutral or canonical word order in Kazakh. Finally, the information-structural properties of SOV sentences are examined, and it is proposed that contrary to previous claims for the immediately preverbal focus position, a (contrastively) focused expression can appear in any position in the preverbal area.

3.2.1. The Notion of Basic Word Order

Usually the tradition of identifying a basic or a canonical word order in a language is dated back to Greenberg (1966), however, as pointed out by Bauer (2009), the ‘natural’ or ‘common sense’ word orders for French and Latin were being discussed as early as 1784 in the work of Antoine de Rivarlor (1757-1801). A comparative study of word order patterns in languages that included English, French, German, Turkish, as well as Ancient Greek and Latin was produced by Weil in 1848.

Weil identifies the word order in modern languages (English, French, German, and Turkish) as “fixed”, while the ancient languages (Ancient Greek and Latin) had “free” word order, “a matter of privilege in inflected languages” ([1848] 1978: 53). Weil also posits the existence of two types of movement (from a ‘basic’ word order) – “the movement of ideas...shown by the order of elements”, and the “syntactic movement” expressed by case-marking ([1848] 1978: 36). What Weil labels as “the movement of ideas” can be compared with the pragmatic word order alternations this chapter examines.

As Dryer (1997) observes for modern linguistics, there is no consensus on the meaning of the term ‘basic’ word order. He proceeds to note, that many linguists equate the basic word order with the most frequent word order, while others identify it as the order that appears in broadest set of syntactic environments, yet another group

might suggest some other criterion or combination thereof. Dryer himself follows the former school of thought, and defines the “basic word order of two or more elements as the most frequent order of those elements in the language” (1997: 71). In his later work Dryer (2013) replaces the expression ‘basic [word] order’ with the expression ‘dominant [word] order’, in order to emphasise that frequency of use is the prioritised criterion.

Dryer (1997) admits that his definition is problematic for a number of reasons, and that there may be “other useful notions that one might apply the expression to, but the different notions should not be confused” (1997: 71). One of the biggest problems that Dryer highlights is of methodological nature: while it might be fairly easy to identify the most frequent order in a given text or even a number of texts, we cannot be certain that our findings are true for all other texts in the language. In order to make an irrefutable conclusion regarding the basic word order based on frequency in a particular language, a wide variety of texts from diverse genres must be examined. Even then, we might not be able to arrive at a conclusive result, and would essentially have to declare that the language under examination has no basic word order.

Siewerska (1988) presents an alternative view of basic word order, as she observes that:

in typological studies, the term ‘basic order’ is typically identified with the order that occurs in stylistically neutral, independent, indicative clauses with full noun phrase participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb presents an action, not a state or an event (1988: 8).

Note some considerable differences from Dryer’s approach: firstly, some syntactic characteristics of a sentence word order which might be considered ‘basic’ are provided; secondly, it is stipulated that only sentences with full noun phrase participants can be labelled as ‘basic’; and thirdly, style is included in the description.

Similar ideas are echoed in Newmeyer’s (2003) definition of basic or canonical word order:

The canonical word order of a language is that order typically found in main clause declaratives when the subjects and objects are encoded by fully referential lexical noun phrases (as opposed to pronominal or clitic elements) (2003: 69).

Like Siewerska, and unlike Dryer (1997), Newmeyer (2003) does not factor frequency of use into his definition of a basic or canonical word order, but instead focuses on the characteristics of a sentence which might be labelled canonical.

The syntactic approach to the notion of basic word order is reflected in Neeleman and van de Koot's (2016) definition, for example. They propose that the basic word order in a language is that where: "(i) arguments are merged in a minimal structure and in accordance with the thematic hierarchy and (ii) no phrasal movements take place other than those required to license the thematic and case properties of arguments" (2016: 384). This definition reflects the approach of the Minimalist framework (and its predecessors) to scrambled word orders, which, according to this view, are arrived at via movement.

Taking all the above mentioned approaches to the notion of 'basic' word order into consideration, it can be concluded that Dryer's approach is problematic for methodological reasons. Highly syntax theory-grounded approaches like that of Neeleman and van de Koot are incompatible with the theoretical framework of DS followed in this thesis, since "all concepts of movement are replaced by the concepts of structural underspecification and update following the parsing dynamics" (Cann et al. 2005: 26). Additionally, as the rest of this chapter demonstrates, in Kazakh, the changes to the basic word order occur for reasons other than to license the thematic and case properties of arguments, namely, for pragmatic/information-structural reasons.

Siewerska's (1988) and Newmeyer's (2003) definitions, on the other hand, are theory-neutral, clear and focused on the characteristics of a basic or canonical sentence, which is why they are adopted in this thesis. Bringing them together, basic or canonical word order is defined as the order observed in a main, independent, declarative sentence, which is information-structurally homogenous (e.g. an all-focus sentence), and in which arguments are expressed by fully referential lexical noun phrases. Due to the additional information-structural component of the definition, and

the overall focus on the information-structural characteristics of word order variations in Kazakh, preference is given to the terms ‘neutral’ or ‘unmarked’ to label this word order.

In the following sub-sections I present several tests that provide grounds for identifying the SOV word order as the neutral or unmarked word order in the Kazakh language. These ‘tests’ used to identify the neutral word order in Kazakh are the minimal context test (sub-section 3.2.2), and the unmarked arguments test (sub-section 3.2.3). In Section 3.2.4 it is shown that identifying SOV as the neutral word order in Kazakh also agrees with typological observations for SOV languages. Section 3.2.5 presents the information-structural characteristics of Kazakh SOV sentences.

3.2.2. Minimal Context Test

There are several reasons for identifying SOV as the neutral word order in the Kazakh language. Firstly, sentences with this word order can felicitously appear in minimal context, such as in answer to the question ‘*What happened?*’ (Pinto 1997), or in discourse-initial positions (Mithun 1992, Erdocia et al. 2012); the examples in (3) and (4) below exemplify these contexts respectively. Preceding or following context is provided in square brackets.

(3) [What happened?]

Bolat mīsīq-tī äkel-di.

Bolat cat-ACC bring-PST(3)

‘Bolat brought the cat.’

(4) Gulnar öz mīsīq-tar-ï-n žaqsi kör-e-di.

Gulnar own cat-PL-POSS.3-ACC love-PRS-3

‘Gulnar loves her cats.’ [She has five of them.]

Any other word order variant from (1) is infelicitous in these contexts, and only the least pragmatically marked sentences which presuppose the least (all-focus

sentences) can be used here. This is not to say that SOV is the only sentence structure to appear in these contexts, however. Sentences consisting of a subject and a verb, just a verb, or verbless sentences (usually of temporal or locative nature) are also acceptable, but they are not of interest for this thesis, since they do not conform to the adopted definition of a sentence with a neutral word order.

Certainly, SOV sentences often contain other constituents, such as oblique arguments of ditransitive verbs, adverbials of time or place, or other adjuncts. The example in (5) shows a modified version of the sentence in (3), which includes an adverbial of time and an adverbial of directionality/location, and can be schematically represented as xSOyV.

- (5) [What happened?]
 Keše Bolat mäsıq-tı üj-ge äkel-di.
 yesterday Bolat cat-ACC home-DAT bring-PST(3)
 ‘Bolat brought the cat home yesterday.’

In (5), the adverbial of time *keše* appears sentence-initially, which is its typical neutral position in an all-focus sentence. The directional/place adverbial *üjge*, expressed by the noun in dative case (recall that the Kazakh grammars refer to this case as dative-directional case, as mentioned in Chapter 1, Section 1.1), is in the immediately preverbal position.

However, if the direct object is unmarked, this unmarked direct object must take the immediately preverbal position and the adverbial is then placed between the subject and the direct object as in (6) – xSyOV.

- (6) [What happened?]
 Keše Bolat üj-ge mäsıq äkel-di.
 yesterday Bolat home-DAT cat bring-PST(3)
 ‘Bolat brought a cat home yesterday.’

Modal adverbials can appear sentence-initially or after the subject, as the examples in (7) and (8) demonstrate; schematically xSOV and SxOV respectively.

- (7) Mümkin ol kitap-ti satip al-a-dı.
 maybe 3SG book-ACC buy-FUT-3
 ‘Maybe he/she will buy the book.’

- (8) Ol mumkin kitapti satip aladi.

Indirect or oblique objects, which are marked with dative case, can be placed immediately before the verb as in (9), unless the direct object is unmarked, in which case the unmarked direct object takes the immediately preverbal position, as in (10).

- (9) [What happened?]
 Bolat misiq-ti Dana-ğa sıjla-dı.
 Bolat cat-ACC Dana-DAT gift-PST(3)
 ‘Bolat gifted/gave the cat to Dana.’

- (10) [What happened?]
 Bolat Dana-ğa misiq sıjla-dı.
 Bolat Dana-DAT cat gift-PST(3)
 ‘Bolat gifted/gave Dana a cat.’

Neutral placement of adverbials, adjuncts and obliques can be summed up in the schema in (11).

- (11) Neutral Word Order in Kazakh – extended.

TIME ADVERBIAL – SUBJECT / MODAL ADVERBIAL – ACCUSATIVE
 DO / IO(OBLIQUE) – OTHER ADVERBIALS – IO(OBLIQUE) /
 UNMARKED DO – VERB

Observations presented in this sub-section demonstrate that SOV word order conforms to the definition of unmarked or neutral word order given in the preceding sub-section. The following sub-section provides another piece of evidence for neutrality of SOV.

3.2.3. Unmarked Arguments Test

Another reason for identifying SOV as the neutral word order in Kazakh comes from Erguvanlı's (1984) observation for Turkish; she notes that SOV is the only interpretation available for sentences such as those in (12) and (13), where the noun phrases are not case marked and the semantics cannot help identify the grammatical roles of different NPs.

- | | | | |
|------|------------------------------------|---------------|----------|
| (12) | Mutluluk | huzur | getirir. |
| | happiness | peace=of=mind | brings |
| | ‘Happiness brings peace of mind.’ | | |
| | *‘Peace of mind brings happiness.’ | | |
| | | | |
| (13) | Huzur | mutluluk | getirir. |
| | peace=of=mind | happiness | brings |
| | ‘Peace of mind brings happiness.’ | | |
| | *‘Happiness brings peace of mind.’ | | |

(Erguvanlı 1984: 5)

Sentence-initial *mutluluk* in (12) has to be the subject, and the interpretation in which it is the object is not available; *huzur*, which takes up the second position in the sentence has to be the object, and cannot be interpreted as the subject. This reverses for the sentence in (13), where *huzur* is now sentence-initial and must be the subject.

Titov (2012) also uses a sentence with unmarked semantically identical (both animate, human) arguments to show that SVO is the unmarked/basic word order in Russian. She posits that scrambling should not be allowed when the grammatical

*‘Peace brings happiness.’

- (16) Tiništiq baqit äkel-e-di.
peace happiness bring-PRS-3
‘Peace brings happiness.’
*‘Happiness brings peace.’

Unlike in most previous Kazakh examples, where direct objects are marked with the accusative case, both noun phrases in the sentences above are not case-marked, nor do they differ in their semantic features, thus making word order the only means of identification of their grammatical roles. Thus, like in the preceding examples from Turkish and Russian, the only available reading for both (15) and (16) is SOV. The alternative OSV reading is not acceptable, which points to SOV being the unmarked or neutral order in Kazakh.

3.2.4. Typological Observations for SOV Languages

Kazakh also conforms to typological observations for SOV languages which were originally made by Greenberg (1966), with further developments in subsequent works by Lehmann (1973), Hawkins (1983, 2004), Siewerska (1988), Dryer (1992, 2007), Song (2001, 2009) amongst many others. Greenberg (1966) discovered and described correlations that typically hold between seemingly diverse word order properties, which he formulated as 45 distinct ‘universal’ statements.

According to these universals, if SOV is a canonical constituent order in declarative sentences, then the language tends to have postpositions, relative clauses preceding the nouns, adverbs before verbs, as well as adjectives and main verbs before auxiliary verbs. All these implications are indeed observed in Kazakh, as the examples below show.

In (17) postposition *soñ* expressing a temporal relation is just one of many Kazakh postpositions (see Section 1.1, Chapter 1); postpositions is the only class of adpositions available in Kazakh, in line with the expectations for a true SOV language.

- (17) Bolat Dana-ni eki saġat-tan soñ küt-ti.
 Bolat Dana-ACC two hour-ABL from wait-PST.3
 ‘Bolat waited for Dana since 2 o’clock.’

The examples in (18) and (19) demonstrate that relative clauses have to precede the nouns in Kazakh, and the reverse order is ungrammatical.

- (18) Munı ajt-qan adam osında otır.
 this speak-PTCP man here sit.PRS(3)
 ‘The man who spoke about this is sitting here.’

- (19) *Adam munı ajt-qan osında otır.
 man this speak-PTCP here sit.PRS(3)
 Intended: ‘The man who spoke about this is sitting here.’

In the example in (20) we see the adverb *qattı* appear before the verb it modifies, and (21) shows that changing their order so that the adverb follows the verb results in ungrammaticality.

- (20) Bolat ana-si-n qattı žaqsi kör-e-di.
 Bolat mother-POSS.3-ACC strongly love-PRS-3
 ‘Bolat loves his mother very much.’

- (21) *Bolat ana-si-n žaqsi kör-e-di qattı
 Bolat mother-3POSS-ACC love-PRS-3 strongly
 Inteded: ‘Bolat loves his mother very much.’

In (22) the auxiliary inflected for person and number follows the main verb used in its converbial form, and together they indicate that the action denoted by the

main verb is a habitual action in the present. Again, a change in the order is unacceptable, as evidenced by the ungrammatical example in (23).

(22) Men universitet-te oqı-p žür-min.
 1SG university-LOC study-CVB AUX-1SG
 ‘I study/am studying at university.’

(23) *Men universitet-te žür-min oqı-p.
 1SG university-LOC AUX-1SG study-CVB
 Intended: ‘I study/am studying at university.’

Following Greenberg, Dryer (2007) notes that in SOV languages, in genitive noun phrases, nouns that modify a noun and express possession or kinship-like relation precede the noun. This indeed is the case in Kazakh, as the examples below demonstrate.

(24) Adam-niñ qarındas-ı ädemi.
 Adam-GEN younger.sister-POSS beautiful
 ‘Adam’s younger sister is beautiful.’

It has thus been shown that from the point of view of typological observations for languages with different basic word orders, Kazakh behaves like a typical SOV language.

3.2.5. Information-structural Properties of SOV

In this section I take a close look at the information-structural properties of the SOV word order in Kazakh. This word order can encode more information-structural relations than the other possible word orders. This specifically refers to the fact that the whole sentence and almost every constituent of it can express new information focus (NIF; recall the distinction from contrastive focus, described in Chapter 2), as

the examples below show. Wh-questions are used to provide contexts in which different parts of a sentence are focused. The syllable in small caps receives focus stress, which in these cases coincides with sentential stress – I discuss this in more detail below.

- (25) [NE bol-di?]
what be-PST(3)
'What happened?'

Bolat *mi*s̈iq-ẗi äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

- (26) [Bolat NE iste-di?]
Bolat what do-PST(3)
'What did Bolat do?'

Bolat *mi*s̈iq-ẗi äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

- (27) [Bolat NE äkel-di?]
Bolat what bring-PST(3)
'What did Bolat bring?'

Bolat *mi*s̈iq-ẗi äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

Questions given in square brackets in (25)-(27) provide context for different sizes of focus. Thus, the answer to the question in (25) is an all-focus sentence (recall

Lambrecht’s classification from Section 2.1 of Chapter 2); the question in (26) provides context for a narrowly focused verb phrase (direct object + verb are in focus) with a topical subject, or predicate focus; the question in (27) is to the direct object, which is narrowly focused in the response – argument focus, and the rest of the utterance is presupposed. It can be seen that in all three cases, regardless of the size of the focus, the focus stress is on the last syllable of the direct object – that is, on the immediately preverbal position.

What we observe in (25)-(27) – the same sentence expressing a variety of focus sizes in different contexts – was described as focus projection, and briefly discussed and exemplified in Section 2.1 of Chapter 2. Thus, similarly to the English example from Section 2.1, repeated in (23) for convenience, the examples (25)-(27) can be collapsed to represent focus projection in Kazakh, as shown in (29) below.

(28) [She [bought a [GIRAFFE.]_F]_F]_F

(29) [Bolat[[mäsïq-tï]_F äkel-di.]_F]_F

The main difference between (28) and (29) is in the location from which the focus projects which is concurrent with the location of the direct object. Since English is an SVO language, focus projects from the direct object which is rightmost in the sentence. In an SOV language like Kazakh, the direct object is located immediately preverbally, and focus projection originates from that position.

Returning to the examples in (25)-(27), we can note that the subject *Bolat* is part of an all-focus sentence in (25), but is a topical expression in (26) and (27). Since in both (26) and (27) the topic referent remains the same in questions and answers, we are dealing with non-contrastive, continuous aboutness topic, as defined in Section 2.1 of Chapter 2; more on topic positions in Kazakh is presented in Section 3.3.

From cursory observations, I can posit that in the absence of a topic-marker, topicality of a sentence-initial element is indicated prosodically, by means of a fall-rise intonation contour. This is a non-controversial comment for topical prosody, as Hedberg and Sosa’s (2008) detailed analysis of the relation between information-

structural categories and prosody in English demonstrates. A similar detailed examination of IS-prosody relation in Kazakh and other Turkic languages of Central Asia would be of great interest.

We can thus preliminarily assume the information-structural characteristics for an SOV sentence in Kazakh as presented in Table 12 below.

Table 12. Information structure of an SOV sentence (to be revised)

Subject	Object	Verb
All focus		
Topic	Focus	
Topic	Focus	Presupposed/Backgrounded

Table 12 concludes this section on SOV word order, which has been demonstrated to be the neutral or canonical word order in Kazakh. Now that we have considered the information-structural properties of an SOV sentence, and before we proceed to examining scrambled word orders, let us take a closer look at the different sentential positions and their typical information-structural characteristics. The following section explores this, and a revised and expanded version of Table 12 is provided at the end of it.

3.3. Sentential Positions and IS

In this section I examine different sentential positions that have been linked with different information-structural notions. In 3.3.1 I consider the immediately preverbal position, which has been reported to be the focus position for many languages. I then move to the notion of focus field – the alternative to the notion of focus position in Turkish proposed by Göksel and Özsoy (2000) – in 3.3.2. In 3.3.3 I consider focus field and focus position in Kazakh, and demonstrate that the immediately preverbal position is the typical NIF position in Kazakh, while the whole of the preverbal area is the (contrastive) focus field. Sub-section 3.3.4 examines the sentence initial position, and, finally, sub-section 3.3.5 considers the post-verbal position. Sub-section

3.3.6 concludes this section and presents a revised version of Table 12 from the preceding section.

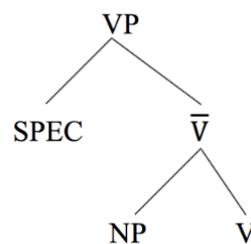
3.3.1. The Immediately Preverbal Position

It is claimed in the literature on Turkish and Kazakh that the immediately preverbal position (IPV; also referred to in the literature as ‘immediately before verb’ or IBV position) is the focus position in these languages (for Turkish: Erkü 1983, Erguvanlı 1984, Butt and King 1996, Kennelly 1997 & 1999, Kornfilt 1997 amongst others; for Kazakh: Balakaev 1959, Muhamedowa 2016).

As well as being widely-used in the literature on IS, the term ‘immediately preverbal position’ received a lot of attention in the literature on differential object marking in languages like Turkish, Urdu, Hindi, Punjabi, Malayalam, Spanish (Mahanjan 1990, de Hoop 1992, Ramchand 1993, Butt 1993, Primus 2012, amongst many others), and other languages, where unmarked direct objects are usually confined to the IPV position; I leave the detailed discussion of differential object marking till Chapter 4.

In most syntactic theories, IPV is a syntactically loaded term, which identifies the complement position of the verb, as shown in (30) (modified from example (11) (Butt 1993: 98)).

(30) SPEC vs complement positions of direct objects



As (30) shows, according to this view, a direct object can occupy two distinct syntactical positions – SO[ØV] (specifier) or S[OV] (complement), and only the latter

– ‘lower’ – one is considered to be the ‘true’ IPV. According to this view, unmarked direct objects in Turkish (and by extension in Kazakh) are confined to the IPV, complement position.

In the DS representation, a direct object would appear on the same node on the final DS tree regardless of its position in a sentence or case-marking. The difference between marked and unmarked direct objects is reflected in the parsing process itself rather than in the position a direct object occupies in a hierarchically structured tree, which is explained in detail in Chapter 4.

I therefore use the term IPV position purely descriptively to identify a constituent that immediately precedes the verb linearly. The DS explanation of the ‘special’ nature of the IPV position is also rooted in the way the parsing process unfolds rather than its place in an assumed hierarchical structure of a verb phrase.

3.3.2. IPV Position to Focus Field in Turkish

Not all linguists are in consent regarding the IPV position being the focus position in Turkish. According to Kural (1992), Göksel (1998), and Göksel and Özsoy (2000), there is no defined focus position in this language. Göksel and Özsoy (2000) propose that instead of ‘focus position’ the term ‘focus field’ must be used, which they identify as “the area between the constituent that takes focal stress and the position that includes the verb complex”, and as “the domain that hosts elements designating non-recoverable information” (2000: 227). They further posit that “stress is the sole indicator of focus in Turkish” and it “can be assigned to any preverbal constituent”, which indicates that “the surface syntax of Turkish does not have a designated focus position” (2000: 224).

The IPV position, according to Göksel and Özsoy (2000), is the position designated for sentential stress, rather than specifically for focused constituents, which, of course, does not preclude focused constituents from appearing there. The authors use Turkish examples given below to demonstrate that focus stress and sentential stress are different; as previously, stress is indicated by capitalisation.

- (31) a. Ev-e GİT-me-di-m.
home-DAT go-NEG-PST-1SG
‘I didn’t go home / *HOME.’
- b. EV-E git-me-di-m.
home-DAT go-NEG-PST-1SG
‘I didn’t go HOME / *home.’
- c. EV-E git-ti-m.
home-DAT go-PST-1SG
‘I went HOME/home.’

(Göksel and Özsoy 2000: 226-227)

For examples (a) and (b) in (31) Göksel and Özsoy observe that the whole proposition is negated in the former, while the latter is a sentence with contrastive focus and thus “has a different set of presuppositions, namely that the speaker has gone somewhere (but not home)” (2000: 227).

They further observe that sentence (a) shows that Turkish negative clitic *-mA* attracts sentential stress (assigned to the preceding syllable), which, according to them, explains why sentence (b) can only be interpreted as having focal stress. The authors state that there is no explanation as to why the sentence (b) in (31) “should not have a non-focal reading”, seeing as the IPV position is the position for sentential stress.

The authors provide sentence (c) in (31) for comparison, and show that it is ambiguous as it can be perceived as both “a neutral sentence with sentential stress, and a sentence which has focal stress” (2000: 227). They claim that the relative degree of stress might help resolve the ambiguity. By “neutral sentence with sentential stress” I assume the authors mean an all-focus sentence or sentence focus in Lambrecht’s terms.

They conclude that this ambiguity is “a natural result of the existence of two types of stress” (2000: 227) – the focus stress and the sentential stress. In the following section an alternative view of the data like those in (31) is given, which leads to a

different set of conclusions in regard to the focus position and focus field in Kazakh. It is evident, however, that the notion of focus field provides for a better description of information-structural characteristics of a Kazakh clause.

3.3.3. Contrastive Focus Field and New Information Focus Position in Kazakh

The examples like those provided by Göksel and Özsoy (2000) for Turkish are also available in Kazakh – they are presented in (32), (33), and (36) below, but the analysis I propose for Kazakh differs from Göksel and Özsoy’s. Using Kazakh examples in (32)-(36), I build on Göksel and Özsoy’s analysis to make more detailed observations for Kazakh.

The sentence in (32), which is identical to sentence (a) in (31), shows that in Kazakh, sentential stress is attracted to negation suffix *-MA*, which is used to construct the negated form of the predicate in *-DI* past or simple past. According to Göksel and Özsoy, the whole proposition is negated in the Turkish equivalent, and this is an example of a ‘neutral’ reading with a ‘pure’ sentential stress, which can serve as the answer to the question ‘*What happened?*’.

- (32) Universitet-ke tüs-ME-di-m.
 university-DAT get.in-NEG-PST-1SG
 ‘I did not get into university./I did NOT get into university.’

However, in Kazakh, some other readings are available for this sentence, as it can be used as an answer to questions ‘*What have you done?*’ or ‘*What happened to you?*’, which induces the topic-comment or predicate-focus structure; the topic expression can be omitted or overtly present as *men* ‘I’.

Another reading is also available for a sentence like that in (32) namely, the verum focus reading (see Section 2.1, Chapter 2), as indicated by the second English interpretation. This shows that the sentence in (32) is not necessarily ‘neutral’, and is not an example of a sentence with a ‘pure’ sentential stress. Depending on the context this sentence can be an all-focus sentence, a sentence with narrow focus on the verb

phrase, or a sentence with verum focus (negative response to something like ‘*You got into university.*’).

The sentence in (33) is similar to that in (31b), and, like (31b) yields a narrow argument-focus or a contrastive focus reading on the oblique argument, which receives the main stress: the speaker either replies to the narrowly focused question ‘*Where did you get into?*’; or indicates that she got into somewhere (maybe a college) but not university. She might also be correcting an erroneous statement like ‘*You didn’t get into college, right?*’. Sentence- or predicate-focus readings are not available for this sentence. Unavailable alternative readings are indicated by a *.

- (33) Universitet-KE tüs-me-di-m.
 university-DAT get.in-NEG-PST-1SG
 ‘I did not get into UNIVERSITY / *university.’

In the example in (34), the sentence-initial subject receives the main stress, which yields the only available reading where the stressed constituent is contrastively focused. This demonstrates that placing the main sentential stress on the constituent that would not have usually received it in a particular sentence creates the narrow contrastive focus effect.

- (34) SEN universitet-ke tüs-me-di-ñ.
 2SG university-DAT get.in-NEG-PST-2SG
 ‘YOU did not /*NOT get into university / *UNIVERSITY.’

Moving to the sentence in (35) we can see that it is identical to the sentence (c) in (31), where the main sentential stress is in its canonical location – the IPV syllable/position. This sentence could be interpreted as an all-focus sentence, a sentence with a NIF on the oblique argument plus verb, or a sentence with a narrow NIF on the oblique argument; additionally, it can have the contrastive focus on the oblique argument reading.

- (35) Universitet-KE tüs-ti-m.
 university-DAT get.in-PST-1SG
 ‘I got into university/UNIVERSITY.’

The sentence in (36), which is identical to (35) in everything but the placement of the main stress, has the exclusive verum focus reading and can only be used to contradict a statement of opposite polarity, something like ‘*You didn’t get into university.*’; (35) would be infelicitous in this context.

- (36) Universitet-ke tüs-TI-M
 university-DAT get.in-PST-1SG
 ‘I DID get into university.’

All these observations can be applied to our sample SOV sentence, from the examples (25)-(27); the negated version of the sentence is given in (37) below.

- (37) Bolat misiq-ti äkel-ME-di.
 Bolat cat-ACC bring-NEG-PST(3)
 ‘Bolat did not bring the cat. / Bolat did NOT bring the cat.’

The main sentential stress in this sentence falls on the negation suffix *-MA*, and not on the IPV position, as it normally would in the affirmative sentence. The sentence with this stress assignment can be an all-focus sentence, a sentence with predicate-focus, or a sentence with verum focus. It can be said that focus projection takes place here: focal reading can be projected outwardly from the part which receives the main stress over the whole of the sentence, as schematically represented in (38).

- (38) [Bolat [misiqtü [äkel-[ME]_F-di.]_F]_F]

When the sentential stress in this negated sentence is shifted to the canonical IPV position, as shown in (39), the previously described all-focus, predicate-focus, and

verum focus interpretations become unavailable, and only a narrow NIF and contrastive focus readings on the direct object are possible.

- (39) Bolat [misiq-ti]_F äkel-me-di.
'Bolat did not bring the cat/ CAT.'

When the stress in the affirmative sentence is shifted to the verb, as shown in (40) only narrow predicate focus and verum focus readings are available. That is to say, the sentence below could be a felicitous response to the question '*What did Bolat do to/with the cat?*', or contradiction of the statement '*Bolat did not bring the cat*'.

- (40) Bolat misiq-ti äkel-DI.
'Bolat *brought / BROUGHT the cat. / Bolat DID bring the cat.'

Finally, if a non-verbal and non-IPV constituent receives the main sentential stress, as in (41), only the reading where this constituent is contrastively focused is available.

- (41) BOLAT misiq-ti äkel-DI.
'BOLAT / *Bolat brought the cat.'

With these examples in mind, Göksel and Özsoy's observations on the focus position in Turkish can be expanded and fine-tuned for Kazakh. Firstly, it is clear, that in order to make accurate observations on focus positioning in Kazakh, it is necessary to distinguish between the new-information, contrastive, and verum focus types, as explicated in Section 2.1 of Chapter 2.

Secondly, it can be said that in a simple affirmative sentence the main sentential stress falls on the IPV position, where a focused element would typically appear, thus satisfying two requirements – for a sentence to have a main stress, and for the focused element to be stressed. Sentences with main stress in the IPV position can yield a wide all-focus reading, focus on the verb phrase (verb plus argument)

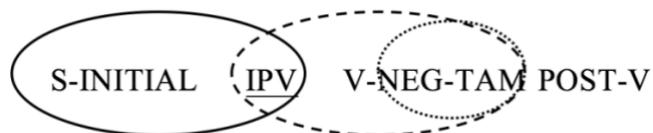
reading, or a narrow focus on the IPV element reading; additionally, an immediately preverbal, stressed element can be contrastively focused. The size and type of focus are understood from the context, as explained in Section 2.1 of Chapter 2, and the preceding section (this chapter) where the notion of focus projection is discussed.

Thirdly, in instances where the main stress is attracted to the verb for morpho-phonological reasons, as in (32) and (37), all-focus, predicate-focus, and verum-focus readings are available. In sentences where the main stress is on a part of the verb for pragmatic reasons, as in (36) and (40), only verum focus reading is available.

Finally, where the stress is not on the verb or on the IPV element as described above, only contrastive focus reading is available vis-à-vis the stressed element, as in (34) and (41).

It can be summarised, then, that contrastive focus reading is available in any preverbal position, including the IPV position, thus making the whole of the preverbal area the contrastive focus field which is indicated by the solid line in the schema in (42). New-information focus field includes the verb and the IPV position, indicated by the dashed line in (42). The verbal component carries the main sentential stress when it is attracted by certain elements (such as negation, for instance), otherwise, the main sentential stress falls on the IPV syllable. Verum focus, indicated by the dotted line in (42) is marked by the stress on either the negation or TAM elements.

(42) Focus types and positions in Kazakh



Thus, we can see that the new-information, contrastive and verum focus types have their own fields, some of which overlap. Recall, however, that verum focus can be considered a type of contrastive focus (Chapter 2, Section 2.1), which means that all of the preverbal field including the verbal complex is available for contrastive focusing. As per the schema in (42), a focused element cannot occur in the post-verbal

position⁶ – this position is reserved to backgrounded/presupposed/given information; this is covered in detail in Section 3.3.5.

As far as differentiation between sentential and focus stress is concerned, the examples (32)-(36) have demonstrated that there is no need to differentiate sentential and focus stress in Kazakh. The main sentential stress is assigned either to the IPV position or to a component of a verbal complex, in which case the sentence-focus and predicate-focus readings are available. If the main stress is assigned to a non-canonical position, the stressed element is understood to be contrastively focused. Whether this calls for identifying a contrastive focus stress as a distinct stress type in Kazakh requires a further in-depth investigation, which is out of the scope of the present study.

To sum up, in Kazakh, the IPV position is the canonical position for a narrowly focused non-verbal element; the size and type of focus are identified from the context. An SOV sentence can be an all-focus sentence, a predicate-focus sentence (focus is on the verb and the direct object), as well as a sentence with narrow or contrastive focus on the direct object (this is demonstrated in the preceding section). In the next sub-section, I turn my attention to the of the sentence-initial position, which is typically associated with topical position in Turkic languages.

3.3.4. Sentence-initial position

As observed by Erkü (1983), Erguvanli (1984), and Göksel and Kerslake (2005), amongst many others, the sentence-initial position in Turkish is typically associated with topics. Muhamedowa (2016) makes the same observation for Kazakh. However,

⁶ Focused *wh*-words occasionally appear post-verbally in highly emotive speech, as in (i).

- (i) Mīna kitap-tī Bolat-qa ber-di KIM?!
 this book-ACC Bolat-DAT give-PST(3) who
 ‘Who’d do such a thing as giving this book to Bolat?!’

A question like this is not a real request for information, but rather a rhetorical question uttered to express discontent. I do not consider these and other highly emotive utterances in my study, as they do not follow the usual patterns.

in this section it is demonstrated that the sentence-initial slot is not exclusively reserved for topics.

Let us return to the examples from (26) and (27) – repeated in (43) and (44) below. It has been observed for these sentences that the direct object and verb are in focus in (43), and only the direct object is focused in (44).

(43) [Bolat NE iste-di?]
Bolat what do-PST(3)
'What did Bolat do?'

Bolat misiq-ti äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

(44) [Bolat NE äkel-di?]
Bolat what bring-PST(3)
'What did Bolat bring?'

Bolat misiq-ti äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

Sentence-initial subject *Bolat* is topical in both response utterances in (43) and (44). More precisely, *Bolat* is the continuous aboutness topic, as opposed to being a contrastive/switch topic (recall the discussion on types of topic – aboutness and contrastive – in Section 2.1 of Chapter 2), since both questions and answers in (43) and (44) are about one and the same topic referent.

This continuous topic expression can be pronominalised or omitted, especially in spoken Kazakh. This is in contrast to cases where the sentence-initial subject is either a part of sentence-focus, or a contrastive topic, as exemplified by (45) and (46) respectively.

- (45) [What happened?]
 Bolat misiq-Tİ äkel-di.
 Bolat cat-ACC bring-PST(3)
 ‘Bolat brought the cat.’
- (46) [What did the boys bring?]
 Bolat misiq-Tİ äkel-di.
 ‘Bolat brought the cat.’

As previously demonstrated, in (45), *Bolat* is part of sentence-focus; in (46), *Bolat* is a topic expression, like in (43) and (44), however, since the referent of this topic expression is different from that of the topic expression in the question (*the boys*), *Bolat* is a contrastive topic in this sentence.

Notably, in (43) and (44), the topical subject *Bolat* can be pronominalised or omitted altogether due to its referent being the same as the topic referent of the preceding question. However, due to the difference in the information-structural status of the same sentence-initial subject in (45) and (46), it cannot be omitted, and can only be pronominalised if the utterance is accompanied by an overt gesture pointing to the intended referent of the pronoun.

This unequivocally demonstrates that: a) not all sentence-initial subjects are topical; and b) not all topics are the same, and that contrastive topics are in some way more focus-like.

Recall also from the previous section, that the sentence-initial constituent can be contrastively focused. This is observed when the main sentential stress, which usually falls on the IPV position, is shifted to the sentence initial constituent, as was shown in (41). These sentence-initial, prosodically prominent subjects are unanimously perceived as contrastively focused by the native speakers.

It can be concluded, then, that the sentence-initial position does not strictly speaking equate to being the topical position, even in a sentence with the canonical

or neutral SOV word order. Context and prosody must be taken into consideration when information-structural role of a sentence-initial constituent is to be determined.

Sentence-initial constituents are typically topical: whether that topic is continuous or contrastive/switch topic is determined by the preceding context, as demonstrated above. However, a sentence-initial constituent can also be a part of sentence-focus, as well as be narrowly contrastively focused. This is especially relevant for spoken Kazakh since contrastive focusing of a sentence-initial constituent is achieved through prosodic highlighting of the sentence-initial element. We now move to considering the information-structural properties of elements occurring post-verbally.

3.3.5. Post-verbal Position in Kazakh

It has been observed for Turkish that the post-verbal area – the area following the predicate – possesses certain prosodic and pragmatic properties which differentiate it from the preverbal area (Erkü 1983, Erguvanlı 1984, Göksel and Kerslake 2005). In this section I check whether these observations also apply to the post-verbal area in Kazakh. Göksel and Kerslake (2005) state that placing constituents in the post-verbal position is restricted to spoken and informal written Turkish, while formal written styles do not typically have post-verbal constituents; this is also applicable to the Kazakh language.

It also has to be noted that in this section I do not deal with verb-initial clauses in which all the non-verbal elements appear post-verbally – sub-Section 3.4.3 of this chapter is specifically dedicated to verb-initial sentences (VSO and VOS). However, the claims made in this section about the post-verbal area are also applicable to these verb-initial sentences.

Göksel and Kerslake (2005) remark that the post-verbal area is the site for backgrounded information, which is either assumed to be shared between the interlocutors, or was recently mentioned in their communication process; in other words, the post-verbal area hosts information which is presupposed or given. As such,

nominal elements appearing post-verbally cannot be indefinite, as noted by Erguvanlı (1984), and illustrated by the examples in (47) and (48).

(47) Adam-ın oda-sın-da bir lamba yanı-yor-du.
 man-GEN room-POSS.3-LOC one lamp burn-PROG-PST
 ‘A lamp was burning in the man’s room.’

(48) *Adam-ın oda-sın-da yanı-yor-du bir
 man-GEN room-POSS.3-LOC burn-PROG-PST one
 lamba.
 lamp

(Erguvanlı 1984: 46)

In (47) the indefinite noun phrase *bir lamba* is the subject of the sentence, and felicitously appears between the locative noun phrase and the verbal predicate; the presence of the numeral *bir*, which also serves as the indefinite article in Turkish (Göksel and Kerslake 2005) as well as in Kazakh (Muhamedowa 2016), confirms the indefiniteness of the noun *lamba*. The ungrammatical example in (48) demonstrates that placing an indefinite noun after the verb/predicate is not acceptable.

The same is true for the Kazakh language, as the examples in (49) and (50) demonstrate.

(49) Qazına-nı bir kız tapt-ті.
 treasure-ACC one girl find-PST(3)
 ‘A girl found the treasure.’

(50) *Qazına-nı tapt-ті bir kız.
 treasure-ACC find-PST(3) one girl

The indefinite subject *bir kız* felicitously appears before the verb in (49), but cannot be moved to the post-verbal position as the ungrammaticality of (50) shows.

Unmarked direct objects, which are indefinite, cannot appear post-verbally either, as the ungrammaticality of (51) shows.

- (51) *Bir kız tap-ti qazina.
 one girl find-PST(3) treasure
 Intended: ‘A girl found a treasure.’

While generally agreeing with infelicity of indefinite nouns appearing post-verbally, Erkü (1983) provides examples in which an indefinite noun appears in the post-verbal position. According to Erkü, placing an indefinite noun post-verbally is possible if the referent of this noun belongs to an activated set, as the example in (52) demonstrates.

- (52) Context: the day after the party given by the speaker at which the hearer was also present...
 dün gece 15 kişi vardı-ya sigarası-nı
 last night 15 people there.was-as.known cigarette-POSS.3
 yanık unut-muş bir misafir
 burning forget-PFV a guest
 ‘there were 15 people last night, as you know, a guest left his/her cigarette burning’

(Erkü 1983: 151)

By ‘activation’ Erkü means “a psychological status that certain concepts have in discourse”; the speaker assumes that an activated concept is “in the addressee’s consciousness at the time of the utterance”, and that the addressee’s attention is focused on this information at that point in the conversation (Erkü 1983: 142). This concept is comparable with that of givenness, as described in Section 2.1 of Chapter 2. However, this does not appear to be applicable to Kazakh, as the sentence from (52) cannot be successfully replicated, meaning that indefinite nouns are unacceptable in declarative sentences in Kazakh.

Following on from the observations made in regard to the pragmatic properties of the elements that can occur post-verbally, post-verbal elements are not expected to be prosodically prominent. Indeed, this is borne out for Turkish, as illustrated by the example in (53) below .

- (53) Dün Ankara-dan dön-dü Murat/*MuRAT.
 yesterday Ankara-ABL return-PST Murat
 ‘Yesterday Murat / *MURAT returned from Ankara.’
 (Erguvanlı 1984: 45)

As (53) shows, stressing the post-verbal subject *Murat* is ungrammatical. I must note, that I am not sure that Erguvanlı’s use of the (*) symbol to indicate ungrammaticality is entirely suitable here. This, however, is dependent on the approach one takes to information structure and its interaction with syntax. For Erguvanlı, topic and focus in Turkish have structural positions – sentence-initial and IPV respectively, therefore a focused term appearing outside of the IPV position would indeed result in ungrammaticality. In the approach I take, it would result in infelicity. The sentences like those in (53) and (55) would still be grammatical on paper/without context, but infelicitous in the context where the subject is narrowly focused or is part of a larger focused group.

Erguvanlı’s commentary on impossibility for a stressed element to appear post-verbally also applies to Kazakh. As was shown in Section 3.3.3, the main sentential stress is found in the preverbal area (including the verb), and the examples in (54) and (55) show that assigning stress to a post-verbal constituent is not possible in Kazakh.

- (54) Keše Astana-DAN qajt-ti Bolat.
 yesterday Astana-ABL return-PST(3) Bolat
 ‘Yesterday Bolat returned from Astana.’
- (55) #Keše Astana-dan qajt-ti BOLAT.

Interrogative pronouns, which attract sentential stress are thus expectedly excluded from the post-verbal area, as shown in (56)

- (56) *Astana-dan qajt-ti kim?
Astana-ABL return-PST(3) who
Intended: Who returned from Astana?

The inability of a stressed constituent to appear post-verbally is in line with our earlier claim that the main sentential stress has to be assigned to a preverbal element or the verb itself.

Returning to the example in (54), it must be noted, that this sentence is the felicitous response to the question ‘*Where did Bolat return from yesterday?*’, which makes *Bolat* the expression of a continuous topical referent. However, in the context of the question ‘*Where did the boys return from yesterday?*’ the sentence from (54) would not be felicitous, as *Bolat* would become a contrastive/switch topic here and would have to appear sentence-initially, as shown in (57).

- (57) Bolat Astana-DAN qajt-ti .
Bolat Astana-ABL return-PST(3)
‘Bolat returned from Astana.’

Summing up, several comments can be made on the information-structural properties of post-verbal elements. Firstly, it can be concluded that stressed elements cannot appear post-verbally, which precludes all types of focus from appearing in that position.

Secondly, post-verbal elements must be presupposed or given (explicitly or implicitly), and as such they can (but do not have to) be topical. It has also been noted, however, that this only applies to continuous aboutness topics, as contrastive/switch topics are excluded from the post-verbal area, and must appear sentence-initially.

3.3.6. Conclusions

This section has demonstrated that different positions in a Kazakh sentence display diverse information-structural properties in that they vary in the kinds of information-structural elements they typically host, or can host.

Firstly, it has been shown that while the IPV position is the canonical focus position, it should not be labelled as ‘the focus position’, since elements outside of this position can also be in focus. It has been proposed that the notion of focus field, which includes the whole of the pre-verbal area up to and including the verb, should be used instead.

Secondly, it has been demonstrated that the sentence-initial position, which is traditionally associated with topicality, does not always host topics, and can also host an element that is a part of an all-focus sentence, or even a contrastively focused element.

Thirdly, it has been established, that presupposed or ‘given’ elements, including continuous aboutness topics, can appear post-verbally. Any type of non-presupposed constituents, focused elements, or contrastive topics are excluded from the post-verbal area.

With these observations in mind, Table 12 from the preceding section can be expanded as follows in Table 13.

Table 13. Information structure of an SOV sentence (revised)

Subject	Object	Verb
Focus		
Topic/Contrastive Topic	Predicate Focus	
Topic/Presupposed	Focus/Contrastive Focus	Presupposed/Background
Contrastive Focus	Presupposed/Background	
Topic	Presupposed/Background	Verum Focus

Table 13 represents all the possible information-structural configurations of a Kazakh SOV sentence, with the first three rows showing the most typical configurations. The

third row represents what is considered the canonical information-structural lay out of a Kazakh sentence, where the sentence-initial constituent is the topic, and the immediately preverbal one is the focus.

3.4. Information-structural Properties of Scrambled Word Orders in Kazakh

In this section I consider non-canonical or scrambled word orders in Kazakh and demonstrate that they are pragmatically motivated. As was shown in the previous section, the IPV position is the typical focus position in Kazakh, while topics typically appear sentence-initially. These observations allow use to make certain predictions about the information-structural properties of Kazakh sentences with scrambled word orders. This section also considers information-structural properties of post-verbal constituents.

3.4.1. OSV

I begin with considering the OSV word order and with drawing some comparisons between it and the neutral SOV word order. Following the conclusions that have been made so far for the SOV word order, we can predict that in the OSV word order the direct object is topical, while the subject is focused. Indeed, the examples below show that this prediction is correct, thus confirming our previous observations and allowing us to make further predictions on the other non-neutral word orders.

Let us return to some of our previous examples and consider the possibility of an OSV sentence appearing in the same contexts as an SOV sentence.

(58) [NE bol-dīʔ]
what be-PST(3)
‘What happened?’

Misiq-ti BolAT äkel-di.
cat-ACC Bolat bring-PST(3)

‘Bolat brought the cat.’

(59) [Bolat NE iste-di?]

Bolat what do-PST(3)

‘What did Bolat do?’

Misiq-ti BOLAT äkel-di.

cat-ACC Bolat bring-PST(3)

‘Bolat brought the cat.’

(60) [Bolat NE äkel-di?]

Bolat what bring-PST(3)

‘What did Bolat bring?’

Misiq-ti BOLAT äkel-di.

cat-ACC Bolat bring-PST(3)

‘Bolat brought the cat.’

As the examples in (58)-(60) show, the OSV sentence, which is grammatical, and is truth-conditionally equal to the SOV sentence *Bolat misiqti äkeldi*, is not felicitous in the same contexts. This demonstrates that an OSV sentence is information-structurally/pragmatically different from a semantically equal SOV sentence, and the context in which it can be used must be different. This clearly shows, then, that the use of the OSV word order is pragmatically motivated which presents difficulties for movement-based accounts as appearance of terms out of their canonical or ‘in situ’ positions without a good syntactic reason (i.e. to license the thematic and case properties of arguments, as posited by Neeleman and van de Koot (2016)) could not be properly explicated.

The example in (61) below shows the appropriate contextual question for an OSV sentence, and the example in (62) provides an illustration of the use of an OSV

sentence in the context of a written narration (the Kazakh version of the Little Raccoon story).

- (61) [Misiq-ti KIM äkel-di?]
cat-ACC who bring- PST(3)
'Who brought the cat?'

Misiq-ti BoLAT äkel-di.
cat-ACC Bolat bring-PST(3)
'Bolat brought the cat.'

#Bolat misiq-ti äkel-di.
Bolat cat-ACC bring-PST(3)
'Bolat brought the cat.'

- (62) Kiškentaj Žanat orman-ğa kele žat-ti.
little raccoon forest-DAT come AUX-PST(3)
Onı majmıl körd-di.
3SG.ACC monkey see-PST(3)
'Little Raccoon came to the forest. A monkey saw him.'

In both the question and the response in (61), the sentence-initial direct object *misiqtı* is pragmatically presupposed and has the information-structural role of topic, while the immediately preverbal subject *Bolat* is focused (it can also be contrastively focused); the verb is part of backgrounded information here. It is shown that the neutral SOV word order is not felicitous in this context.

In the second sentence in (62), the sentence-initial direct object, expressed by the pronoun referring to the subject of the previous sentence, is the topic expression of the sentence. The subject and verb form the non-presupposed, or pragmatically asserted part of the sentence, thus becoming the comment.

It must be noted, that unlike SOV sentences, in which a direct object can be marked or unmarked with the accusative case-marker, direct objects must be accusatively-marked in OSV sentences, as ungrammaticality of the example in (63) shows.

- (63) *Misîq BoLAT äkel-di.
cat Bolat bring-PST.3
Intended: ‘Bolat brought a/the cat.’

This is discussed in detail in Chapter 4 on differential object marking. Simply noting the obligatoriness of the accusative marker on direct objects in OSV clauses is sufficient for this section.

Looking at the infelicitous examples in (58)-(60), the question and response pair in (61), and the second sentence in (62), we can make some initial observations about the information-structural properties of the OSV word order. An OSV sentence cannot be an all-focus sentence (as exemplified by (58)), it cannot have focused direct object + verb, or narrowly focused direct object. In question-answer pairs the sentence-initial direct object is the topic, while the immediately preverbal subject is narrowly focused; the verb in these cases is usually a part of the background.

Where an OSV sentence is used in a narrative, like in (62), the direct object is topical, while the subject and verb form the comment. This technically qualifies for Lambrecht’s (1994) definition of predicate focus (Section 2.1 of Chapter 2.), which is also referred to as the event-reporting or presentational sentence type.

Additionally, the sentence-initial topical element can also be contrastive – recall the characteristics of it given in Section 2.1 of the preceding chapter, and the observations made for SOV sentences in the preceding section. This is exemplified by (64) where the question contains the topical referent *the animals*, and the answer is given in regard to a different, contrastive topical referent *misîq* ‘cat’.

- (64) [Who brought the animals?]
Misîq-tî BoLAT äkel-di.

cat-ACC Bolat bring-PST(3)
 ‘Bolat brought the cat.’

So far, then, we have established that the OSV word order can have the information-structural components as presented in Table 14.

Table 14. Information structure of an OSV sentence (to be revised)

Direct Object	Subject	Verb
Topic	Focus	Background
(Contrastive) Topic	Comment/Predicate Focus	

In the traditional generative approach, the OSV word order for a typically SOV language is described as scrambling of object over subject. Kučerová (2007) observes that, in Czech, this is motivated by givenness, as the examples in (65) demonstrate.

- (65) a. Chlapec našel lízátko.
 boy.NOM found lollipop.ACC
 ‘A/the boy found a/the lollipop.’
 new > new
 given > new
 given > given
 #new > given
- b. Lízátko našel chlapec
 lollipop.ACC found boy.NOM
 ‘A boy found the lollipop’
 given > new

(Kučerová 2007: 10)

The same observation applies to Kazakh, as shown in (66), and confirms our findings presented in Table 14, seeing as a topical element is usually given (recall the discussion in Chapter 2).

- (66) a. Bala tort-ti tap-ti.
 child cake-ACC find-PST(3)
 ‘A/the boy found the cake.’
 new > new
 given > new
 given > given
 #new > given
- b. Tort-ti bala tap-ti.
 cake-ACC child find-PST(3)
 ‘A boy found the cake.’
 given > new

These examples corroborate the conclusions we have so far made about the information-structural properties of SOV and OSV sentences.

A small remark must be made here about the stress in OSV sentences. As indicated by small caps in the OSV sentences in (61), the main sentential stress falls on the canonical IPV position as previously described. However, if the main sentential stress is assigned to the sentence-initial direct object instead, that direct object receives a contrastively focused reading. This, of course, results in the sentence no longer having the information-structural configurations as presented in Table 14, thus making an OSV sentence like that infelicitous in the context given in (61), as shown in (67) below.

- (67) [Misiq-ti KIM äkel-di?]
 #Misiq-ti Bolat äkel-di.
 cat-ACC Bolat bring-PST(3)
 ‘Bolat brought the cat.’

A felicitous context for an OSV sentence with a contrastively focused direct object is presented in (68).

- (68) [Keše Bolat it-ti äkel-di.]
 yesterday Bolat dog-ACC bring-PST(3)
 ‘Yesterday Bolat brought the dog.’
 (Žoq), misiq-ti Bolat äkel-di.
 No cat-ACC Bolat bring-PST(3)
 ‘No, it was the cat that Bolat brought.’

We can see that in the case where the direct object is contrastively focused in an OSV sentence, the information structure of the clause changes completely. Now the sentence-initial direct object is contrastively focused, while the subject and the verb form the background, *Bolat* being the continuous aboutness topic.

This demonstrates again that although the sentence-initial position is typically reserved for topics, topics are by no means confined to this position, especially in spoken language where prosody plays an important part in marking the information-structural statuses of constituents. The occurrence of a contrastively focused element sentence-initially is not surprising, taking into account sub-Section 3.3.3 and the schema in (42). We can now expand Table 14, and the revised version is given below.

Table 15. Information structure of an OSV sentence (revised)

Direct Object	Subject	Verb
Topic	Focus	Background
(Contrastive) Topic	Predicate Focus	
Contrastive Focus	Topic	Background

3.4.2. Verb-medial Word Orders – SVO and OVS

This section is dedicated to two word orders in which the verb takes the medial position – SVO and OVS word orders. In both these word orders one term is in the IPV position, and the other term is located post-verbally, which means that the verb is no longer in its canonical sentence-final position, as shown in the examples in (69) and (70).

(69) BOLAT äkel-di mäsïq-tï.
Bolat bring-PST(3) cat-ACC
'Bolat brought it, the cat.'

(70) Mäsïq-tï äkel-di Bolat.
cat-ACC bring-PST(3) Bolat
'It was the cat that he brought, Bolat.'

As discussed previously, the IPV position is the canonical NIF position in Kazakh, and it retains this property in the verb-medial sentences, which can be confirmed by considering contexts in which SVO and OVS sentences are felicitous and infelicitous. Let us start with the SVO word order and consider the contexts in which the sentence from (69) is felicitous.

(71) What happened?
#Bolat äkeldi mäsïqtï.

(72) What did Bolat do?
#Bolat äkeldi mäsïqtï.

(73) What did Bolat bring?
#Bolat äkeldi mäsïqtï.

- (74) Who brought the cat?
 BOLAT äkeldi mäsıqtı.

The questions in (71)-(73) provide contexts for all-sentence focus, focus on the direct object and verb, and narrow focus on the direct object respectively, and the SVO sentence from (69) is not felicitous in any of these contexts. However, as expected, it is felicitous in the context of narrow focus on the subject. Both the verb and the object in this case are pragmatically given and belong to the backgrounded part of the sentence.

The post-verbal direct object *mäsıqtı* is a continuous topic expression, and its appearance post-verbally can create the effect of an afterthought, especially, since the topic referent is already clear from the question/context.

As before, a narrowly focused term in the IPV can yield a contrastive focus reading, as demonstrated by (75).

- (75) [Marat brought the cat yesterday.]
 Žoq, Bolat äkeldi mäsıq-tı.
 NEG Bolat bring-PST(3) cat-ACC
 ‘No, BOLAT brought the cat.’

The typical information structure of an SVO sentence can be represented as follows:

Table 16. Information structure of an SVO sentence (to be revised)

Subject	Verb	Direct Object
Focus/Contrastive Focus	Background	Topic

OVS sentences follow the same pattern, as the question-answer pairs below demonstrate.

- (76) What happened?
 #Mäsıqtı äkeldi Bolat.

- (77) What did Bolat do?
#Mīsīqtī ākeldi Bolat.
- (78) What did Bolat bring?
Mīsīqtī ākeldi Bolat.
- (79) Who brought the cat?
#Mīsīqtī ākeldi Bolat.

The OVS sentence from (70) is not felicitous in any context other than that in which the direct object is narrowly focused, as in (78); the post-verbal subject is the topic. Again, the immediately preverbal direct object can be contrastively focused, as (80) shows.

- (80) [Bolat brought the dog.]
Žoq, Mīsīq-tī ākel-di Bolat.
NEG cat-ACC bring-PST.3 Bolat
'No, it was the cat that Bolat brought.'

Thus, we can see that the information structure of an OVS sentence is as presented in Table 17.

Table 17. Information structure of an OVS sentence (to be revised)

Direct Object	Verb	Subject
Focus/Contrastive Focus	Background	Topic

Let us consider some further examples given in (81) and (82) below. The example in (81) contains an exchange between Speaker A and Speaker B, where Speaker A makes a negated statement, and speaker B contradicts it by stressing a part of the verb thus creating the verum focus effect.

- (81) A: Aliya-ni Marat kör-di de, al
 Aliya-ACC Marat see-PST(3) CONJ but
 Dana-ni kör-gen žoq.
 Dana-ACC see-PTCP NEG
 ‘Marat has seen/saw Aliya, but (as for) Dana, he hasn’t seen her.’
- B: Dana-ni kör-DI Marat.
 Dana-ACC see-PST(3) Marat
 ‘As for Dana, Marat *has* seen her.’

In this case, then, both the direct object and the subject are presupposed, and the predicate is stressed which in this case indicates verum focus. The same pattern can be observed for an SVO sentence, as shown in Speaker B’s utterance in (82).

- (82) A: Bolat kör-di Dana-ni, al Marat
 Bolat see-PST(3) Dana-ACC but Marat
 Dana-ni kör-gen žoq.
 Dana-ACC see-PTCP NEG
 ‘Bolat has seen Dana, but Marat has not seen Dana.’
- B: Marat kör-DI Dana-ni
 Marat see-PAST(3) Dana-ACC
 ‘As for Dana, Marat *has* seen her.’

The main difference between an SVO and an OVS sentence in these examples lies in the information-structural properties of the arguments. The argument preceding the verb is topical, whereas the one following it is backgrounded. Alternatively, it can be said that the argument preceding the focally stressed verb is the primary topic, and the argument following the verb is the secondary topic.

We can therefore expand Tables 16 and 17 into Tables 18 and 19 respectively.

Table 18. Information structure of an SVO sentence (revised)

Subject	Verb	Direct Object
Focus/Contrastive Focus	Background	Topic
Topic	Verum Focus	Background/Secondary Topic

Table 19. Information structure of an OVS sentence (revised)

Direct Object	Verb	Subject
Focus/Contrastive Focus	Background	Topic
Topic	Verum Focus	Background/Secondary Topic

3.4.3. Verb-initial Word Orders – VSO and VOS

In this section I move to the verb-initial word orders and their information-structural properties. I first refer to Göksel and Kerslake (2005) for a description of verb-initial sentences in Turkish, and then show that their observations are also true for Kazakh.

First, a remark on imperative sentences which are often verb-initial is in order. Following Erguvanlı (1984) I do not take imperative sentences into consideration in this description, since despite surface similarity to verb-initial declarative sentences, the difference in the illocutionary force results in significant differences in the syntax. For example, as (83) shows, that an unmarked indefinite direct object can appear after a verb in an imperative clause, which is unacceptable in a declarative clause, as shown in sub-Section 3.3.5 of this chapter.

- (83) Äkel mağan bir gazet, men de
bring.IMP me.DAT one newspaper 1SG too
oq-i-mïn.
read-FUT-1SG
'Bring me a newspaper, I'll read as well.'

Luraghi (1995) notes that many languages display a tendency to place imperative verb forms into the sentence-initial position, and views these sentences (along with verb-initial questions in English, for example) as grammaticalised types of VSO sentences since they have the function of indicating a certain sentence type. For this reason, Luraghi also excludes imperative verb-initial sentences from her study into pragmatics of verb-initial sentence types in some ancient Indo-European languages.

For Turkish verb-initial sentences proper, Göksel and Kerslake observe that “the predicate is obligatorily stressed (and focused) if it is at the beginning of a sentence” (2005: 344). They state that if all sentential constituents are placed post-verbally, their relative positioning is insignificant, and illustrate this with the examples in (84)-(86) which, according to the authors, are “informationally identical” (2005: 346).

(84) Bitir-Dİ Fatma üniversite-yi bu yıl.
 finish-PFV Fatma university-ACC this year
 ‘Fatma FINISHED university this year.’

(85) BitirDİ üniversiteyi bu yıl Fatma

(86) BitirDİ bu yıl Fatma üniversiteyi.

(Göksel and Kerslake 2005: 346)

English approximation of the sentences above reflects focused status of the verb. Göksel and Kerslake do not provide the context for these examples, however, native speakers of Turkish informed me that sentences like those in (84)-(86) would be felicitously uttered in the context of a preceding statement of opposite polarity *Fatma did not finish university this year*. We can thus conclude that the examples in (84)-(86) contain verum, rather than corrective predicate focus which we would have got as a response for a sentence like *Fatma started university this year*.

Verb-initial sentences do not occur frequently in either spoken or written Kazakh, however, native speakers accept the sentences in (87) and (88) as possible sentences of the language.

(87) Äkel-di Bolat misiq-ti.
 bring-PST(3) Bolat cat-ACC
 ‘Bolat BROUGHT/ DID bring the cat.’

(88) Äkel-di misiq-ti Bolat.
 bring-PST(3) cat-ACC Bolat
 ‘Bolat BROUGHT/DID bring the cat.’

Like in Turkish, sentence-initial verbs in Kazakh are necessarily stressed and focused, and all the observations on the properties of post-verbal constituents (Section 3.3.5) hold for the constituents following the sentence-initial verb. The inability of the post-verbal elements to bear the main sentential stress leaves the verb itself the only available stress-bearing element in verb-initial sentences.

As mentioned earlier, Göksel and Kerslake (2005) state that the relative positioning of sentential constituents is not important. I found that this is also true for the Kazakh verb-initial sentences, however, it appears that the preceding context can determine a preferred ordering of the elements in the post-verbal area of verb-initial sentences. Consider the example in (89) below, where in a short exchange between three speakers, the sentence from (87) is repeated in sentence C; sentences A and B provide context.

(89) A: Misiq-ti KIM äkel-me-di?
 cat-ACC who bring-NEG-PST(3)
 ‘Who did not bring the cat?’

B: Misiq-ti BOLAT äkel-me-di.
 cat-ACC Bolat bring-NEG-PST(3)
 ‘Bolat did not bring the cat!’

C: Äkel-DI Bolat misiq-ti.
 bring-PST(3) Bolat cat-ACC
 ‘Bolat BROUGHT/ DID bring the cat.’

In sentence B, *misiqti* is the topic expression, and *Bolat* is the focus; in sentence C both *Bolat* and *misiqti* are given/presupposed and appear post-verbally in a SO sequence, which appears to be preferred over the OS order in this case, although neither option was ruled out completely and categorically by my language consultants. It seems that if one of the terms is narrowly focused in the preceding sentence/context, in a following verb-initial sentence it would appear immediately after the verb, and would be somewhat more information-structurally prominent than the term that was topical in the preceding context. However, if both terms are topical or focused in the preceding context, then they can occur in any order in the following verb-initial sentence, as shown in (90), where both sentences C and C’ are felicitous in the context of an all-focus sentence in B.

(90) A: What is the matter? Why are you upset?

B: Bolat misiq-ti äkel-me-di!
 Bolat cat-ACC bring-NEG-PST(3)
 ‘Bolat did not bring the cat!’

C: Äkel-di Bolat misiq-ti.
 bring-PST(3) Bolat cat-ACC
 ‘Bolat BROUGHT/ DID bring the cat.’

C’: Äkel-di misiq-ti Bolat.
 bring-PST(3) cat-ACC Bolat
 ‘Bolat BROUGHT/ DID bring the cat.’

It would appear, then, that the information-structural status of terms in the context immediately preceding the verb-initial sentence can influence the preferred ordering of the terms in the post-verbal area. The term which is more information-structurally prominent (i.e. a narrowly focused) in the preceding context would immediately follow the verb in the verb-initial sentence, while information-structurally ‘equal’ terms do not have a preferred order in the post-verbal area.

To sum up, we can represent the information-structural properties of verb-initial sentences in the following way:

Table 20. Information structure of verb-initial sentences

Verb	Subject/Direct Object	Direct Object/Subject
Verum/Contrastive Focus	Topic	Background

3.4.4. Conclusions

This section has examined five scrambled word orders available in Kazakh: OSV, verb-medial SVO and OVS, and verb-initial VSO and VOS orders. It has been demonstrated that none of these word orders can be used to express sentence-focus, which once again proves SOV the neutral word order in Kazakh.

It has been shown that all scrambled word orders are pragmatically motivated, and are only felicitous in certain sometimes very restricted contexts, which has been determined through numerous contextualised examples. The presented examples help us reach beyond the usual limited observations which claim that topics appear sentence-initially, and focused elements appear in the IPV position. This section fills this gap in the literature, and describes information-structural characteristics of all possible permutations of a Kazakh sentence comprising a subject, a direct object and a verb.

3.5. Kazakh Word Order in Dynamic Syntax

In this section I present the DS approach to word order variation in Kazakh, and demonstrate that the information-structural effects described above for the neutral and scrambled word orders can be easily accounted for if the dynamics of the parsing process are taken into consideration.

First, the step-by-step parsing process of a sample Kazakh sentence is given in 3.5.1; 3.5.2 focuses on the neutral SOV word order and its information-structural configurations. In 3.5.3, the DS account of the OSV word order is sketched out, followed by 3.5.4 which considers verb-medial and verb-initial word orders.

3.5.1. Modelling Kazakh in DS

This sub-section presents the parsing process of a sample Kazakh sentence and its DS representation. As mentioned in Section 1.1 of Chapter 1, the Kazakh language has a rich case system. This is significant for the parsing process, as this allows the parser to identify and fix the node onto which a certain noun phrase must be parsed, as was shown for the Korean language in Section 2.2 of Chapter 2.

As also outlined in Section 2.2 of Chapter 2, prosody plays an important role in the parsing process. It was shown that an IP boundary identifies subjects in Kazakh as such in the same way an overt case marker would. DS allows us to incorporate this observation into a formal representation of the parsing process, which sets this theoretical framework apart from most other formal syntactic frameworks.

The sample Kazakh sentence under consideration is given in (91) below.

- (91) Bolat kitap-ti Dana-ğa berdi.
 Bolat book-ACC Dana-DAT gave
 ‘Bolat gave Dana a/the book.’

Unlike in the Korean example examined in Section 2.2.7, the subject in our Kazakh example is not overtly case marked. As mentioned in Section 1.1 of Chapter 1,

traditional grammars refer to this unmarked form as the nominative case form, however, I refer to it as the unmarked or bare form. Unmarked nouns typically function as subjects, although they can also function as direct objects⁷ when appearing immediately before the verb, as described in detail in Chapter 4.

As demonstrated in sub-Section 2.2.8 of Chapter 2, in Kazakh, an unmarked sentence-initial nominal may or may not be followed by an IP boundary. Where there is an IP boundary after an unmarked noun, that noun is parsed as the subject of the sentence. In a sense, then, an IP boundary acts as a case marker, and fixes the node at which it appears. The lexical entry for an IP boundary in Kazakh is repeated in (92) below.

(92) Lexical entry for an IP boundary % (simplified)

IF	Fo(α), Ty(e)
THEN	IF $\langle \uparrow_0 \uparrow_i^* \rangle$ (Tn(a), ?Ty(t))
	THEN put(? $\langle \uparrow_0 \rangle$ Ty(t))
	ELSE Abort
ELSE	Abort

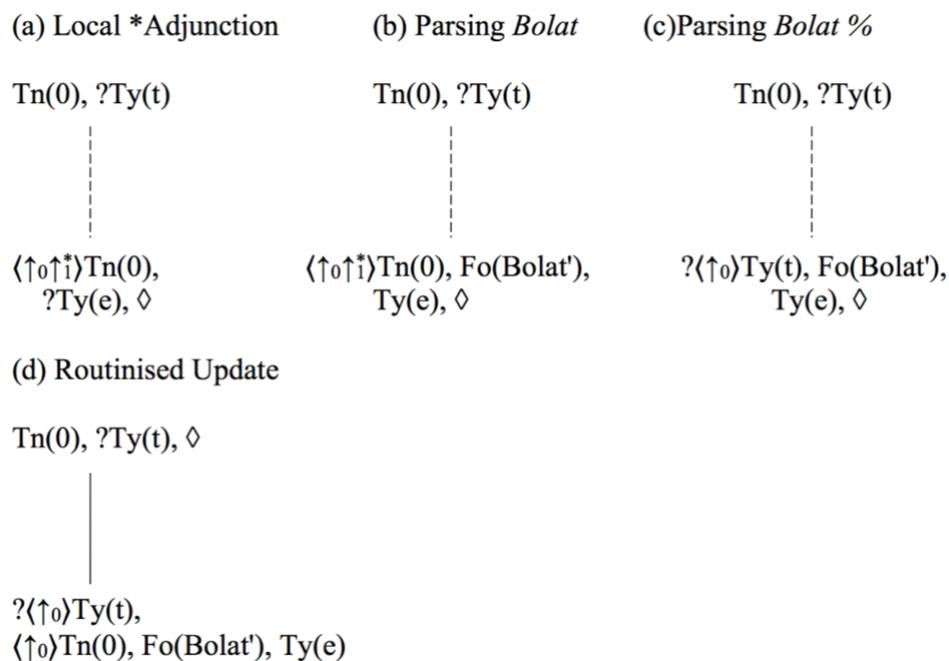
According to the first IF-statement of this lexical entry, the IP boundary must be parsed at an unfixed node decorated with a formula of Ty(e). If this condition is satisfied, another IF-statement checks that this unfixed node is dominated by the root node, which ensures that we are dealing with a locally unfixed node. The following THEN-statement decorates the node with a requirement to be fixed as an argument daughter of that root node, thus effectively fixing it as the subject node. If either condition from IF-statements is not satisfied, the parse is aborted.

⁷ It must be noted that unmarked nouns can also function as adverbial elements (e.g. *keše* ‘yesterday’), as well as appear within certain prepositional phrases (e.g. *žumis turalı* ‘about work’). In this thesis, however, the focus is on unmarked nouns function as predicate terms.

I do not exclude the possibility that an IP boundary might perform some other functions in Kazakh, but this is a matter for a much larger prosodic investigation, which is out of the scope of this work. The lexical entry presented above solely reflects the function an IP boundary plays in the parsing of a non-case-marked nominal in a simple Kazakh sentence. Note, however, that this lexical entry is expanded upon in Section 6.7 of Chapter 6, where sentences with relative clauses are examined.

Looking at our sample sentence, then, *Bolat* is first parsed onto an unfixed node via the Local *Adjunction rule. The IP boundary (indicated %) then signals to the parser that this node should be fixed as the subject of the sentence, as per the lexical entry above. A routinised update cleans up the tree and brings the parser up to the root node. This parsing process is shown in (93) below.

(93) Parsing *Bolat* %



Recall Kiaer's (2014) observation (see sub-Section 2.2.7 of Chapter 2) that in Korean, structures can be built without case markers, and that the nominative case marker is frequently dropped in spoken Korean. She further states that it is the sentence-initial positioning of the noun phrase, and not the nominative case marker

that creates expectation for this phrase to be the subject by default. It would be of interest to consider the prosody of Korean sentences without an overt nominative marker with a view to determining whether an IP boundary replaces the case marker in these instances or not.

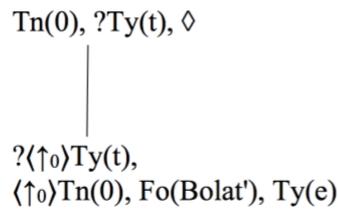
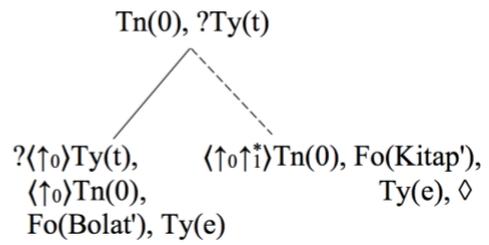
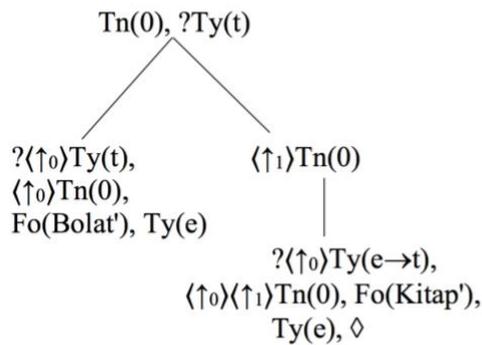
Moving on to the next constituent from (91) – the direct object *kitap-ti*. As can be seen from the gloss, there is an overt accusative marker indicating direct objecthood. The lexical entry for the accusative case marker is presented in (94) below.

(94) Lexical entry for *-DI/-n* (accusative)

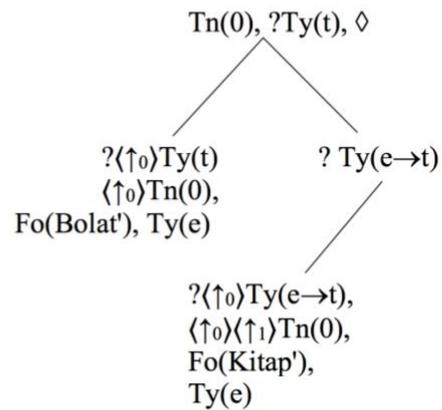
IF	$Fo(\alpha), Ty(e)$	
THEN	IF	$\langle \uparrow_0 \uparrow^* \rangle (Tn(\alpha), ?Ty(t))$
	THEN	$put(? \langle \uparrow_0 \rangle Ty(e \rightarrow t))$
	ELSE	Abort
ELSE	Abort	

This lexical entry is identical to that for the Korean accusative marker shown in (84). The first IF statement identifies the node at which the case marker should be parsed at an unfixed node with a type e formula decorating it. The following THEN statement contains another subset of IF-THEN-ELSE statements. The second IF statement stipulates that node at which the accusative case marker is parsed should be a locally unfixed node, that is, it should have been introduced via the Local *Adjunction rule. The THEN statement decorates the node with a requirement for it to be the argument daughter of a type $(e \rightarrow t)$ node, thus, essentially, identifying it as the direct object. A routinised update then takes place and fixes the node in the same way as shown for Korean in (85). The tree construction process below demonstrates the parsing of the first two constituents of the sentence from (91).

(95) Parsing *Bolat % kitap-ti...*

(a) After *Bolat* has been parsed(b) Parsing *Bolat kitap...*(c) Parsing *Bolat kitap-ti...*

(d) Routinised Update



As can be seen from (95), by the end of the parse of *kitapti* a partial structure is built up, where proper name *Bolat* decorates the subject node, and the accusatively marked noun *kitap* decorates the direct object node.

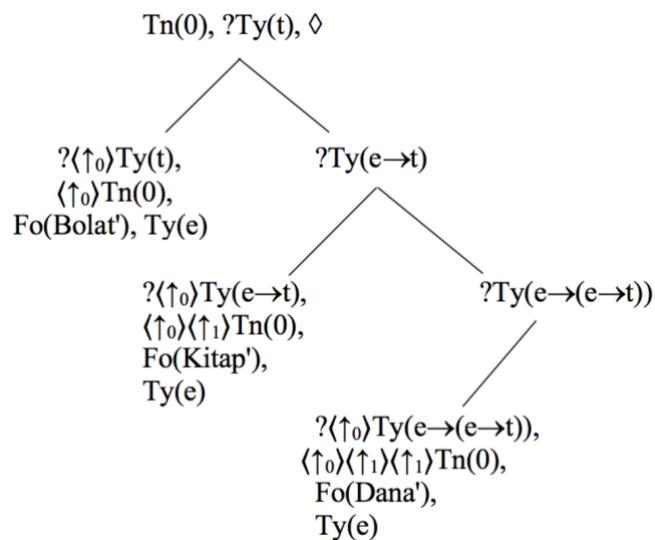
Moving on to the next constituent from the sample sentence in (91) – *Danağa* – it is marked with the dative case marker indicating its indirect objecthood. The lexical entry for dative case markers is presented in (96) below.

(96) Lexical entry for $-\check{G}A/-nA/-A$ (dative)

IF		$\text{Fo}(\alpha), \text{Ty}(e)$
THEN	IF	$\langle \uparrow_0 \uparrow_1^* \rangle (\text{Tn}(a), ?\text{Ty}(t))$
	THEN	$\text{put}(?\langle \uparrow_0 \rangle \text{Ty}((e \rightarrow (e \rightarrow t))))$
	ELSE	Abort
ELSE		Abort

The main difference between the lexical entries of the accusative and dative case markers is in the second THEN statement – the node carrying the datively marked constituent has the requirement to be underneath the decoration of type $(e \rightarrow e \rightarrow t)$, thus ensuring its position as the indirect object. The partial tree in (97) illustrates the progress of the parsing process after the parse of *Danağa* has been completed.

(97) Parsing *Bolat kitap-ti Danağa...*



The final element of the sample sentence is the verb *berdi* ‘gave’. The lexical entry for it is presented in (99) below. Similarly to the lexical entry of the Korean three-place predicate given in sub-Section 2.2.7 of Chapter 2, the Kazakh verb *ber-* ‘give’ projects a full argument-predicate structure with metavariables decorating the argument nodes. This is due to the full pro-drop nature of the Kazakh language, which is confirmed by the example in (98) below.

(98) Q: Bolat kitap-ti Dana-ğa ber-di me?
 Bolat book-ACC Dana-DAT give-PST(3) Q
 ‘Did Bolat give Dana the book?’

A: Ber-di.

give-PST(3)

‘He gave it to her./He did.’

The second sentence in (98) consists only of the verb *berdi* ‘gave’ and constitutes a fully grammatical response to the posed question. Unlike its English translation, which minimally must include at least the pronominal form of the subject, none of the terms have to be overtly expressed in the Kazakh sentence. This confirms that the lexical entries of Kazakh verbs project full argument-predicate structures with metavariables decorating the argument nodes. These metavariables are subsequently substantiated from the context, and the semantic tree constructed by the end of the parse of a sentence like the response utterance in (98) would not look any different from the tree at the end of the parse of a full sentence *Bolat kitapti Danağa berdi*. The same would hold for the sentence with pronominal forms of the arguments. With this in mind, the simplified lexical entry for *berdi* is given below.

(99) Lexical entry for *berdi* ‘gave’ (simplified)

IF	?Ty(t)
THEN	put(Tns(PAST)) ⁸ ; make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(U), Ty(e), ? $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(?Ty(e \rightarrow t)); make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(V), Ty(e), ? $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(?Ty(e \rightarrow (e \rightarrow t))); make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(W), Ty(e), ? $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$); make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(Fo(Ber'), Ty(e \rightarrow (e \rightarrow (e \rightarrow t))), [\downarrow] \perp);

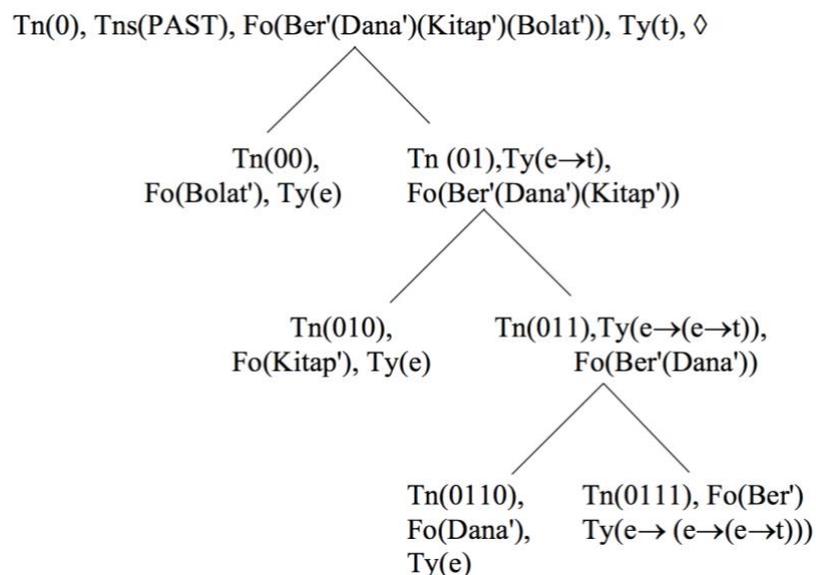
⁸ To be precise, this decoration along with the bottom restrictor is projected specifically by the tense marking suffix *-DI*, however, this point is not vital for our discussion. Therefore, a simplified version of the lexical entry which combines the contributions of both the lexical verb root and the suffix is presented. A detailed DS account of Kazakh (and generally speaking Turkic) tense markers is a vast topic for future research.

ELSE Abort

As mentioned in Section 2.2.7 of Chapter 2, there is no issue with the same argument nodes being constructed twice. If two nodes with the same location are constructed in the same tree, then they are simply identified as one and the same node, without the need for Merge to take place.

Once the verb is parsed and all the Completion and Elimination processes have taken place, the final tree for our sample sentence would look as shown in (100) below. Parsing of an utterance can be considered complete and successful if there are no outstanding requirements on any tree nodes – this is precisely what is observed in the complete tree in (100).

(100) Final tree for *Bolat kitapti Danağa berdi* (simplified)



The tree presented above is constructed as the result of a successfully completed parse for the string *Bolal kitapti Danağa berdi*. Since the above tree is not a hierarchical syntactic structure, but a semantic representation of the utterance, any configuration of the same elements would result in the same tree as shown in (100). That is to say, if the string under consideration were *Danağa Bolal kitapti berdi* or *Kitapti Bolal*

Danağa berdi, the tree growth processes would differ, but the final semantic tree would be the same. This accurately reflects the identical truth-value semantics of the three strings mentioned, as well as the relatively free word order of Kazakh and other similar languages. No additional movement rules need to be introduced to explain or ‘license’ the appearance of an indirect object in the clause-initial position, for instance.

To summarise, this section presented a (simplified) DS parsing process for a sample Kazakh sentence consisting of a full-form overt subject, direct and indirect objects and a three-place predicate. It has been demonstrated that a partial semantic tree structure is built up incrementally from the very start of the parsing process thanks to prosodic and morphological clues that enable the parser to identify different arguments as a subject, direct object or an indirect object.

Each argument is initially parsed onto a locally unfixed node, which becomes fixed before the parsing of the next argument begins. Overall Kazakh shows a strong preference for immediate fixing of nominal constituents and leaving unfixed nodes as such is usually avoided. However, in the next sub-section I discuss some cases where a node remains unfixed until the following constituent is parsed.

3.5.2. SOV in DS

Following on from the previous section, this section presents possible developments of the parsing process for an SOV sentence with different information-structural characteristics: sentence-/all-focus, topic-comment, and topic-focus-background (see Section 3.2.5 of this chapter). I do not examine clauses with contrastive topics or focuses, since these are essentially the same as their non-contrastive counterparts but with the additional semantics of diverging in some way from the preceding utterance. Formal semantic representation of this is left for further research and is not the focus of this study.

Taking into consideration the DS parsing process shown in the preceding section, we can easily sketch out the DS parsing process for our sentence *Bolat mäsıqtı äkeldi* where the whole sentence is in focus – as demonstrated in (101); the lexical entry for the verb *äkeldi* is given in (101).

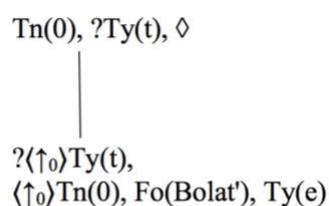
(101) Lexical entry for *äkeldi* ‘brought’ (simplified)

IF ?Ty(t)
 THEN put(Tns(PAST));
 make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(U), Ty(e), ? $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$);
 make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(?Ty(e \rightarrow t));
 make($\langle \downarrow_0 \rangle$); go($\langle \downarrow_0 \rangle$); put(Fo(V), Ty(e), ? $\exists x.Fo(x)$); go($\langle \uparrow_0 \rangle$);
 make($\langle \downarrow_1 \rangle$); go($\langle \downarrow_1 \rangle$); put(Fo(Äkel'), (Ty(e \rightarrow (e \rightarrow t)))); [\downarrow] \perp);
 ELSE Abort

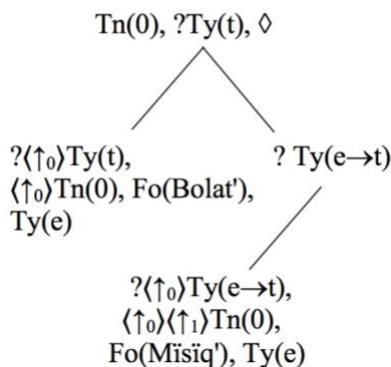
This lexical entry is very similar to that for the verb *berdi* given in (99) in the preceding section, apart from the number of argument nodes it builds and decorates with metavariables. According to this lexical entry, the verb *äkeldi* projects the argument structure consisting of a subject node and a direct object node. As previously, by the time the parser processes the verb, the argument structure is already constructed with the help of prosody and case marking, and the argument structure projected by the verb simply collapses with the existing argument structure.

(101) Parsing [*Bolat misiqti äkeldi*]_{Focus}

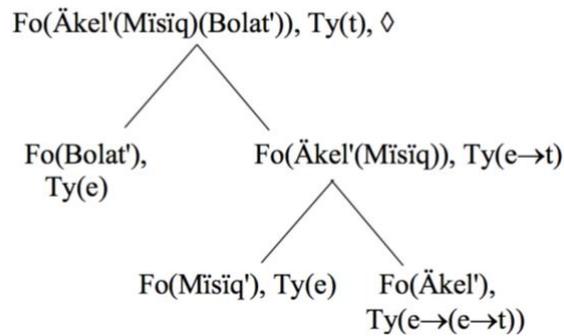
(a) Parsing *Bolat...*



(b) Parsing *Bolat misiqti...*



(c) Parsing *Bolat misiqti akeldi*

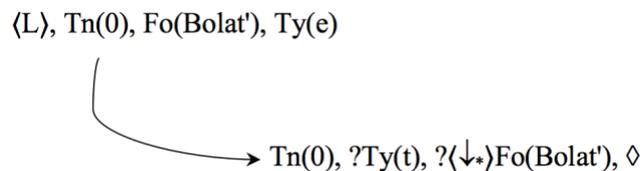


The final semantic tree above has no outstanding requirements, and represents the complete parsing process for our string *Bolat misiqti akeldi*.

Let us now consider how the parsing process develops for the same string with different information-structural characteristics. Recall, that when this sentence is used as the response to the question *What did Bolat do?*, it displays the topic-comment structure. Recall also (Section 2.2 of Chapter 2) that in DS, Linked structures which are used to formally represent topics as partial trees that provide minimal context against which the rest of the utterance is parsed. Additionally, via a Linked structure, a topic provides an antecedent which is subsequently identified by some anaphoric element – this, according to Kempson et al. (2006) is exactly what creates the topic effect.

Thus, the sentence-initial topical subject *Bolat* in our topic-comment string *Bolat misiqti akeldi* is parsed onto a Linked structure, as shown in (102) below.

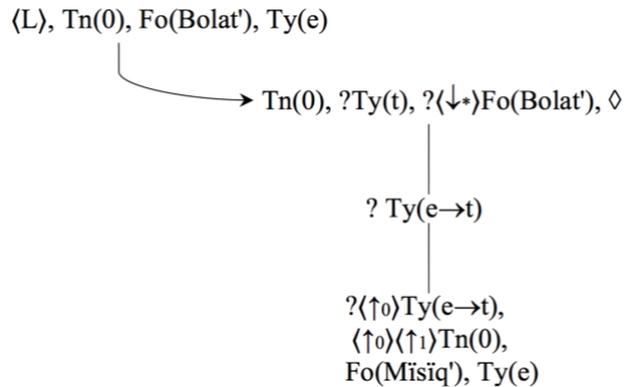
(102) Parsing [*Bolat*]_{Topic}



At the end of this parse, the topical element *Bolat* is parsed onto a node of a partial tree, which is Linked to the root node of another partial tree, which is decorated with a requirement for a copy of *Fo(Bolat')* to be present somewhere in that tree. This partial

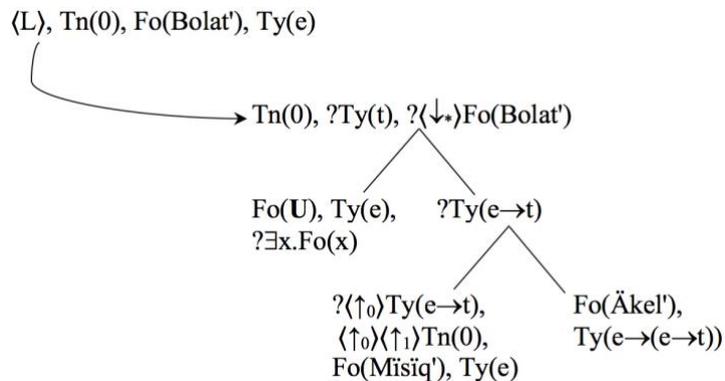
tree provides the minimal context against which the rest of the utterance – information-structural comment – is parsed. This creates the effect of new, focused information being added to the CG in relation to some known referent. Next, the accusatively marked direct object *misiqti* is parsed, as demonstrated in (103) below.

(103) Parsing [*Bolat*]_{Topic} [*misiqti...*]_{Comment}



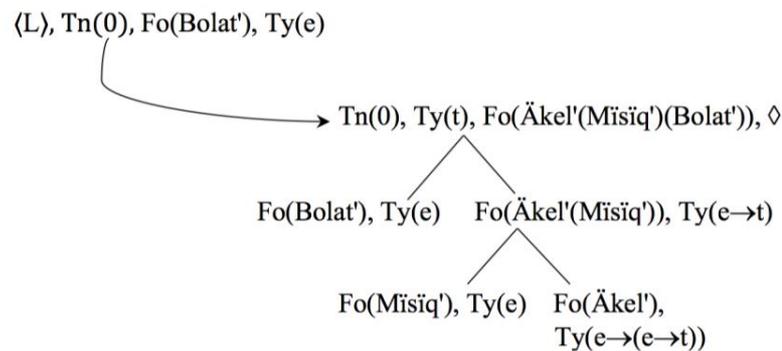
Thanks to the accusative case marking, the direct object *misiqti* is parsed onto a fixed node. Next in the string is the verb, which projects the full argument structure, as per the lexical entry in (101) above. Since the direct object node is already present in the structure, the same node projected by the verb harmlessly collapses with it. The subject node, on the other hand, is not present in the structure prior to the parsing of the verb, and is, therefore, projected by the verb and decorated with a metavariable, as shown in (104).

(104) Parsing [*Bolat*]_{Topic} [*misiqti akeldi*]_{Comment}



Now that there is a fixed node with a requirement for a formula in the tree structure, the requirement for a copy of $Fo(\text{Bolat}')$ to be present somewhere in that tree can be satisfied; this is the antecedent-anaphora relation which creates the topicality effect, as described above and in Chapter 2. With these, and all other requirements fulfilled, the information is compiled up the tree, resulting in the structure shown in (105) below.

(105) Parsing [*Bolat*]_{Topic} [*misiqti äkeldi*]_{Comment} – final



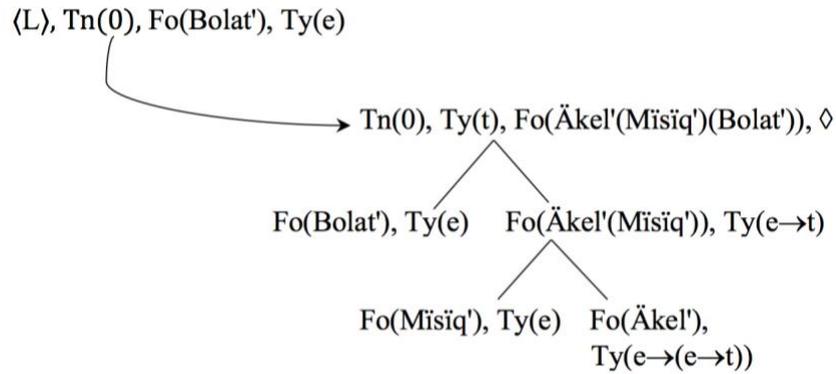
I now move to the final information-structural configuration for an SOV sentence – topic-focus-background – in which everything but the focused part of the sentence is presupposed. This provides basis for a clear illustration of the DS approach to focus and, more specifically, to narrow focus on an argument. Recall (sub-Section 2.2.9 of Chapter 2) that in DS, the focus effect is considered to arise from the process of updating a proposition in a way that results in its completion.

This means that the response *Bolat misiqti äkeldi* from the question-answer pair from (27), repeated in (106) below for convenience is essentially almost fully constructed by the question, and only the update is required in order for the proposition to be complete. This is reflected by the tree in (107) where the metavariable decoration requiring an update is underlined.

(106) [*Bolat* NE *äkel-di*?]
 Bolat what bring-PST(3)
 ‘What did Bolat bring?’

Once all requirements are satisfied and the information is compiled up the tree, the resulting final tree looks no different from that in (105) above:

(109) Parsing [*Bolat*]_{Topic} [*misiqti*]_{Focus} [*äkeldi*]_{Background} – final



Comparing the final trees from (101), (105), and (109) above, we can see that the only difference between these trees lies in the presence of a Linked structure in the latter two; the root node decorations of all three trees are identical. This, together with the step-wise parsing process represented in the semantic trees, formally expresses and explicates the difference in the IS of these sentences.

Summing up, the DS representation of SOV sentences with diverse information-structural configurations confirmed: a) that the information-structural effects can be elegantly explicated through the step-wise representation of the parsing process within context; b) the lack of necessity to identify specific information-structural positions or assign information-structural features to any sentential positions; and c) that despite the differences in IS, parsing of one and the same string results in the same main tree root node decoration which demonstrates semantic/truth-conditional equivalence.

3.5.3. OSV in DS

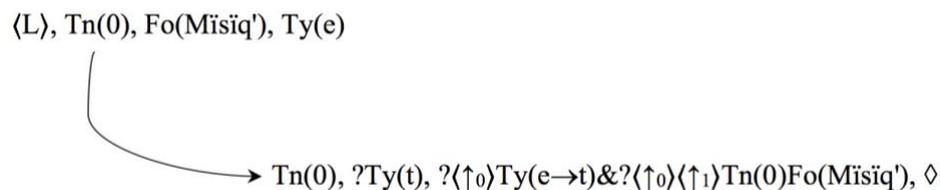
In this section the DS sketches of different information-structural configurations of an OSV sentence is presented. As previously, I consider the ‘canonical’ information-

structural characteristics of these word orders, and leave contrastive topics and focuses out of the scope of this analysis.

Let us return to the OSV sentence from Section 3.4.1 – *Mis̄iqti Bolat äkeldi* – for which it was explicated in the same section that unlike its SOV counterpart, it cannot be an all-focus sentence. The standard information-structural configurations available for an OSV sentence are: topic-comment, and topic-focus-background. The DS representation of the parsing process for both of these configurations for an SOV sentence was presented in the preceding section, and it is very similar for an OSV sentence, as the trees below demonstrate.

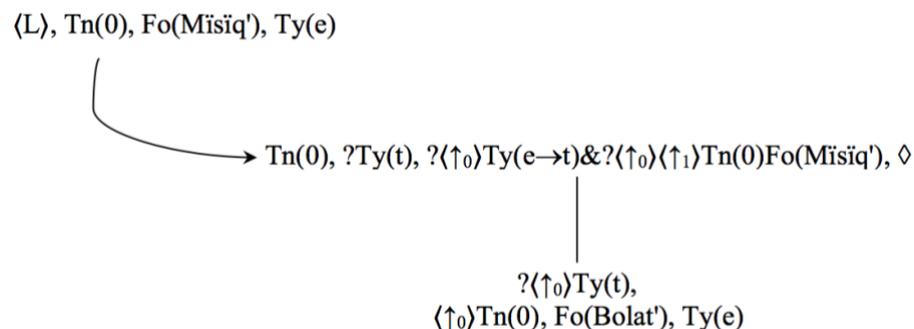
Beginning with the topic-comment information structure for the sentence *Mis̄iqti Bolat äkeldi*, the parsing process starts off with the construction of a Linked structure, as shown in (110) below.

(110) Parsing [*Mis̄iqti*]_{Topic}



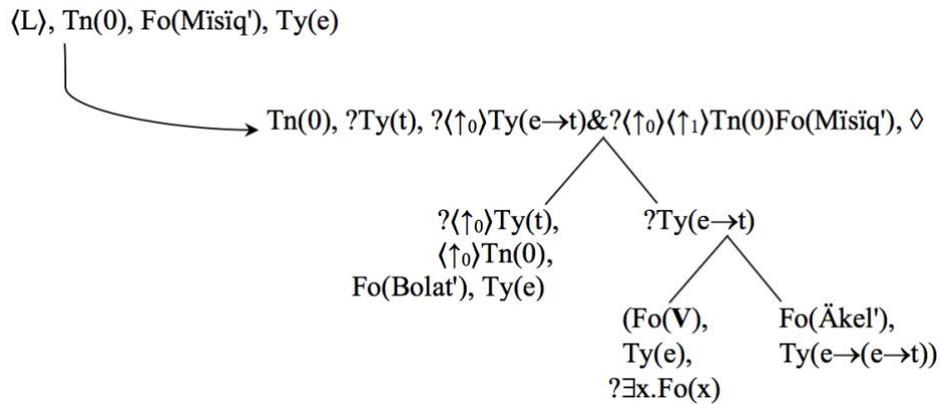
Note, that thanks to the accusative case marker, the requirement for the copy of the formula from the Linked structure to be found on a node of the developing tree structure became more specific: the copy of $Fo(Mis̄iq')$ must be located on the direct object node of the tree under development. Next, the subject *Bolat* is parsed:

(111) Parsing [*Mis̄iqti*]_{Topic} [*Bolat...*]_{Comment}



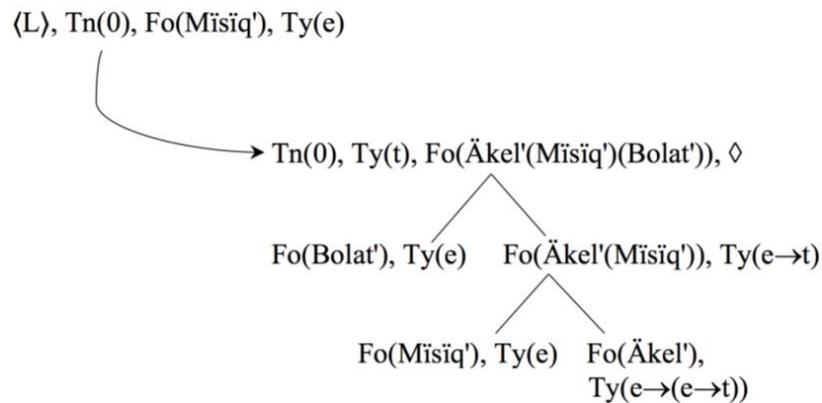
Finally, the verb *äkeldi* is parsed; as previously, it projects the full argument structure, out of which the subject node is already present in the tree, and the direct object is node is constructed afresh and decorated with a metavariable and a requirement for a semantic formula of type *e*, as shown in (112) below.

(112) Parsing [*Misïqti*]_{Topic} [*Bolat äkeldi*]_{Comment}



The final tree, in which all the requirements have been satisfied and all the information has been compiled up to the root node, is given in (113) below.

(113) Parsing [*Misïqti*]_{Topic} [*Bolat äkeldi*]_{Comment} – final

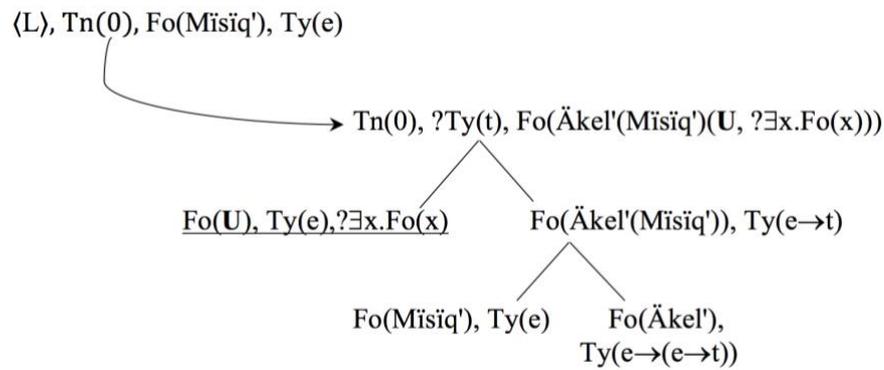


As can be observed, this tree is almost identical to that from (105) for the string [*Bolat*]_{Topic} [*misïqti äkeldi*]_{Comment}, apart from the topical element that is parsed on to

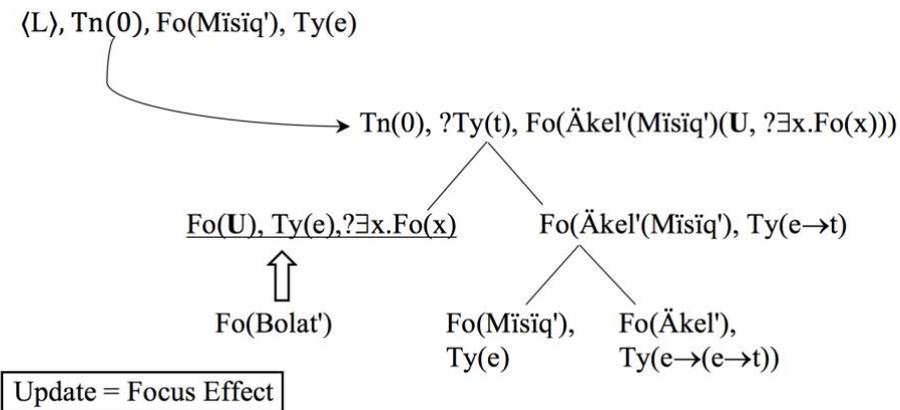
a Linked node. The root node decorations are the same in these trees, which, again, points to the fact that the strings they represent are semantically equal, despite diverse constituent ordering.

I propose that the tree for the OSV string informationally structured as $[Misiqti]_{\text{Topic}} [Bolat]_{\text{Focus}} [äkeldi]_{\text{Background}}$ is partially built up by the question (*Who brought the cat?*) in the same manner as proposed for the string $[Bolat]_{\text{Topic}} [misiqti]_{\text{Focus}} [äkeldi]_{\text{Background}}$. The partial tree structure shown in (114) is already present in the CG, and the update provided by *Bolat* creates the focus effect, as illustrated in (115).

(114) Parsing $[Misiqti]_{\text{Topic}} [Bolat]_{\text{Focus}} [äkeldi]_{\text{Background}}$ – context

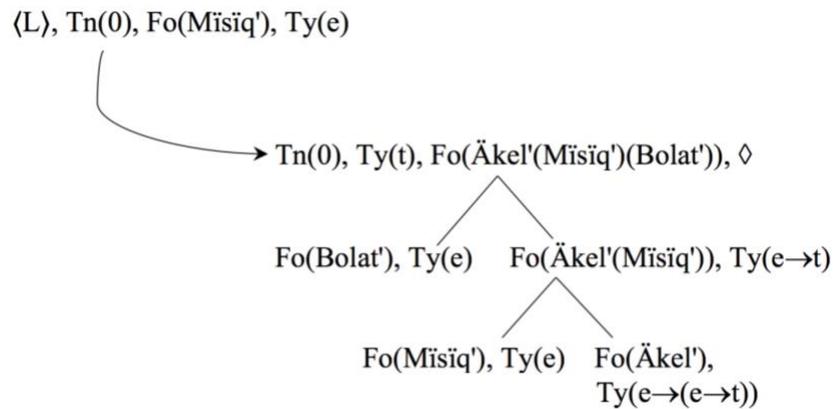


(115) Parsing $[Misiqti]_{\text{Topic}} [Bolat]_{\text{Focus}} [äkeldi]_{\text{Background}}$ – update



After all the updates take place and all the information is compiled up, the final tree is identical to that in (113) above:

(116) Parsing [*Mis̄iqti*]_{Topic} [*Bolat*]_{Focus} [*äkeldi*]_{Background} – final



Summing up, this section has demonstrated, that the differences in the possible information-structural configurations of an OSV string can be formally represented with the help of DS semantic trees and their incremental development within context. No recourse has been made to additional levels of grammar or to supplementary information-structural features to help account for these differences.

It has also been shown that despite the different ways in which the parsing processes unfold, the final semantic trees are identical. Moreover, the root nodes of the final trees from this section are the same as the root nodes of the final trees presented in the previous section, which, again, clearly signals their semantic and truth-conditional equivalence.

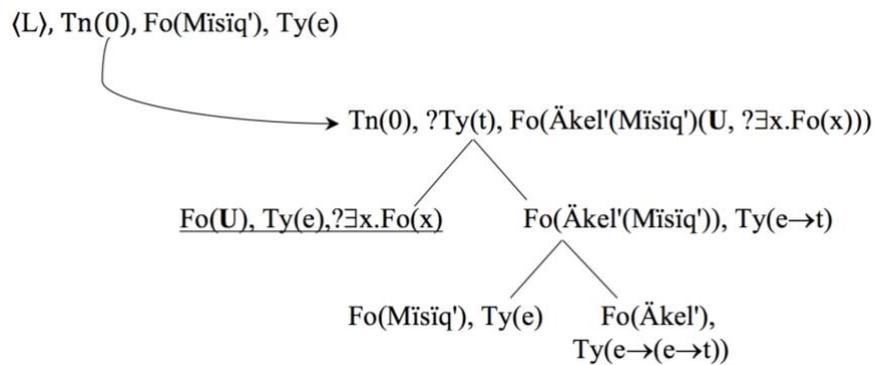
3.5.4. Verb-medial and Verb-initial Word orders

In this section the DS approach to verb-medial, and verb-initial word orders is presented. DS sketches showing the parsing process of one verb-medial and one verb-initial sentence are given with a view to demonstrating how the information-structural effects most associated with these word orders are achieved.

Recall, that typically, in verb-medial sentences – SVO and OVS – the argument preceding the verb is focused, the verb is backgrounded, and the post-verbal argument is topical. Recall also, that these sentences would only appear in very restricted contexts; for instance, the SVO sentence *Bolat äkeldi misiqti* would be felicitous in the context of the question *Who brought the cat?*, similarly to the OSV sentence considered in the preceding section.

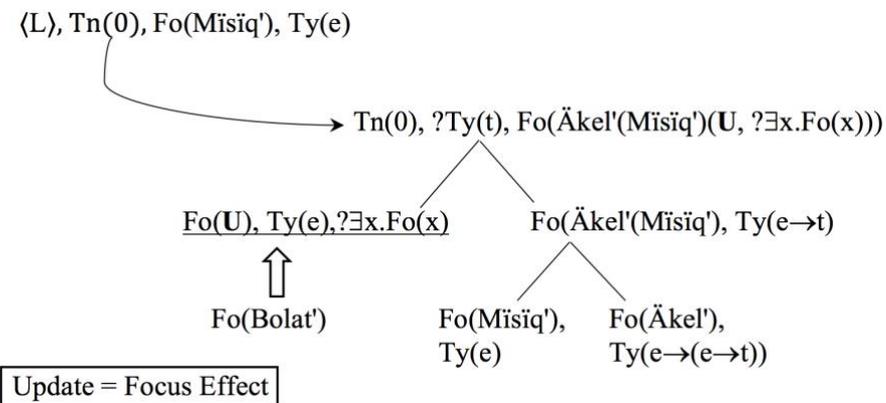
I propose, then, that in the similar way as was previously shown for an OSV sentence with narrow focus on the subject, the SVO string *Bolat äkeldi misiqti* gets parsed in the context of partially constructed tree as shown in (117) below.

(117) Parsing [*Bolat*]_{Focus} [*äkeldi*]_{Background} [*misiqti*]_{Topic} – context



When *Bolat* is uttered, it provides an immediate update to the ‘missing’ information in this tree, as shown in (118).

(118) Parsing [*Bolat*]_{Focus} [*äkeldi*]_{Background} [*misiqti*]_{Topic} – update

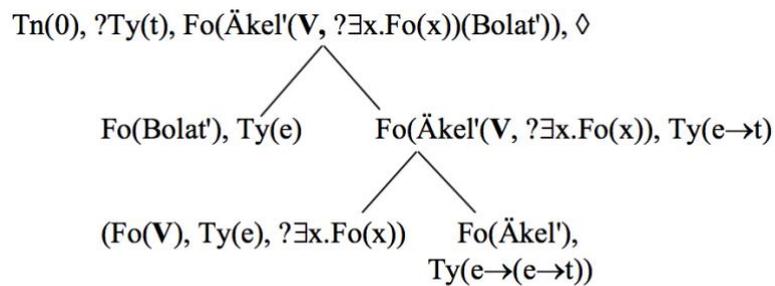


When the rest of the string – *äkeldi misiqti* – is uttered, the nodes contrasted by it seamlessly collapse with the nodes already present in the structure, thus creating the effect of backgrounding.

The appearance of the topical expression *misiqti* post-verbally does not result in the construction of a Linked structure, since it is already present in the tree, but creates the effect of an afterthought. The final tree for this string is identical to that given for the OSV string in (116) above. The parsing process develops in the same way for an OVS sentence, but with the direct object being the update/focus, and the subject being the topic/afterthought.

If we were to sketch the parsing process for our OVS sentence without context, then the subject and verb would be parsed first, which would result in the tree as shown below:

(119) Parsing Bolat äkeldi ...

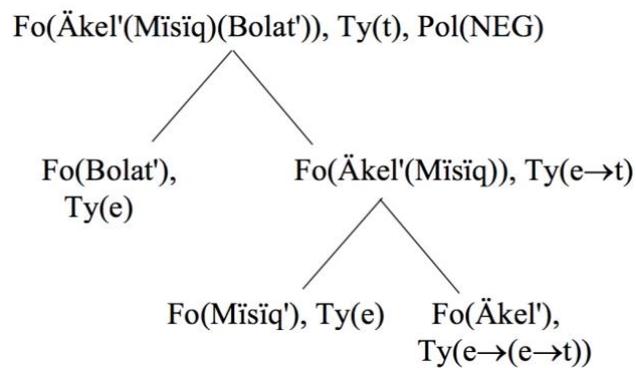


Since the string is being considered out of context, the metavariable decorating the direct object node cannot be updated at this point, which is not typical for a non-contextualised sentence: in a stand-alone sentence the parsing process is typically complete (there are no outstanding requirements in the tree) by the end of the parse of the verb (as we saw for an all-focus SOV string).

This explains the reaction of the language consultants to strings like this given out of context – they immediately wanted to change the word order to SOV; it is only within appropriate context that such strings were fully acceptable. Thus, we can see that the DS representation of the parsing process can help account not only for ungrammaticality, but also for infelicity of a string.

Moving to verb-initial strings, recall that they are highly restricted, and have the information structure of verum focus-background. Recall, also, that verum focus is a type of contrastive focus, namely, contrastive focus on the polarity of an utterance in the context of a semantically identical proposition of opposing polarity. The verb-initial string like *Äkeldi Bolat mäsıqtı* ‘Bolat DID bring the cat’, then, can only be uttered as a response to the negated proposition *Bolat mäsıqtı äkelmedi* ‘Bolat did not bring the cat’; the simplified DS tree for the latter is shown in (120).

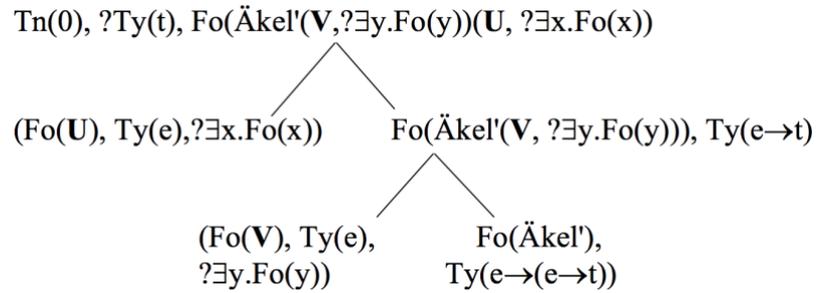
(120) Parsing Bolat mäsıqtı äkelmedi



Following Lucas (2014), a simple label *Pol(NEG)* is added to the root node to indicate the negative polarity of the string under consideration. When the verb-initial sentence *Äkeldi Bolat mäsıqtı* is uttered in the context of the tree above, the affirmative polarity of the verb *äkeldi* pronounced first contrasts with the negative polarity of the tree in (120), and creates the strong verum focus effect. When the post-verbal elements are uttered, they simply collapse with the existing nodes which results in their appearing backgrounded.

Considering a verb-initial string out of context results in the verb projecting a full argument structure decorated with metavariables requiring to be updated, as shown in the tree below.

(121) Parsing sentence initial *Äkel'di...*



Having two nodes requiring formula updates after the verb has been parsed contradicts the usual development of the parsing process in verb-final Kazakh, which explains why non-contextualised verb-initial strings were even less acceptable to the language consultants than non-contextualised verb-medial strings.

3.5.5. Conclusions

This section has presented the DS sketches of the parsing process development for Kazakh strings with neutral and scrambled word orders. It has been demonstrated that the differences in the information-structural configurations of diverse word orders or even of one and the same word order can be explicated through the dynamics of the contextualised parsing process. Additionally, DS semantic trees have been used to explain infelicity of some word orders when attempted to be parsed out of context.

The important observation that can be made from examining the semantic trees resulting from the parsing of sentences with diverse information structure, is that the formula at the root node of each of these trees is exactly the same, regardless of how the tree itself has been constructed. This formally confirms that the sentence with the neutral SOV word order, and the sentences with scrambled word orders are semantically and truth-conditionally equal.

3.6. Conclusions

This chapter has considered various word orders and their information-structural characteristics. Firstly, the canonical or neutral word order – SOV – has been demonstrated to be able to encode a wide range of context-dependent information-structural configurations, including sentence-focus.

The other five possible word order permutations – OSV, SVO, OVS, VSO, VOS – have also been examined. It has been shown that these scrambled word orders are pragmatically motivated, and unlike the neutral SOV word order, cannot encapsulate sentence focus; they are restricted to appearing in certain, sometimes very narrow, contexts.

It has also been demonstrated that in order to arrive at a thorough description of information-structural relations in a Kazakh sentence the notions of topic and focus must be expanded to include their contrastive variants. This provides for a more detailed and comprehensive description of information-structural characteristic of a Kazakh sentence, which go beyond the traditional limited claims for sentence initial topics and immediately preverbal focuses.

The DS analysis of the neutral and scrambled word orders has been sketched out, and has demonstrated that all the information-structural effects that might be associated with different word orders, or even with one and the same word order, can be elegantly accounted for without the recourse to special syntactic positions in a hierarchical tree, movement, or feature assignment procedures.

The diverse ways in which the parsing process unfolds for strings with different information structure has been reflected in the growth processes of the DS semantic trees. However, crucially, the same semantic formula decoration at the root nodes of the trees resulting from the parsing of strings with different word orders formally confirms the semantic equivalence of these strings.

In the next section I move to the phenomenon of differential object marking, which is directly linked with some of the word order variation restrictions mentioned throughout this section. Additionally, some linguists propose that this phenomenon is information-structurally motivated, while many others consider it to be connected with

the pragmatic notion of specificity. No description of Kazakh IS and word order would be complete without an examination of differential object marking.

4. Differential Object Marking in Kazakh

In this chapter I describe and analyse the phenomenon of differential object marking (DOM) in Kazakh, and show that the accusative case marking on a direct object in the IPV position yields a specific reading for that object's referent; unmarked direct objects have an indefinite and non-specific reading.

It is proposed that these pragmatic effects are directly linked with the way marked and unmarked direct objects are parsed, and a DS sketches of both parsing processes are presented to formalise and support this analysis.

Section 4.1 introduces the phenomenon of DOM; Section 4.2 focuses on Kazakh and Turkish DOM: previous descriptions of Kazakh DOM are amalgamated and systematised, followed by previous analyses of Turkish DOM to which I resort due to the lack of analytical work on Kazakh DOM. In Section 4.3 the DS approach to Kazakh DOM is introduced – it is shown that this phenomenon can be analysed and explicated through the DS notions of fixed and unfixed nodes. This analysis also includes an insight into the 'special' nature of the IPV position.

4.1. Introduction

The phenomenon of DOM is an instance of a wider phenomenon of differential case marking (DCM) or differential argument marking, which are also known as case alternation, and case diathesis. DCM phenomena are observed in many languages of the world, where non-uniform marking of terms is present. 'Dative alternation' in English is one of the examples of DCM.

(1) Monik gave a book to Francisca.

(2) Monik gave Francisca a book.

The goal argument in (1) is marked by the preposition *to* and realised after the theme argument, while in (2), the same goal argument precedes the theme and is unmarked.

These alternations have been observed to yield a difference in interpretation albeit quite a subtle one in the case of English.

It is argued that the prepositional variant, as in (1), involves movement of theme to goal while the double object construction, as in (2), involves a possession relationship between the goal and theme. This difference becomes more obvious with abstract or inalienable objects, and objects that cannot easily stand in a possession relation to each other; the prepositional variant involving the former and the double object variant involving the latter are highly marked (although not entirely ungrammatical in all contexts (Krifka 2004)), as the examples in (3) and (4) show.

(3) #Monik gave a bad headache to Francisca./Monik gave Francisca a bad headache.

(4) Monik sent Paris a package./ Monik sent a package to Paris.

Another example of DCM is differential subject marking (DSM), as observed in the examples from Hindi in (5) and (6), where the subject is unmarked (or is in nominative case), and marked with ergative case respectively.

(5) ram khās-a
Ram.M.NOM cough-PFV.M.SG
'Ram coughed.'

(6) ram=ne khās-a
Ram.M=ERG cough-PFV.M.SG
'Ram coughed (purposefully).'

(Tuite et al. 1985:264)

As can be seen from the English approximations, the unmarked (or nominative) form is neutral in interpretation, while the ergative marking signals agentivity of the subject – this is what de Hoop and Malchukov (2007) refer to as 'fluid DCM in Hindi'.

Moving to DOM, Bossong (1982, 1985) is attributed with coining this term; three types of this phenomenon have been identified in the literature: direct objects marked with different cases; presence/absence of agreement with direct objects on the verb; and the unmarked/marked direct objects alternations. Witzlack-Makarevich and Seržant (2018) label the first type of DOM symmetric, and the latter two – asymmetric.

The first type of differential object marking can be illustrated by the examples from Estonian (Finnic), in which the alternation between genitive and partitive marking on the direct object indicates either the degree of affectedness of the object by the verbal action or the grammatical aspect of that action, as shown in (7) and (8) respectively.

(7) Poiss sõi supi (ära).
 boy.NOM.SG eat.PST.3SG soup.GEN.SG up
 ‘The boy ate the soup (up).’

(8) Poiss sõi suppi.
 boy.NOM.SG eat.PST.3SG soup.PRT.SG
 i. ‘The boy was eating soup.’
 ii. ‘The boy ate (some) soup.’

(Cann and Miljan 2012: 587)

This type of DOM is also observed in other Finnic languages, as well as in Russian (Slavic); an accusative-dative case marking alternation is observed in Icelandic (Germanic). Dalrymple and Nikolaeva (2011: 19) refer to this type of DOM as ‘variable marking of objects’ and exclude it from their comprehensive cross-linguistic study of DOM. The authors only consider languages with overt presence or absence of grammatical marking – either case or agreement – on direct objects. These are the types of DOM I turn to now.

DOM in the form of differential object agreement is observed in Palauan (Austronesian), where the direct object either does or does not trigger agreement on the same verb, as demonstrated in the examples in (9) and (10) respectively. According

- (13) Ilaa=ne bacce=ko (*baccaa) uTaayaa
 Ila=ERG child=ACC child.NOM lift.PFV
 ‘Ila lifted a/the child.’

(Mohanani 1994: 104)

Let us now consider some examples of DOM from Turkish, which is very similar to DOM in Kazakh, especially in that an unmarked direct object can only appear in the IPV position, and is ungrammatical in all other positions, as the examples in (14) and (15) demonstrate.

- (14) Murat kitap ok-uyor.
 Murat book read-PROG
 ‘Murat is reading a book. (Murat is book-reading.)’

- (15) *Kitap Murat ok-uyor.

(Erguvanli 1984: 21)

An accusatively-marked direct object is able to appear both in the IPV position, and anywhere else in a sentence, as the examples in (16) and (17) illustrate.

- (16) Murat kitap-ı ok-uyor.
 Murat book-ACC read-PROG
 ‘Murat is reading the book.’

- (17) Kitab-ı Murat ok-uyor.

(Erguvanli 1984: 22)

In (16) and (17) the accusatively marked direct object receives a definite interpretation, however, as the example in (18) shows, an indefinite direct object can

also be accusatively marked. In the example below the direct object is preceded by the indefinite article *bir* (from the numeral *bir* ‘one’)

- (18) Ben bir kitab-ı oku-du-m.
1SG a book-ACC read-PST-1SG
‘I read a certain book.’

The examples such as (18) led Lewis (1967), Erguvanlı (1984), Kornfilt (1997), Enç (1991), Aydemir (2004), amongst many others, to conclude that the accusative case suffix expresses specificity rather than definiteness – recall the discussion on the distinction between these two notions in sub-Section 2.1.4 of Chapter 2. Von Heusinger and Kornfilt (2005) conclude, however, that specificity is the determining semantic property in Turkish DOM. I return to the proposed analyses of Turkish DOM in the next section, after presenting previous descriptions of Kazakh DOM, to which I now move.

4.2. Differential Object Marking in Kazakh

As mentioned before, DOM in Kazakh is of the same type as in Turkish: it is asymmetric, and a direct object is either unmarked or marked with the accusative case, as in (19) and (20).

- (19) Bolat kitap satıp al-dı.
Bolat book buy-PST(3)
‘Bolat bought a book.’

- (20) Bolat kitap-tı satıp al-dı.
Bolat book-ACC buy-PST(3)
‘Bolat bought the book.’

Similarly to Turkish, at first glance it might seem that the accusative case marking is to do with definiteness, however, like in Turkish, an indefinite direct object can also be case-marked, as in (21) below.

- (21) Men bir kitap-tı satıp al-dı-m.
 1SG a book-ACC buy-PST-1SG
 ‘I bought a certain book.’

The indefinite specific interpretation the direct object receives in the example above points to specificity being the crucial factor for DOM in Kazakh.

As mentioned in the preceding sections on word order, an accusatively marked direct object is not restricted to the immediately preverbal position, and can be placed anywhere in the preverbal area, as in (22) where it is sentence-initial, and can also appear post-verbally, as in (23).

- (22) Kitap-tı Bolat satıp al-dı.
 Book-ACC Bolat buy-PST(3)
 ‘Bolat bought the book.’ (or: ‘As for the book, Bolat bought it.’)

- (23) Bolat satıp al-dı kitap-tı.
 Bolat buy-PST(3) book-ACC
 ‘Bolat bought the book.’ (or: ‘Bolat bought it, the book.’)

As demonstrated in preceding chapter, these word order variations are pragmatically motivated and marked, which means that they are only felicitous in certain contexts.

An unmarked direct object can appear only in one other sentence type apart from the SOV sentence-type – OVS, since this is the only other word order where the direct object appears in the immediately preverbal position, as in (24).

- (24) Kitap satıp al-dı Bolat.
 book buy-PST(3) Bolat
 ‘Bolat bought a book.’

The examples in (25)-(28) show ungrammaticality of unmarked direct objects appearing anywhere other than in the IPV position.

(25) *Kitap Bolat satıp aldï.

(26) *Bolat satıp aldï kitap.

(27) *Bolat kitap keşe satıp aldï.
Bolat book yesterday buy-PST(3)
Intended: ‘Bolat bought a book yesterday.’

(28) *Bolat kitap düken-nen satıp aldï.
Bolat book shop-ABL buy-PST(3)
Intended: ‘Bolat bought a book in the shop.’

Traditionally, in Kazakh grammars of early and mid 20th century written in the USSR (in Russian), marked and unmarked direct objects are referred to as *oformlennie* (literally: ‘framed’ or ‘decorated’), and *neoformlennie* (‘unframed’ or ‘undecorated’) respectively. Balakaev’s 1959 tome –*Sovremennyj kazakhskij jazyk – sintaksis* ‘Modern Kazakh Language – Syntax’ – provides the most detailed description of DOM in the literature of that era, and presents a functional approach to analysing the phenomenon.

Balakaev (1959) states that all explanations of differential object marking in Turkic languages proposed by both Soviet and pre-Soviet linguists point to DOM being dependent on the “grammatical definiteness or indefiniteness of the direct object’s referent. The speaker’s awareness/knowledge or unawareness/lack of knowledge of the referent are the determinant features of definiteness and indefiniteness respectively” (1959: 162).

However, Balakaev does not support this point of view and proposes a functional approach to DOM. He separates nouns into ‘potential objects’ and

‘potential subjects’, where the former respond to question ‘*What?*’, and the latter to ‘*Who?*’. According to him, then, the ‘potential objects’ can be either marked or unmarked when functioning as direct objects, while the nominals that are not ‘potential objects’ have to be marked. Essentially, he claims that all animate direct objects must be marked as they are too similar to typical subjects.

Balakaev provides an extensive list of the direct objects that must be accusatively-marked which I present below in almost its entirety for descriptive purposes. I rephrase and expand on some of the points and use citation marks to indicate direct quotations; the examples are from Balakaev, unless stated otherwise. According to Balakaev (1959), the following categories must be marked with accusative case:

1. People’s proper names when functioning as a direct object, as in (29).

(29) [Men] Baršagül-di šaqirt-qan-min.
 1SG Barshagül-ACC invite-PFV-1SG
 ‘I invited Barshagül.’

2. Personal and demonstrative pronouns when functioning as a direct object, as in (30) below. Pronouns have special case forms which differ from their unmarked or bare form (see Section 1.1 of Chapter 1), and the accusative case form must be used in this instance; the unmarked form is ungrammatical.

(30) Men sağan sonı / *sol ğana ajt-qı-m
 1SG 2SG.DAT this.ACC this only say-NMLZ-1SG
 keledi.
 want
 ‘I only want to say this to you.’

3. Substantivised verbal and adjectival elements when functioning as a direct object. According to Balakaev, “[A]ll non-nominals (adjectives, numerals, participles, etc.) are similar to demonstrative pronouns in that they are ‘potential modifiers’, according to their lexical meanings and syntactic functions. To make them perform the function of a direct object they must be used in an appropriate grammatical form of a direct object, which is the accusative case. Otherwise these non-nominals might be mistaken for modifiers or become a part of an incomprehensible combination of words” (1959: 164).

The example in (31) contains a substantivised adjective *asau* ‘wild’ which cannot remain unmarked in this sentence.

- (31) Qazaq bala-si asau-di/*asau üjret-e
 Kazakh guy-POSS.3 wild-ACC tame-CVB
 bil-e-di.
 know-PRS-3
 ‘Kazakh guys know how to tame a wild horse.’

4. Direct objects expressed by nouns with possessive affixes, as in (32).

- (32) Nikolaj bas-i-n/*bas-i ize-di.
 Nikolay head-POSS.3-ACC nod-PST(3)
 ‘Nikolay nodded his head.’

5. “Nouns with locative or temporal modifiers, which may be expressed by demonstrative pronouns, or adjectives formed with suffixes *-ǵı/-gi*, *-qı/-ki*” (Balakaev 1959: 164).

The example in (33) shows that the direct object modified by a demonstrative pronoun *mına* must be accusatively marked; omitting the case-marking results in ungrammaticality. In (34) the direct object is

modified by an adjective formed from the noun in locative case with the help of the suffix *-gi* and must be marked with the accusative case.

- (33) Men mīna kitap-tī / *kitap oq-i-mīn.
 1SG this book-ACC read-FUT-1SG
 ‘I’ll read this book.’

- (34) Ana-m qol-in-da-ğī kitap-tī / *kitap
 mother-POSS.1SG hand-POSS.3-LOC-ADJ book-ACC
 aš-tī.
 open-PST(3)
 ‘(My) Mother opened the book that was in her hands.’

6. Direct objects expressed by the same word as the subject, as in (35).

- (35) Adam adam-dī / *adam sījla-j-dī.
 human human-ACC respect-PRS-3
 ‘A human respects a human.’

7. Direct objects expressed by nominals with reduplicated roots, as in (36).

- (36) Añši tau-tau-dī kez-ip žūr.
 hunter mountain-mountain-ACC wonder-CVB AUX(3)
 ‘The hunter wonders all over the mountains.’

Contrary to Balakaev’s functional approach to DOM, Muhamedowa (2016) states that the accusative case suffix “indicates whether the direct object is definite or non-definite. The definite direct object must bear the accusative suffix” (2016: 228). While her list of obligatorily case-marked direct objects is not as extensive as Balakaev’s, it includes some of the items not mentioned in the latter.

According to Muhamedowa, then, the accusative suffix is obligatory with: 1) headless relative clauses, as in (37) below; 2) complement clauses containing the

verbal noun ending with *-(U)w*,⁹ as in (38); 3) definite objects, as in previously given examples; and 4) noun phrases containing the possessive suffix, as in (32) above.

(37) *Kör-gen-i-m-di* / **körgenim ajt-ti-m*.
 see-PTCP-POSS-1SG-ACC say-PST-1SG
 ‘I told what I saw.’

(38) *Obama kelissöz-der-di qajta basta-w-di* / **bastaw*
Obama negotiation-PL-ACC again begin-VN-ACC
sura-di.
 ask-PST(3)
 ‘Obama asked to begin with negotiations again.’

(Muhamedowa 2016: 95)

Muhamedowa also remarks that the accusative suffix is obligatory when indefinite nouns are topicalised, and provides the examples below for comparison. In (39) the direct object *araq* is in the immediately preverbal position and is unmarked.

(39) *Sağat eki-den basta-p araq iše basta-di-q*.
 hour two-ABL begin-CVB vodka drink begin-PST-1PL
 ‘We began drinking vodka from two o’clock.’

(40) *Araq-ti sağat ekiden bastap iše*
vodka-ACC hour two-ABL begin-CVB drink
basta-di-q.
 begin- PST-1PL
 ‘We began drinking vodka from two o’clock.’

(Muhamedowa 2016: 11)

⁹ What Muhamedowa refers to as ‘verbal noun’ is essentially a substantivised infinitive.

Note that although Muhamedowa talks about topicalisation, her English interpretations for both sentences are the same. I would argue, however, that the example in (40) should be translated as: ‘As for vodka, we started drinking it from two o’clock’.

It must be highlighted that despite her original statement linking accusative case marking with definiteness of a direct object, Muhamedowa’s list of obligatorily case-marked direct objects is predominantly based on grammatical rather than pragmatic features. This casts doubt on her conclusions as regards to the function of the accusative case marker.

Some further additions can be made to Balakaev’s and Muhamedowa’s lists. For instance, *wh*-word *kim* ‘who’ must always be accusatively marked when it functions as a direct object, as in (41).

- (41) Sen kim-di kör-di?
 2SG who-ACC see-PST(3)
 ‘Whom did you see?’

Wh-word *ne* ‘what’ can be marked or unmarked, in the same way as inanimate, non-human direct objects can be. There is, however, a significant difference in the use of the unmarked *ne*, and the marked *neni*.

- (42) Ne žaz-ip žatir-siñ?
 what write-CVB AUX-2SG
 ‘What are you writing?’

- (43) Ne-ni žaz-ip žatir-siñ?
 what-ACC write-CVB AUX-2SG
 ‘What (exactly/precisely/specifically) are you writing?’

As evident from the interpretations of (42) and (43), the question with the unmarked variant *ne* is a request for new information, and the questioner does not have any

presuppositions for or expectations from the response; the use of the marked variant *neni* implies that the questioner expects the response to fall within a specified set (perhaps the questioner knows that her interlocutor had plans to write an essay and a letter, and she is now finding out which piece exactly is being worked on). Göksel and Kerslake (2005) make similar observations for Turkish.

Keeping the notions of definiteness and specificity aside, what Balakaev's and Muhamedowa's lists primarily show is that the accusative case-marking is necessary and obligatory when the direct objecthood of the element functioning as a direct object needs to be overtly confirmed. In a way, this brings us back to the functional approach proposed by Balakaev. However, I do not claim that a direct object needs to be accusatively marked in order to be distinguished from the subject, as by and large, this is not the case in all instances of obligatory marking of a direct object, as many of the examples provided above show.

As for the association of accusatively marked direct objects with the notions of definiteness and specificity, I posit that it is the latter that is most pertinent of the case-marked direct objects (especially those in the immediately preverbal position). However, as mentioned above, specificity should not be considered to be the 'meaning' of the accusative case. The examples below prove this point by demonstrating that when one and the same specific referent functions as the direct object, it obligatorily receives the accusative case marker, however, when it functions as the subject, it cannot be accusatively case marked.

(44) Men bul kitap-ti /* kitap ötken žili žaz-di-m.
 1SG this book-ACC book past year write-PST-1SG
 'I wrote this book last year.'

(45) Bul kitap / *kitap-ti ötken žili žaz-il-di.
 This book book-ACC past year write-PASS-PST(3)
 'This book was written last year.'

Despite appearing to be trivial and obvious, these examples provide an additional piece of evidence against attributing any semantic or pragmatic meanings to the accusative case marker *-DI*. Specificity is simply the pragmatic effect of accusative case marking on the direct object. That is to say, it is not the case that specific direct objects receive accusative marking, but that accusatively marked direct objects receive a specific reading.

Now that Kazakh DOM has been sufficiently described, I move to its analysis. Due to lack of thorough analytical work on Kazakh DOM, in the following sub-section I consider the main questions examined in the previous analyses of Turkish DOM.

4.2.1. Previous Analyses of Differential Object Marking in Turkish

As mentioned earlier, since no analytical work has been published on DOM in Kazakh, in this sub-section I consider previous analyses of DOM in Turkish. Most of the observations made for Turkish DOM are also applicable for Kazakh DOM, and provide a suitable basis for developing my analysis for the latter.

As far as DOM in Turkish is concerned, the discussions in the literature focus on two main questions: 1) the syntactic status of marked vs. unmarked direct objects; and 2) the difference in the semantics and pragmatics of marked and unmarked direct objects. I briefly touched upon some of the proposed answers to the second question in the preceding sections of this chapter – this will be expanded upon in this sub-section. I begin, however, with considering the first question.

The main discussion regarding the difference in the syntactic status of marked and unmarked direct objects in Turkish revolves around the concept of object incorporation. Generally speaking, incorporation is the compounding of a word (usually a verb or a preposition) with another element (usually a noun, pronoun or an adverb). The resulting compound “serves the combined syntactic function of both elements” (Gerds 1998: 84). In object incorporation specifically, the verb forms a compound with its direct object, while preserving its syntactic function; the direct object appears inside the verb phrase, as opposed to forming an independent constituent (Baker 1988).

It is posited for Turkish that unmarked direct objects are incorporated, and positioned within the VP; while marked, non-incorporated direct objects are VP-external. Kornfilt (1997, 2003), Gerdts (1998), Kuribayashi (1989), Mithun (1984), Erguvanlı (1984), Aydemir (2004), amongst others maintain that unmarked direct objects in Turkish are incorporated, while Kornfilt (1984), Orgun and Inkelas (2004), amongst others are against this view.

Let us turn to some of the evidence used to support the incorporation analysis of unmarked direct objects. The incorporation claims are mainly based on the lack of accusative case marking, and on the observation that these unmarked objects are confined to the IPV position, as examples given below demonstrate.

(46) Ahmet dün akşam pasta-yı ye-di.
 Ahmet yesterday evening cake-ACC eat-PST(3)
 ‘Yesterday evening, Ahmet ate the cake.’

(47) Ahmet pasta-yı dün akşam
 Ahmet cake-ACC yesterday evening
 ye-di.
 eat-PST(3)
 ‘Ahmet ate the cake yesterday evening.’

(48) Ahmet dün akşam pasta ye-di.
 Ahmet yesterday evening cake eat-PST(3)
 ‘Yesterday evening, Ahmet ate (a) cake.’

(49) *Ahmet pasta dün akşam ye-di.
 Ahmet cake yesterday evening eat-PST(3)
 Intended: ‘Ahmet ate (a) cake yesterday evening.’

(Kornfilt 2003: 127-128)

Examples (46) and (47) show that accusatively marked direct objects are not restricted to appearing immediately before the verb, and can easily move out of this position – subject to information-structural constraints, similar to those described in the previous chapter. The unmarked direct object in (48), however, is unable to move out of the IPV position, as illustrated by ungrammaticality of (49).

According to Kornfilt (2003), the accusative case marker is either overtly realised on specific direct objects or dropped from the non-specific ones, although she also states that she does not “imply that there is an actual rule of morpheme deletion” (2003: 127). The claim seems to be, then, that the accusative case has two realisations: overt, via case-markers on direct objects, and covert or null, which Kornfilt (2003) labels the “accusative drop”. Kornfilt (2003) arrives at the generalisation that non-specific direct objects with “dropped” structural case are confined to the immediate left of the verb, while their specific counterparts with overt case marking are not.

Kornfilt (2003) takes this generalisation further, and applies it to the subjects of non-nominalized, tensed clauses, which, according to her, can have “strong” or “weak” nominative case, despite there being no corresponding morpheme (like in Kazakh). The examples the author provides to support this statement are given in (50)-(52) below.

(50) (Bir) arı çocuğ-u sok-tu.
 a bee child-ACC sting-PST(3)
 ‘A/the bee [specific] stung the child.’

(51) Çocuğ-u arı sok-tu.
 child-ACC bee sting-PST(3)
 ‘Bees stung the child.’

(52) Çocuğ-u bir arı sok-tu.
 child-ACC a bee sting-PST(3)
 ‘A bee [non-specific] stung the child.’

(Kornfilt 2003: 128-129)

The example in (50) shows a specific (definite or indefinite) nominative subject in canonical, clause-initial position; in (51) the subject is non-specific, generic, and in the IPV position; while in (52) the pre-verbal subject is non-specific and indefinite.

It must be noted, however, that all these readings are highly context dependent, and the so-called “strong” reading in (50) can easily be interpreted as non-specific and indefinite if it were to appear as an opening sentence of a story, for example. The interpretation then would be: ‘Some bee, a bee (non-specific) stung the child. The child started crying and ran to his mother...’.

An alternative interpretation of the subject in (51) is also available in the situation where the sentence is uttered in response to the question *Why is the child crying?*, while the interlocutors are standing in the proximity of some beehives. The interpretation of (51) can then be *The bees stung the child*. It must be noted, however, that Kornfilt’s plural interpretation of *arı* in this sentence is not entirely clear; the expected plural form is *arılar*.

The non-specific and indefinite interpretation of (52) is also highly context-dependent, as well as reliant on the general facts about the world, since we do not tend to identify a specific bee out of a swarm or a hive, in the same way as we are unlikely to identify an individual ant, for instance, unless we are holding it in our hand, or are pointing to it. It is thus not surprising that the noun phrase *bir arı* has a non-specific reading. Changing the indefinite *bir* for the demonstrative *bu*, would make the same “weak” nominative ‘bee’ definite and specific, as shown in (53) below; minimal context is provided in square brackets.

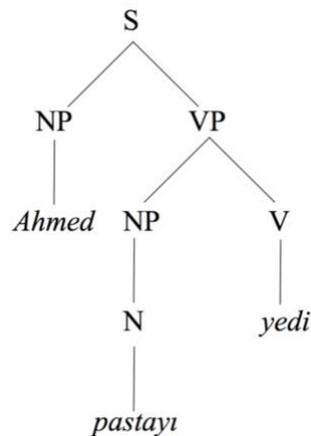
- (53) [‘What happened to the child, why is he crying?’]
 Çocuğ-u bu arı sok-tu.
 child-ACC this bee sting-PST(3)
 ‘This bee stung the child.’

It is also noteworthy that the immediately preverbal position in Turkish is strongly associated with the focal position, which means that ‘new’, non-specific and indefinite referents are more likely to appear in this position. This does not appear to

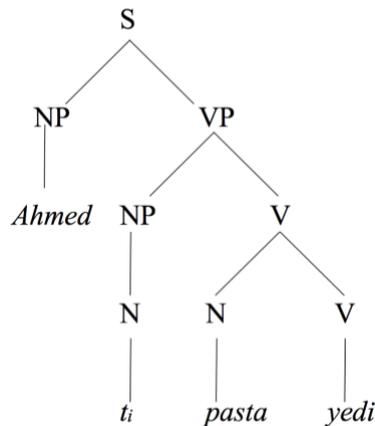
be taken into consideration in Kornfilt's proposal for 'strong' and 'weak' nominative case.

Returning to the syntactic position of direct objects, the difference between the non-incorporated and incorporated direct objects is traditionally represented as shown in (54) and (55) respectively, simplified after Baker (1988: 80).

(54) Non-incorporated direct object



(55) Incorporated Direct Object

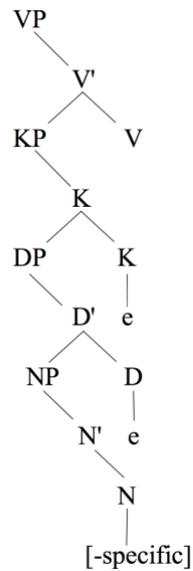


As can be seen from the structures above, the assumption is that unmarked, incorporated direct objects are structurally much closer to the verb than marked, non-incorporated direct objects, which form an independent syntactic phrase. In (55), where the direct object is assumed to be incorporated, the moved noun root leaves a

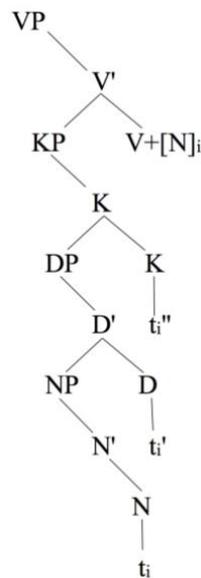
trace (*t_i*) which heads a direct object phrase that receives a theta role from the verb thus satisfying the verb's subcategorisation requirements.

Kornfilt (2003) develops this analysis by assuming that DPs are in fact embedded within Case Phrases (KPs), as shown in the structures below from (Kornfilt 2003: 143).

(56) Before incorporation



(57) After incorporation



Where the K-head is phonologically filled with an overt case-marker, the N-head is unable to move into the V, whereas if the K-head position is empty, the N can move into that position and then further into the V.

As mentioned earlier, there is no consensus on the incorporation analysis, and Orgun and Inkelas (2004) present phonological, lexical, and syntactic evidence against this view. Following Gerdts (1998), the authors state that although theoretical approaches to incorporation might differ in many ways, one central assumption is the same – the structure resulting from noun incorporation is a (compound) word.

From the phonological point of view compounds typically stress their first member, like in the examples in (58), where stress is indicated by acute accent.

- (58) a) babá-anne – literal meaning: ‘father-mother’, compound
 meaning: ‘paternal grandmother’
 b) kará-tahta – literal meaning: ‘black-board’, compound
 meaning: ‘blackboard’

(Orgun and Inkelas 2004: 266)

At first blush the unmarked DO-verb sequences seem to have the same stress pattern as the lexical compounds above, which differs from the stress pattern with case-marked DOs, as in (59) and (60).

- (59) Zeynep kitab-ı oku-dú.
 Zeynep book-ACC read-PST(3)
 ‘Zeynep read the book.’

- (60) Zeynep kitap oku-du.
 Zeynep book read-PST(3)
 ‘Zeynep engaged in book-reading.’

(Orgun and Inkelas 2004: 266)

Orgun and Inkelas demonstrate, however, that the stress patterns shown in (59) and (60) can change under the influence of information-structural effects, as shown in (61) and (62); recall the same observations for Kazakh in Chapter 3.

(61) Zeynep kitabı oku-du.
 ‘Zeynep read the BOOK (not something else).’

(62) Zeynep kitap oku-dú.
 ‘Zeynep DID engage in book-reading.’

(Orgun and Inkelas 2004: 267)

The change of stress placement in (61) and (62), as compared to their pragmatically neutral counterparts in (59) and (60) respectively, indicates contrastive focus on the direct object in (61), and verum focus in (62), reflected in the English interpretations. Orgun and Inkelas observe that such pragmatically determined change in the stress pattern is not characteristic of lexically-determined word stress, which points to phrasal rather than incorporated nature of unmarked DO-V combinations.

Lexical integrity is another criterion of word-hood considered by Orgun and Inkelas in relation to the direct object incorporation analysis. By lexical integrity they mean the impossibility of another syntactic element to intrude within the word, or of syntactic permutation of word-internal elements (for example, pre- or post-posing one morpheme). According to the incorporation analysis, unmarked DO+V sequences should exhibit lexical integrity, however, Orgun and Inkelas show that they do not, since a wide range of syntactic elements can intervene between an unmarked direct object and a verb. Amongst such elements are: question particle *mI*, particle *DA* ‘too’, particles like *bile* ‘even’, certain adverbial clitics, and reduplicated stems. The examples in (63)-(65) below demonstrate this point.

(63) Gül bugün kitap mı oku-du?
 Gül today book Q read-PST(3)
 ‘Is it a book/books that Gül read today?’

(64) Mustafa kitap bile oku-r.
 Mustafa book even read-IPFV
 ‘Mustafa reads even books.’

(65) Ali kitap falan oku-maz.
 Ali book etcetera read-NEG-IPFV
 ‘Ali does not read books and the like.’

(Orgun and Inkelas 2004: 269)

As can be seen from the examples above, particles and full lexical words can sometimes intervene between an unmarked direct object and a verb. True compounds, on the other hand, do not allow this, as shown in the examples in (66), where insertion of the same elements as shown in (63)-(65) between the elements of the compound *baş-makan* ‘head-minister = prime minister’ results in ungrammaticality.

(66) a. *baş mı bakan
 b. *baş bile bakan
 c. *baş falan bakan

Orgun and Inkelas additionally demonstrate that syntactic permutation motivated by pragmatic factors is possible in the sentences with unmarked direct objects, as shown in the example in (67). Recall a similar, imperative Kazakh example (83) from the preceding section.

(67) Hocam, yaz-sana bana bir ‘S’ be!
 teacher write-IMP 1SG.DT a letter.S EXCL
 ‘Teacher, write me an ‘S’ already!’

(Orgun and Inkelas 2004: 269)

The non-case-marked direct object ‘letter S’ not only follows the verb, but is also separated from it by the postponed indirect object pronoun in Dative case. Orgun and Inkelas re-iterate that true compound words cannot be so permuted.

The final piece of evidence against the incorporation analysis presented by Orgun and Inkelas is to do with possible syntactic complexity of an unmarked direct object. The authors observe that according to the incorporation hypothesis the direct object should be a stem or word, but not a syntactic phrase. However, this is not the case since non-case-marked direct objects can be syntactically complex, contain determiners, quantifiers, adjectives, relative clauses, or a combination of these elements. The examples in (68)-(72) below illustrate this point.

(68) Ali şiir yaz-dı.
 Ali poem write-PST(3)
 ‘Ali wrote poems (engaged in poem-writing).’

(69) Ali bir şiir yaz-dı.
 Ali a poem write-PST(3)
 ‘Ali wrote a poem.’

(70) Ali üç şiir yaz-dı.
 Ali three poem write-PST(3)
 ‘Ali wrote three poems.’

(71) Emine kırmızıelma ist-iyor.
 Emine red apple want-PROG
 ‘Emine wants red apples.’

(72) Emine organik yetiş-tir-il-miş ol-an
 Emine organic grow-CAUS-PASS-PFV AUX-PTCP
 elma ist-iyor.
 apple want-PROG

‘Emine wants organically grown apples.’

(Orgun and Inkelas 2004: 271)

Orgun and Inkelas refer to similar cases from Mohawk, which have been analysed as incorporation of head noun, with modifier stranding (Baker 1988, Mithun 1984), and demonstrate that this analysis fails to capture the cases of coordinated unmarked direct objects, like that given in below.

(73) Ali kitap da dergi de okur.
Ali book too magazine too read
‘Ali reads both books and magazines.’

(Orgun and Inkelas 2004: 271)

The incorporated analysis and the proposal to treat the unmarked direct object as the head noun would require *dergi* or even *dergi de* to be treated as the head noun, and *kitap da* as its stranded modifier, which is clearly impossible. Additionally, like in some of the previous examples, the particle *de* intervenes between the unmarked direct object and the verb here too.

Gapping constructions of the type shown in (74) below provide another piece of syntactic evidence against the incorporation account.

(74) Ali kitap, Mustafa dergi okur.
Ali book Mustafa magazine reads
‘Ali reads books, Mustafa reads magazines.’

According to the incorporation analysis, *dergi okur* would have to be considered a single word, which would remove the identity of the verb *okur* across conjuncts, thus making the gapping construction impossible.

It is noteworthy, that Orgun and Inkelas also argue against the semantic incorporation analysis for Turkish unmarked DO+V sequences, which, according to Van Geenhoven (1998), can happen independently from syntactic incorporation. In

semantic incorporation, the verb and the incorporated argument create a unit that identifies an activity. The examples such as that in (75) fit this description of semantic incorporation.

- (75) Ahmet hergün pipo iç-iyor.
Ahmet every.day pipe smoke-PROG¹⁰
'Ahmet smokes a pipe (engages in pipe-smoking) every day.'
(Mithun 1984)

Orgun and Inkelas underline that although examples like that in (75) might manifest semantic incorporation, not all sequences of an unmarked direct object and verb display semantic incorporation. The authors illustrate this point with the examples in (76) below.

- (76) a) bir doktor ar-ıyor-um
one doctor seek-PROG-1SG
'I am looking for a doctor.'
- b) üç doktor tan-ıyor-um
three doctor know-PROG-1SG
'I know three doctors.'
- c) birçok doktor tan-ıyor-um
many doctor know-PROG-1SG
'I know many doctors.'
- (Orgun and Inkelas 2004: 272)

¹⁰ Orgun and Inkelas (2004) do not mention that direct translation of the verb *iç-* is 'to drink', and it is used together with nouns such as *pipo* 'pipe' or *sigara* 'cigarette' to denote smoking of some sort. The same is observed in Kazakh, where the verb *tartu* 'to pull' is used in combination with the words *temeki* 'tobacco', or *trubka* 'pipe' to identify smoking activities.

According to Orgun and Inkelas, the objects are referential in these examples, as the sentences do not describe situations where the subject engages in the activities of ‘doctor-knowing’ or ‘doctor-seeking’, but rather introduces previously unspecified doctors as discourse participants. This also conforms with the canonical information-structural pattern of Turkish (and Kazakh), where the immediately preverbal position is strongly associated with focus, which is usually associated with non-presupposed information (the update).

Additionally, since semantic incorporation analysis means activity naming, pluralising an unmarked direct object – which is possible and grammatical – would imply a contrast in the meaning of activity types depending on the quantity of the direct object’s referent (Orgun and Inkelas 2004). Moreover, pluralising an unmarked indirect object can turn the proposition from an activity to an accomplishment, especially in the past tense, as the examples in (77) demonstrate.

- (77) a) Körpü kur-du-k.
 bridge build-PST-1PL
 ‘We (have) engaged in bridge-building.’
- b) Körpü-ler kur-du-k.
 bridge-PL build-PST-1PL
 ‘We (have) built bridges.’

(Orgun and Inkelas 2004: 272)

Direct objects are unmarked in both examples in (77); in the first example in (77) the direct object is in singular form, while in the second example the same direct object is in the plural form. The difference in the interpretation of these direct objects is such that the unquantified (singular) direct object yields the semantic incorporation reading, while its quantified (plural) counterpart does not.

To sum up, as far as the syntactic status of unmarked direct objects is concerned, two main positions have been presented: the syntactic incorporation analysis, and the partial semantic incorporation analysis. It has been demonstrated that

the syntactic incorporation analysis of the unmarked direct object + verb sequences does not withstand the scrutiny proposed by Orgun and Inkelas. The authors arrive at the conclusion that marked or unmarked indefinite complements are attracted to the immediately preverbal position; and that complements without case-marking are also attracted to the immediately preverbal position. The same conclusion holds for Kazakh.

Let us now move to the second main point of discussion in the studies of Turkish DOM – the difference in the semantics and pragmatics of marked and unmarked direct objects. The central notions under discussion here are those of definiteness and specificity (as described in sub-Section 2.1.4, Chapter 2). It is posited that unmarked direct objects are indefinite and/or non-specific, while their marked counterparts are definite and/or specific.

While it is sometimes proposed that case-marked direct objects in Turkish are definite (Orgun and Inkelas 2004, amongst others), von Heusinger and Kornfilt (2005: 41) convincingly demonstrate that specificity is “the most important semantic property that determines overt objective, i.e. accusative case marking in Turkish”. The examples in (78)-(81) below support this point.

(78) Murat kitap ok-uyor.

Murat book read-PROG

‘Murat is book-reading.’

(79) Murat bir kitap ok-uyor.

Murat a book read-PROG

‘Murat is reading a book.’ (‘Murat is reading the book.’)

(80) Murat kitab-1 ok-uyor.

Murat book-ACC read-PROG

‘Murat is reading the book.’

(81) Murat bir kitab-1 ok-uyor.

Murat a book-ACC read-PROG

‘Murat is reading a certain book.’

(von Heusinger and Kornfilt 2005)

From English approximations of sentences in (78)-(81) we can see that an unmarked direct object in (78) receives incorporated interpretation, and it can also receive indefinite and non-specific interpretations; addition of the indefinite article *bir*, as in (79), removes the possibility of incorporated interpretation, and the direct object in this example can only be understood as indefinite and non-specific.

The case-marked direct object in (80) has the definite and specific interpretation, and the case-marked direct object preceded by the indefinite article *bir* in (81) is interpreted as indefinite, but specific. This fine-grained distinction is based on von Heusinger’s opposition to treating specific expressions as a subclass of indefinite NPs, and on his alternative proposal to view specificity as a referential property of NPs, which cuts across the definite-indefinite distinction – as outlined in sub-Section 2.1.4 of Chapter 2.

While von Heusinger’s (2002) and von Heusinger and Kornfilt’s (2005) conclusions regarding possible interpretations of accusatively marked direct objects in the IPV position are convincing and comprehensive, they do not provide an explanation of how the pragmatic interpretation of specificity is achieved. An assumption seems to be made that an accusative case marker contributes the pragmatics of specificity, which, in turn, implies that accusative is not a structural/morphological case.

In the same vein, Orgun and Inkelas explicitly state that since accusative case marking is used only for definite objects, the accusative case “is not in fact purely structural” (2004: 276). This, of course, is a problematic statement, since the only supporting evidence for it lies in pragmatic interpretations of marked direct objects. Also, these approaches to accusative case marking of immediately preverbal direct objects do not explain the ability of case-marked direct objects to appear anywhere in a sentence, and the inability of unmarked direct objects to do so.

As was mentioned at the beginning of this sub-section, and has been remarked in it, all presented observations and analyses can be applied to Kazakh DOM. The same questions and reservations would subsequently apply too. In what follows I propose an alternative approach to DOM in general, and to DOM in Kazakh and Turkish in particular.

The DS-based analysis does away with a need to posit certain syntactic features pertinent to one or another sentential position, or to propose that one of case markers carries some additional pragmatic meanings. All pragmatic effects and syntactic restrictions are explicated via difference in parsing of marked and unmarked direct objects. The following sub-section introduces a ‘non-traditional’ approach to DOM, upon which the DS approach to DOM – presented in the following section – is then developed.

4.2.2. An Alternative Approach to Differential Case Marking

So far we have considered traditional approaches to DOM, which attempt to establish one-to-one connections between form, function and meaning, and which consider a case-marker (or lack thereof) to be a realisation of some morpho-syntactic or pragmatic feature of the constituent to which it is attached.

Cann and Miljan (2012) move away from this Graeco-Roman tradition, and propose that case-marking “should be taken seriously as directly providing information that determines the interpretation of clauses in which they appear” (Cann and Miljan 2012: 590). Cann and Miljan present their take on partitive case alternation in Estonian, and on the use of the case-marker *-ko* in Hindi/Urdu, and demonstrate that explanations of puzzling syntactic facts lie in examining the interactions of a case-marker with its linguistic context, and with other case-markers. As Hindu/Urdu marker *-ko* displays some similar usage with Kazakh accusative marker, for the purposes of this study I concentrate on Cann and Miljan’s approach to it.

According to Ahmed (2006) (cited in Cann and Miljan 2012), the case-marker *-ko* marks temporal, spatial, and purpose adverbials; certain clauses; dative subjects and objects; accusative causees; and accusative objects. Ahmed (2006), and Butt and

Ahmed (2010) suggest that *-ko* indicates a possibly unattained or abstract endpoint marker, as illustrated by Old Urdu examples below.

(82) ek vilayat mē poāce
 one city in reached
 ‘reached a city’

(83) is manzil ko kab poāco-ge
 this destination ACC/DAT when reach.2-FUT.PL
 ‘When will (you) reach this destination?’

(Butt and Ahmed 2010)

Ahmed examines and explains various uses of *-ko* in Modern Hindi/Urdu on the basis of that meaning, and it appears to work well for expressing spatial and temporal location, uses of *-ko* with indirect objects, experiencer subject use, as well as for affected agents in causative constructions and purpose.

However, it becomes problematic once the use of *-ko* with direct objects is taken into consideration. Case-marker *-ko* must be used with all human direct objects as shown in (84); when used with non-human direct objects, *-ko* indicates specificity or definiteness as shown in (86) juxtaposed with (85).

(84) anjum=ne saddaf=ko dekhaa.
 Anjum=ERG Saddaf=ACC see.PFV.M.SG
 ‘Anjum saw Saddaf.’

(85) anjum=ne kashtii dekhi.
 Anjum=ERG boat.F.SG see.PFV.F.SG
 ‘Anjum saw a/the boat.’

(86) anjum=ne kashtii=ko dekhi.
 Anjum=ERG boat.F.SG=ACC see.PFV.F.SG

‘Anjum saw the boat.’

(Cann and Miljan 2012: 601)

Cann and Miljan agree that the semantic of ‘(potential) endpoint’ cannot be extended to cover the uses of *-ko* demonstrated above. They propose an alternative approach that explains the pragmatic effects resulting from the use of *-ko* not via assigning some specific semantics to it, but via considering its place in the grammatical system.

First, Cann and Miljan consider all types of arguments marked by *-ko* in Modern Hindi/Urdu: indirect objects, experience subjects, and affected agents of causatives, and conclude that all these arguments could be considered unusual or ‘peculiar’ in some way. Indirect objects, for example, are ‘unusual objects’ since they are additional to direct objects and are not proto-patients (Dowty 1991). Experiencer subjects are usual subjects in that they lack agentivity expected from a typical subject which is volitional or active in the event. Affected agents of causatives are not typical arguments as they display agentive properties in relation to the caused event, but patient-like ones in relation to the causation.

Cann and Miljan posit that the use of *-ko* with direct objects can be derived from a pragmatic extension of this marker’s indication of an unusual or in some way unexpected argument. The hearer has to identify what is unusual about a direct object marked with *-ko*. Recall that *-ko* is obligatorily used with human direct objects, which are considered to be less usual than non-human ones, as humans are more likely to be agents of the event described by the predicate.

As for the use of *-ko* with inanimate specific direct objects, there is also an element of unusual behaviour here, since inanimate specific direct objects are less typical than the indefinite ones, because typologically objects are less likely to be topics (Bossong 1985, Filimonova 2005, amongst others). What we observe, then, is a grammaticalisation process whereby the ‘goal-like’ interpretation of *-ko* has been extended to direct objects that are pragmatically marked or in some way unexpected. Cann and Miljan (2012: 602) state that:

The fact that a pragmatic extension inviting hearers to infer a peculiarity with respect to some type of marked object over a grammatical system may itself become grammaticalized and restricted to particular constructions is neither surprising nor syntactically significant.

The analysis of Kazakh DOM developed in the following section arrives at conclusions along the same line of thought, while also attempting to provide a formal DS sketch of the phenomenon.

Cann and Miljan (2012) also present an example of *-ko* being used to signal the difference in the modality; compare (87) and (88) below.

(87) nadya=ne zu ja-na he.
 Nadya.F=ERG zoo go-INF be.PRS.3SG
 ‘Nadya wants to go to the zoo.’

(88) nadya=ko zu ja-na he.
 Nadya.F=DAT zoo go-INF be.PRS.3SG
 ‘Nadya has to/wants to go to the zoo.’

(Cann and Miljan 2012: 602, citing Ahmed 2006)

The use of the ergative marker *-ne* observed on the subject of (87) has only recently been extended from marking the transitive subject of a perfective verb to signalling volition of the subject. In this usage, then, the ergative marker emphasises volitionality of the subject. The marker *-ko*, being the typical experiencer marker, might potentially signal nothing more than weak volitionality.

However, Cann and Miljan (2012) suggest considering this use of *-ko* as the paradigmatic contrast to the developing usage of the ergative marker *-ne*: by omitting the latter the speaker prompts the hearer to infer the reason for the omission, leading to the implication of non-volitionality, which is associated with the use of *-ko* to indicate an experiencer subject. Non-volitionality, in turn, is typically associated with coercion, which triggers a pragmatic implicature of deontic modality, as indicated by the possible interpretation of (88). This contextualised Relevance Theoretic inference

provides for a simple explanation of the modal interpretation *-ko* activates, which does not necessitate positing any additional semantics for the marker.

One of the main conclusions Cann and Miljan draw is that extensions of the use of a case-marker may develop from “pragmatic inference over the grammatical systems that emerge from semantic extensions of use” (2012: 603). As was demonstrated for Hindi/Urdu marker *-ko*, these extensions can arise from reinterpreting a marker as indicating a more general property, which connects different grammatical uses of that marker. Thus, it was shown that analysing the Hindi/Urdu *-ko* as a marker for ‘unusual’ arguments can explicate its uses with grammatically diverse types of objects. Additionally, considering the use of *-ko* in contrast with the ergative marker *-ne* provided for an explanation of its contribution to modality of the clause.

The account of DOM in Kazakh I propose in the section below follows the same general approach as that by Cann and Miljan (2012) presented in this section. It is demonstrated that no additional semantic or syntactic features need to be posited to cover all uses of Kazakh accusative marker *-DI*. The pragmatic effect of specificity associated with this marker in Kazakh (and in Turkish) can be successfully explained by considering the overall grammatical system of case-marking, and the parsing process of utterances (within context) in which accusatively-marked constituents, especially in the immediately preverbal position.

While the final semantic trees resulting from the parsing sentences with and without an accusatively-marked direct object look the same, the difference in the ways these trees were constructed explicates the difference in the pragmatics of marked and unmarked direct objects.

4.3. Kazakh DOM - A Dynamic Syntax Analysis

Having considered the DS representations of the parsing process for several sentences in the preceding chapter, we can move straight to the DS analysis of Kazakh DOM. It is proposed that the difference in the parsing of marked and unmarked direct objects yields the interpretational, pragmatic effects associated with these constituents. This

analysis supports Orgun and Inkelas's rejection of incorporated analyses for unmarked direct objects (in Turkish, and by extension in Kazakh), and provides a formalisation for the differences in interpretation of marked and unmarked direct objects in regard to their specificity.

As mentioned in the preceding sections, no formal analyses have been published for Kazakh DOM, and previous analyses of Turkish DOM focus their attention on the differences in the syntactic, semantic and pragmatic features of marked and unmarked direct objects. These analyses have yet to reach a consensus, and to produce an insight that goes beyond the descriptive terms of topicality, incorporation, animacy, definiteness, or specificity, none of which apply to all instances of DOM in neither Turkish nor Kazakh. In this section I present the first formal analysis of Kazakh DOM, which can be extended to Turkish DOM, and potentially to DOM in other related or unrelated languages.

Applying the DS approach to the phenomenon of DOM in Kazakh means switching from the traditional 'static' view of syntax to examining DOM within the dynamic process of incremental parsing, as it unfolds step-by-step. This delivers a new understanding of the phenomenon and its mechanisms. This, in turn, allows us to move away from describing the syntactico-semantics characteristics of differentially marked direct objects in a static sentence, which fundamentally, leads to a cyclical explanation of the phenomenon along the lines of: the direct object is marked because it is specific (or definite/ animate/ human/ topical), and it is understood to be specific (or definite/ animate/ human/ topical) because it is marked.

The proposed analysis is also in line with Cann and Miljan's (2012) appeal to consider the overall grammatical system, and the place the marker under examination occupies within it. The opposition of marked versus unmarked direct objects is extended to the opposition of marked versus unmarked constituents, and the role accusative case marking is explored within this wider context.

Some observations regarding the role of accusative case in Kazakh can already be made on the basis of the lexical entry for accusative case marking, presented in the preceding chapter, and repeated below for convenience.

(89) Lexical entry for *-DI/-n* (accusative)

IF	Fo(α), Ty(e)
THEN	IF $\langle \uparrow_0 \uparrow_i^* \rangle$ (Tn(a), ?Ty(t))
	THEN put(? $\langle \uparrow_0 \rangle$ Ty(e \rightarrow t))
	ELSE Abort
ELSE	Abort

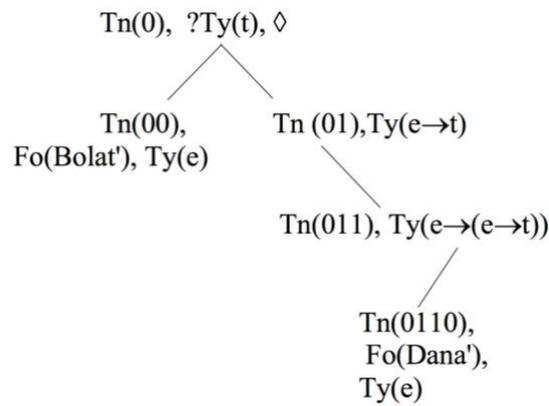
As can be seen from this lexical entry and its application within the parsing process, I consider the accusative case marker to be the marker of direct objecthood only. The lexical entry above does not suggest that any additional pragmatic features of definiteness or specificity are contributed by this case marker. The ‘specific’ reading that accusatively marked direct objects are considered to manifest are simply contextual, pragmatic effects of the use of the accusative case marker especially in contrast to unmarked direct objects.

Let us now consider the parsing process for a sentence with an unmarked direct object, like that in (90) below, which is a slight alternation of the sample sentence from Section 3.5 of Chapter 3.

- (90) Bolat Dana-ğa kitap berdi.
Bolat Dana-DAT book gave
‘Bolat gave Dana a book.’

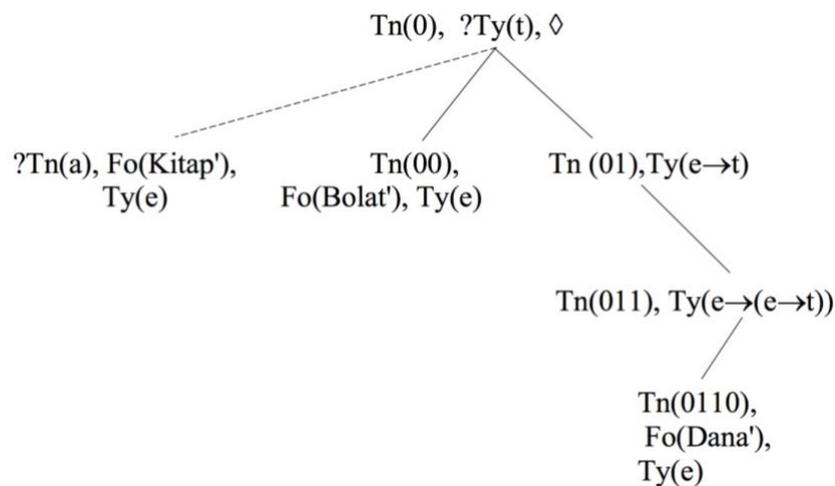
The first two nominals are parsed as previously shown, and fixed as the subject and the indirect object of a partial tree, as shown in (91) below.

- (91) Parsing *Bolat Danağa...*



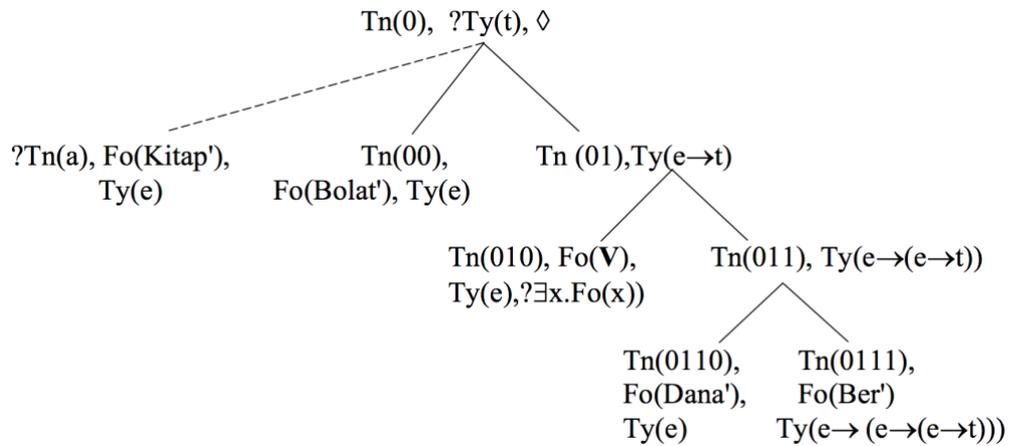
Next, the unmarked noun *kitab* must be parsed. Recall, that in the preceding chapter, where the parse of a sample S-DO-IO-V sentence was presented, an IP boundary indicated that an unmarked sentence-initial nominal was the subject. In the case of an unmarked direct object in the IPV position, the lack of an IP boundary or a case marker indicates that this constituent forms one intonation phrase with and is to be parsed in relation to the following constituent. In DS terms, this means that a node onto which such an unmarked nominal is parsed remains unfixed, as shown in (92).

(92) Parsing *Bolat Danağa kitap...*



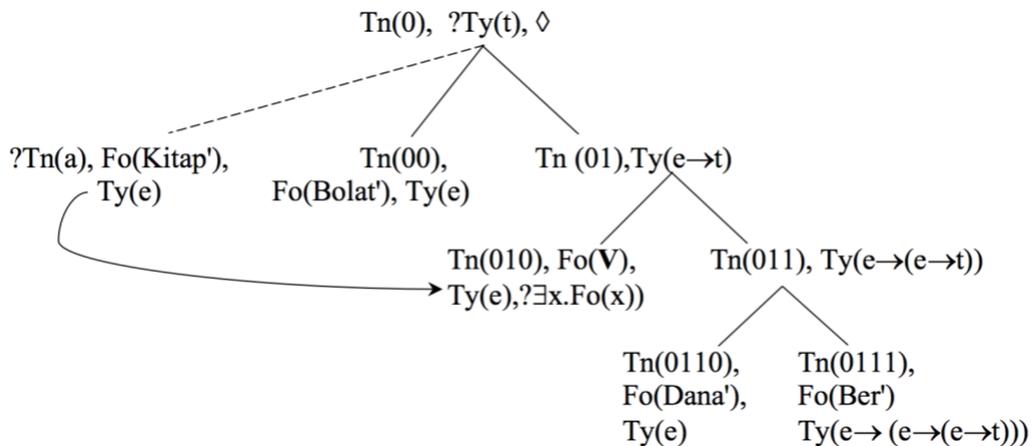
The verb *berdi* follows the unmarked direct object and projects the full argument structure, as per the lexical entry given in (99) in Section 3.5 of Chapter 3.

(93) Parsing *Bolat Danağa kitap berdi...*



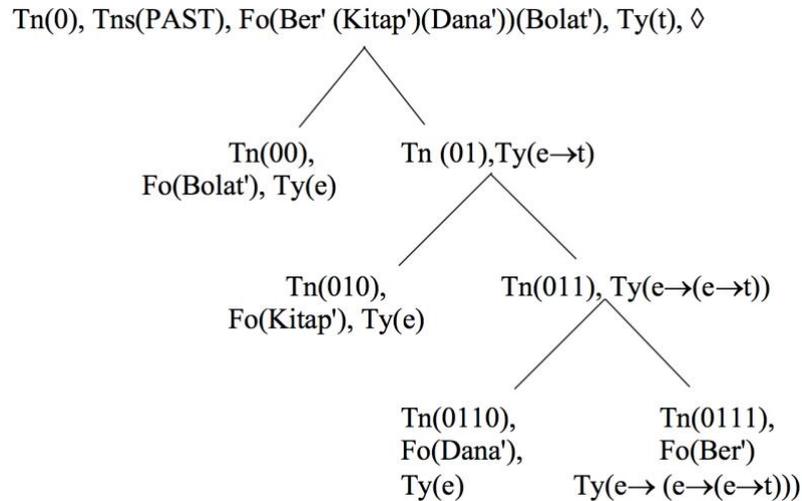
At this stage we have an incomplete tree with an unfixed node, and with a fixed Tn(010) node with a variable and a requirement for it to be substantiated with a semantic formula. Both these nodes are of type e , and have no decorations that might prevent their Merge, which is precisely what occurs:

(94) Parsing *Bolat Danağa kitap berdi* – merge



In (95), we see the final tree after the Merge, and once all the clean-up and completion process have taken place.

(95) Parsing *Bolat Danağa kitap berdi* – final tree



As can be observed, this final tree is identical to that given in section 3.5 of Chapter 3 for the sentence *Bolat kitapti Danağa berdi*, which once again shows that the DS analysis of DOM (or indeed any other phenomenon) does not rely on any differences in theory-specific hierarchical syntactic structuring of a sentence.

Having considered parsing processes for trees with a marked and an unmarked direct object, we can make several observations explaining the nature of Kazakh DOM. Firstly, going back to Cann and Miljan’s remark on the need to take the whole (case marking) system into consideration, and not just a narrow phenomenon within it, we can note that non-case-marked nominals are generally speaking more unusual for Kazakh than marked ones. Recalling the rich case system presented in Section 1.1 of Chapter 1 – a nominal is much more likely to be case marked than not in Kazakh.

Building on that, we can identify two further types of non-case-marked constituents: subjects and direct objects. It was shown, however, that in lieu of a case-marker, subjects are identified with the help of an IP boundary, therefore, they cannot be seen as completely unmarked – the parser receives a discernible signal which allows these non-case-marked nominals to be parsed as subjects. In DS terms this means that these non-case-marked nominals appear on a fixed node before the parsing

of the next element in the string begins, as is typical for and preferred in Kazakh with its ‘fix-as-you-go’ parsing strategy.

Non-case-marked direct objects, on the other hand, remain unfixed until the predicate following them is parsed, since neither a case marker nor an IP boundary is able to fix an unmarked direct object. This distinguishes unmarked direct objects from other types of arguments, since they are the only type of argument to be parsed onto and to remain on an unfixed node until the following constituent is parsed.

This underspecified, unfixed position an unmarked direct object takes in the semantic tree is symbolic of non-specific, generic reading unmarked direct objects are so strongly associated with in Kazakh and Turkish (see earlier sections of this chapter). The DS representation of this, given in (93) above, provides for an intuitive and accurate explication of the interpretations unmarked direct objects receive, and of their relation with the predicate.

In contrast to unmarked direct objects, accusatively marked direct objects obtain a fixed position in the semantic tree before the parsing of a following constituent begins. This divergent behaviour signals a difference in the interpretation – a case-marked direct object is fixed in the semantic tree, and, by metaphoric and cognitive extension, in the world. This yields the pragmatic effect of specificity strongly identified with accusatively marked direct objects. This distinction is particularly strong for direct objects in the IPV position, since this is the only position in which an unmarked direct object can appear.

In DS terms, the IPV position is special due to its ability to host an unfixed node, since it guarantees that this node will be fixed at the next step of the parsing process. This removes the need to posit a special syntactic relation of an unmarked direct object with a predicate, as per some of the incorporation analyses for Turkish DOM. Both marked and unmarked direct objects eventually take up the same Tn(010) position in DS semantic trees, with the former getting there faster than the latter, and independently from the predicate.

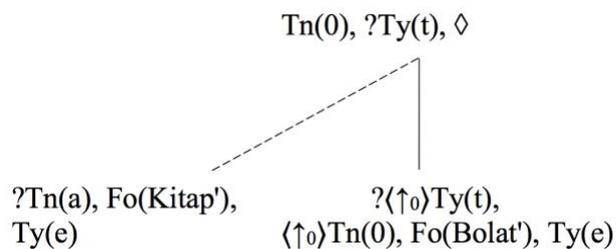
This approach to the difference between marked and unmarked direct objects also explains the ungrammaticality of unmarked direct objects appearing outside of the IPV position, as shown in (122).

(122) *Bolat kitap Danağa berdi.

Intended: 'Bolat gave a book to Dana.'

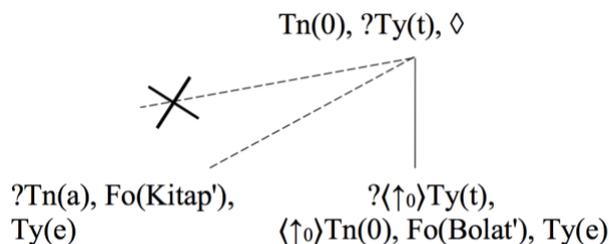
Once the subject and the unmarked direct object are parsed, we end up with the following semantic tree:

(123) Parsing Bolat kitap...



Then, another nominal constituent needs to be parsed, before we reach the predicate that is able to 'fix' the unfixed direct object. Recall that each nominal starts off being parsed onto an unfixed node (via Local *Adjunction) before being fixed into place by either a case-marker or an IP boundary. This is when we encounter an issue since we cannot have two unfixed nodes of the same type (locally unfixed in our case) in a tree at the same time:

(124) Attempting to parse *Bolat kitap Dana...*



Restricting unmarked direct objects to the IPV position prevents this from occurring.

To sum up, then, I posit, that in DS terms, the phenomenon of DOM can be explained through the notions of fixed and unfixed nodes. An accusatively marked direct object is fixed onto a Tn(010) node within a semantic tree before the parsing of a following constituent begins, whereas an unmarked direct object remains unfixed until the predicate is parsed.

The presented analysis does away with the need to assign additional pragmatic meaning of specificity to the accusative case in Kazakh. The contrast in the ways marked and unmarked direct objects are parsed creates the pragmatic effects of specificity and genericity associated with the former and the latter respectively. No assumptions need to be made as to the syntactic statuses of marked or unmarked direct objects either.

The IPV position is the only position that can host an unmarked direct object as this is the only position in which an unfixed node is guaranteed to be fixed as soon as the verb is parsed. Disallowing unmarked direct objects to appear outside of the IPV position prevents ‘illegal’ situations where two unfixed nodes of the same type are present in a semantic tree at the same time. Obligatory accusative case marking on all direct objects outside of the IPV position ensures the same.

4.4. Conclusions

This chapter has considered the phenomenon of DOM in Kazakh, and provided a detailed description of it. It has been shown that the accusative case marking on a direct object in the IPV position yields a specific reading for that object’s referent, and that unmarked direct objects have an indefinite and non-specific reading.

The DS analysis of DOM has also been presented in this chapter. Contrary to many previous descriptions of Kazakh DOM and analyses of Turkish DOM which assign additional meanings of specificity or definiteness to the accusative case marker, the DS analysis of DOM demonstrates that the specificity is simply the pragmatic effect arising from the way a marked direct object is parsed, especially in comparison with parsing of an unmarked direct object. The ‘special’ nature of the IPV position to which unmarked direct objects are confined is also explained through DS terms.

This analysis of Kazakh DOM can also be applied to DOM in Turkish, and many other related languages, with possibility of extension to non-related languages as well – I leave this to be tested in future research.

5. Topic Markers

As became evident in Chapter 3, Kazakh is a discourse-prominent language with a tendency towards topic-prominence. It was shown that topics typically occupy the sentence-initial and sentence-final positions, with new and contrastive topics always being sentence-initial; however, it was also shown that not all sentence-initial elements are necessarily topical.

In this chapter I demonstrate that placing a topical element into a particular sentential position is not the only way to signal topichood in Kazakh, as I examine three Kazakh topic markers – *še*, *bolsa*, and *degen*.

This chapter proceeds as follows: Section 5.1 briefly re-introduces the notion of topic markers which was considered in detail in Section 2.1 of Chapter 2; in Section 5.2 topic marker *še* is considered; Section 5.3 is dedicated to topic marker *bolsa*; and Section 5.4 examines topic marker *degen*. Section 5.5 presents a sketch of the Dynamic Syntax approach to Kazakh topic markers, while Section 5.6 concludes this chapter.

5.1. Introduction

A number of world's languages have special morphemes that mark the topic or the focus of a sentence – topic and focus markers respectively. According to Gundel (1988) a language might have both topic and focus markers (e.g. Quechua and Mongolian), only focus markers (e.g. Duala, Marathi, Yukaghir), only topic markers (e.g. Burmese, Hmong, Hua, Lisu), or neither topic nor focus markers (Russian, Spanish, English, German).

As mentioned in Section 2.1 of Chapter 2, topic markers can mark a diverse range of topical constituents, which can but do not have to be an argument of the main predicate. Topic markers can also follow temporal and locative elements, which, according to Radetzky (2002) is not unexpected from the diachronic perspective. It was also shown that topic markers can mark relative clauses and protases of conditional clauses, which, according to Gundel (1988), follows from the presupposed

relationship that topics, relative clauses, and protases have with the rest of the sentence.

Kazakh belongs to the group of languages with only topic markers – no focus markers have been identified in this language in the literature, or in my own research. It was shown in Chapter 3 that the focal status of a constituent is usually marked by placing it in the immediately preverbal position, by making it prosodically prominent, or via a cleft-type construction explored in Chapter 6.

Gorelova (2005) states that all Altaic languages have topic markers, which have sometimes been analysed as subject markers or particles. She observes that these markers usually have origins in verbal forms, more specifically, in the verbal forms “derived from existential verbs ‘be’, ‘exist’ and ‘become’ [...], as well as from the verb of speech ‘say’ (2005: 149)”; the latter are referred to as *verba dicendi*.

Both types of topic markers are present in the Kazakh language. The first type is represented by the markers *še* and *bolsa*, which are discussed in Sections 5.2 and 5.3 respectively, while Section 5.4 is dedicated to the topic marker of the second type – *degen*. Each section begins with a sub-section on previous research, if it is available, and continues with syntactic distribution and meaning of the topic marker under discussion.

5.2. Topic Marker *Še*

This section provides a description of the Kazakh item *še*, which performs functions of a topic marker and a question particle, with the latter function currently being the dominant one for *še*. I propose that the underlying meaning of *še* is contrast, which makes it a contrastive topic marker and a contrastive question particle.

The section proceeds as follows: first, in 5.2.1 I present an overview of previous research on *še*. In 5.2.2 I consider a related Turkish marker *ise*, and hypothesise that *še* originated from the same source as *ise*, although further diachronic research is required. This is followed by 5.2.3 which is dedicated to results of a corpus study, which clearly indicate that the use of *še* as a question particle unequivocally prevails its use as a topic marker. In 5.2.4 and 5.2.5 I focus on *še* as a topic marker and

a question particle respectively; their distribution patterns and meanings are examined. Finally, 5.2.6 concludes this section.

5.2.1 Kazakh Particle *še* – Previous Research

In different Soviet linguistic/philological sources, *še* is also identified as *ši* and *šī*, while the usage described is the same for all the forms. I assume this difference to be of dialectal nature and do not aim to provide a full explanation for it in this work, however, further research would be of interest. I identify and preserve the forms given in the literature by the authors, however, I only refer to this pragmatic marker as *še* (or *-še* for grammaticalised forms). This is motivated by the fact that only this free form is found in the Almaty Corpus of Kazakh Language (ACKL) (introduced in Section 1.3 of Chapter 1).

Balakaev (1959) refers to *ši* and *šī* as question particles, however, differentiates them from the question particle *MA* (described in Chapter 1, Section 1.1.) by stating that *ši* and *šī* are only used with nominal predicates, as in the examples below.

- (1) Al, Žamal ši? Özin ši?
 and Žamal Q yourself Q
 ‘And what about Žamal? And yourself?’

- (2) At-īm Qajırǵali, öziñ-niñ at-iñ
 name-POSS.1SG Qajırǵali yourself-GEN name- POSS.2SG
 ši?
 Q
 ‘My name is Qajırǵali, and what is your name?’

(Balakaev 1959: 111)

Balakaev does not explicitly point out that the form of the particle to be used depends on vowel harmony, but it can be seen from the sentences in (1) and (2) where the noun

with front vowels is followed by *ši* in the first sentence, and the noun with back vowels is followed by *šī* in the second sentence.

Balakaev et al. (1962) also classify *še* as a question particle, but do not mention the forms *ši* and *šī*. It is said that the particle *še* is used when “the speaker wishes to find out something about the subject in general” (1962: 415). It is noted that this specialised use of *še* is the main difference of this particle from the question particle *MA*, and that it also explains the very limited use of *še* with predicates. The following examples are provided to illustrate the use of *še* as a question particle following nominals about which a ‘general’ question is asked.

- (3) Sen *še*?
 2SG Q
 ‘And (what about) you?’

- (4) Balžan *še*?
 Balžan Q
 ‘And (what about) Balžan?’

(Balakaev et al. 1962: 415)

Note that the examples in (3) and (4) are identical to that in (1), which confirms that *še*, *ši* and *šī* are indeed one and the same item.

Balakaev et al. also observe that *še* can be used with verbs with suffixes *-SA* and *-GandA*; these suffixes express conditional and subjunctive moods respectively (see Section 1.1 in Chapter 1). The examples in (5) and (6) demonstrate this point.

- (5) Sibir-diñ kejbir žildarī bol-atin üzdiksiz
 Siberia-GEN some time be-PTCP continuous
 žaňbīr-ī bastal-sa *še*?
 rain-POSS.3 begin-COND Q
 ‘And what if those unending Siberian rains start?’

- (6) Žol žaj-ï esime tüs-ken-de,
road condition-POSS.3 memory arrive-PTCP-LOC
köñl-im žüde-gen sijaqtan-a-dï.
mood-POSS.1SG suffer-PTCP resemble-PRS-3
Žüde-me-gende še?
suffer-NEG-SBJV Q

‘When I recall the state of the roads I become very upset. And how could I not?’

(Balakaev et al. 1962: 415)

Baskakov (1966) states that the question particle *še* has an allomorph *ša* – used after words with back vowels –, and that this particle contributes the meaning ‘and’ to a question, as seen in the English approximations in (3) and (4) above. However, the authors contradict their own claim for the existence of the ‘back’ version of the particle by using *še* after a word that contains the back vowel *a*, as shown in (7).

- (7) Ospan še?
Ospan Q
‘And what about Ospan?’

(Baskakov 1966: 433)

It must be noted, however, that there is a suffix *-še/-ša* in the Kazakh language which is used to form adverbs, as shown in the examples in (8). It appears that Baskakov simply assumed the same vowel harmony for the standalone item *še*.

- (8) a. žaña-ša b. eski-še
new-ADV old-ADV
‘in a new way/manner’ ‘in an old way/manner’

Bizakov (2014) also classifies *še* as a question particle and puts it in the same category as the question particle *MA* without mentioning any differences in usage. The

examples he provides are identical to those in (3) and (4). Bizakov mentions *ši* and *šī* under the subtitle ‘intensifying particles’ and provides the following example of its use.

- (9) Sura-sa-j-šī.
ask-COND-PTCP-šī
‘Do ask. / Go ahead and ask.’

(Bizakov 2014: 181)

No context or further explanations are provided by the author as to the distribution or functions of *ši* and *šī*. Neither does the author justify his classifying these items as particles while only giving an example in which *šī* is suffixed on to the verb.

The most detailed description of *še* available in the literature is found in Muhamedowa (2016), who identifies *še* as a clitic, and states that there are two main forms for it: *-šI* and *-še*. Similarly to the claims provided above, Muhamedowa asserts that these items are used to form questions corresponding to the English question *And what about...?*. According to the author, *-šI* and *-še* “occur on the last element of a clause and can be cliticised to nouns, personal pronouns, the determinative *öz* (‘own’) and headless relative clauses” (2016: 19).

The first three distribution options are shown in the examples (1), (2) and (3). Muhamedowa provides the example in (10) below as an illustration of the use of *-šI* with headless relative clauses.

- (10) 10 šahter aman-esen oral-dī. Qal-ğan-dar-ī
10 miner healthy return-PST.3 stay-PTCP-PL-POSS.3
šī?
CL
‘Ten miners came back healthy. And what about the others?’

(Muhamedowa 2016: 19)

Muhamedowa proceeds to stating that the clitic *-šI* has two surface realisations: front vocalic *-ši* and back vocalic *-šī*, while *-še*, does not follow vowel harmony and

is always realised as *-še* regardless of the quality of the vowel in the preceding syllable; no explanation is provided as to how a speaker makes a choice between *-šI* and *-še*.

Citing Balakaev et al.'s (1962) observation on the use of *-še* after the verbs with suffixes *-sA* and *-GAndA*, Muhamedowa asserts that in these cases *-še* “expresses speculative questions” (2016: 19), which can be translated into English as *And what if...?*. She also adds that “the conditional clause is left without the main clause” (2016: 19) wherever *-še* is used, by which, I assume, she means that an interrogative conditional sentence consists only of a protasis, and does not include a consequent.

This is not an unexpected development since an interrogative protasis is used by one speaker for the purpose of inducing a consequent from her interlocutor. Muhamedowa's example of the use of *-še* in this function of forming “speculative questions” is given in (11) below.

- (11) Bir bala asīra-p al-sa-q še?
one child bring.up-CVB take-COND-1PL CL
‘And what if we adopt a child?’

(Muhamedowa 2016: 19)

Although Muhamedowa refers to *-šI* and *-še* as ‘clitics’, the only example where *-šI* is actually affixed to the host word (as a clitic typically expected to be) is given in (12) below, where *-šI* is used to “make requests more polite when added to imperative forms” (2016: 26).

- (12) Kel-ši!
come.IMP.2SG-CL
‘Come!’¹¹

(Muhamedowa 2016: 26)

¹¹ A more appropriate translation, reflecting the use of *-šI* is ‘Would you (please) come!’ or ‘Please come!’

As well as mentioning *-šI* in the section on interrogation, Muhamedowa (2016) considers it in a separate section titled ‘The Clitic *-šI*’. In this section, the author states that this clitic can be used in interrogative, imperative and declarative sentences. The use of *-šI* in interrogative and imperative sentences has already been shown above. Regarding its use in declarative clauses, Muhamedowa (2016) claims that:

In declarative sentences *-šI* is used as a discourse particle and has a purely pragmatic function. *-šI* is an addressee-oriented particle and appeals to the hearer’s common knowledge. It is comparable to the English *you know*. The main area of use of *-šI* in spoken discourse is narratives...(2016: 264).

The following example is provided to illustrate this point:

- (13) Men ši negizi Germanija-ğa
 1SG PART¹² generally Germany-DAT
 bar-ğĩ-m kel-e-di ši.
 go-NMLZ-1SG¹³ AUX-PRS-3 PART
 ‘I want, you know, to go to Germany, you know.’

(Muhamedowa 2016: 264)

Citing this example Muhamedowa claims that *-šI* can occur “after almost every word class” (2016:264), which is not quite clear, since *šI* is only used after a pronoun and a verb in this example. Muhamedowa (2016) also states that *-šI* can be removed without affecting the meaning of the sentence in any way, which contradicts some of her previous statements regarding the pragmatic contribution of this item.

¹² In the gloss of this sentence Muhamedowa marks *šI* as ‘particle’, despite generally classifying it as a clitic. In fact, the inconsistency of referring to *šI* is observed throughout the work as it is also referred to as a suffix (2016:195) and emphatic particle (2016:205).

¹³ Here, the morpheme glossed by Muhamedowa as *GI* - which is just the surface form of this morpheme - is in fact a nominalising suffix and is productively used to form nouns and adjectives from verbs.

Having consulted several native speakers of Kazakh, the following observations can be made for the sentence in (13): a) using *šI* in this sentence is acceptable for the speakers from the South of Kazakhstan; and b) the variant *še* can be used here too without any change to the meaning. However, the speakers did not agree with Muhamedowa's translation of *šI* as 'you know' – an alternative translation is shown in (14) below. The informants were not certain about the use of *še* at the end of the sentence, and considered it highly colloquial sloppy speech; the example in (14) below was strongly preferred by all language consultants.

- (14) Men še Germanija-ğa bar-ğï-m
 1SG TM Germany-DAT go-NMLZ-1SG
 kel-e-di.
 AUX- PRS-3
 'As for me, I would like to go to Germany.'

To summarise, it has been shown that there is no clarity as to the meaning and function of *še* or *šI* in the Kazakh language. In literature, it has been classified as a clitic, a particle, and a suffix, without much concern for the characteristics, functions and distribution patterns of this item and overtly similar morphemes. This item has been attributed with the function of creating 'general' or 'speculative' interrogations, as well as with contributing the meanings of 'and' or 'you know'.

Before proposing a new detailed description of the distribution and meaning of *še*, to which Sections 5.2.3 and 5.2.4 are dedicated, I present research on the Turkish topic marker *ise*, which is related to the Kazakh *še*.

5.2.2. Turkish Topic Marker *ise*

According to Gorelova (2005), the Turkish topic marker *ise* developed from the conditional converbial form of the ancient defective verb **er-* 'to be' (*ise* < *i-* (**er-*) 'to be' + *-se* = CVB-COND (*-sA*)). Some examples of the use of *ise* are presented in (15) and (16) below; in (15) *ise* acts as a contrastive topic marker, and in (16) its exact

function cannot be identified due to the lack of context – it can be a contrastive or non-contrastive topic marker.

- (15) Ben amele-yim kardeş-im ise doctor-dur.
1SG worker-1SG brother-POSS.1SG TOP doctor-3SG
‘I am a worker, as for my brother, he is a medical doctor.’

- (16) Ben ise bunu-n bir sakınca-sı yok.
1SG TOP this-GEN one drawback-POSS.3 NEG
‘As for me, it does not have any drawbacks.’

(Gorelova 2005: 161)

Gorelova (2005) states that *ise* has been considered a particle in the Turkish linguistic tradition, however, unfortunately, no references are provided for this, and I have not been able to find a confirmation of this in the literature on the subject in English.

Göksel and Kerslake (2005: 101) refer to *ise* – which also occurs as *-(y)sA* – as a clitic and note that out of the two forms of this topic marker *ise* is the less common form, and that the form *-(y)sA* is the preferred option. It is mostly added to 1st and 2nd person pronouns or proper nouns, and can follow case markers, and temporal adverbials, as the examples in (17)-(20) show. This distribution is in accordance to what was noted for topic markers cross-linguistically as outlined in Section 2.1 of Chapter 2.

- (17) Ben-se...¹⁴
1SG-TOP
‘as for me...’

- (18) Ahmet[?]-se / Ahmet ise

¹⁴ The examples in Göksel and Kerslake (2005) are not glossed; all glosses for examples from that work are mine.

Ahmet-TOP Ahmet TOP

‘As for Ahmet...’

(19) Zehra’-yı-ysa

Zehra-ACC-TOP

‘As for Zehra (ACC)...’

(20) Öğle-den sonra-ysa yürüyüşe çık-tı-k.
afternoon-ABL after-TOP walking leave-PST-1PL

‘And in the afternoon, we went for a walk.’

(Göksel and Kerslake 2005: 101-102)

As far as the pragmatic features of *ise/-(y)sA* are concerned, Göksel and Kerslake characterise this item as “a discourse connective with a topic-shifting and contrastive functions” (2005: 101), which sends a strong signal “indicating that a significantly different point is going to be made about the new topic” (2005: 348). This is demonstrated in the example in (21), where *-(y)sA* is added to the first person singular pronoun *ben* to identify it as the new topic, different from the previous topic – *sen* ‘you’; as expected in topic-switching contexts, what is said about the new topic marked by *-(y)sA* is in contrast with what was said about the previous/old topic.

(21) Sen hep kolayına kaç-ı-yor-sun, ben-se her
2SG always take.the.easy.way-PRS-2SG 1SG-TM every
iş-i en iyi şekilde yapmaya çalış-ı-yor-um.
task-ACC most good way do work-PRS-1SG
‘You always take the easy way out, whereas I try to do everything in
the best way possible.’

(Göksel and Kerslake 2005: 348)

However, Göksel and Kerslake also state that what follows a marked new topic can be “a rather striking, perhaps even surprising, further development of the same

theme” (2005: 348) as the examples in (22) and (23) show; note that the topic marker in (23) is attached to the noun in the genitive case form.

- (22) Kağıt-lar-ı ancak toparla-ya-bil-di-m.
 paper-PL-ACC only gather-CVB=know-PST-1SG
 Kitap-lar-sa hala kutu-lar-da dur-uyor.
 book-PL-TM still box-PL-LOC stand-PRS(3)
 ‘I have only managed to tidy up the papers. As for the books, they are still in their boxes.’

- (23) Ahmet Semra-’yı hiç ara-mı-yor-muş.
 Ahmet Semra-ACC never call-NEG-PRS(3)-EVID
 Semra-’nın-sa buna hiç aldırdığı yok.
 Semra-GEN-TM this never caring NEG.EXST
 ‘Apparently Ahmet doesn’t call Semra anymore. Semra, for her part, couldn’t care less.’

(Göksel and Kerslake 2005: 348)

Kabak and Schiering (2004) offer a deeper insight into the nature of *ise/-(y)sA*, as they examine this topic marker in the context of grammaticalisation along the cline of *function word > clitic > affix* proposed by Hopper and Traugott (1993), where *ise*, *-(y)sA*, and the conditional suffix *-sA* occupy the first, second, and third positions of the cline respectively. All three of these items are active in the domain of conditional marking, but for the purposes of this work I only focus on the topic marking properties they describe for *ise* and *-(y)sA*.

Kabak and Schiering note that in the topic marking function *ise/-(y)sA* can only follow/attach to nominals, such as nouns, pronouns and nominalised verbs – examples (24), (25), and (26) respectively; using *ise/-(y)sA* with verbs results in the conditional meaning, as the example in (27) shows.

- (24) Ayrılma-mız-sa çok zor.

separation-POSS.1PL-TM very difficult

‘As for our separation, it’s very difficult.’

(25) Ben ise patates hiç ye-mi-yoru-m.

1SG TM potato never eat-NEG-PRS-1SG

‘As for me, I don’t eat potatoes at all.’

(26) Ayrılmak-sa çok zor.

separate-TM very difficult

‘As for separation, it’s very difficult.’

(27) Ayrılır-sa-k çok zor.

‘If we separate, it’s very difficult.’ / *‘As for separating/our separation, it’s very difficult.’

(Kabak and Schiering 2004: 7)

Let us now consider the conclusions Kabak and Schiering draw from their corpus study of the distribution of free *ise* and bound $-(y)sA$; I only focus on the conclusions relevant to the current discussion of topic marking. As the basis for their research the authors use the corpus of “Morphologically Analyzed and Disambiguated Turkish Texts” which comprises of 800,000 words from electronic Turkish newspapers. This allows the authors to control for the parameter of formal vs. non-formal speech register, as the formal register is the general norm for the newspaper texts in the corpus.

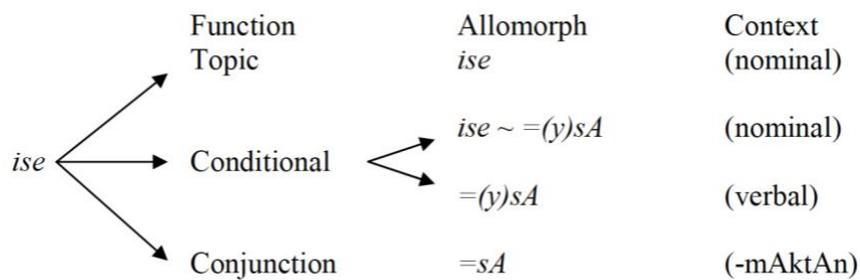
In total, 2301 instances of free *ise* and bound $-(y)sA$ are found, 831 of which are of *ise*, and the remaining 1470 – of $-(y)sA$. Out of the 831 cases, *ise* appears in the nominal context (i.e. following a non-verbal or adjectival, adverbial and other non-nominal constituents) 770 times. Bound $-(y)sA$ appears in the nominal context 135 out of 1470 times. Thus, it is clear, that while there is an overall preference for the bound over the free form, the free form is strongly preferred for nominal contexts. For verbal contexts the bound allomorph $-(y)sA$ is strongly preferred – it is used in 99% of the

total 1007 cases. Kabak and Schiering conclude that there is a “correlation between nominal context and the choice of the free allomorph, and between verbal context and the choice of the bound allomorph” (2004: 9).

Looking closer at the occurrences of both allomorphs in nominal contexts, the authors restrict the data further to only nominal contexts in the nominative case, and find that *ise* is used as a topic marker in 474 out of 491 cases, while $-(y)sA$ is used in the same function in 40 out of 90 cases. Out of the total 514 times *ise* and $-(y)sA$ were used as a topic marker (474 and 40 times respectively), the use of *ise* in 92% of these cases clearly indicates a strong preference of this form for marking topics. The conditional use of *ise* and $-(y)sA$ in nominal contexts follows the overall pattern of $-(y)sA$ being preferred for marking conditionals, with it being used in 62% of cases (25 out of the total 40 cases).

Kabak and Schiering represent their findings in the schema given in (28) below (their Figure 1, 2004: 11). The first two lines show that the free form *ise* is predominant in nominal contexts in the function of the topic marker. Both *ise* and $-(y)sA$ are used in the conditional function in nominal contexts, while the latter is strongly preferred in the same function in verbal contexts. The third line is dedicated to the conjunctive function of the allomorph $-sA$, but I do not discuss it since it is not observed in Kazakh.

(28) Allomorphs of the conditional formula in different contexts/functions



Based on Kabak and Schiering’s discussion of the Turkish *ise*, I develop a hypothesis regarding the comparable and related Kazakh forms; it is presented in the following sub-section.

5.2.3 Hypothesis on the Diachrony of *še*

I hypothesise that the conditional form *ese* of the now non-productive/defunct (but still utilised as an auxiliary in different forms) verb **e-* ‘to be’ developed in Kazakh in similar three ways to that described by Kabak and Schiering (2004) for Turkish. The presumed form *ese* has produced the conditional suffix *-sA*, the politeness marker *-šI*, and the topic marker *še*. The examples in (29), (30) and (31) below demonstrate these forms respectively.

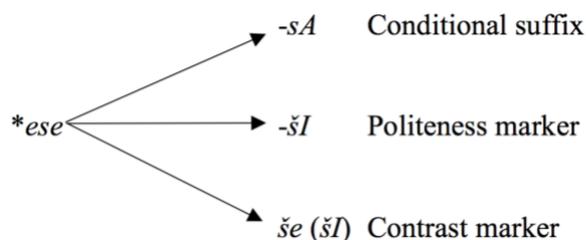
- (29) (Eger) men kel-se-m, ol da kel-e-di.
 if 1SG come-COND-1SG 3SG too come-FUT-3
 ‘If I come, she’ll come too.’

- (30) Qız-ım, mağan šaj al-ši.
 daughter-POSS.1SG 1SG.DAT tea bring.IMP-POL
 ‘Daughter, could you bring me tea. / Daughter, bring me tea please.’

- (31) [Our parents went away on holiday...]
 biz še, dos-tar-ımız-dī qonaq-qa šaqır-dī-q.
 1PL TM friend-PL-POSS.1PL guest-DAT invite-PST-1PL
 ‘...as for us / and we invited friends over to our house.’

The schematic representation of this proposal is given in (32); it must be noted that this is a tentative proposal that requires much further diachronic research.

- (32) Possible development pattern from *ese*



Although this schema is only speculative and tentative, it helps us to start recognising different forms and functions of identical-looking items, such as the suffix *-šI* and the particle *še*, which is sometimes realised as *šI*. This short description does not cover all the uses of the suffix *-šI* which is often used together with other suffixes to build different verb forms, as, for example in (33).

- (33) Astana-da men Bolat-men kezdes-pek-ši-min.
 Astana-LOC 1SG Bolat-INSTR meet-FUT-ŠI-1SG
 ‘I intend to meet with Bolat in Astana.’

According to Valjaeva (2018), adding the suffix *-šI* after the intended future suffix *-MAK*, as in (33), serves to highlight the necessity of the future action indicated by the verb; omitting the suffix *-šI* here does not result in ungrammaticality. This and other uses of *-šI* need to be researched further and are out of scope of this work. I now turn to *še*, its distribution, meaning and functions.

5.2.4. Identifying Two Types of *Še*

In this section I present the results of a corpus study, on the basis of which a clear distinction between two types of *še* is made. It is demonstrated that *še* can function as a topic marker and a question particle, both of which convey contrastive meaning. On the one hand, this study confirms some of the observations from previous research summarised in 5.2.1. On the other hand, it broadens and deepens the understanding of *še* by: a) positing a previously unreported function for this word form; b) giving detailed distribution patterns for both items; and c) exploring the meaning both these items contribute to strings in which they appear.

The examples below (all from the ACKL) show some of the typical uses of *še*, and demonstrate that it can follow different word classes. It is seen used after pronouns - in (34), (35) and (36), nouns - in (37), nominalised verb forms - in (38), adverbs - in

(39), verbs in the conditional and subjunctive forms - in (40) and (41); some of the preceding and/or following context is presented in square brackets.

(34) Barliq-ta, bajliq-ta. Al sen še?
 Plenty-LOC rich-LOC And 2SG TM
 ‘I live in wealth and comfort. And you? / And what about you?’ [‘You spend your days in torment and torture, scraping by, without enough food and clothes.’]

(35) -Iä, žerles-ter-iñ-e maqala-ñ
 yes country.man-PL-POSS.2SG-DAT article-POSS.2SG
 una-ma-dī.
 appeal-NEG-PST
 ‘Yes, your article did not appeal to your fellow countrymen.’
 -Sağan še?
 2SG.DAT TM
 ‘And what about you?’
 [-Don’t mind me.]

(36) [‘I wonder how their characters have not changed, as their looks changed with the years.’]
 Al biz še, zaman-niñ ağıs-ī, uaqit-tiñ
 and 1PL TM century-GEN flow-POSS.3 time-GEN
 iğ-ī-na qara-j, ämanda özger-ip
 influence-POSS-DAT look-CVB always change-CVB
 otir-atin “modali” minez tau-ip
 AUX-PTCP fashionable character find-CVB
 al-ğan žoq-piz ba osi?
 take-PTCP NEG-1PL Q this
 ‘And as for us, don’t we always change and take on the “fashionable” character under the influence of the flow of time.’

- (37) [‘Every Khan and sultan found his own path/road/way: some went to serve the Russian tsar, some were in cahoots with the Žongar leader.]
 Al qara haliq še?
 and black people TM
 ‘And what about the common people/commoners?’
 [‘Sensing that they will soon be losing their land, they became anxious.’]
- (38) [‘The messenger arrived quickly by horse, jumped out of the saddle, and kneeled in front of the king.’
 - Aldijar, Žalajirdan bizge tek tört-bes žüz ğana žigit qosilatın türi bar...
 ‘Your highness, from the Žalajir dynasty we might only have 400-500 people.]
 - Qal-ğan-dar-i še?
 remain-PTCP-PL-POSS.3 TM
 ‘And what about the remaining (of the Žalajirs)?’
 [‘Trying to escape from the Qoqans they are packing up their homes.’]
- (39) Bul keše edi. Al bügin še?
 this yesterday was and today TM
 ‘This was yesterday. And what about today?’
 [‘Today Äbilqajie was also very worried about yesterday/what has passed.’]
- (40) [‘It’s the same here: 150 grams of vodka is a car and one glass of beer is a trailer added to it.’]
 - Al eger araq žüz elu emes, žüz
 and if vodka hundred fifty NEG hundred
 gram bol-sa še?
 gram be-COND TM
 ‘And what if it’s not 150 but 100 grams of vodka?’

[‘Then just one glass of beer cannot be a trailer (you need more).’]

- (41) [Institut bitirgen astanada qaluñ kerek edi. Sen Bäkizatqa üjlendiñ de...
 ‘After finishing university, you should have stayed in the capital. You married Bäkizat and...’]
 - Žä, meni qajt-e-siñ.
 enough 1SG.ACC not.mind-PRS-2SG
 ‘Enough, don’t worry about/mind me.’
 - Qajt-pe-gende še.
 not.mind-NEG-SBJV TM
 ‘And how can I not worry about you?’

As these examples show, *še* has a wide distribution, and can follow items from various grammatical categories, which are described in detail below.

Še is often followed by a question mark, which probably prompted previous researchers to consider it a question particle. As mentioned above, I used the ACKL to help identify the most widely-used form of the particle. Searching for free forms *ši* or *šī* in the ACKL did not return any results, however, the search for *še* returned 100 matches from 25 sources. Having examined these matches, I discarded 8 of them as unsuitable – the form *še* in them was not a true free-standing morpheme: it was a part of the word beginning with the syllable *še-* where the latter was separated from the rest of the word by a hyphen.

The remaining 92 cases were analysed regarding the position of *še* in the sentences and the punctuation following it in clause-final positions. The results are presented in the table below.

Table 21. Clause positional distribution of *še*

Position	Non-clause-final	Clause-final followed by ‘?’	Clause-final not followed by ‘?’	Total
Number of instances	3	84	5	92
Percentage	3.3%	91.3%	5.4%	100%

The results presented in Table 21 demonstrate that in the clear majority of cases (91.3%) *še* is found in the clause final position and is followed by a question mark, although there are also cases where it is not clause final, like in (31) and (36), or is not followed by a question mark in the clause-final position. It must be noted, however, that even for the cases where the clause-final *še* is not followed by a question mark, native speakers read it with interrogative intonation, and provided interrogative translation, as shown in (41) above. I therefore treat such cases as instances of contrastive question particle use, rather than topic marker.

Looking at the data from the point of view of grammatical distribution, three main categories after which *še* appears can be identified: nominal, which includes proper and common nouns in different cases and numbers, pronouns, and nominalised forms; verbal, which consists of verbs ending in *-sA* and *-GandA*; and adverbial. The results are presented in Table 22 below.

Table 22. Distribution of *še* with grammatical categories

Grammatical Category	Nominal	Verbal	Adverbial	Total
Number of instances	51	31	10	92
Percentage	55.4%	33.7%	10.9%	100%

As can be seen from Table 22, *še* follows a nominal constituent in more than half the instances, a verbal constituent in about a third of cases examined, and an adverbial constituent in about ten percent of cases.

Looking at the Tables 21 and 22 together, it can be noted that the most common case of the use of free standing *še*, at least in the ACKL, is where it follows a nominal constituent and appears in a clause-final position of an interrogative clause. This explains the overwhelming labelling of *še* as a question particle in previous research. However, as the examples in (34)-(41) and the analysis of corpus data presented above show, *še* does not exclusively appear in questions but can also follow non-clause-final elements to indicate a change of topic. In order to capture this difference

in the behaviour of this particle I propose to analyse it as a contrast or disjunction marker.

That is to say, the underlying meaning of *še* is contrastivity, and it allows this item to perform two different functions, those of a contrastive topic marker and a contrastive question particle. Recall, that I define a contrastive topic as a topic divergent from the preceding one (see sub-Sections 2.1.1 and 2.1.3 of Chapter 2); this can be extended to a contrastive question – a question regarding a topic different from the previous one.

To sum up, the corpus research presented in this sub-section has demonstrated that *še* can follow constituents belonging to various grammatical categories, although in over half the cases found in the ACKL it follows a nominal element. It has also been shown that in a vast majority of cases *še* appears clause-finally and followed by a question mark, which explains previous references to it in the literature as a question particle. However, in some cases *še* does not appear in the sentence final position, but follows the sentence initial constituent – these cases are considered in more detail in the following sub-section.

5.2.5. Meaning and Distribution of the Contrastive Topic Marker *Še*

When *še* appears after a sentence-initial nominal constituent, like in (14) and (31) for example, it functions as a contrastive topic-marker. The sentence from (14) is repeated below in (42) for convenience.

- (42) Men *še* negizi Germanija-*ğa* bar-*ğĩ-m*
1SG TM indeed Germany-DAT go-NMLZ-POSS.1SG
kel-e-di.
AUX-PRS-3
'As for me, I would like to go to Germany.'

In this example, *še* performs the function of a contrastive topic marker, as this sentence would only be possible in a context where the speaker is contrasting her desire to go

to Germany with either someone else's unwillingness to go there, or their wish to go to some other country; the latter context is shown in (43) below, where the speaker contrasts her mother's preference of a European country to visit with her own. The new contrastive topic in the second sentence in (43) is followed by the contrastive topic marker *še*.

(43) [Talking about an upcoming trip to Europe.]

Ana-m		tek kana		Francija-ğa	bar-ğī-sī
mother-POSS.1SG		only		France-DAT	go-NMLZ-3SG
kel-e-di.	Men	še	Germanija-ğa	bar-ğī-m	
AUX-PRS-3	1SG	TM	Germany-DAT	go-NMLZ-1SG	
kel-e-di.					
AUX-PRS-3					

‘My Mother wants to go only to France. As for me/ And I want to go to Germany.’

The sentence with *še* would not be felicitous as an out of the blue or discourse-opening sentence unless *še* is removed, which demonstrates the contrastive or disjunctive nature of this element. In the same way, the use of *še* is not felicitous where a switch of the topic occurs, but nothing contrastive is said about that topic, as shown in (44), where the proposition construed about the first topic – *anam* – is also given about the contrastive topic – *men*.

(44)	Ana-m		Francija-ğa	bar-ğī-sī	
	mother-POSS.1SG		France-DAT	go-NMLZ-3SG	
	kel-e-di.	#Men	še	Francija-ğa	bar-ğī-m
	AUX-PRS-3	1SG	TM	France-DAT	go-NMLZ-1SG
	kel-e-di.				
	AUX-PRS-3				

‘My Mother wants to go only to France. As for me/ And I want to go to France.’

Another piece of evidence for the contrastive meaning of *še* is the possibility of using the contrastive conjunction¹⁵ *al* together with, or in place of *še* in the topic marking function. Although Muhamedowa (2016) claims that *še* can be removed from a sentence without affecting its meaning, it is important to note that it is the truth-conditional meaning that does not get affected, while the pragmatic meaning, and discourse appropriateness do indeed change. If *še* is removed from the second sentence in (43), the sentence would not become ungrammatical, but it would be less pragmatically felicitous, unless either the sentence-initial pronoun *men* is pronounced with a contrastive intonation (fall-rise), or the contrastive conjunction *al* is placed sentence initially, as shown in (45).

- (45) Ana-m tek kana Francija-ğa bar-ğī-sī
mother-POSS.1SG only France-DAT go-NMLZ-3SG
kel-e-di. Al men Germanija-ğa bar-ğī-m
AUX-PRS-3 CONJ 1SG Germany-DAT go-NMLZ-1SG
kel-e-di.
AUX-PRS-3

‘My Mother wants to go only to France. And I want to go to Germany.’

Interestingly, when talking about contrastive topic markers, Muhamedowa (2016) mentions *al* as one of them (along with *bolsa* and *degen*, which are discussed in following sections), but not *še*. Muhamedowa states that *al* differs from the other contrastive topic markers in that it precedes the contrasted element, rather than follows it. She also claims that it is possible to combine two different contrastive topic markers - *al* and *degen*, or *al* and *bolsa* - with the contrasted element occurring between them (2016: 15).

¹⁵ For a detailed semantic characteristic of contrastive, adversative, and other types of coordination markers, the reader is referred to Malchukov (2004).

Surprisingly, the use of *al* together with *še* is not mentioned by Muhamedowa (2016) at all, although this is a very common co-occurrence, especially where *še* functions as a question particle. In fact, out of the 84 occurrences of *še* where it is clause-final and followed by a question mark, the clause begins with *al* in 33 cases (39.3%); in the ACKL such ‘*al...še?*’ clauses can include up to 8 constituents between the clause initial *al* and the clause final *še*. I do not follow Muhamedowa’s approach to *al*, and consider it a contrastive conjunction.

Since contrastive topics typically appear in the sentence initial position, it follows that the contrastive topic marker *še* appears after the sentence initial constituent. As far as the grammatical function of this sentence initial constituent is concerned, it is usually the topical subject of the sentence expressed by a noun, a proper name, or a pronoun. The contrastive question particle *še* can follow constituents performing many other grammatical functions, as demonstrated in the following sub-section.

This highly restricted distribution of the topic marker *še*, especially compared with that of the question particle *še* discussed in the following section, seems to point to the fact that *še* is on the way out as a topic marker. The results of a corpus study presented in this section corroborate this observation, as only in 3 out of 92 occurrences of *še* in the corpus was it used as a topic marker.

In summary, this sub-section described the distribution of the contrastive topic marker *še* which is rather limited. It follows the sentence initial contrastive topic (expressed by a noun or a pronoun), and is not expected to appear in any other position. Let us now move to examining the contrastive question particle *še*.

5.2.6. Meaning and Distribution of the Contrastive Question Particle *Še*

The contrastive question particle *še* has a much wider distribution than its topic marking variant. As mentioned in Section 5.2.3, it can follow nouns, pronouns, adverbs, nominalised verb forms, as well verbs in conditional and subjunctive moods. As far as the grammatical functions are concerned, this question particle can follow subjects, adjuncts of time, predicates of conditional and subjunctive clauses.

Let us now consider how the contrastive question particle *še* differs from the ordinary question particle *MA*. Some examples of *še* as a question particle are given in (34)-(41), excluding (36). The example in (46) below is from a non-elicited natural conversation which was overheard by me; I gloss *še* as CQ - contrastive question particle.

(46) Speaker A: Men erteñ Astana Opera-ğa
 1SG tomorrow Astana Opera-DAT
 bar-a-miñ. Sen še / *be?
 go-FUT-1SG 2SG CQ Q
 ‘I am going to Astana Opera tomorrow. And what about you?’

Speaker B: Men de bar-a-miñ.
 1SG too go-FUT-1SG
 ‘I am going too.’

In the dialogue above, Speaker A first utters a statement about herself. She then uses the contrastive question particle *še* to ask her interlocutor whether the same statement applies to him. *Še* acting as a contrastive question particle signals a change of topic and asks whether the preceding statement, made about the preceding topic also applies to this new topic, which functions as the subject here. With this in mind, the question with *še* can be roughly paraphrased as ‘does the preceding statement also apply to this new topic which is contrastive to the one before?’.

Another way of constructing this sort of question would be to use the contrastive conjunction *al* in the combination with interrogative prosody on the contrastive topic, as shown in (47).

(47) Men erteñ Astana Opera-ğa bar-a-miñ.
 1SG tomorrow Astana Opera-DAT go-FUT-1SG
 Al sen?

CONJ 2SG

‘I am going to Astana Opera tomorrow. And you?’

It must be noted, however, that, although this variant is acceptable for the native speakers, most of my language consultants had a strong preference for adding *še* to this question even when *al* was present. It is highly likely that the variant ‘*Al sen?*’ is only acceptable due to inference from the Russian language. In Russian, a question of this form is the norm since there are no question particles in this language.

Note the ungrammaticality of the standard question particle *MA* in this sentence, which demonstrates a significant difference between *še* and *MA*. While *še* can refer to and, in a way, ‘copy’ the preceding proposition, *MA* is unable to do so. To form a grammatical question with *MA* we must minimally include the predicate, as shown in (48).

- (48) Men erteñ Astana Opera-ğa bar-a-min. Sen
1SG tomorrow Astana Opera-DAT go-FUT-1SG 2SG
bar-a-sin ba / *še
go-FUT-2SG Q
‘I am going to Astana Opera tomorrow. Are you going?’

Now that the question is no longer just to the topical subject, but to the whole proposition, the question particle *še* cannot be used.

Why is it, then, that *še* can form questions to the conditional and subjunctive clauses, like it does in (40) and (41)? This goes back to the discussion in Section 2.1 of Chapter 2, and Haiman’s (1978) observation that the protases of the conditional clauses are essentially topics. Where the question particle *še* follows a verb in the conditional or subjunctive mood, it constructs a question to a part of a complex clause that would serve as the minimal context for the main proposition. In fact, the same observation applies to questions in which *še* follows an adverbial element. It is noteworthy, that these topical elements are also contrastive, that is, they present a context that is different from some context expressed in the preceding clause.

These general observations can be applied to all the examples containing the question particle *še* that have so far been presented in this section. In (49) below, a schematic representation of the use of this question particle is given. Square brackets contain the necessary context for a *še*-question to be felicitous.

- (49) [Topic + proposition.] Contrastive Topic/Context + *še*? = ‘What proposition follows for the contrastive topic/context?’

The crucial part of this schema is the presence of the preceding topic or context – a question formed with the contrastive question particle *še* cannot appear as an out-of-the-blue, discourse-initial utterance. This demonstrates that in its function as a contrastive question particle *še* is used to form a distinct type of questions to the contrastive topic/context.

As the data presented in Table 21 show, the use of *še* in the function of a question particle drastically prevails its use as a topic marker. This allows us to talk about an almost completed process of *še* becoming exclusively a question particle. This observation is in line with the diachronic and synchronic analyses of standard polar question particles as disjunction elements (cf. Jayaseelan 2008, Bailey 2010, 2013), as it would be logical for a contrastive question particle to develop from a contrastive element.

This grammaticalisation process would also help explain the emergence of another Kazakh contrastive topic marker – *bolsa* – which, like the Turkish *ise*, and the possible Kazakh form *ese*, originated from the conditional form of the verb ‘to be’. *Bolsa* is increasingly used as the contrastive topic marker, but, unlike *še* it is not used as a question particle. It is discussed in more detail in Section 5.3.

5.2.7. Conclusions

It has been shown in this section *še* has two main active functions in the Kazakh language - that of a contrastive topic marker and of a contrastive question particle. Corpus research and data analysis have shown that *še* is most commonly used in the

function of a question particle. This explains the significantly reduced use of it as a contrastive topic marker, although this function is still available in speech and writing, and is accepted by the native speakers. It is possible, however, to make a prediction, that this contrastive particle will cease to function as a contrastive topic marker, and will only be used as a contrastive question particle.

5.3. Topic Marker *Bolsa*

In this section I examine the topic marker *bolsa*. As mentioned in the previous section, this contrastive topic marker appears to be replacing *še* in this function. This section proceeds in the following way: firstly, in 5.3.1 I briefly discuss similar items from related languages. Since not much previous research has been done on *bolsa* as the topic marker in Kazakh, there is no separate sub-section covering this; all information on previous research is included into 5.3.2, which is dedicated to distinguishing the topic marker *bolsa* from the conditional form of the verb *bol-* ‘to be’ and distribution of the topic marker *bolsa*, while 5.3.3 is dedicated to its meaning. The section’s conclusions are presented in 5.3.4.

5.3.1. Topic Marker *Bolsa* in Related Languages

Etymologically *bolsa* originates from the conditional form of the verb *bol-* ‘be’, which is constructed by adding the conditional suffix *-sA* to the root. Gorelova (2005) reports similar topic markers for other Turkic languages, such as Turkmen and Tuvinian, for example.

According to Gorelova, the Turkmen topic marker *bolsa*, originated from the existential verb *bol-* ‘to become’ with the added conditional suffix *-sA*. Turkmen examples in (50) and (51) show this topic marker in use.

- (50) Mugallim bolsa, te olar daga-yanca sered-ip
 teacher TM PCL they go.away-CVB look-CVB
 dur-an.

stand-PTCP.PST(3)

‘As for the teacher, he was standing and watching until they went away.’

- (51) Men bolsa hič yer-e git-žek del.
1SG TM ever place-DAT go.away-FUT NEG

‘As for me, I will not go anywhere.’

(Gorelova 2005:161, citing Dmitriev 1962: 407)

Gorelova reports that the Tuvinian topic marker *bolza* also developed from the existential *bol-* ‘to be’ with the addition of the conditional suffix *-zA*. The examples of this topic marker are presented in (52) and (53).

- (52) Havan bolza hörzün-den cemnen-ir.
pig TM soil-ABL feed-PTCP.FUT

‘As for a pig, it feeds on roots.’

(Gorelova 2005:161)

- (53) Men bolza ton satıp al-ir men.
1SG TM coat buy- PTCP.FUT 1SG

‘As for me, I will buy the coat.’

(Gorelova 2005:161, citing Shamina 2001:104)

It is unfortunate that these examples are provided without context, which would allow us to establish with certainty the precise nature of the Turkmen and Tuvinian topic markers formed from *bol-*, namely, whether they are contrastive topic markers or not. We can suspect, however, that *bolsa* and *bolza* in (50)-(53) are indeed contrastive topic markers, like *bolsa* in Kazakh.

Gorelova also reports similar topic markers, that is topic markers which originated from the conditional form of an existential verb, for Mongolian, Buryat, and many other Altaic languages. A more thorough and in-depth examination of topic

markers in Turkic and Altaic languages is of great interest for further research. Let us now turn to the Kazakh *bolsa*.

5.3.2. Kazakh *Bolsa* – Identifying the Topic Marker and its Distribution

As mentioned previously, the topic marker *bolsa* originates from the conditional form of the verb *bol-* ‘be’. The full conjugation paradigm for *bol-* in the conditional mood is presented in Table 23 below.

Table 23. Conjugation of *bol-* in the conditional mood

Person	Singular	Plural
1	<i>bolsam</i>	<i>bolsaq</i>
2 informal/formal	<i>bolsañ / bolsañız</i>	<i>bolsañdar / bolsañızdar</i>
3	<i>bolsa</i>	<i>bolsa</i>

This verb form is widely used in its primary function, that is, in the protases of conditional sentences. It usually appears as part of a nominal predicate, as shown in the examples in (54) and (55).

- (54) Sen qajtadan keş bol-sa-ñ, men žalğız
 2SG again late be-COND-2SG 1SG alone
 attan-a-mın.
 set.off- FUT-1SG
 ‘If you are late again, I’ll leave on my own.’

- (55) Eger asqan nadan bol-sa, onda ol
 if exceptional ignorant.man be-COND(3) then 3SG
 onı tipti laqtır-ıp tasta-j-di.
 3SG.ACC even throw-CVB throw.away-PRS-3
 ‘And if the man is exceptionally ignorant, he would even throw it away.’

In the example in (54), *bolsa* appears in the second person singular form - *bolsañ* - to agree with its subject, and is used at the end of the protasis of a conditional sentence. The protasis may be optionally introduced by *eger* ‘if’, as in the example in (55), where *bolsa* follows the third person singular nominal subject.

Moving to *bolsa* in the topic marking function, I have only been able to find one description of it in the literature. Muhamedowa (2016) states that *bolsa* follows a contrasted element which must be the subject of the sentence expressed by a noun or a pronoun; her example is given in (56) below.

- (56) Ol bolsa važnij bol-ïp kel-e-di.
 3SG TM important be-CVB come-PRS-3
 ‘As for her/him, s/he comes as if s/he is a very important person.’
 (Muhamedowa 2016:13)

Unfortunately, no context is provided for this sentence to confirm the contrastive topical nature of the element followed by *bolsa*.

The elicited example in (57) below demonstrates the use of *bolsa* as a contrastive topic marker within context: the first sentence contains a proposition about horses, while the second sentence contains a contrastive topic – camel – about which an opposing statement is made.

- (57) Žilqï su-di köp iš-e-di. Tüje bolsa šölge
 horse water-ACC a.lot drink-PRS-3 camel TM thirst
 šïdamdï žanuar.
 tolerant animal
 ‘Horses drink a lot of water. As for camels, they are tolerant to thirst.’

mother-POSS.1SG 1SG.DAT a book bring-PST(3)
 Asem-ge bolsa ešteñe äkel-me-di.

Asem-DAT TM nothing bring-NEG-PST(3)

Intended: ‘My mother brought me a book. As for Asem, she didn’t bring her anything.’

- (61) Keşe biz köp žumıs iste-di-k. Bugün bolsa
 yesterday 1PL much work-PST-1PL today TM
 biz erin-ip tur-mız.
 1PL be.lazy-CVB AUX-1PL

Intended: ‘Yesterday we worked a lot. As for today, we are being lazy.’

As far as the sentential position of the contrastive topic followed by *bolsa* is concerned the same observation can be made as in the preceding section for the contrastive topic marker *şe*. Namely, that contrastive topics occur sentence-initially, which means that the topic marker *bolsa* would always follow a sentence initial contrastive topic.

Let us now consider some further examples, which are problematic for our labelling of *bolsa* as a topic marker in Kazakh. In (62), (63), and (64), topic marker *bolsa* follows the same paradigm as given in Table 15 for the conditional form of the verb *bol-*, and agrees in person and number with the topical subject it follows – second person singular, first person singular, and first person plural respectively.

- (62) [Two company owners talking about a female employee, and one says to the other:]

BauİR-İM, ol bul is-ti bes kün-de
 brother-POSS.1SG 3SG this work-ACC five day-LOC
 tındır-dİ. Solaj emes pe? Sen bolsa-ñ
 finish-PST(3) like.this NEG Q 2SG TM-2SG
 onİ renžit-ip al-dİ-ñ.
 3SG.ACC offend-CVB give-PST-2SG

‘Brother, she finished this work in five days. Isn’t it so? As for you, you offended her.’

(MBT)

- (63) [Isn’t it true that even if a not such a beautiful girl looks at you, your poor heart, struggles and frets like a thief that was caught?]

Al men bolsa-m, samğa-p kel-ip, bir
 CONJ 1SG TM-1SG soar-CVB come-CVB one
 kurs-ta möldire-p otır-ğan otiz qız-ğa
 year-LOC be.pure-CVB seat-PTCP thirty girl-DAT
 tap bol-dī-m otiz tor-ğa širmal-dī-m.
 find.oneself-PST-1SG thirty net-DAT be.caught-PST-1SG

‘And as for me, I found myself caught in thirty nets of thirty pure and tender looking girls amongst which I landed in that year (of studies).’

(ACKL)

- (64) Sol temir-den zeñbirek pen soqa qu-ja-dī.
 this iron-ABL canon and plow forge-PRS-3
 Al biz bolsa-q mal-imiz-diñ
 and 1PL TM-1PL cattle-POSS.3PL-GEN
 soñi-nan qula tüz-de qanğ-ip žür-miz.
 after-ABL flat valley-LOC wander-CVB AUX-1PL

‘They forge canons and ploughs out of this iron. And as for us, we just wander around the vast flat valleys behind our cattle. [And brag that our steppe is so vast.]’

(ACKL)

The agreement of *bolsa* with the subject observed in these examples is problematic, since fully fledged topic markers are not expected to agree with the topic they mark. This indicates that the process of grammaticalisation is not yet complete. We can, however, confidently state that *bolsa* is used as a topic marker, and not the conditional form of the verb *bol-* in the examples above for at least three reasons.

Firstly, if, in the examples above, *bolsa* were to be processed as the conditional form of the verb *bol-*, then the English translations would have to reflect that: instead of ‘as for you’ or ‘as for me’, these phrases would need to be translated as ‘if it is/were you’ or ‘if it is/was me’. None of my language consultants suggested this kind of translation for any of the examples above.

The same can be reported for Uyghur – a Turkic language spoken in the Xinjiang Uyghur Autonomous Region of China –, where the process of grammaticalisation also has not yet finished, and agreement of the topic marker *bolsa* with the topical subject is also still observed, as shown in the example in (65).

- (65) Men bolsa-m hičner-ge bar-ma-j-men.
 1SG TM-1SG nowhere-DAT go-NEG-FUT-1SG
 ‘As for me, I shan’t go anywhere.’

According to my Uyghur language consultant¹⁶ an interpretation where *bolsam* is translated as the conditional verb form is not available here. This is in contrast to the Turkmen example in (51), where, according to a native speaker consultant¹⁷, the conjugated form *bolsam* would be ungrammatical:

- (66) Men bolsa / *bolsam hič yer-e git-žek
 1SG TM ever place-DAT go.away-FUT
 del.
 NEG
 ‘As for me, I will not go anywhere.’

¹⁶ I am grateful to Mirshad Ghalip, a PhD student in anthropological linguistics at Indiana University, Bloomington, for his time and for sharing his native Uyghur language expertise with me during our conversations at the 3rd Conference of Central Asian Language and Linguistics at Indiana University, and a Skype elicitation session.

¹⁷ Many thanks to Leyli Berdiyeva, a native speaker of Turkmen, who kindly agreed to help me verify some examples via an online messenger.

Secondly, where *bolsa* functions as a topic marker, it can be removed without affecting the grammaticality of the string, as the sentence in (67) shows. Here we see the example from (62) repeated without *bolsañ* after *sen* – the sentence is still completely grammatical.

(67) Bauirim, ol bul isti bes künde tindir-di. [...] Sen onı renžitti aldın.

It must be noted, however, that although the sentence does not become ungrammatical without *bolsañ*, it is somewhat not as pragmatically felicitous as with the contrastive topic marker. To make this sentence more pragmatically fitting, the contrastive conjunction *al* can be added, as shown in (68); alternatively, contrastive topicality can be indicated by prosody, but the exact prosodic devices for this require further research.

(68) Bauirim, ol bul isti bes künde tindir-di. [...] Al sen onı renžitti aldın.

The same observations can be made for the examples in (63) and (64).

Let us compare this with a string where *bolsa* functions as a verb, like that in (54), a simplified version of which is given in (69) for convenience.

(69) Sen keš bol-sa-ñ, men žalğiz attan-a-min.
 2SG late be-COND-2SG 1SG alone set.off- FUT-1SG
 ‘If you are late, I’ll leave on my own.’

Unlike in our previous examples, *bolsañ* cannot be omitted from this sentence without causing ungrammaticality, as (70) demonstrates.

(70) * Sen keš, men žalğiz attan-a-min.
 2SG late 1SG alone set.off- FUT-1SG
 Intended: ‘If you are late, I’ll leave on my own.’

The addition of *al* does not rectify the situation, as the ungrammaticality of the string in (71) confirms.

- (71) *Al sen keš, men žalǵız attan-a-mın.
 Intended: ‘If you are late, I’ll leave on my own.’

The third piece of evidence for *bolsa* functioning as a topic marker, and not the conditional form of the verb *bol-* is demonstrated in the examples in (72) and (73) below.

- (72) Eger sen keš bol-sa-ñ, men žalǵız
 if 2SG late be-COND-2SG 1SG alone
 attan-a-mın.
 set.off- FUT-1SG
 ‘If you are late, I’ll leave on my own.’

- (73) *Eger sen bol-sa-ñ, onı renžitti aldın.

As was mentioned earlier, the conjunction *eger* ‘if’ is optional in the protases, however, it can always be added without affecting the meaning or the grammaticality of the sentence. This is illustrated by the example in (72), where *eger* has been added to the conditional string from (69). In contrast, the addition of *eger* to the string from (62) results in an ungrammatical and incomprehensible utterance in (73).

These pieces of evidence suggest that *bolsa* in Kazakh (and by analogy Uyghur) has not yet fully grammaticalised as a topic marker as far as its form is concerned. However, despite still following the same agreement paradigm as the verb form from which it originated, topic marker *bolsa* has a distinct function and distribution pattern, which allow us to identify it as such.

5.3.3. Meaning of the Topic Marker *Bolsa*

This matter has already been touched upon throughout this section, and it has been proposed to treat *bolsa* as a contrastive topic marker. That is to say, it follows a topic expression whose referent is different from that of the topic expression in the preceding proposition.

Consequently, we cannot expect *bolsa* to mark topics of a discourse initial or out of the blue utterance, and indeed this is the case, as infelicity of the sentence in (74) within the identified context demonstrates.

- (74) [Beginning of a lesson about different animals of Kazakhstan.]
Tüje bolsa šölge šīdamdī žanuar.
camel TM thirst tolerant animal
'As for camels, they are tolerant to thirst.'

Neither is *bolsa* felicitous after non-contrastive topics, as shown in (75) below. The context of this utterance is a general question about the hearer's knowledge about camels, which, provided the hearer acts in accordance with Grice's cooperative principle, would warrant continuation of *camels* as the topic of the answer. Since no change of topic occurs, a contrastive topic marker following the topical element is out of place, making the utterance infelicitous.

- (75) [What do you know about camels?]
#Tüje bolsa šölge šīdamdī žanuar.
camel TM thirst tolerant animal
'As for camels, they are tolerant to thirst.'

Changing the context slightly to make 'camels' a contrastive topic results in felicity of the sentence from (75), as demonstrated in (76).

(76) [Can you compare horses with other even-toed ungulates of Kazakhstan?]

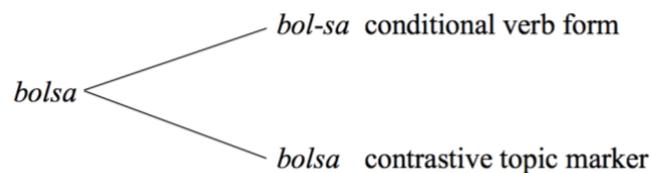
Horses drink a lot of water. Tüje bolsa šölge šīdamdī žanuar.

Ultimately, *bolsa* is a contrastive topic marker, which signals that the referent of the topic expression it follows differs from the referent of the preceding topic expression.

5.3.4. Conclusions

This section has demonstrated that the conditional form of the verb *bol-* ‘to be’ can be used both as a verb in the conditional mood, and as a contrastive topic marker, as represented by the schema in (77).

(77) Functions of *bolsa*



It has been observed that *bolsa* has not yet fully grammaticalised as a topic marker, as it still agrees in person and number with the topical subjects it follows. Three pieces of evidence have been presented to confirm that in these instances we are indeed dealing with a topic marker rather than the conditional form of the verb *bol-*.

Distribution of the topic marker *bolsa* has been discussed, and it has been shown that this topic marker follows contrastive topics, which always appear sentence-initially. These contrastive topics can be arguments or adjuncts; and might be expressed by nouns (common and proper), pronouns, substantivised infinitives, or adverbs. In short, *bolsa* can follow any contrastive topic.

Considering examples from other Turkic languages, such as Turkmen and Tuvinian, where *bolsa* appears to have fully grammaticalised as a topic marker and

does not display agreement with the topic it follows, we can predict that this will eventually also occur in Kazakh. A follow-up study of this matter in some three to five decades, or even later, would be of great interest.

5.4. Topic Marker *Degen*

In this section the Kazakh topic marker *degen*, originating from the past participial form of the verb *de-* ‘say’, is examined. This section proceeds as follows: 5.4.1 introduces topic markers originating from verbs of speech, while 5.4.2. focuses on previous research on *degen*. In 5.4.3 I present an alternative description of word-form *degen* and identify its non-topic-marking uses, and 5.4.4 then concentrates on the distribution patterns and meaning of the topic marker *degen*; sub-Section 5.4.5 concludes this section.

5.4.1. *Verba Dicendi*-related Topic Markers

As mentioned in Section 5.1, Gorelova (2005) notes that many topic markers cross-linguistically have been observed to originate from *verba dicendi* – verbs denoting a speaking activity. One of the examples she provides is from the Manchu language (Tungusic), in which the topic marker *serengge* developed from the verb *se-* ‘to say’ with the addition of the imperfective participle suffix *-re* and the nominalising suffix *-ngge*. The example in (78) below demonstrates this topic marker in use.

- (78) *menggun* *jiha* *serengge* *ergen* *be*
 silver money(coin) TM life ACC
 uji-re *sekiyen*
 nourish-PTCP.IPFV source

‘As for silver and coins, [they are] the source of nourishing a life.’

(Gorelova 2005: 159)

Matić and Pakendorf (2013) also report the use of verbs of speech in topic marking function for North Mongolian, Buryat, Sakha, Dolgan, Shor, and Tuvan. The Tuvan example in (79) demonstrates the use of the present participial form of the verb *de-* ‘say’ in the topic marking function.

- (79) Hamik küdee-ler de-er-ge čïg-l-ïp
all suitor-PL say-PRS.PTCP-DAT gather-PASS-CVB
kel-gen (...)
come-PST(3SG)

‘As for all those suitors, they have already gathered (...).’

(Matić and Pakendorf 2013: 383, citing Harrison 2005)

In Uyghur, the participial form of the verb *de-* ‘say’ is also used as a topic marker, as the example below demonstrates. This example was provided by the same Uyghur language consultant as previously – Mirshad Ghalip.

- (80) Pizza degen, bu resturan-da bek oxshaydu.
pizza TM this restaurant-LOC very tasty

‘As for pizza, it is very tasty in this restaurant.’

Let us now move to the Kazakh *degen*, and consider previous research on this item in the following sub-section.

5.4.2. Previous Research on *Degen*

Very little research has been carried out on the topic marker *degen*; this sub-section is based on Muhamedowa (2011) and Muhamedowa (2016), as these appear to be the only mentions of this topic marker in the literature to date.

The word form *degen* is the past participial form of the Kazakh verb *de-* ‘say’, formed by the addition of the past participial suffix –*GAn* to the stem: *de-gen* ‘saying’

or ‘said’. This verb form is most commonly used as shown in the examples in (81) and (82).

(81) [An angry father replies to the pleas to help his son Ahmet whom he had disowned:]

Meniñ	Ahmet de-gen	ul-im	žoq!
1SG.GEN	Ahmensay-PTCP	son-POSS.1SG	NEG

‘I do not have a son called Ahmet!’

(MBT)

(82) Ajgerim “Kim köp bāliš že-j-di” de-gen
 Ajgerim who a.lot pie eat-FUT-3 say-PTCP
 sajīs-qa qatīs-a-dī.
 competition-DAT participate-FUT-3

‘Ajgerim will participate in the competition “Who eats most pies”.’

(KT)

In both (81) and (82) *degen* performs a linking function: it links the head nouns (*ulim* and *sajīs* in (81) and (82) respectively) with the modifying phrases. The use of *degen* is not optional here and the sentences become ungrammatical if it is removed. Note, however, that *degen* is not necessarily overtly present in the translations of these types of constructions into other languages, as in (82).

Muhamedowa (2011) refers to this as the ‘basic function of the participle *degen*’, and shows that the same is observed in such Turkic languages as Krymchak and Tatar, as in (83) and (84) respectively.

(83) Gäl-di-ler Gärar de-gen šeher-gä.
 arrive-PST-3PL Gerar say-PTCP town-DAT
 ‘They arrived in the town called Gerar.’

(Muhamedowa 2011: 264 citing Erdal 2002: 123)

- (84) Avıl-ibüz-da Äxmät di-gän ber batır
village-POSS.1PL-LOC Äxmät say-PTCP one brave
eget bar.
young.man EXST
‘In our village, there is a brave young man called Äxmät.’
(Muhamedowa 2011: 264)

Matić and Pakendorf (2013) describe noun phrases with these internal, non-canonically used verbs of speech as appositive constructions, where the head noun is preceded by another noun plus a verb of speech; this other noun denotes the name of the referent of the head noun. They report these constructions for several Siberian languages, such as Dolgan, Eastern Evenki, Tuvan and Shor. An example from Eastern Evenki is given in (85) below.

- (85) Tyvevul gune-r i: ami:ka:n-mi
Tyvevul say-IPFV.PTCP grandfather-1SG
ulguče:n-če:-n.
tell-PST-3SG
‘This story was told by my grandfather (called) Tyvevul.’
(Matić and Pakendorf 2013: 380 citing Romanova and Myreeva 1964)

Göksel and Kerslake (2005) report the same function for the converbial form *diye* of the Turkish verb *de-* ‘say’. They refer to this use of *diye* as the ‘adjectival function’, thus differentiating it from the function of a subordinator which they also assign to this word form. Göksel and Kerslake underline that adjectivals constructed with *diye* are used exclusively with indefinite noun phrases, and are not found in formal registers; the authors provide the examples in (86) and (87) to illustrate this use of *diye*.

- (86) Bizim mahalle-de Güneş Gıda di-ye
1PL.GEN neighbourhood-LOC Güneş Gıda say-CVB

bir market var.
 one shop EXST

‘In our neighbourhood there’s a self-service shop called Güneş Gıda.’

(87) Mehmet di-ye bir oğlu daha var.
 Mehmet say-CVB one son more EXST

‘S/he’s got another son, called Mehmet.’

(Göksel and Kerslake 2005: 175)

Where *diye* in Turkish performs subordinating or complementizing functions, the equivalent *dep* (de-CVB) is used in Kazakh. Thus, Turkish *diye* performs the functions of both *dep* and *degen*. Verbs of speech functioning as complementizers are not uncommon and have been reported, for instance, for many Siberian languages, such as Evenki, Udihe, Yakut, Buriat, Yukaghir, Nanai, Shor, and others (see Nikolaeva, 2005; Matič and Pakendorf, 2013 for further details).

Muhamedowa (2011) provides another example where *degen* is used in what she refers to as the ‘basic function’ - it is presented in (88) below.

(88) “Äže” ata-n-sa-m de-gen
 grandmother name-PASS-COND-1SG say-PTCP
 arman-ïm bar.
 dream-POSS.1SG EXST

‘I have the dream that is to be referred to as ‘grandmother’.’

(Muhamedowa 2011: 265)

In this example *degen* once again links the head noun with a modifier, the difference with the previous examples being that the modifier in (88) is expressed by a finite clause *äže atansam*. Muhamedowa (2011: 265) claims that “the syntactic function of *degen* is akin to the function of a conjunction that links the finite clause *äže atansam* with the head *armanim*”. In the same way as in the examples in (81) and (82), the use

of *degen* in (88) is syntactically motivated/non-optional. Muhamedowa’s reference to it as a ‘conjunction’ here is not quite clear, however.

Apart from this ‘basic’ function of *degen*, Muhamedowa claims that this word form functions as an emphatic particle, topic particle and a contrast marker. Let us consider all these reported functions in more detail.

Muhamedowa provides the examples in (89) and (90) to illustrate the use of what she refers to as the emphatic particle *degen*.

- (89) Oqu-ğa degen ïnta-m-dï
 read-DAT degen interest-POSS.1SG-ACC
 žoğalt-pa-dï-m.
 lose-NEG-PST-1SG
 ‘I haven’t lost my interest in reading (books).’

- (90) Marat-tiñ degen mäšine-si su žaña.
 Marat-GEN degen car-POSS.3SG water new
 ‘Marat’s car is brand new.’

(Muhamedowa 2011: 266)

In (89), *degen* is inserted between two nouns out of which *ïnta* ‘interest’ is the head noun, and *oqu* ‘read / reading’ is its complement marked with the dative case; other Kazakh nouns that require dative-marked nominal complements are, for example: *täbet* ‘appetite’, *qažettilik* ‘necessity’, *reniš* ‘offence’, amongst many others.

According to Muhamedowa, where *degen* appears between the dative-marked complement and the head noun, as in (89), it is used as an emphatic particle. She further observes that *degen* is not syntactically obligatory here and can be omitted without making the sentence ungrammatical. She continues to state that “the function it [*degen*] fulfils is a pragmatic one: emphasizing the dative complement” (2011: 266). It is difficult to gauge the emphasis *degen* is claimed to add to the sentence in (89) from the translation provided by Muhamedowa and lack of context; contextualised examples of this use of *degen* are provided in Section 5.4.2 below.

The example in (90) represents another instance of the use of *degen* as an emphatic particle, which Muhamedowa describes as “very remarkable” (2011: 266). In (90), *degen* is inserted between the possessor *Marat* in genitive case, and the possessed *māšine* marked with the possessive suffix which agrees with the possessor in person and number. Muhamedowa states:

It is hard to find an adequate translation of *degen* here. It can probably be compared with the adverb *very* in colloquial English; hence (90) would be rendered in English as ‘The car of this very Marat is brand new’. However, in contrast to the English *very*, *degen* does not convey negative connotations. As a desemanticized item it rather emphasizes the possessor *Marat* without expressing any positive or negative attitude of the speaker to the proposition (2011: 267).

Another example Muhamedowa provides to demonstrate the use of *degen* as an emphatic particle in possessive constructions is given in (91) below.

- (91) Türkistan-niñ degen bedeldi kelin-der-i-niñ
Turkestan-GEN degen influential bride-PL-POSS.3-GEN
bir-eu-i bol-di qazir.
one-COLL-POSS.3 be-PST(3) now
‘Now she became one of the influential brides of the very Turkestan.’
(Muhamedowa 2011: 267)

Similarly to the examples in (89) and (90), *degen* can be removed from (91) without affecting its grammaticality. Muhamedowa states that “the reason for its [*degen*] use is pragmatic. With the help of *degen*, the speaker highlights the possessor *Türkistanniñ*” (2011: 267).

One more example of *degen* used as an emphatic particle is given in (92), where it follows a noun in genitive case which has a part-whole relationship with its head. Muhamedowa (2011) underlines that only a genitive marked noun with a generic reference can be followed by *degen*.

- (92) Tau-diñ degen bök-ter-i žap-žasıl.
 mountain-GEN degen slope-PL-POSS.3 very.green
 ‘Slopes of mountains are very green.’

While she does not discuss any constraints on the use of *degen* after dative nominal complements (as in (89)), Muhamedowa claims that usage of emphatic *degen* with genitive attributes is “tricky and constrained”, and that “not every possessor can be emphasized by *degen*” (2011: 267).

According to Muhamedowa, then, the emphatic *degen* cannot be used in “possessive constructions that form a compound or a fixed expression” (2011: 267), or with “possessors that are not referred to by a proper name” (2011: 268), although if the possessor has a generic reference, i.e., it refers to the whole class of entities, the use of *degen* is acceptable.

Having consulted with at least 10 native speakers of Kazakh, I found that the examples in (90)-(92) were only marginally acceptable for 3 out of 10 speakers, who emphasised that these sentences pertained to the highly colloquial speech register. The use of *degen* after a dative complement, as in (89), is acceptable for all the speakers in appropriate contexts.

ACKL research corroborated the language consultants’ comments. The use of *degen* after nouns in genitive case as shown in (90)-(92) has not been attested by the corpus data, while numerous examples of the use of *degen* after a dative complement have been found in the ACKL; this is discussed in more detail in Section 5.4.2.

Moving to *degen* functioning as a topic particle, Muhamedowa provides the following examples to demonstrate this use.

- (93) Ol degen kögerip-sazar-ıp qaš-ıp
 3SG degen go.green.with.anger-CVB dash.away-CVB
 šiq-ti.
 AUX-PST(3)
 ‘He turned blue in the face from anger and dashed away.’

- (94) Qazir žumīs tab-u degen öte qiin ğoj.
 today work find-INF degen very hard really¹⁸
 ‘Today it is really very hard to find a job.’

(Muhamedowa 2011: 269)

In the example in (93) *degen* follows the subject of the sentence expressed by the pronoun *ol*; in (94), it follows the subject expressed by the infinitive clause *žumīs tabu*. Muhamedowa observes that “[A]s in the case with emphatic *degen*, we are apparently dealing with the semantically empty *degen* here. [...] *degen* is syntactically not obligatory and can be omitted without violating the structure of the sentence” (2011: 269). Muhamedowa notes that the Turkish *diye* cannot be used in the same way as *degen* in (93) and (94).

In her attempt to establish the function of *degen* in sentences like those in (93) and (94), Muhamedowa refers to Erkü’s (1983) work on information structure of Turkish, and claims to apply two of Erkü’s topichood tests to the Kazakh examples. The first test Muhamedowa mentions involves the discourse context, that is the context in which an utterance is anchored. For some reason, this test is only mentioned by Muhamedowa, but not actually applied to the examples in (93) and (94); the context in which the examples from (93) and (94), as well as some other examples containing the topic marking *degen* are given in Section 5.4.3.

The second test for determining a topical expression involves rephrasing a sentence following the pattern presented in (95) below, where *X* will be the topical constituent.

- (95) X turali äñgimeles-se-k...
 X about talk-COND-1PL
 ‘If we talk about X... / talking about X...’

¹⁸ I keep the original glossing of *ğoj* as given in Muhamedowa (2011), although I do not agree with it - the detailed analysis of *ğoj* is provided in Chapter 6.

Rephrasing both (93) and (94) following the schema in (95) results in (96) and (97) below, and demonstrates that *ol* and *žumīs tabu* are topics according to this test.

(96) Žumīs tabu turalī āṅgimelessek...

(97) Ol turalī āṅgimelessek...

Muhamedowa concludes that since *degen* follows the topics in (93) and (94), and “has neither syntactic nor semantic functions here, its use is only pragmatically determined: *Degen* signals the topic of the utterance” (2011: 270). In its function as a topic marker *degen* scopes over the constituent to its immediate left. Muhamedowa links this function of *degen* with Amirov’s (1972) observation that the main function of *degen* is to put an emphasis on the ‘subject of thought’.¹⁹ Muhamedowa uses the example in (98) below as further evidence that *degen* is a topic marker, as it can only appear after the topical *Säule*, and not after the focused *pasportin*.

(98) [What has Saule lost?]

Säule	degen	pasport-ï-n	*degen	žoḡalt-ti.
Saule	degen	passport-POSS.3-ACC		lose-PST(3)
‘Saule has lost her passport.’				

Interestingly, Muhamedowa also uses the example from (92) to confirm the topic marking function of *degen*, even though *degen* was classified as the emphatic particle in that example. She states that *degen* can be used in the example in (92) only because the phrase *taudiñ bökteri* ‘mountain slopes’ is the topic of that sentence, and that when the same phrase is focused, the addition of *degen* between its constituents is unacceptable.

¹⁹ The term *predmet mysli* ‘subject of thought’ was used in Soviet linguistics interchangeably with the term ‘topic’.

Muhamedowa provides the example in (99) to illustrate the use of *degen* as a contrast marker; she defines contrasting as “when two pieces of information are contrasted within a single utterance” (2011: 271).

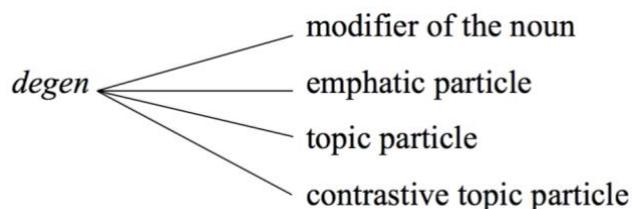
- (99) Almatī taulī, Pekin degen žazīq.
 Almaty mountainous Beijing degen flat
 ‘Almaty is mountainous, Beijing is flat.’

According to Muhamedowa, the topic *Pekin* ‘Beijing’ is contrasted with *Almatī* ‘Almaty’, and *degen* follows and takes scope over *Pekin*. The author proceeds to compare the contrast marker *degen* with the Turkish *ise / -(y)sA*, which were discussed in Section 5.2.2. The Turkish equivalent of the sentence in (99) is given in (100).

- (100) Almatı dağlık, Pekin ise düzlük.
 (Muhamedowa 2011: 271)

The schema in (101) below illustrates the four functions *degen* performs according to Muhamedowa’s analysis of this word form.

- (101) Four functions of *degen*, according to Muhamedowa (2011)



Muhamedowa (2011) concludes that the main differences between the four functions of *degen* are to do with the syntactic and semantic obligatoriness, the meaning of the verb *de-*, and the availability of what she refers to as the passive reading (that is, when *degen* is translated as ‘called’ as in (81), for example). In the sub-section below, I

present an alternative approach to analysing *degen*, and propose that there are only two rather than four types of *degen*: the participial form, and the topic marker.

5.4.2. Non-topic-marking Uses of *Degen*

In this sub-section I consider the word form *degen* from a wider perspective – that is in its canonical and non-canonical uses. Following Matic’ and Pakendorf (2013), I define canonical use of a word as directly relating to its primary meaning, and the non-canonical use as not relating to it.

The primary meaning of *degen* is in relation to the verb *de-* ‘say’, from which this word form originates, and is observed in its use as the predicate. One of non-canonical uses of *degen* is observed in constructions where it is used as a linking element between a head noun and its modifier, which can be expressed by another noun, or a finite clause. Another non-canonical use of *degen* is its use as a topic marker, which is discussed in detail in a separate sub-Section 5.4.3.

The primary canonical use of *degen* is presented in (102) and (103) below, where it acts as the predicate of the main clause.

(102) [I thought I’d be working near tractors in the spring and will be able to pick up the skill of driving it myself.]

Audan-da	mağan	solaj	de-gen.
district-LOC	1SG.DAT	like.this	say-PST(3)

‘That’s what they told me in the district.’

(103) Ana-miz biz-ge bilaj de-gen: “Köp
 mother-POSS.1PL 1PL-DAT so say-PST(3) many
 söz – kumis, az söz – altın.”
 word silver few word gold
 ‘Our Mother told us this: “Speech is silver, silence is golden”.’

(ACKL)

Degen in these examples appears in the third person singular or plural distant past tense form; the full conjugation paradigm of the verb *de-* in the distant/*-ĜAn* past tense is given in the table below.

Table 24. Conjugation of the verb *de-* in the *-ĜAn* past tense

Person	Singular	Plural
1	<i>de-gen-min</i>	<i>de-gen-biz</i>
2	<i>de-gen-siñ / de-gen-siz</i>	<i>de-gen-siñ-der / de-gen-siz-der</i>
3	<i>de-gen</i>	<i>de-gen</i>

As mentioned in the preceding section, *degen* can be used as a linking element between a head noun and its modifier in the sentences in (82), (83), and (88); further examples are presented in (104) and (105). I gloss this form of *degen* as a participle, as it appears to have originated from the participial form, however, this requires further analysis and clarification, which I leave for further research.

- (104) Men “Šiğis” de-gen žornal-dī oq-ïp
 1SG Šiğis say-PTCP magazine-ACC read-CVB
 otır-mīn.
 AUX-1SG
 ‘I am reading Šiğis magazine.’

- (105) Äueli London-da žülde al-sa-m de-gen
 firstly London-LOC prize take-COND-1SG say-PTCP
 maqsat-im-niñ kül-talqanı šiq-ti.
 goal-POSS.1SG-GEN shatter.to.pieces-PST(3)
 ‘Firstly, my dream of winning/to win a prize in London shattered to pieces.’

(ACKL)

In (104) and (105), *degen* is used to link the head nouns *žornal* and *maqsat* with their modifiers *šigīs* and *Londonda žülde alsam* respectively. Note, that *degen* is not directly reflected in the English approximation of the sentence, as can be expected for an auxiliary element, but not for a lexical item.

Removing *degen* from these sentences would result in ungrammaticality, as the head nouns and phrases modifying them would not be appropriately syntactically connected. If *degen* were to be removed from (104) and (105), the sentences would need to be adjusted accordingly in order to maintain grammaticality, as shown in (106) and (107).

(106) Men “Šigīs” žornal-*i*-n oqip
 1SG Šigīs magazine-POSS.3-ACC read-CVB
 otir-min.
 AUX-1SG
 ‘I am reading Šigīs magazine.’

(107) Äveli London-da žülde al-u turali
 firstly London-LOC prize take-INF about
 maqsat-*im-niñ* kül-talqan-*i* šiq-ti.
 intention-POSS.1SG-GEN ash-dust-POSS exit-PST(3)
 ‘Firstly, my dream of winning/to win a prize in London shattered to pieces.’

In (106) the possessive suffix *-i* has been added to the head noun *žornal* to indicate its link with the preceding proper name; in (107), the conditional form of the verb *al-* is no longer grammatical, and the infinitival form must be used instead followed by a postposition which links the non-finite clausal modifier with the head noun *maqsat*.

Although the ‘linking’ *degen* itself does not mark a particular information-structural status of the constituent it follows, the phrases that follow the pattern noun-*degen*-noun would usually introduce a referent that is expected by the speaker to be entirely new for her interlocutor. Using this formulation to refer to a well-known entity

comes across as patronising and condescending, and would provoke a certain reaction, as illustrated in the short exchange in (108).

(108) A: Where in England do you live?

B: London de-gen qala-da tur-a-min.
London say-PTCP city-LOC live-PRS-1SG
'I live in the city called London.'

A: Why are you saying it like that? Everyone knows about London!

Speaker A in (108) takes exception to Speaker B's use of the noun-*degen*-noun phrase to talk about London as this sounds as if Speaker B does not expect Speaker A to have ever heard that such a city exists. If, however, the name of not so well-known English city or town (Malvern, for example) were to be used here instead, the sentence would not sound patronising at all, on the contrary, Speaker B would come across as a polite and considerate conversationalist. Notably, *degen* is not optional in these noun-*degen*-noun phrases, as demonstrated further in this sub-section.

Contrary to Muhamedowa (2011, 2016), I propose that what she refers to as the emphatic use of *degen* – *degen* used between the dative complements of some nouns and the nouns themselves nouns (as in (89) and (90)) – follows the same pattern as shown in the examples in (82), (83), (88), (104) and (105) above. That is to say, *degen* is used as an attributiviser, which turns a dative noun phrase into a modifier. Some more examples of this non-canonical use of *degen* are presented further below followed by a discussion.

Before we proceed to that, it must be noted, that this is not uncommon cross-linguistically, as the Hungarian examples in (109)-(111) below demonstrate.²⁰

²⁰ I am grateful to András Bárány for his time, and for sharing his native speaker knowledge in our elicitation session at SOAS in May, 2018, during which these examples were collected.

(109) Élvez-t-em a beszélgetés-t Mari-val.
 enjoy-PST-1SG the conversation-ACC Mari-COM
 ‘I enjoyed the conversation with Mari.’

(110) Élvez-t-em a Mari-val való
 enjoy-PST-1SG the Mari-COM be.PTCP
 beszélgetés-t.
 conversation-ACC
 ‘I enjoyed the conversation with Mari.’

(111) *Élvez-t-em a Mari-val beszélgetés-t.
 enjoy-PST-1SG the Mari-COM conversation-ACC
 Intended: ‘I enjoyed the conversation with Mari.’

In (109), proper noun *Mari* in comitative case form follows the head noun *beszélgetés* ‘conversation’, which it modifies. In (110), the same proper noun precedes the head noun it modifies, with a participial form of the verb ‘to be’ appearing between the modifier and the head in much the same way as *degen* appears between two nouns in (89), (90), and (108) for instance.

Ungrammaticality of (111) confirms that the use of the participle in the constructions where the modifier precedes the head noun is not optional in Hungarian, in the same way as in Kazakh examples of noun-*degen*-noun phrases. I hypothesise that this will eventually become the case for Kazakh noun-DAT-*degen*-noun phrases, to which I return in the examples below.

(112) Orta ğasir-lar-dıñ soñında Batis žäne
 middle century-PL-GEN after Western and
 Ortaliq Europa-da širkeu-ge de-gen qarım-qatınas
 CentralEurope-LOC church-DAT say-PTCP attitude
 özger-di.
 change-PST(3)

‘After the Middle Ages, the attitude to church changed in Western and Central Europe.’

(ACKL)

- (113) Bizdiñ žüreg-imiz-de Otan-ğa,
 1PL.GEN heart-POSS.1PL-LOC Fatherland-DAT
 ana-ğa, süjgen žar-ğa de-gen šeksiz
 mother-DAT beloved wife-DAT say-PTCP endless
 mahabbat bol-dï.
 love be-PST(3)

‘In our hearts, there was endless love to our Fatherland, mother, and beloved wife.’

(ACKL)

- (114) Jağni til-di meñger-u üšin ülken inta,
 that.is language-ACC acquire-INF for big passion
 zejın, talap, oquşı-niñ til-ge
 dedication aspiration student-GEN language-DAT
 de-gen bejimdilig-i men qızıǵuşıliǵ-ı
 say-PTCP propensity-POSS.3 with interest-POSS.3
 kerek.
 need

‘That is to say, that to learn a language, passion, dedication, aspiration, as well as the student’s interest and propensity to the language are required.’

(ACKL)

- (115) [The title of a short story:]
 Pizza-ğa de-gen quštarlıq.
 pizza-DAT say-PTCP passion
 ‘Passion for Pizza.

(KT)

All the examples above have been translated and discussed with 7-10 native speakers of Kazakh. The English interpretations provided for the sentences were arrived at with the help of the language consultants (via Russian), and, as can be seen, none of them include any form of emphasis on the dative complement. All the language consultants agree that *degen* is optional in these sentences, that is, it can be removed without making the sentences ungrammatical. Many of the language consultants noted, however, that they preferred *degen* to be included. One consultant observed that even when *degen* is removed, she felt like implicitly it was still there, and like there should be something linking the nominal dative modifier and the head noun.

Another observation that has led me to proposing that *degen* performs a linking, rather than an “emphatic” function in these types of structures comes from corpus (ACKL) research. Recall that Muhamedowa claims that *degen* functions as an emphatic particle not only after the dative marked complements but also after the genitive case marked possessors, as in (90) and (91). The sentences in (90) and (91) have not been accepted by most of the native speakers I consulted, and most of them suggested removing *degen* altogether, or placing it after the genitive-possessive phrase – thus, effectively making it a topic marker –, as shown in (116) below, on the example from (90).

- (116) Marat-tiñ mäšine-si degen su žaña.
 Marat-GEN car-POSS.3SG degen water new
 ‘Marat’s car is brand new.’

Searching the ACKL for instances where *degen* follows a dative case marked noun and a genitive case marked noun returned the following results: 404 matches for the former search, and 7 matches for the latter. All 7 strings where a Genitive marked noun was followed by *degen* were of the type shown in (117) below.

- (117) Türik-te: “Qarn-iñ toj-ğan žer -
 Turk-LOC stomach-POSS.2SG eat.to.satiety-PTCP
 Otan-iñ” de-gen šöz bar.

Fatherland-POSS.2SG say-PTCP word EXST
 ‘The Turks have a saying: “Your homeland is where your belly is full”.’

In (117) and 6 other sentences where *degen* followed a noun in the genitive case, that noun appears as part of a clause all of which serves as a modifier to a head noun; once again, in all these instances *degen* performs a linking or attributivising function, as shown and discussed above. As the use of *degen* as an emphatic particle after the genitively marked possessor and the possessive marked head noun is: a) not accepted by the vast majority (8 out of 10) of the native speakers I have consulted; and b) is not attested in the ACKL, I do not examine this proposed usage of *degen* any further.

However, it is clear that *degen* is indeed widely used between a dative case marked nominal complement and a head noun. The examples like those in (112), (113), and (114) are representative of the majority of the 404 cases of *degen* following a dative marked noun in the ACKL.

As mentioned earlier, I do not follow Muhamedowa’s analysis of this use of *degen* as an emphatic particle and propose instead that *degen* in these cases also performs the function of an attributiviser – a link between the head noun and its dative marked modifier. Recall, that in a canonical case of a noun modifying another noun (the former usually being a proper name), the head noun receives a possessive marker, as in the examples (a) and (b) in (118).

- | | | | |
|----------|--|----|---|
| (118) a. | Astana qala-si
Astana city-POSS.3
‘Astana city’ / ‘the city of Astana’ | b. | Dostiġ köŝe-si
Dostiġ street-POSS.3
‘Dostiġ street’ |
| c. | Astana degen qala | d. | Dostiġ degen köŝe |

As described previously, these noun phrases can also be expressed with the help of the participle *degen*, as in the examples (c) and (d) in (118). It seems, then, that the speakers are increasingly following the pattern of using *degen* in the noun phrases

where the head noun is unmarked (to show the connection with its complement), to compensate for this lack of marking which they are used to seeing in the other noun phrases with nouns as modifiers. Although the removal of *degen* from the ‘dative nominal modifier-unmarked head noun’ phrases is not considered ungrammatical, the majority of my language consultants expressed a strong preference for the presence of *degen* in such phrases.

A process of grammaticalisation seems to be ongoing, and the participle *degen* is becoming an attributiviser which connects a noun in dative case acting as a modifier with an unmarked head noun it modifies. Note, that this also explains the fact that *degen* placed between the genitive marked possessor and the possessive marked head noun is largely unacceptable for the native speakers – there simply is no need or space for an attributiviser within a possessive construction.

The results of the corpus search and consultations with native speakers show that the use of *degen* to which Muhamedowa refers as the “emphatic particle”, is, in fact, the participial form of *degen* used as a linking element/attributiviser. A prediction can be made that *degen* will become an obligatory element to appear between dative complements and unmarked head nouns.

5.4.3 Topic Marker *Degen* – Distribution and Meaning

As a topic marker, *degen* can mark old and new topics, which can, but do not have to be contrastive; contrary to Muhamedowa’s (2011) observations, I demonstrate that *degen*, unlike *bolsa*, is simply a topic marker, not a contrastive topic marker.

Some examples of topic marking use of *degen* were given earlier in (93) and (94), and some further examples are presented in (119) and (120) below.

(119) A: Marat-tiñ mašine-si qandaj?
 Marat-GEN car-POSS.3 like.what
 ‘What’s Marart’s car like?’

B: Marat-tiñ mašine-si degen su žaña.
 Marat-GEN car-POSS.3 TM water new

‘As for Marat’s car, it’s brand new.’

- (120) [In a conversation between a hotel owner and a guest about the bill, the owner says the guest has to pay for the fruit that were present in the guest’s room every day even if the guest did not touch them. The guest replies that the hotel owner has to pay for kissing his (the guest’s) wife, to which the hotel owner replies:]

H.O.: Biraq men sizdiñ äjel-iñiz-di
But 1SG 2SG.GEN wife-POSS.2SG-ACC
süj-gen žoq-pin!
kiss-PTCP NEG-1SG
‘But I did not kiss your wife!’

G.: Bul sizdiñ šarua-ñiz. Ol degen
this 2SG.GEN business-POSS.2SG 3SG TM
kün sajın nömır-de bol-di emes
day every hotel.room-LOC be-PST(3) NEG
pe?

Q

‘That’s your business. She was present in the hotel room every day, isn’t it so?’

Speaker B’s response in (119) is the altered sentence from Muhamedowa’s example originally provided in (90). This alternation was suggested by some of my language consultants when the example in (90) was presented to them. They were then asked to come up with an appropriate context for the altered sentence (that is, after they indicated their preference for *degen* to appear after the genitive-possessive phrase and not inside of it), and they all suggested something like what is given as Speaker A’s question in (119). This makes for a clear, contextualised example of *degen* being used as a topic marker.

In the second sentence of the Guest's response in (120), the pronoun *ol* which is the topic of the sentence is followed by the topic marker *degen*. The context, and especially the preceding Hotel Owner's remark, in which the Guest's wife is explicitly mentioned, help to clearly identify *ol* as a continuous topic in the Guest's second sentence.

Additionally, the constituents followed by *degen* in both (119) and (120) pass Erkü's (1983) topichood test, which is outlined in the previous section and demonstrated in (95), (96), and (97). Following the schema from (95), we can rephrase the sentences with *degen* from (119) and (120) as shown in (121) and (122) respectively.

(121) Marat-tiñ mašine-si turali äñgimeles-se-k /
 Marat-GEN car-POSS.3 about talk-COND-1PL
 ajt-sa-k...
 speak-COND-1PL
 'Talking/speaking about Marat's car...'

(122) Ol turali äñgimeles-se-k / ajt-sa-k...
 3SG about talk-COND-1PL speak-COND-1PL
 'Talking/speaking about her...'

Returning to Muhamedowa's examples in (93) and (94), recall that those were presented out of the discourse context, which the author claimed to be the first test of topichood. The felicitous contexts for (93) and (94) are presented in (123) and (124) below, and they also confirm the topical nature of the constituents followed by *degen*.

(123) A: Sodan kejin ne bol-di? Ol ne
 afterwards what be-PST(3) 3SG what
 iste-di?
 do-PST(3)
 'What happened after? What did he do?'

B: Ol degen kögerip-sazarip qašip šiqtī.
 ‘He turned blue in the face from anger and dashed away.’

(124) A: What do you think about the current job market? Is it easy to find a job?

B: Qazir žumis tabu degen öte qiin.
 ‘It is really hard to find a job these days.’

Unlike attributiviser *degen*, topic marker *degen* can follow constituents from different grammatical categories (including pronouns and non-finite clauses) as long as they are topical. This demonstrates the clear distinction between the participial *degen* and the topic marking *degen*: the distribution of the former is restricted syntactically, and of the latter – pragmatically.

As far as the grammatical functions of topical elements that can be followed by *degen* are concerned, once again, there are no syntactic restrictions: if a constituent can be a topic, it can be followed by *degen*. Thus, *degen* can follow a topical subject, as shown in the examples above, a topical direct or indirect object, as in (125) and (126) respectively, and adverbial element as in (127).

(125) [Have you read this book?]

Mīna kitap-tī degen oqī-ğan žoq-pīn.
 this book-ACC TM read-PTCP NEG-1SG
 ‘This book, I’ve not read it yet.’

(126) [Will you go to the theatre tonight?]

Teatr-ğa degen bargī-m kel-me-j-di.
 theatre-DAT TM go-NMLZ-1SG want-NEG-PRS-3
 ‘As for the theatre, I don’t want to go there.’

(127) [Are we meeting tomorrow?]

Erteñ degen kel-e al-ma-j-mīn...

tomorrow TM come-CVB can-NEG-FUT-1SG

‘As for tomorrow, I cannot come...’

As discussed in Chapter 3, Kazakh is generally a topic prominent language and the topical element tends to appear sentence-initially, hence topic marker *degen* usually follows the sentence initial topical element. Matic’ and Pakendorf (2013) observe the same for the Siberian languages they examine; they note that where forms of verbs of speech are used as topic markers they are “attached directly to the topic expression, which is moved to the left edge of the clause” (2013: 383).

I analyse what Muhamedowa (2011) refers to as the contrast marking function of *degen* as the same topic marking function as described above. I posit that *degen* itself does not carry a contrastive meaning, but is simply a topic marker, which might occasionally appear in contrastive contexts, as shown in Muhamedowa’s example in (99). It must be noted though, that there was a strong preference among the language consultants to replace *degen* with *bolsa* when presented with the example in (99), or at least add the contrastive conjunction *al* before the contrastive topic of the second clause of the sentence.

One of the pieces of evidence against analysing *degen* as a contrastive topic marker, lies in its ability to mark a topic in a discourse-initial or an out of the blue sentence, where no change of topic occurs; this is exemplified by the example in (128). As we saw in sub-section 5.3.3, this is not possible for *bolsa*.

(128) [Discourse-initial utterance:]

Ömir degen ülken supermarket, ne qala-j-siñ,
life TM big supermarket what want-PRS-2SG
bäriñ al.
all take.IMP

‘Life is a big supermarket, take whatever you want.’ [But do not forget, that there is a check out ahead, and you have to pay for everything.]

(KT)

Another piece of evidence for *degen* being a non-contrastive topic marker is provided in the example in (129), where the sentence from (128) serves as the answer to the question about ‘life’. Recall, that the contrastive topic marker *bolsa* was not felicitous in such cases of continuous topic.

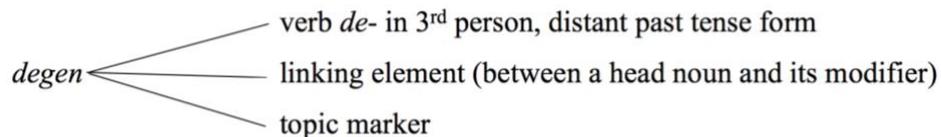
- (129) [What do you think about ‘life’?]
 Ömir *degen* *ülken* supermarket...
 ‘Life is a big supermarket...’

Corpus data present yet another piece of evidence against analysing *degen* as a contrastive topic marker. As the contrastive conjunction *al* is often followed by a contrastive topic, I searched the corpus for the combinations of *al* with *degen* where the distances between these two words was no longer than 3 words. This search returned 21 results in total, in none of which *degen* was used as a topic marker – contrastive or not. In comparison, the same search with *bolsa* instead of *degen* returned 76 matches, and only in 16 of those matches *bolsa* was not a contrastive topic marker.

5.4.4. Conclusions

The schema in (130) summarises the functions of the word form *degen* identified in this section.

- (130) Canonical and non-canonical uses of *degen*



As can be seen from the schema above, I identify three uses of the word form *degen*: the canonical, primary use as a predicate, which is directly relevant to the semantics of the verb *de-* ‘say’; the non-canonical use as a linking element, which appears to be

the grammaticalised form of the past participle of the same verb; and another non-canonical use as a topic marker.

It has been demonstrated, that contrary to prior claims in the literature, in its topic marking function *degen* should not be labelled a contrastive topic marker. Unlike true contrastive topic markers, such as *bolsa*, for instance, *degen* is able to follow and mark non-contrastive topics, such as topics of discourse initial utterances, or continuous topics that appear sentence-initially in the answers to questions about those topics.

The other non-canonical uses of *degen*, as well as other forms of *de-*, such as the converbial *dep*, require much further research, since: a) they have not been examined in the literature; and b) some interesting grammaticalisation processes that appear to be currently in progress have not been mentioned in the literature.

5.5. Dynamic Syntax Approach to Kazakh Topic Markers

In this section I present a sketch of the DS approach to Kazakh topic markers discussed in the preceding sections of this chapter. While the role of topic markers in the parsing process is a rather uncontroversial matter, this section starts to fill a big gap in the literature, as no attempts of formal analysis of Kazakh topic markers have been made to date. Additionally, in this section I provide an explication of the connection between conditionals and topic markers grounded in the Dynamic Syntax formal representations of both.

5.5.1. DS Analysis of Topic Markers

As outlined in Section 2.2 of Chapter 2, no additional layers of syntax are posited in Dynamic Syntax, and the information-structural notions of topic and focus are viewed as pragmatic effects that emerge in the process of contextualised parsing. A topic, in particular, is considered to be the minimal context on the basis of which the rest of the utterance constructs an update. If a topical element is linked to the preceding context, it constitutes a backgrounded or continuous topic. If, however, a topical element

signals a departure from the preceding context or the preceding topic, it is considered to be contrastive or shifted.

Linked structures, introduced in sub-Section 2.2.5 of Chapter 2, are used to formally represent topics as partial trees that are constructed within one parsing process, and whose function is to provide an antecedent for subsequent identification by some anaphoric device – usually a pronoun, as in (131). It is this antecedent-anaphora relation that is said to produce the topicality effect (Kempson et al. 2006).

(131) Speaker A: How is Mary doing these days?

Speaker B: Mary, she is much better, thank you.

However, the anaphoric element is often absent, especially in pro-drop languages. Recall, for instance, some of the Korean examples given in Section 2.1 of Chapter 2, one of which is repeated below in (132), as well as Kazakh examples given in Section 3.5 of Chapter 3.

(132) John nin i salam-lil mannaassta.
John TOP this man-OBJ met
'(As for) John, (he) met this man.'

Recall, that in DS, the difference between 'old' or continuous, and contrastive/shift topics is considered to lie in the relationship of the topical element and the preceding context (Section 3.5, Chapter 3). While 'old' topics are linked to the preceding context, contrastive/switch topics diverge from it thus creating the contrastive effect. This does not affect the behaviour of the topical element within the utterance in which it appears, and contrastive/switch topics are still parsed onto a Linked structure in the same way as non-contrastive topics are; recall the Link Adjunction and Topic Structure Introduction rules in sub-Section 2.2.5 of Chapter 2, and the semantic trees in Section 3.5 of Chapter 3.

A topic marker, then, is an item that explicitly identifies a constituent as the topic; in Dynamic Syntax terms this means that a topic marker contains the instruction to parse the element it follows onto a Linked structure. This applies to all three Kazakh topic markers described in this chapter: *še*, *bolsa*, and *degen*.

Let us examine how the parsing process develops for a sentence with the topic marker *degen* as shown in (133).

- (133) [What is Nadia doing these days?]
 Nadia degen dissertacija-si-n žaz-ıp žatır.
 Nadia TM thesis-POSS.3-ACC write-CVB AUX(3)
 ‘As for Nadia, she is writing her thesis (now/these days).’

Before proceeding to the outline of the parsing process, a lexical entry for topic markers is required. A simplified, sketchy version of it is presented in (134).

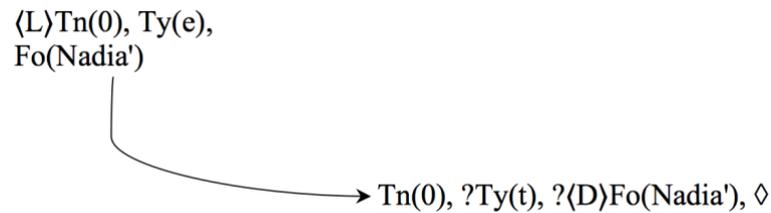
- (134) Lexical Entry for Kazakh Topic Markers (Simplified)

IF	Ty(e), Fo(α)
THEN	make(L), go(L); put(Tn(0), ?Ty(t), ?(D) Fo(α));
ELSE	Abort

This lexical entry essentially provides a lexical trigger for initiation of the Topic Structure Introduction and Requirement Rules given in sub-Section 2.2.5 of Chapter 2. The IF statement insures that a topical constituent has been parsed. The THEN statement builds the Link from the partial tree, which only consists of one node carrying Ty(e) and Fo(α) decorations, to the root node of the main tree, which gets annotated with a requirement for Ty(t), and the requirement for a copy of the formula from the partial tree to be present somewhere in this tree (= ?(D) Fo(α)).

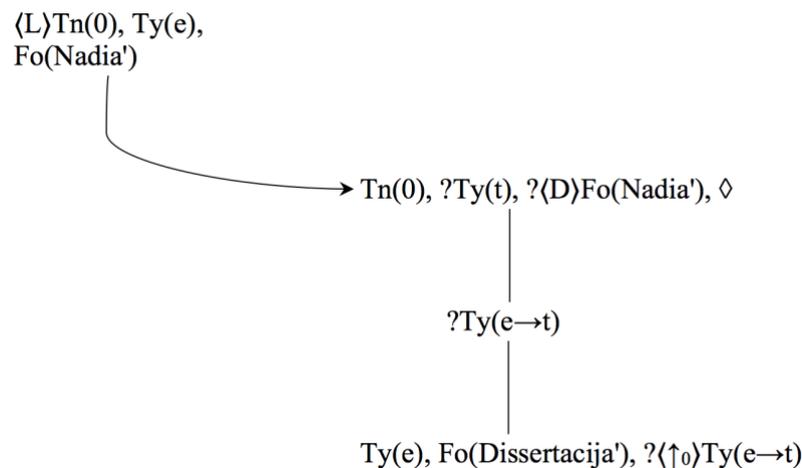
Taking this lexical entry into consideration, the tree in (135) shows the parallel partial trees developed after the sequence *Nadia degen* has been parsed.

(135) Parsing *Nadia degen...*



Next, the direct object *dissertacijasın* is parsed, as represented by the tree in (136) after a routinised update, which builds the $?Ty(e \rightarrow t)$ node above the node onto which the accusatively marked noun phrase has been parsed (recall sub-Section 3.5.1 of Chapter 3 containing the parse of a sample Kazakh sentence).

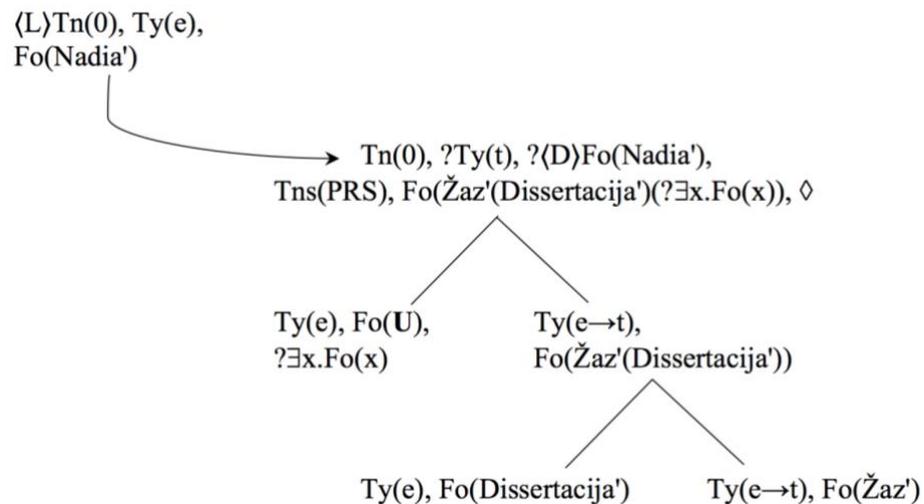
(136) Parsing *Nadia degen dissertacijasın ...*



Next, the complex verb form is parsed; all TAM properties are simplified since they have no direct relevance for this analysis. Recall, that the verb in Kazakh projects the full argument structure with variables decorating the argument nodes. If some argument node already exists in the tree, and the same argument node is built by the lexical action of the verb, these nodes are simply identified as one and the same node and harmlessly collapse.

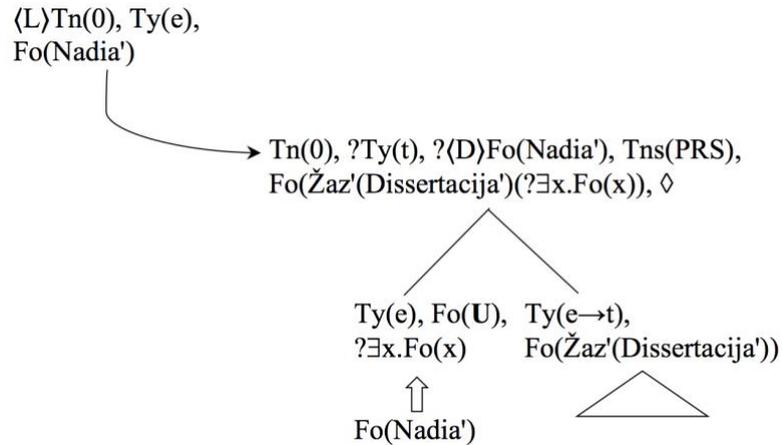
Thus, since the direct object node is already present in our tree, the direct object node constructed by the lexical action verb collapses with it straightaway. As there is no subject node in the tree when the verb is parsed, the lexical action of the verb constructs one and decorates it with a formula variable, as shown in the tree in (137).

(137) Parsing *Nadia degen dissertacijasın žazıp žatir...*



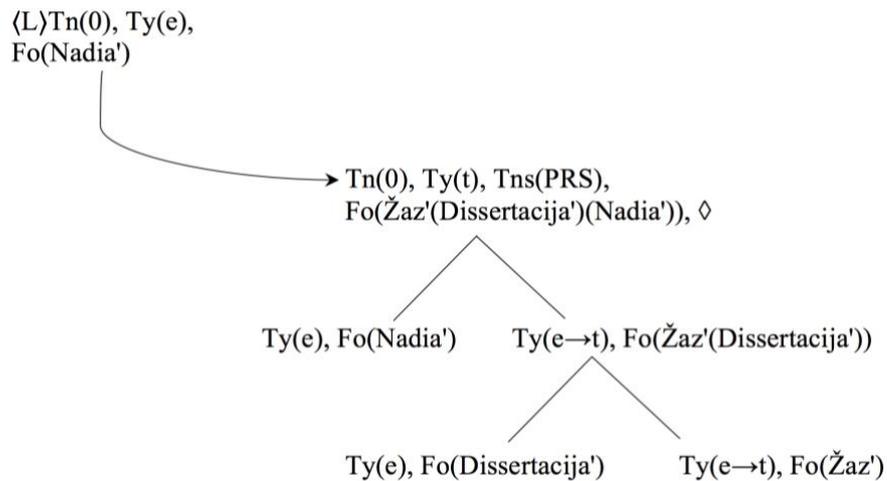
Although the verb has been parsed and the full argument structure has been developed in the tree, the tree is not yet complete as there are some outstanding requirements at the root and the subject nodes. The outstanding requirement at the root node – $?\langle D \rangle Fo(Nadia')$ –, and the outstanding requirement at the subject node – $?\exists x.Fo(x)$ – can be satisfied by updating the variable decorating the subject node with $Fo(Nadia')$, as illustrated in (138).

(138) Parsing *Nadia degen dissertacijasın žazıp žatir* – update



Once this update occurs, the tree is cleaned up via Elimination, all the requirements are satisfied, and all the information is compiled up the tree, the parsing is complete; the final tree is given in (139).

(139) Parsing *Nadia degen dissertacijasın žazıp žatır* – final



The final tree clearly demonstrates that the utterance at hand has an explicitly identified topical element which has been parsed onto a Linked structure, and provided context for the development of the main tree. The fact that there was no overt anaphoric element (as in the English approximation of (133)) which took the topical element as the antecedent cannot be established from the final tree alone, but can be perceived from the step-by-step tree-growth process presented above.

The parsing process for our sample sentence in (133) would be the same if *še* or *bolsa* followed the sentence-initial topic expression *Nadia*. Recall, however, that I classify *še* and *bolsa* as contrastive topic markers, which means that they indicate a divergence of the current topic referent from the preceding topic referent. I do not offer a way of formally representing this in DS, and I do not believe this information would appear as part of tree-node decorations, however, it would be of interest to investigate this further in future research. Crucially, this difference does not affect or alter the parsing process, and the tree-growth stages represented above apply for all three topic markers.

The main conclusion to be made from this tree-growth process is to do with the function topic markers play in tree building – they trigger the construction of a Link from the node onto which the topical element is parsed to the main tree, where the rest of the utterance gets parsed in the context of that topical element. A requirement is added to the root node of the main tree such that a copy of the formula from the partial tree is to be found somewhere in the main tree. In many languages this copy is expressed as an overt anaphoric, pronominal element. Since Kazakh is a pro-drop language, an overt pronominal element is typically not present, however, the requirement for a copy of the formula is satisfied by a variable at the subject node projected as part of the verb's lexical actions, which gets updated with the formula from the Linked topical node.

In the next sub-section, I specifically focus on the topic marker *bolsa* (as well as *še*, assuming my hypothesis regarding its origin is correct), and investigate the connection between the two functions of this word form via DS formalisms.

5.5.2. Conditionals as Topic Markers

In this section I employ the formalisms of DS to explore the connection between the conditional form of the verb *bol-* and topic marking. As noted in Section 2.1 of Chapter 2, the use of conditional forms as topic markers is not unusual and has been attested cross-linguistically. Additionally, as mentioned in Sections 5.2 and 5.3 of this chapter,

the use of the conditional form of the verb ‘to be’ as a topic marker has been observed in a number of Turkic languages.

Although there are objections to Haiman’s equating topics and conditional clauses on the grounds that they both are “presuppositions of their sentences” (1978: 585), the fact of the use of conditional forms as topic markers in many languages remains. In order to draw a parallel between the topic expressions and the protases in DS terms, let us recall that topics in DS are considered to provide immediate contexts for the rest of the utterance in which they appear. This contextual relation is expressed via Linked structures, as demonstrated in the preceding section.

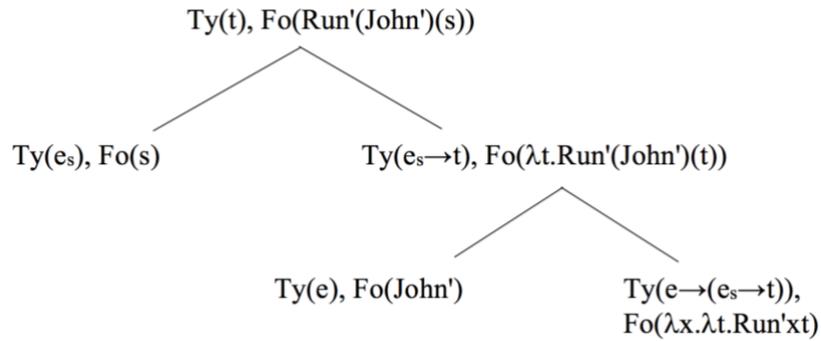
Notably, Gregoromichelaki (2006, 2010) demonstrates that the English protases introduced by *if* also form a Link relation with the main clauses of the conditional sentences in which they appear. Gregoromichelaki (2006) refers to the *if*-clauses as antecedents, and the main clauses as consequents. In order to provide a comprehensive account for the semantics of conditional clauses, Gregoromichelaki makes use of an additional argument – the situational argument, which is used for propositional representations standing for the situation of evaluation. Gregoromichelaki follows Farkas’s (1997) proposal that for each world w an existential model is defined as: $M_w = \langle S_w, U_w, V_w \rangle$. S_w here is a set of situations in w , U_w represents the set of individuals in w , and V_w performs the function of assigning values to the constants of the language in relation to the situations in S_w . Following Kratzer (1986), Gregoromichelaki assumes that situations are parts of worlds, each situation being a part of a unique world.

For the purposes of her analysis Gregoromichelaki explicitly represents the situation argument as an argument of a predicate, which combines with it via the usual route of function application. The situation argument is considered to be of $Ty(e)$, and is identified from the other subtypes of $Ty(e)$ – $Ty(e_i)$ for individuals, and $Ty(e_w)$ for worlds – by a subscript: $Ty(e_s)$. The situation argument can be a variable, an epsilon term or a tau term.

I omit all the technical details Gregoromichelaki provides on the ways whereby a situation argument might be introduced into the tree, and present in (140) the complete tree which includes a situation argument node for a string like *John runs*.

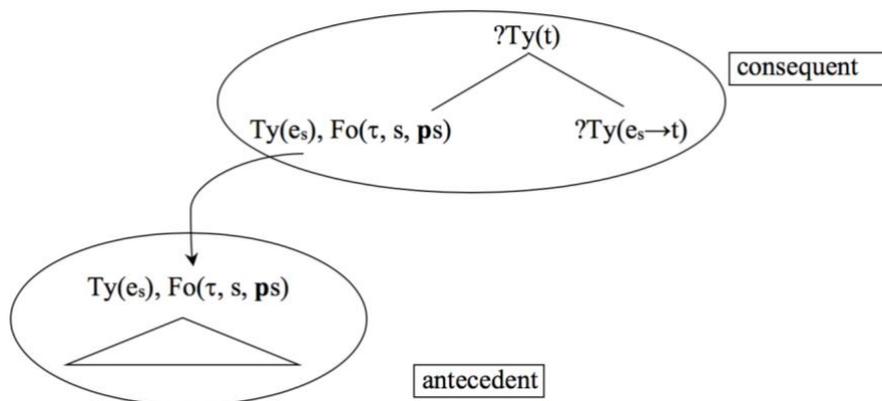
This is a slightly simplified version of Gregoromichelaki's example (97) (2006: 199) – I exclude some references to the world of evaluation and scope statements since they are not directly relevant to the current discussion.

(140) Parsing *John runs*



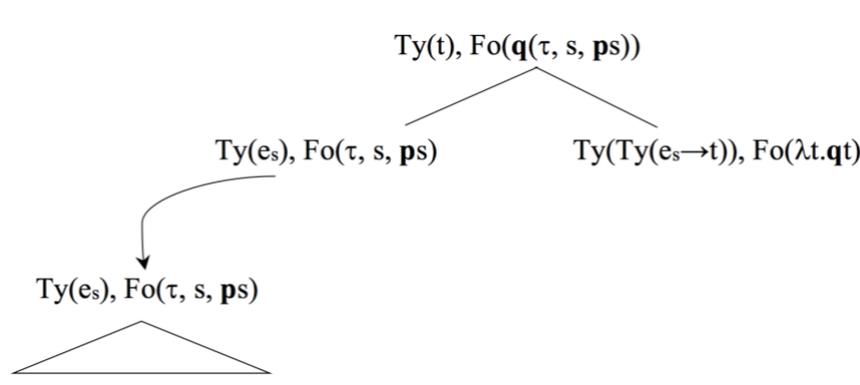
As mentioned earlier, the sentences containing *if*-clauses are assumed to consist of two Linked trees: one reflecting the parse of the main clause – the consequent, while parsing of the *if*-clause – the antecedent – results in another tree. According to Gregoromichelaki's analysis, then, the antecedent is connected with the situation node of the main tree via the Link relation, as illustrated in (141), which has been closely modelled after Gregoromichelaki's example (114) (2006: 205).

(141) Parsing *if p*, from the string *if p, q*



We can see that the tau term (τ -term) produced during the parsing of the antecedent is at the root node of that tree, and its copy is at the situation node of the main tree. In the main tree this τ -term is treated like an ordinary argument of the predicate and is incorporated into the propositional formula at the root node via the standard function-application process. The tree in (142) represents the structure built up by the end of parse of a string like *if p, q*; Gregoromichelaki's example (116) (2006: 206).

(142) Parsing *if p, q* – final



Further developments of Gregoromichelaki's analysis of conditional sentences are not directly relevant for this discussion. The main takeaway from this analysis is the fact that a protasis of a conditional sentence is Linked to the main clause via a Link structure that goes from a $Ty(e)$ node to a $Ty(e)$ node – in the same way as a topical node is Linked to the rest of the clause in which it appears.

In the English language a protasis of a conditional sentence is signalled by *if* (characterised by Gregoromichelaki as a particle), the lexical actions of which construct the Link. *Eger* (mentioned in Section 5.3) is the Kazakh equivalent to *if*, as per the example in (143), where *eger* indicates the condition, and *onda* – the consequence.

- (143) Eger olar kel-me-se, onda biz özimiz
if 3PL come-NEG-COND(3)then 1PL self.1PL
bar-a-miz.

go-FUT-1PL

‘If they don’t come, then we’ll go ourselves.’

However, there is one important reason not to consider *eger* a full equivalent of *if*: it can be easily omitted without affecting the grammaticality or the ‘conditionality’ of the protasis, as (144) demonstrates.

- (144) Olar kel-me-se, biz özimiz bar-a-miz.
3PL come-NEG-COND(3) 1PL self.1PL go-FUT-1PL
‘If they don’t come, we’ll go ourselves.’

In fact, omission of *eger* is much more common than its inclusion, presumably due to the availability of a dedicated conditional verb form in Kazakh. The difference between *if* and *eger*, then, is in that *if* initiates the construction of a Linked structure, while *eger* simply adds a requirement for a conditional verb form to appear in the same tree. It is the conditional verb suffix that builds the Linked structures in Kazakh, in the same way as, according to Gregoromichelaki, *if* does in English. The ungrammatical example in (145) corroborates this conclusion.

- (145) *Eger olar kel-me-j-di, biz özi-miz
if 3PL come-NEG-FUT-3 1PL self-1PL
bar-a-miz.
go-FUT-1PL
Intended: ‘If they don’t come, we’ll go ourselves.’

I propose that this Link-building property of the conditional verb suffix resulted in its utilisation in the topic marking function, starting from the conditional form of the Turkish ancient defective verb **er-* ‘be’ (see sub-Section 5.2.2), and continuing into the present-day Turkic languages where the conditional form of the verb *bol-* ‘be’ is used as a topic marker. The use of the verb ‘to be’ and not any other verb, can potentially be explained by its neutral semantics that is easily bleached.

The algorithm of constructing a Linked structure from a Ty(e) node in one tree to a Ty(e) node in another tree, and copying the formula across contained in the lexical entry of the conditional suffix follows the same steps as the process of topic construction, as represented by the DS semantic trees, and demonstrated in the preceding sub-section.

Instead of creating a novel item to perform the function of topic marking, speakers simply recycled an item whose lexical entry already contained the necessary algorithm, which could be simplified and adapted for the new function. This is a common occurrence intra- and cross-linguistically.

5.5.3. Conclusions

This section has provided a sketch of a DS analysis for Kazakh topic markers. A general lexical entry for topic markers has been presented; this lexical entry reflects the function of topic markers as inducers of Link building. This Link construction creates the pragmatic effect of topicality which, in DS terms, simply means provision of an immediate context against which the rest of the utterance can be parsed. While this is not a controversial or novel observation, this is the first attempt at formalising Kazakh topic markers within any theoretical framework.

DS formalisms were also used in this section to explicate the utilisation of conditional forms of the verb *bol-* ‘be’ for topic marking. While the connection between conditional and topic markers has been observed in many languages of the world, no formal explanation of this instance of grammaticalisation was offered. In this section it has been demonstrated that a DS analysis of conditionals proposed by Gregoromichelaki (2006) can move us towards a formal account of the link between the conditional forms and topic markers. It has been shown that protases are parsed onto a tree separate from the consequent in much the same way as the topics appear on a separate partial tree. Both the former and the latter provide an immediate context, against which the rest of the clause is to be parsed.

It is necessary for a protasis to contain an element which would comprise lexical actions inducing the construction of a Linked structure from the current to a

new tree and copying the formula across to the new tree. According to Gregoromichelaki (2006), in English, this function is performed by *if*. It was shown, however, that the Kazakh equivalent of *if* – *eger* – does not behave in the same way since it can be easily omitted. What cannot be omitted in a Kazakh protasis is the conditional suffix on the verb *-sA*, which essentially performs the Link construction function. It is this Link-constructing property that made the conditional form of the verb ‘to be’ a suitable candidate for the topic-marking function.

5.6. Conclusions

This chapter pursued three main goals: a) to provide descriptions for Kazakh topic markers, which have been little or not at all researched before; b) to identify some ongoing grammaticalisation processes, which have not been mentioned in the literature; and c) to offer some initial formal DS analyses for some of the Kazakh topic marking phenomena.

Three topic markers – *še*, *bolsa*, and *degen* – have been identified and described. Numerous contextualised examples have been provided for all three items, which allowed for their meanings and distribution patterns to be established and detailed.

Some interesting grammaticalisation processes have been highlighted. One of these processes is to do with the transformation of *še* from a contrastive topic marker, to a contrastive question particle. Another grammaticalisation process discovered thanks to the original, previously unreported data is to do with *bolsa*. It has been shown that in some instances where *bolsa* functions as a topic marker it still agrees with the topical subject it follows, which points to an ongoing process of grammaticalisation of the conditional verb form into a topic marker.

Yet another grammaticalisation process appears to be taking place in the realm of noun modification, which is not directly related to topic marking, but is connected to the topic marker *degen* considered in the chapter. Contrary to previous analyses of *degen* reporting it as functioning as an ‘emphatic particle’ in instances where it follows

an oblique nominal modifying an unmarked head noun, I propose that *degen* performs as an attributiviser, compensating for the lack of marking on the head noun.

The sketch of a formal Dynamic Syntax analysis that has been presented in this chapter is the first attempt at any sort of formal representation of Kazakh topic markers. Additionally, the DS tools have been applied to produce a formal explication for the use of conditionals in topic marking.

6. Particle *ǰoj*

6.1. Introduction

This chapter is dedicated to the little researched particle *ǰoj* and to providing a unified analysis for its multiple uses. Many uses of *ǰoj* are associated with the information-structural notions of contrast and givenness (see Chapter 2, Section 2.1 for definitions). Indeed, similar particles from other languages have been analysed as having contrastivity or givenness as their core meanings.

Moving away from using pragmatic or information-structural labels to explicate the multifunctional nature of particles like *ǰoj*, I propose to apply Matić and Nikolaeva's (2014) analysis of the Tundra Yukaghir particle *mə(r)=* as an existential operator to *ǰoj*. The validity of this approach is confirmed in 6.6.4 where it is demonstrated that the pragmatic effects associated with the presence of *ǰoj* in a sentence result from its existential semantics.

This chapter proceeds as follows: Section 6.2. presents previous descriptions of *ǰoj* available in the literature; in Section 6.3. I discuss the use of *ǰoj* in imperative clauses, and argue that what has been labelled as the particle *ǰoj* is in fact the imperative form of the verb *qoju*, thus excluding the so-called imperative use of *ǰoj* from further examination.

Section 6.4. presents contextualised elicited, natural, and corpus data, which demonstrate the distribution pattern of *ǰoj*, while Section 6.5. focuses on its pragmatic contribution. In Section 6.6. the analyses of similar-functioning particles from Russian and Sanzhi Dargwa are presented, as well as the analysis for the Tundra Yukaghir particle *mə(r)=*. It is posited, that applying Matić and Nikolaeva's (2014) analysis for the Tundra Yukaghir particle *mə(r)=* to *ǰoj* yields the most comprehensive explanation of the multi-functionality of this particle. A sketch of a formal Dynamic Syntax analysis of *ǰoj* is given in Section 6.7; Section 6.8 concludes this chapter.

6.2. Previous Research

Little has been written about the particle *ğoj* – realised as *qoj* after voiceless consonants – in the English, Russian or Kazakh linguistic literature. This might be due to this particle being perceived by the native speakers as an item exclusively appearing in spoken language, and, as such, only carrying some sort of emotive or emphatic meaning without interacting with the grammatical structure of the language.

This attitude is reflected in Straughn (2011), who refers to *ğoj* as a sentence-final particle, which expresses emotivity. He claims that *ğoj* is completely optional and is never a part of the verbal complex, which is why its presence “should not be seen as any sort of formal marking” (2011: 134). Straughn gives the example in (1) to demonstrate that “the addition of *ğoy* merely indicates that the speaker is expressing an emotive attitude toward the content of this utterance” (2011: 135); in his gloss Straughn labels *ğoj* as ‘EXCL’ for ‘exclamative particle’.

- (1) Qıtay-lar-dinǵ žaǵa žil-i eken ğoy.
Chinese-PL-GEN new year-POSS.3 EVID EXCL
‘It’s (apparently) the Chinese New Year!’

(Straughn 2011: 135)

Straughn concludes that the particle *ğoy* indicates “the speaker’s emotive stance” and is “essentially limitless in distribution”, thus it, along with the Uzbek equivalent which Straughn also considers in his thesis, “should not be seen as primary verbal markers of emotivity, but instead as discourse particles” (2011: 136). In Section 6.3 it is shown that *ğoj* is not a limitless in distribution exclamative particle, and that its functions go beyond merely expressing the speaker’s emotive attitude towards the content of his utterance.

Muhamedowa (2016) summarises her analysis of meaning of *ğoj* as follows:

It is difficult to find an appropriate translation for this particle in English. *Yoy* is similar to the English *you know*, as it appeals to shared information between the speaker and the hearer (2016: 163).

Additionally, Muhamedowa states that this particle has a “fixed position in a sentence and must occur after the predicate” (2016: 15). As Section 6.4 shows, Muhamedowa’s description of the placement of *ǰoj* after the predicate is more accurate than Straughn’s, as the sentence-final position for *ǰoj* proposed by Straughn does not cover cases where one of the terms appears after the predicate (as an afterthought, for instance). Muhamedowa also states that “the particle *yoy* attached to the –A converb softens an imperative” (2016: 27), as shown in the example in (3) below; this is discussed in detail in Section 6.3, where an alternative analysis of such imperative utterances is given.

Abish (2014) presents a more detailed description of *ǰoj* in her doctoral thesis *Modality in Kazakh as spoken in China*. The author refers to *ǰoj* as a modal particle and introduces it as follows:

The particle *I²oy* does not have any lexical meaning. Its basic contribution is to mark some type of epistemic evaluation, an assessment of the propositional content. It can express a commitment to the truth of the proposition, i.e. to its certainty, probability, possibility, etc. The source of the epistemic evaluation can be the opinion of the addresser or some other person. The basic meaning varies according to the communicative functions of different types of usages [...] (2014: 75).

Abish (2014) identifies two variants of *ǰoj* – the accented and the unaccented one. The unaccented variant is viewed by the author as an enclitic, which expresses presumption, while the accented variant is used to express repudiation, emphasise shared knowledge, or to form tag-questions (in which case it is pronounced with rising interrogative intonation), as well as to function together with existentials and the conditional mood.

I was not able to observe the difference in the accenting of the particle as described by Abish, and some of the examples provided by the author were not

accepted as grammatical by the Kazakh speakers in Kazakhstan. This may be due to the differences in how *ğoj* is used in the Kazakh language spoken in China and in Kazakhstan. Exploring these differences is outside of the scope of this work, but is an interesting future research subject.

Turning to the descriptions of *ğoj* written in Russian, Balakaev et al. (1962) categorise *ğoj* as an emphatic-limiting and a modal-expressive particle (the authors do not provide criteria or features for either of these particle classes); the description of functions of *ğoj* is only provided under the former heading. Thus, according to Balakaev et al. (1962: 417), the emphatic-limiting particle *ğoj* is said to express the speaker's confirmation of his own words. It is used when the speaker wants to remind his interlocutor of an event or action already known to her; additionally, *ğoj* is used to 'logically underline' or stress a word.

Bol'shoj Kazahsko-Russkij Slovar' (henceforth, *The Big Kazakh-Russian Dictionary*) defines *ğoj* as "a particle used to add emphasis or expressivity to the content of an utterance; translated into Russian as *ved'* or *že*", as in (2), or as "a particle used to soften a request or a command" (1998: 366), as in (3).

(2) Ajt-tï-m ğoj!
 say-PST-1.SG ğoj
 'I did say!/I said, didn't I?'

(3) Kele ğoj!
 come.IMP ğoj
 'Come here then!'

Note that the examples in (2) and (3) are provided without contexts, which makes it difficult to establish the conditions under which these utterances are felicitous; this is especially relevant for the utterance in (2), as it clearly presents the speaker's reaction to another utterance or an event. I discuss the example in (3) in detail in Section 6.3 and present an alternative analysis of these imperative constructions.

Most Kazakh grammars intended for language learners do not provide a description or explanation of the use of *ǵoj*, and those that do only highlight one aspect of its use. These explanations can be contradictory not only between authors, but also between publications by one and the same author. For example, Romanenko (2011: 72) states that *ǵoj* is a particle of “emotional fortification”, while Romanenko (2015: 109) asserts that it is a “clarifying” particle, which “confirms the verity of an utterance”. Bizakov (2014: 181) refers to *ǵoj* as a “veracity particle”, and Valjaeva (2018) claims that *ǵoj* is a “specifying” particle, which gives an utterance “the meaning of completeness and assertiveness”.

Thus, it is clear that there is no consensus on neither the status, nor the meaning of *ǵoj*. It has been referred to as a particle, a modal particle, a discourse particle, an emphatic-limiting particle, a modal-expressive particle and a veracity particle. Its distribution has been described as limitless, and as strictly post-predicative. It has been claimed that *ǵoj* refers to shared information, adds emotivity, confirmativity, assertion, clarity, veracity, specificity, as well epistemic evaluation to an utterance. All these meanings and functions ascribed to *ǵoj* have been illustrated with non-contextualised examples, which do not allow for a full understanding of the contribution the particle makes to a proposition.

Before moving to the next section, in which I challenge some of the analyses of *ǵoj* presented above, a brief terminological remark is in order. In the recent years, a lot of literature (e.g. Degand et al., 2013; Diewald, 2013; Bayer and Struckmeier, 2017, amongst many others) has been devoted to teasing apart modal particles from discourse particles/markers, with much discussion revolving around the questions of identifying these notions, and whether these are distinct notions, or should one be considered a subcategory of the other.

Researchers are far from a consensus on how these categories are to be identified and delimited. While this is an important linguistic question, I do not consider it to be central for the research on *ǵoj* presented in subsequent sections. It is, therefore, concluded, that the simple label ‘particle’ is sufficient for present work, and would allow me to avoid using a controversial and debateable label, which could potentially prove harmful for future typological research.

6.3. *ǰoj* in Imperative Clauses

As mentioned in the previous section, *ǰoj* has been described as a particle that softens the imperatives. In this section I demonstrate that this description is not accurate and propose an alternative approach to analysing imperative sentences with the word form *ǰoj*. More specifically, it is proposed that the *ǰoj* that appears in the imperative clauses is the imperative form of the polysemous verb *qoju*, rather than the particle *ǰoj*.

Muhamedowa (2016: 27) states that “the particle *ǰoy* attached to the –A converb softens an imperative”, and provides the following example:

- (4) Kel-e ǰoy!
 Come-CVB PART
 ‘Please come!’

The same example and explanation for this use of *ǰoj* are given by the *Big Kazakh-Russian Dictionary* (1998:366), as mentioned in Section 6.2 above. Notably, however, the –A converbial form does not coincide with the imperative form of verbs in Kazakh, and the use of *ǰoj* with the imperative verb form results in ungrammaticality as shown in (5) below.

- (5) *Kel ǰoj!
 come.IMP.2SG ǰoj
 ‘Come then! / Come, come!’

It thus appears that the use of the particle *ǰoj* with the imperative form of the verb is ungrammatical, but the use of *ǰoj* with a converbial form results in a ‘soft’ imperative reading. No explanation is provided as to how this combination of a converb – a non-finite verb form (Haspelmath 1995: 3) – and a particle, would produce the ‘soft’ imperative effect in this language.

- (9) Ol anda-sanda kel-e qoj-a-dī.
 3SG from.time.to.time come-CVB AUX-PRS-3
 ‘He would just come over (unexpectedly)/show up from time to time.’
- (10) Oniñ söz-i-ne eškim sen-e
 3SG.GEN word-POSS.3-DAT no.one believe-CVB
 qoj-ğan žoq.
 AUX-PFV NEG
 ‘No one has believed/believed his words.’
- (11) Men et-ti dos-īm-a saqta-p
 1SG meat-ACC friend-POSS.1SG-DAT save-CVB
 qoj-dī-m.
 AUX-PST-1SG
 ‘I saved the/some meat for my friend.’
- (12) Men oniñ qupija-si turali ajt-ip
 1SG 3SG.GEN secret-POSS.3 about say-CVB
 qoj-dī-m.
 AUX-PST-1SG
 ‘I’ve let out his secret (accidentally).’
- (13) Men osi kitap-ti oqi-p qoj-dī-m.
 1SG this book-ACC read-CVB AUX-PST-1SG
 ‘I’ve already finished reading this book.’
- (14) Mağan mīna kitap-ti saqta-p qoj.
 1SG.DAT this book-ACC save-CVB AUX.IMP.2SG
 ‘Save this book for me.’

In (9) the use of *qoju* adds the effect of unexpectedness of the action described by the lexical verb in the converbial form. In the English language this effect is added by the use of *up* in such expressions as *show up* or *turn up*. Interestingly, both the English *up* and the Kazakh *qoju* have the ‘vertical’ semantics in their primary uses, which perhaps aids the creation of imagery of someone appearing suddenly and unexpectedly – one minute they are not there, and the next they are standing (up) in front of you. Investigating this further is out of the scope of this thesis, but could be of interest for a cross-linguistic semantic piece of research.

In (10) *qoju* is used to indicate that the action expressed by the lexical verb is accomplished and irrevocable; in (11) *qoju* shows that the completed action was deliberate and premeditated; while in (12), on the contrary, the auxiliary is used to denote an action that happened unintentionally and unexpectedly, and is irreversible. In (13) *qoju* identifies an action that was completed before the expected completion time; note the use of ‘already’ in the English interpretation of this sentence to indicate the earlier than expected completion of the action. The example in (14) shows the use of the imperative form of the auxiliary *qoju* to intensify the request expressed by the converbial form of the lexical verb.

Returning to the sentence in (4), I propose that this is also a compound verb form, like in the sentences in (9)-(14) above, consisting of a lexical/main verb and the auxiliary *qoju*; the reanalysis of the example in (4) is presented in (15) below, where *qoj* is glossed as the auxiliary verb.

- (15) Kel-e qoj!
 come-CVB AUX.IMP.2SG
 ‘Come then! / Do come!’

In speech, the voiceless [q] is realised as [ɣ] in the intervocalic position; intervocalic voicing is a common occurrence in Kazakh, although the domain of its application is yet to be thoroughly researched for this language (cf. Yu Cho (1990) for Korean). Since [q] always appears between vowels in constructions like that in (15), the

auxiliary verb *qoju* in its imperative form has been erroneously reanalysed as particle *ğoj*, and the function of ‘softening’ of commands was assigned to the latter.

The examples in (16)-(18) below demonstrate that the combination of a converb and the imperative form of the auxiliary *qoju* can appear in different contexts, but always indicate a command or instruction. More importantly, the examples in (17) and (18) show agreement in person and number of the imperative form of the verb *qoju* with the omitted subject – we would not expect particle *ğoj* to follow this agreement pattern.

- (16) Bar-a qoj, men soñinan
 go-CVB AUX.IMP.2SG 1SG afterwards
 bar-a-mïn.
 come-FUT-1SG
 ‘You go then, I’ll come later.’
 (ACKL)

- (17) Erteñ sağat bes-te žeñgej-di
 tomorrow hour five-LOC sister-in-law-ACC
 ert-ip, biz-diñ üj-ge
 bring.with-PTCP 1PL-POSS house-DAT
 kel-e qoj-iñiz.
 come-CVB AUX.IMP-2PL.FRM
 ‘At 5 o’clock tomorrow, please come to our house together with the
 sister-in-law.’
 (ACKL)

- (18) [The wife asked her husband for permission to visit a relative, to which
 the husband replies:]
 Bar-a qoj-iñiz.
 go-CVB AUX.IMP-2SG.FRM
 ‘Do go. / Of course, You can go.’

It can also be speculated that the erroneous analysis of the imperative form of the verb *qoju* as the particle *ǰoj* might have been the result of the Russian language influence which dominated in Kazakhstan for several decades. Russian particle *že*, which is very similar to *ǰoj* in many of its applications as I suggest below in Section 6.6.1, can indeed be used in the imperative sentences, as shown in (19).

(19)	a.	Idi!	b.	Idi	že!
		come/go.IMP.2SG		come/go.IMP.2SG	PART
		‘Come! / Go!’		‘Come then!/Go then!’	

It has to be noted, however, that in the Russian examples the particle *že* follows the imperative form of the verb, unlike in the Kazakh examples, where the imperative meaning was assumed to originate from a combination of a converb and the particle.

To sum up, in this section it has been demonstrated that the particle *ǰoj* that has been claimed to appear in the imperative sentences is in fact *qoj* - the imperative form of the verb *qoju* - which functions as an auxiliary. The combination of the converbial form of the main/lexical verb and the imperative form of *qoju* produces the imperative sentences with the additional meaning of intensity, or permission. In the intervocalic position the initial voiceless uvular plosive of *qoj* [q] is realised as a voiced uvular fricative [ɣ] which appears to be one of the reasons for the erroneous analysis of this imperative construction.

Another reason is the influence of the Russian imperative sentences containing the particle *že*, which is very similar to the Kazakh *ǰoj*. Having demonstrated that the *ǰoj* that appears in the imperative sentences is not the particle *ǰoj* which is my main concern here, I exclude these sentences from further description and analysis of the particle *ǰoj*.

6.4. Syntactic Distribution of Ğoj

As mentioned earlier, the particle *ğoj* is widely used in spoken Kazakh and in informal written styles, and is excluded from formal writing styles. Having demonstrated in the preceding section that this particle is not used in imperative sentences, in this section I focus on the syntactic distribution of this item.

6.4.1. Post-predicative Ğoj

This particle typically follows predicates, which can be expressed by a lexical or modal verb, an existential or evidential copula, an adjective, or a noun. The examples below show *ğoj* following a finite verbal predicate in (20), a modal word in (21), an existential copula in (22), and an evidential copula in (23). The example in (20) additionally demonstrates that the predicate does not have to be in its canonical clause-final position, thus making the particle not obligatorily sentence-final, contrary to Straughn's (2011) and Muhamedowa's (2016) observations. The context – either situational or linguistic – is given in square brackets.

- (20) [One friend utters to another as they hug after not having seen each other for a long time.]

Fusun, äbden saġin-dï-q qoj seni.
Fusun very.much miss-PST-1PL ğoj 2SG.ACC
'Fusun, we did miss you a lot.'

(MBT)

- (21) [After finding out that one of the female employees who is not married has a son, one manager says to another:]

Äkesi bol-u kerek qoj?!
father-POSS.3SG be-INF must ğoj
'There's got to be a father, hasn't there/right?!'

(MBT)

- (22) [A colleague is late for a morning meeting and the boss says: ‘Do you not have an alarm clock?’ She replies:]

Zoq, bar ğoj.

no EXST ğoj

‘Yes, of course I do.’

(MBT)

- (23) [A man wakes up looking disheveled, puffy, red-eyed and unshaven after a long night of New Year celebrations and says to himself in the mirror:]

Oj-baj, üšinši mĩñžıldıq-tiñ adam-ı, seniñ

oh.dear third millennium-GEN man-POSS your

türiñ osındaj eken ğoj.

appearance like.this EVID(3) ğoj

‘Oh dear, so this is what a man of the third millennium looks like then, it would seem.’

The example in (24) shows *ğoj* used after an adjectival predicate.

- (24) [The opening line of a folk-tale about camels.]

Tüje qazırdiñ özinde ädemi ğoj.

camel now itself beautiful ğoj

‘Camels nowadays are beautiful, aren’t they. / As is well known, camels are beautiful at present time. [But a long time ago... (and the tale continues to say that camels used to be more beautiful and how they lost that beauty)].’

(ACKL)

It must be noted that *ğoj* can be removed from the examples in (20)-(24) without affecting their grammaticality. That is to say, the particle does not participate in the

syntactic structure of the sentences in which it appears post-predicatively. It does, however, make a pragmatic contribution, which is discussed in detail in Section 6.5.

Ĝoj can successfully appear post-predicatively in yes-no interrogative sentences, as shown in (25), but not in wh-questions, as demonstrated by ungrammaticality of (26).

- (25) Sen bügin universitet-ke bar-a-siñ ĝoj?
 2SG today university-DAT go-FUT/PRS-2SG ĝoj
 ‘You are going to university today, aren’t you?’
 ‘You are going to university today, right?’

- (26) *Sen bügin qajda bar-a-siñ ĝoj?
 2SG today where go-FUT/PRS-2SG ĝoj

The role of *ĝoj* in interrogative sentences is discussed in detail in Sections 6.5 and 6.6.

To sum up, the particle *ĝoj* can appear in the post-predicative position, and can follow predicates expressed by a variety of syntactic categories. In this position it can appear in affirmative and interrogative yes-no questions, but not in wh-questions.

6.4.2. Post-nominal *Ĝoj*

Let us now consider *ĝoj* in the non-post-predicative position, as shown in Speaker B’s utterance in (27).

- (27) Speaker A: Keše Bolat düken-ge bar-di.
 yesterday Bolat shop-DAT go-PST(3)
 ‘Bolat went to the shop yesterday.’
- Speaker B: Arman ĝoj bar-ĝan / *bar-di!
 Arman ĝoj go-PTCP / *go-PST(3)
 ‘(Of course) it was Arman who went!’

Unlike in previous examples, where *ğoj* follows the predicate, in (27) it appears immediately after the subject expressed by the proper name *Arman*. Notably, the subject is contrastively/correctively focused in this instance, that is to say, it is the constituent that directly rejects the alternative expressed by one of the interlocutors (Gussenhoven 2007, see Chapter 2, sub-Section 2.1.3). The examples in (28) and (29) demonstrate that the subject must be contrastively focused in order for *ğoj* felicitously to follow it.

(28) Speaker A: Keše kim düken-ge bar-di?
 yesterday who shop-DAT go-PST(3)
 ‘Who went to the shop yesterday?’

Speaker B: #Arman ğoj barğan.

(29) Speaker A: Arman keše düken-ge bar-di
 Arman yesterday shop-DAT go-PST(3)
 ma?

Q

‘Did Arman go to the shop yesterday?’

Speaker B: #Arman ğoj dükenge barğan.

In (28) *ğoj* follows a non-contrastive new information focus, which makes the sentence infelicitous. However, a small caveat is in order – the response containing *ğoj* is infelicitous here as a first-time response to the question; it may be used by the speaker to signal her annoyance at having already answered the same questions a number of times. It would then be interpreted along the lines of: ‘It was Arman who went, wasn’t it (I’ve told you this already)’. In (29), *Arman* is the topic and cannot be felicitously followed by *ğoj*.

Another observation that must be made about Speaker B’s utterance in (27) is that only the past participial form of the verb is grammatical in this sentence, while a

finite past tense verb form is not. This is true for all sentences where *ğoj* follows the correctively focused element as the examples in (30) and (31) below illustrate.

- (30) Speaker A: Arman went to the shop yesterday.
 Speaker B: Biz *ğoj* bar-ğan / *bar-ğan-biz /
 1PL *ğoj* go-PTCP go-PTCP-1PL
 *bar-dī-q.
 go-PST-1PL
 ‘It was us who went.’

- (31) [Two friends discussing two brothers called Bolat and Arman. The first speaker claims that Bolat is the one who goes to the library the most. The second speaker objects:]

Qoj-ši, Bolat *qoj* eñ köp bar-atin /
 stop-POL Bolat *ğoj* most a.lot go-PTCP
 *bar-a-dī / *bar-īp žatir.
 go-PRS-3SG go-CVB AUX(3)
 ‘Oh stop! Of course it is Bolat who goes to the library the most!’

As can be seen from the examples above, the attempts to use either the finite past tense forms, as in (30), or the finite present tense forms, as in (31), instead of the past and present participle forms result in ungrammaticality.

The sentences with post-nominal *ğoj*, in turn, become ungrammatical if *ğoj* is removed, as the examples in (32) and (33) show.

- (32) *Biz bar-ğan.
 1PL go-PTCP
 Intended: ‘We went.’

- (33) *Arman eñ köp bar-atin.

Arman most a.lot go-PTCP
 Intended: ‘Arman goes the most.’

The ungrammaticality of a sentence like that in (33) can be easily fixed if a predicate is added, as demonstrated by (34).

- (34) Arman eñ köp bar-atın adam.
 Arman most a.lot go-PTCP man
 Lit: ‘Arman is the most going man.’
 ‘Arman is the one who goes the most.’

The nominal constituent followed by *ğoj* in these types of constructions does not have to be the subject or indeed a term of the sentence - *ğoj* can also follow a contrastively focused direct or indirect object as in (35) and (36) respectively, or an adjunct, as in (37).

- (35) Speaker A: Keşe Bolat mīna kitap-ti
 yesterday Bolat this book-ACC
 satıp al-dī.
 buy-PST(3)
 ‘Bolat bought this book yesterday.’
- Speaker B: Zoq, ol mīna žurnaldī ğoj
 no 3SG this magazine-ACC ğoj
 satıp al-ğan.
 buy-PTCP
 ‘No, it was this magazine that he bought.’
- (36) Speaker A: Keşe Bolat düken-ge bar-dī.
 yesterday Bolat shop-DAT go-PST(3)
 ‘Bolat went to the shop yesterday.’

Speaker B: Zoq, ol kitaphana-ga ğoj bar-ğan.
 no 3SG library-DAT ğoj go-PTCP
 ‘No, it was the library that he went to.’

(37) Speaker A: Bolat aldiñģi küni Astana-ģa
 Bolat day.before.yesterday Astana-DAT
 kel-di.
 come-PST(3)
 ‘Bolat arrived in Astana the day before yesterday.’

Speaker B: Žoq, ol keše ğoj kel-gen.
 No 3SG yesterday ğoj come-PTCP
 ‘No, it was yesterday that he arrived.’

The syntactic and information-structural observations made so far, together with the interpretations the sentences with the post-nominal *ģoj* receive, point to the possible cleft nature of these clauses. In the literature, the notion of cleft is identified as a “specific syntactic pattern which serves to separate a discourse prominent constituent structurally from the rest of the clause” (Hartmann and Veenstra 2013: 1); or as a construction where the clefted constituent expresses either a contrastive or ‘exhaustive’ focus (Hedberg 2008, 2013).

According to Hartmann and Veenstra (2013), in its canonical form, a cleft is a bi-clausal copular construction that consists of an impersonal pronoun (the cleft pronoun), a copular verb, the informationally prominent phrase (the cleft phrase) and an embedded relative clause (the cleft clause). The example the authors provide as an illustration for this description is given in (38) below.

(38) It was chicken wings that Peter ordered for lunch.
 (Hartmann and Veenstra 2013: 1)

In this example, *it* is the cleft pronoun, *was* is the copular verb, *chicken wings* is the cleft phrase, and *that Peter ordered for lunch* is the cleft clause. Certainly, not all languages follow this Germanic cleft sentence structure, as many languages would not require a cleft pronoun, which is essentially a dummy subject not needed in pro-drop languages such as Slavic and Turkic, for instance. I would, therefore, suggest, that the minimal cleft sentence structure consists of a cleft phrase, a copula, and a cleft clause, as shown in (39) on the example of a subject cleft from Wolof.

- (39) [Jigéen ji]i [a]j [lekk gato bi]k.
 woman the COP eat cake the
 [Cleft Phrase]i [Copula]j [Cleft Clause]k
 ‘It’s the woman who ate the cake.’
 (Torrence 2013: 188)

Clefting is typologically wide-spread as a focus marking device, and is the main focus marking device in some languages, as is the case for Bura – an Afro-Asiatic language. According to Hartmann and Zimmermann (2012), focusing on non-subject constituents is achieved via a cleft in this SVO language; this is shown in (40) below.

- (40) dívà-r mphyí àn tí Kùbílí tà.
 mush-LINK sorghum FM REL Kubili prepare
 ‘It’s sorghum mush that Kubili cooked.’
 (Hartmann and Zimmermann 2012)

In this example the informationally prominent phrase – the cleft phrase – is fronted, and is followed by a copula/focus marking element and a relative clause, as previously also seen in Wolof.

Mathew (2013) provides examples of cleft constructions in Malayalam – a Dravidian SOV language from South India. According to Mathew, the cleft construction with the focus marker *a:nu*, whereby any argument, adverb or adjunct can be focused, is widely used. The examples in (41)-(43) demonstrate this.

- (41) karambi-e a:nu Paily kand-aDu
 Karambi-ACCFM Paily saw-3SG
 ‘It is Karambi whom Paily saw.’
- (42) innale a:nu Paily karambi-e kand-aDu
 yesterday FM Paily Karambi-ACC saw-3SG
 ‘It is yesterday that Paily saw Karambi.’
- (43) sankadam vann-ittu a:nu avan karanj-aDu
 sorrow came-PFV FM he cried-3SG
 ‘It is because (he) became sad that he cried.’

(Mathew 2013: 258)

Interestingly, Mathew (2013) refers to cases where the focus marker *a:nu* appears post-predicatively as clefting of the entire clause, and provides the following example for this:

- (44) Paily karambi-e kand-aDu a:nu
 Paily Karambi-ACC saw-3SG FM
 ‘Paily did see Karambi.’

Mathew’s (2013) reasons for classifying the sentence in (44) as a cleft sentence are not clear; however, notably, the interpretation for this sentence is very similar to the interpretations of the Kazakh sentences with post-predicative *ǵoj* in the examples in (20)-(24).

Having considered cleft constructions from other languages, we can return to the Kazakh sentences with the post-nominal *ǵoj*, and observe that they follow the proposed cleft structure. The example from (27) is repeated in (45) below with some additional details.

- (45) [Arman]_i [ğoj]_i [bar-ğan]_k.
 Arman COP go-PTCP
 [Cleft Phrase]_i [Copula]_j [Cleft Clause]_k
 ‘It was Arman who went.’

All the elements of a cleft construction are present: the informationally prominent phrase is the contrastively/correctively focused element appearing clause-initially; *ğoj* performs the function of a copular verb/copula, which is not controversial from the typological perspective; and the embedded relative clause forms the cleft clause.

Additionally, it is worth noting that only a participial verb form must follow a post-nominal/copular *ğoj*, as was demonstrated earlier in (30) and (31), and as can be seen from the ungrammaticality of Speaker B’s utterance in (46) below.

- (46) Speaker A: Arman öte mejirimdi.
 Arman very kind
 ‘Arman is very kind.’

Speaker B: *Zoq, Bolat ğoj öte mejirimdi.
 No Bolat ğoj very kind
 Intended: ‘No, it is Bolat who is very kind.’

This example shows that an attempt to construct a cleft sentence without a participle fails, since the adjective following *ğoj* cannot form a relative/cleft clause on its own. The situation can be remedied by the addition of a participial verb form, as shown in (47).

- (47) Speaker A: Arman öte mejirimdi bol-dī.
 Arman very kind be-PST(3)
 ‘Arman was very kind.’

Speaker B: Zoq, Bolat ğoj öte mejirimdi bol-ğan.

No Bolat *ğoj* very kind be-PTCP
'No, it is Bolat who was very kind.'

Recall, that in sub-Section 1.1.7 of Chapter 1, it is demonstrated that Kazakh relative clauses do not contain a relative pronoun, are non-finite, and are always headed by participial verb forms. The requirement that post-nominal *ğoj* is followed by a participial verb form, and consequently, a relative clause explains ungrammaticality of any other predicate forms in cleft sentences.

To sum up, having considered the syntactic distribution of *ğoj*, it has been observed that it can either appear post-predicatively or post-nominally. In post-predicative position it can follow any type of predicate – verbal or nominal – as long as it is either narrowly focused, or is a part of a wider focus phrase. In this position *ğoj* does not interact with the syntactic structure of the sentence in which it appears, that is, the sentence does not become ungrammatical if *ğoj* is removed. The contribution the post-predicative *ğoj* makes is of pragmatic nature and is discussed in more detail in the following section.

In post-nominal position *ğoj* also follows a focused element - a correctively/contrastively focused one, to be precise. The nominal element can be a correctively focused argument or adjunct expressed by case-marked or unmarked nouns. Unlike post-predicative *ğoj*, post-nominal *ğoj* is involved in the syntactic structure of the sentences in which it appears. The sentences with post-nominal *ğoj* are cleft sentences, which become ungrammatical if *ğoj* is removed. These cleft construction also display restrictions in regards to the form of the verb that must follow *ğoj* – only participial verb forms are permitted since they are the only verb forms capable of heading a relative clause.

Such differences in the syntactic behaviour of *ğoj* appearing post-predicatively and post-nominally inevitably lead to questioning whether this is one and the same *ğoj* or two different ones. The post-predicative *ğoj* appears to be a particle, while the post-nominal *ğoj* displays the syntactic behaviour of a copula. As becomes clear in the following section, however, despite the syntactic differences, the pragmatic

contribution of both these elements is very similar, which allows us to talk about shared semantics for them.

6.5. Pragmatic Contribution of *Ǧoj*

This section is concerned with the pragmatic contribution of both the particle and the copula *ǧoj*. It is demonstrated that despite significant differences in their syntactic distribution, the pragmatic contributions they make are very similar in that they both add implications of givenness or shared information. That is to say, they indicate that information carried by propositions in which they appear must be treated as present in the interlocutors' CG whether it has actually been previously explicitly shared or not.

In 6.5.1 I provide some contextualised examples of use of the particle *ǧoj*, and discuss the pragmatic effects it evokes, and in 6.5.2 the pragmatic effects of the copula *ǧoj* are considered.

6.5.1 Pragmatic Contribution of Post-predicative *Ǧoj*

The examples below show *ǧoj* used in a variety of contexts. In all examples from (48) to (52) the use of the particle indicates that the information contained by the proposition it follows is either given, or communicatively/pragmatically presupposed. The utterances containing *ǧoj* either point out this 'given' status or re-activate the information in the CG shared by the interlocutors.

- (48) [Parents check up on their sleeping child. Mother says to Father:]
Kör-di-ñ be? Ujıqtap žatir dep
see-PST-2SG Q sleep-CVB AUX(3) COMP
ajt-ti-m ǧoj.
say-PST-1SG ǧoj
'Did you see? I did tell you he was sleeping.'

(MU)

- (49) [On a school trip the teacher finished counting the children as they get off the bus. She counts 19, but there should be 20 students.]

Sonda, bireu žoq qoj.

then someone NEG ğoj

‘Someone is missing, aren’t they!’

(MU)

- (50) Speaker A: Qajrat keše düken-ge bar-ğan
 Kairat yesterday shop-DAT go-PTCP
 žoq.

NEG

‘Kairat did not go to the shop yesterday.’

Speaker B: Ol bar-di ğoj!

3SG go-PST(3) ğoj

‘Of course he went! / He did too!’

- (51) [After asking a question and not receiving an answer from his wife, the man utters with annoyance:]

Men sura-p tur-di-m ğoj!

1SG question-CVB AUX-PST-1SG ğoj

‘I asked (you) a question, didn’t I?’

(MBT)

- (52) [After having been told that Alice was on her way and having waited for her to arrive, the man exclaims:]

Ol kel-me-j-di ğoj?!

3SG come-NEG-FUT-3 ğoj

‘She is not coming, is she?’

(ATLG)

Let us also consider the additional, narrower meanings, which can be attributed to the presence of *ǰoj*, as in the examples (48) - (52). In (48), the Mother uses *ǰoj* after her utterance to re-iterate that she has indeed previously said that the child would be sleeping. In a way, we have double re-iteration here: first, the Mother quotes herself saying that the child is sleeping – *ujïqktap žatir* –, then uses *ǰoj* after the predicate of the main clause to re-confirm that the act of producing the quoted utterance did indeed take place earlier on.

In the example in (49) there is an effect of exclamation and surprise at the speaker's own realisation that someone was missing which is re-enforced by the addition of *ǰoj*. The modal particle scopes over the proposition *bireu žoq* and marks it as 'given', thus re-confirming it. The situation is described twice: first, the speaker describes the state of affairs in her proposition *bireu žoq*; then, she gives the proposition an epistemic marking with the particle *ǰoj* to show that the state of affairs described by the proposition is obvious or given in the situational context. This utterance is of particular interest since it is not aimed at another interlocutor, or does not provide a correction to a proposition produced by someone else, but rather to the expected state of affairs.

In the second utterance in (50) the additional effect of using *ǰoj* is the expression of reproach; not only is *ǰoj* used to point to the fact that the information provided by Speaker B was accessible to Speaker A, but also to show disappointment or disapproval that this information had not been used. This adversative use of *ǰoj* with verum focus is fairly frequent in corrective utterances. The omission of *ǰoj* would result in the same utterance semantically, but the pragmatic effect of givenness and reproach would not be conveyed.

In (51) we see *ǰoj* being used to create the effect of a rhetorical question, which is uttered to show the speaker's annoyance. Interestingly, in this example the speaker does not refer to the shared knowledge per se, but rather to his own action (asking a question) which happened only a few moments prior; the addition of *ǰoj*, which adds the effect of givenness, highlights the fact that the hearer was present when the original question was uttered but chose not to reply to it. The speaker could have simply

restated his question, but by uttering (51) he shows his annoyance at how the exchange has unfolded so far.

In (52) we see a tag-question created by the use of *ğoj* in a similar way as was shown earlier in (25). The utterance in (52) is uttered by someone who had been told that Alice was on her way, however, having waited for quite some time, it became obvious that she was not coming at all. By producing (52) the speaker states the obvious – the fact that she was not coming –, and indicates the obviousness of this proposition by using *ğoj*; the tag question effect is created by rising intonation on *ğoj*.

Although the responses to tag-questions like that in (25) and (52) can be the same as the responses to canonical yes/no questions formed with the help of the question particle *MA*, it has to be noted that the interrogative illocutionary force in questions like in (25) and (52) is contributed by the interrogative prosody (rising intonation contour), and not by the particle *ğoj* – this is discussed in more detail in Section 6.6.4. As (53) shows, the same word sequence as in (25) can be used in an affirmative sentence; in this case it is pronounced with a falling intonation on the particle.

- (53) Speaker 1: Why did you say you can't leave the house today?
- Speaker 2: Sen bugün universitetke barasıñ ğoj, [and there is no one else to look after the children].
- Speaker 2': #Sen bugün universitet-ke
 2SG today university-DAT
 bar-a-siñ ba?
 go-FUT/PRS-2SG Q
 'Are you going to university today?'

Interestingly, the same is observed for the English tag-questions (see, e.g. Rando 1980, Huddleston and Pullum 2002, Reese and Asher 2006), which are the closest approximations for utterances like those in (25), (52) and (53). Abish (2014)

refers to this use of *ğoj*, which she identifies as a modal particle, as ‘non-modal’, although no explanation is provided for this labelling.

The example in (24), repeated in (54) below for convenience shows the use of *ğoj* in the very first sentence of a folk-tale, that is, in a sentence which is not preceded by any context.

- (54) [The opening line of a folk-tale about camels.]
 Tüje qazirdiñ özinde ädemi ğoj.
 camel now itself beautiful ğoj
 ‘Camels nowadays are beautiful, aren’t they. / As is well known, camels are beautiful at present time. [But a long time ago... (and the tale continues to say that camels used to be more beautiful and how they lost that beauty)].’

The use of *ğoj* in (54) also creates the effect of givenness, even if the proposition has not been previously shared by the author with his readers. The addition of *ğoj* instructs the reader to accept the proposition as ‘given’ even if she has never received this information before. Another similar usage of *ğoj* is shown in (55) below.

- (55) [Talking about upcoming celebrations of Naurız:]
 Khan Šatır-diñ žanında alań bar ğoj, onda
 Khan Shatyr-GEN near square EXST ğoj there
 erteń koncert bol-a-di.
 tomorrow concert be-FUT-3SG
 ‘There is this square near Khan Shatyr [shopping centre], right / isn’t there, there will be a concert there tomorrow.’ / ‘You know that square near the Khan Shatyr shopping centre, there will be a concert there tomorrow.’

In this example *ğoj* follows the existential *bar* – a construction widely used in spoken Kazakh²¹ as a referent introducing construction, which establishes the topic of the following utterances, as mentioned in Section 1.1 of Chapter 1. In the example above, the speaker does not use the combination of *bar* and *ğoj* to inform her hearer of the existence of the square near the shopping centre – as would have been the case if the same utterance was produced without *ğoj*.

The addition of *ğoj* indicates the speaker’s assumption that the information contained in the proposition preceding *ğoj* is present in the hearer’s mind, that it is ‘given’, even if it has never been previously explicitly shared between these two interlocutors. However, if the information is in fact not present in the hearer’s mind, the addition of *ğoj* instructs her to treat it as ‘given’, as a fact of the world.

Frequent co-occurrence of *bar* and *ğoj* in these type of utterances appears to have led Abish (2014) to analysing *bar ğoj* as a complex particle, in which *ğoj* is accented. She also claims that this complex particle “can follow any constituent of a sentence” (2014: 83). I do not agree with this analysis on a number of points. Firstly, I see no reason to treat *bar ğoj* as one complex particle, as *ğoj* can be omitted from the sentence without causing ungrammaticality, while *bar* cannot be, as the modified examples of (55) given in (56) and (57) demonstrate.

(56)	Khan Šatır-dīñ	žanında	alañ	bar,	onda	erteñ
	Khan Shatyr-GEN	near	square	EXST	there	tomorrow
	koncert	bol-a-dī.				
	concert	be-FUT-3SG				

²¹ Interestingly, this construction has been calqued into the variety of Russian spoken in Kazakhstan as *est’ že*. Although this widely used Kazak Russian construction is not used in Standard Russian, once native speakers of Standard Russian come across it in conversations, they are able to process it in the intended meaning without difficulty. I would go as far as to propose that this construction might be common in the variety of Russian spoken near the border of Russia and Kazakhstan. Further research on the varieties of the Russian language in the post-Soviet states is sure to reveal many more fascinating observations.

‘There is a square near Khan Shatyr, there will be a concert there tomorrow.

- (57) *Khan Šatır-dīñ žanında alaň ğoj, onda erteñ
 Khan Shatyr-GEN near square EXST there tomorrow
 koncert bol-a-di.
 concert be-FUT-3SG

As (57) shows, omitting *bar* is unacceptable and leads to ungrammaticality, while (56), where *bar* is not followed by *ĝoj*, is grammatical. Comparing (55) and (56), in which *ĝoj* is present and absent respectively, we can note the change in the interpretation of these sentences: the latter utterance does not assume any previous knowledge or carry an instruction on how the information contained in the proposition should be processed. In fact, when uttering (56) the speaker assumes no previous knowledge of the square, and simply informs the hearer that such a square exists.

To contextualise these examples even further, we can say that (55) would be inappropriate for a hearer who has just arrived in Astana, and cannot be expected to know even what Khan Shatyr is, let alone the square near it; while (56) would not be uttered in a conversation with someone who lives in Astana, as the knowledge of at least the reference point – Khan Shatyr – can be assumed. In fact, (56) would sound condescending and patronising if uttered to someone who lives in Astana – as if the speaker is trying to imply that the hearer does not know the capital, or perhaps is a newcomer.

Secondly, as mentioned in Section 6.2, it may be the case that the variety of Kazakh spoken in China differs from Standard Kazakh, however, in Standard Kazakh the particle *ĝoj* is not stressed in this construction – the main stress is on the existential *bar*. In fact, from cursory observations on the prosody of *ĝoj*, it is not stressed in any of its uses in Standard Kazakh.

Another argument against positing that *bar ĝoj* is a complex particle lies in the simple fact that *bar* can easily be replaced by its negative counterpart *žoq*, which can still be successfully followed by *ĝoj*, as shown in the example in (58).

- (58) Khan Šatır-dīñ žanında alañ žoq ğoj,
 Khan Shatyr-GEN near square NEG.EXST ğoj
 sol sebipten erteñ koncert Bäjterek-tiñ
 that because tomorrow concert Baiterek-GEN
 žanında bol-a-di.
 near be-FUT-3
 ‘There is no square near Khan Shatyr, right/is there, this is why
 tomorrow the concert will be near the Baiterek.’

If Abish (2014) were to insist on the complex particle analysis, it would need to be modified to reflect the possibility of *ğoj* combining with both the affirmative and negative existential forms.

And lastly, the claim that *bar ğoj* can follow any constituent is also not applicable to Standard Kazakh. As was demonstrated above, *ğoj* can be removed from utterances where it follows *bar* without affecting their grammaticality; the combination of *bar* and *ğoj* can only successfully follow those constituents, that can be successfully followed by *bar* in the first place, as outlined in sub-Section 1.1.11 of Chapter 1, and as shown in the examples in (59) and (60).

- (59) Meniñ kölig-im bar / žoq
 1SG.GEN car-POSS EXST / NEG.EXST
 Lit.: ‘My car exists / does not exist.’
 ‘I have/do not have a car.’

- (60) London-da köp universitet bar.
 London-LOC many university EXST
 ‘There are many universities in London.’

Ğoj can follow sentences like those in (59) and (60) and add the pragmatic effects discussed above, however, unlike *bar* and *žoq*, it plays no role in the syntax of these

utterances. Since *bar* and *žoq* must be preceded by certain types of constituents, it cannot be posited that the ‘complex particle’ *bar ġoj* can follow any constituent. It would also be strange to claim that only the first element of the proposed complex particle participates in the grammar of the sentence in which it appears.

To sum up, then, we have seen so far that the particle *ġoj* brings the effect of givenness or pragmatic presupposition to the clause in which it appears. By using *ġoj* the speaker instructs the hearer to either retrieve the information carried by the proposition, or to treat it as given in any case. In other words, the proposition *p* followed by *ġoj* is assumed to belong to the CG, whether it had been explicitly added there during the preceding exchange or not. The givenness effect introduced by *ġoj* and its multiple applications and realisations stem from the existential semantics of this particle, which is considered in more detail in Section 6.6. Let us now move to the pragmatics of the post-nominal, copular *ġoj*.

6.5.2. Pragmatic Contribution of the Post-nominal *ġoj*

As mentioned earlier, the nominal constituent immediately followed by the predicative particle *ġoj* must be contrastively focused. Krifka (2007) also notes that such so-called cleft focus constructions “often signal an exhaustive interpretation that in-situ focus lacks” (2007:7). As Hedberg further (2013) observes for clefts:

...a cleft sentence packages a proposition in such a way that the two principal semantic parts of a cleft—an exhaustive focus and a pragmatic presupposition—are mapped transparently onto two syntactic constituents— a clefted constituent and a cleft clause—and are equated with each other via a copula (2013: 6).

As can be seen from Hedberg’s description of cleft sentences, their information-structural properties are set by default, which restricts the contexts in which these sentences can be used. As far as the Kazakh *ġoj*-cleft constructions are concerned, their usage is restricted to the contexts in which a nominal element is

correctively/contrastively focused, and the rest of the utterance is presupposed. Let us consider the example from (27) again, repeated in (61) below for convenience.

(61) Speaker A: Keše Bolat düken-ge bar-di.
yesterday Bolat shop-DAT go-PST(3)
'Bolat went to the shop yesterday.'

Speaker B: Arman ğoj bar-ğan!
Arman ğoj go-PTCP
'(Of course) it was Arman who went!'

By choosing to use the *ğoj*-cleft construction, Speaker B indicates that: a) some of the information provided by Speaker A is being corrected, and b) the correct information provided by Speaker B has previously been shared with or available to Speaker A.

This givenness effect rises from the use of *ğoj* in this sentence, as the example of the alternative corrective/contrastive response without *ğoj* in (62) demonstrates.

(62) Speaker B': Žoq, Bolat emes, Arman bar-di.
No Bolat NEG Arman go-PST(3)
'No, Arman went, not Bolat.'

In (62) *Arman* is contrastively/correctively focused, but without the indication that this information has been previously made available, or should somehow have been known to Speaker A.

The same observations apply to the other *ğoj*-cleft sentences, like those presented in (31), (35) and (36). In all those sentences, *ğoj* follows a nominal element and not only creates a cleft construction with an exhaustive, corrective/contrastive, non-verbal focus, but also adds the effect of givenness, the sense that the correct information has been previously shared with or has somehow been available to the hearer. This sometimes results in the additional connotations of reproach, impatience, or reprimand to the interlocutor who had to be corrected due to not using the correct

information despite it being available to him (according to the person producing the *ǰoj*-cleft utterance).

It has become clear that despite the differences in the syntactic properties of the post-predicative and post-nominal *ǰoj*, the pragmatic effects which arise from their use are the same. The speaker using either particle in her utterance indicates to her hearer that (some of) the information she provides has been previously shared, or should be treated as such. The effects under discussion can be said to be pragmatic since we discussed the appropriateness and felicity of utterances with *ǰoj* in various contexts, as well as their infelicity in some situations. In the next section, I move to examining the common semantics for both types of *ǰoj*.

6.6. Semantics of *ǰoj*

In this section I consider the semantics of both post-nominal and post-verbal *ǰoj*. As was shown in the preceding section, the inclusion of *ǰoj* into a sentence in either position produces the pragmatic effect of givenness. That is, (at least some of) the information carried by the proposition with *ǰoj* is assumed by the speaker to be already present in the CG or the interlocutor's mind. We have also seen that a post-predicative *ǰoj* can be used to re-confirm givenness of the state of affairs to the speaker himself, so that a double verbalisation of sorts takes place: by adding *ǰoj* to a self-directed/rhetorical utterance the speaker confirms that the proposition contained in that utterance is indeed given. In other words, the speaker confirms that the situation described in his utterance exists in the real world.

I begin this section by considering two particles – the Russian *že*, and the Sanzhi Dargwa *=q'al* – which function very similarly to *ǰoj*. Feldman's (2001) analysis of the former, and Forker's (2017) analysis of the latter are presented in 6.6.1 and 6.6.2 respectively. Sub-section 6.6.3 presents Matic and Nikolaeva's (2014) analysis of the Tundra Yukaghir particle *mə(r)=* as an existential operator, which is then applied to *ǰoj* in 6.6.4.

6.6.1. Russian *že*

In this section I focus on Feldman's (2001) account of the multifunctional Russian particle *že* and the unifying analysis proposed for it. This particle is directly comparable to *ğoj* in many of its uses, and, as mentioned in Section 6.2. of this chapter, the *The Big Kazakh-Russian Dictionary* (1998) states that *že* is one of the particles (the other being *ved'*) that can be used to translate *ğoj* into Russian. To confirm this, let us consider two of the Kazakh examples from the previous sections and their Russian equivalents with *že*.

(63) (a) Kazakh:

Sen	bügin	universitet-ke	bar-a-siñ	ğoj?
2SG	today	university-DAT	go-FUT/PRS-2SG	ğoj

(b) Russian:

Ty	že	idjosh	segondnja	v	universitet?
2SG	že	going	today	to	university

'You are going to university today, aren't you?'

'You are going to university today, right?'

(64) (a) Kazakh:

Kör-di-ñ	be?	Ujıqtap	žatir	dep
see-PST-2SG	Q	sleep-CVB	AUX(3)	COMP
ajt-ti-m	ğoj.			
say-PST-1SG	ğoj			

(b) Russian:

Videl?	Govorila	že	čto	on	spit.
see	said.PST.F	že	that	he	sleeping

‘Did you see? I did tell you he was sleeping. I told you he was sleeping, didn’t I?’

In the examples above, the Kazakh (a) sentences with *ğoj* and the Russian (b) sentences with *že* are used in the same contexts and receive the same interpretation.

As Feldman observes, *že* has been commonly considered to occur after the first prosodic word, and treated as a second position clitic – a commonly attested occurrence in almost all Slavic languages, as well as in some other languages; this is seen in the examples in (63) and (64) above. While it is true for the majority of utterances with *že*, it is not true for all of them, as Feldman’s examples given below show.

(65) My dolžny tam byt’ segodnja že večerom.
we must there be today že evening
‘We have to be there tonight!’

(66) U menja est’ takaja že kniga.
to me EXST such že book²²
‘I have exactly the same book.’

(Feldman 2001: 188)

Besides, as Feldman points out, treating *že* as a second position clitic implies that it can appear in any sentence as long as it occupies the second position, which is not supported by the data, as the example below confirms.

(67) #Ty že podpisывaeš’sja na ‘Pravdu’ ili
you že subscribe to Pravda or
‘Jerusalem Post’?

²² Relying on my native knowledge of the Russian language I amended Feldman’s inaccurate gloss for this example ‘I have such *že* book’. Some of the transliteration has also been altered.

Jerusalem Post

‘Do you subscribe to ‘Pravda’ or to ‘the Jerusalem Post?’

(Feldman 2001: 188)

It is clear, then, that *že* is not a second position clitic, and an alternative analysis is needed.

In the traditional Russian grammars (e.g. Vasilyeva 1972, Rozental and Telenkova 1985), *že* is considered to be a polysemous particle that is used in to indicate emphasis, contrastivity, some sort of justification, and even similarity. Recall, that a very similar situation is described for *žoj* in Section 6.2 of this chapter.

In the examples in (65) and (66) *že* is used to indicate emphasis and similarity respectively, while the examples in (68) and (69) below demonstrate the so-called ‘contrastive’ and ‘justificational’ uses of *že*. Feldman notes that where *že* is used to indicate ‘justification’ it is comparable with the English *after all*, or Hebrew *harey*; it is also equivalent to the Kazakh *žoj* here.

(68) On ostajotsja, ona že uežžat.
he stays she že leaves
‘He is staying, but she is leaving.’

(69) Čto ty stoiš’? Sadis’ v mašinu. Ona že
what you stand sit in car it že
naša.
ours
‘What are you waiting for? Get in the car. It’s ours, isn’t it! / It’s ours after all!’

(Feldman 2001: 189)

Feldman opposes to this polysemous account of *že*, since in many cases the particle can be removed without the loss of the meaning it supposedly contributes; this indeed is true for all the examples containing *že* provided thus far. The removal of *že* from the sentence in (68), for instance, does not result in the loss of contrastivity

between the two parts of that sentence, since the contrast is present in their semantics. Interestingly, McCoy (2003) takes the ‘contrastivity’ approach further, and analyses *že* as a ‘kontrastive marker’ or a ‘k-marker’, in Vallduví and Vilkuna’s (1998) understanding of ‘kontrast’ as the ability of certain linguistic expressions to generate a set of alternatives. While this analysis might work for some of the uses of *že*, it fails to capture all of its uses.

Feldman rejects the ‘second position clitic’ and the contrastive analyses of *že*, and proposes an alternative, Relevance-theoretic account of it. Relevance Theory (Sperber and Wilson 1986) is grounded on the assumption that a hearer’s interpretation of an utterance is directed by relevance considerations, that is to say, hearers inevitably interpret utterances as maximally relevant. Relevance here is considered to be a trade-off between contextual effects – informativity –, and processing costs.

An important distinction between conceptual and procedural elements was introduced into the Relevance Theory framework by Blakemore (2002): the distinction between the conceptual and procedural meanings. Conceptual meaning is carried by content words which belong to open word classes (e.g. nouns, verbs, or adjectives) and encode concepts. Since conceptual meaning contributes to the propositional content of utterances, it is truth-conditional.

Procedural meaning, on the other hand, is not considered to be directly consciously accessible, resistant to conceptualisation, and, thus, not easily (or indeed at all) paraphrasable or translatable. Discourse markers, some text-connective markers, and modal particles are some of the examples of linguistic items with procedural meaning. The function of these items is not to express concepts, but to “guide the inferential comprehension process by imposing procedural constraints on the construction of intended contexts and cognitive effects” (Wilson 2011: 6).

Feldman proposes to treat *že* as a procedural element, namely, as a discourse marker that signals to the hearer that the speaker considers certain information in the utterance to be hearer-old; it also instructs the hearer to perceive the information from the utterance as activated or re-activated information. Feldman follows Prince’s (1981) distinction between ‘discourse-old’ or ‘discourse-new’, and ‘hearer-old’ or ‘hearer-new’ types of information, where discourse-old information is that evoked or

activated in the current discourse, and hearer-old information is information that the speaker trusts to be present in hearer's knowledge. It follows from these descriptions that a 'discourse-new' piece of information is not necessarily 'hearer-new', however, 'discourse-old' information is necessarily 'hearer-old'. Prince's 'hearer-old' information is comparable with Lambrecht's presupposed information.

Feldman juxtaposes the example from (69) with an identical example bar the presence of *že*, as shown in (70), to demonstrate the differences in their contextual effects.

- (70) Čto ty stoiš? Sadiš' v mašinu. Ona naša.
 What you stand sit in car it ours
 'What are you waiting for? Get into the car. It's ours.'

(Feldman 2001: 191)

In (70), the information that the car belongs to the speaker is 'hearer-new' – the speaker is simply informing the hearer of this. It could be uttered in a situation where having bought a new car without letting her husband know, the speaker points to it and informs her husband of the purchase by uttering (70). Adding *že* to the final part of that utterance, like in (69), would be incompatible with this context, since the information about the car ownership is completely 'hearer-new' here.

In (69), however, the fact that the car belongs to the speaker and hearer is 'hearer-old' or presupposed, that is, the speaker assumes this information to be in the hearer's store of knowledge. This utterance would be felicitous in the situation where two people have eventually found their car in the parking lot, but one of them does not quite recognise it as theirs. The speaker would be able to utter (69) in this case, since he assumes that the information about the car the couple own is known to the hearer. Feldman notes that *že* might be used in utterances aimed at retrieving information from long-term memory, as well as from linguistic or situational contexts.

Feldman demonstrates the application of her analysis in various contexts, where *že* might be used. She adds, that as well as signalling hearer-old information, in some contexts, *že* restricts a given set to a given proper set. Namely, this is claimed to occur in questions, and conditionals, as shown in (71) and (72).

- (71) Čej že den' roždenija segodnja?
 whose že day of.birth today
 'So, whose birthday is it today?'

(Feldman 2001: 196)

- (72) Esli že ja vyp'ju sejčas kofe, ja ne budu
 if že 1SG drink now coffee 1SG NEG will
 spat' noč'ju.
 sleep at.night
 'If I drink coffee now, I will not sleep at night.'

(Feldman 2001: 197)

For (71), Feldman posits that by adding *že* the speaker assumes that there is a restricted set of people whose birthday might be being celebrated that day, and that this proper given set is available to the hearer. In (72) the speaker uses *že* to instruct the hearer to infer the existence of a set of conditions, and that the condition mentioned in the utterance is one of the applicable alternatives.

Feldman does not explicitly identify the mechanism whereby the extension of the function of *že* from simply referring to hearer-old information to identifying a proper set happens. She simply states that this is the result of the interaction of the function of *že* with the "regular effect of the wh-question, where the speaker assumes some set of information options to be given" (2001: 196), for (71) and similar examples; or with the "general function of conditional sentences" (2001: 197), for examples like that in (72).

The proposed analysis correctly predicts the infelicity of the examples like that in (67), where *že* cannot appear in a true alternative question, where the questioner genuinely does not know which of the alternatives will be chosen by the respondent. According to Feldman, this freedom of choice between the proposed alternatives "blocks the function of *že* as a marker of known/inferable information which should be processed vis-à-vis activated or reactivated information" (2001: 199).

now the.more widower-ADVZ=MOD that already
 hana
 now
 ‘The more, he is already a widower.’

(Forker 2017: 3)

It can also function as a predicative particle – a special word class identified in Dargwa languages. These particles typically function in a similar way to copulas or auxiliaries, as shown in the examples in (75) and (76). In (75), =*q'al* is the head of a nominal predicate, while in (76) it is combined with a non-finite verb form to form a full independent clause predicate.

(75) ʔu^ʕrus ʁaj-la=q'al il
 Russian language-GEN=MOD that
 ‘This (word) is from Russian.’

(76) ij lak er Ø-ik'-uk=q'al
 this up look M-look.at.IPFV-CVB=MOD
 ‘He is looking upwards.’

(Forker 2017: 4)

Omission of =*q'al* from both (75) and (76) would result in their ungrammaticality as stand-alone independent main clauses, which points to their predicative and finite nature. As Forker notes, the particle has the ability to turn a subordinate clause into a main clause in the absence of other morpho-syntactic markers with the same ability. Otherwise =*q'al* is used exclusively as a modal particle, which does not directly interact with the syntax of the clause in which it occurs, that is, if the particle can be removed from a clause without affecting its grammaticality.

Similarly to *ğoj* and *že*, =*q'al* is multifunctional, and does not contribute any truth-conditional content. Forker provides numerous examples of the use of =*q'al*, and posits that =*q'al* indicates that the information contained in the utterances with =*q'al* is pragmatically presupposed. Due to this, the closest approximation of many

(79) [Omar returns home. He does not expect other people. When he enters the corridor he sees the cloak of Ali and says:]

ʔa^hli sa-jʁ-ib-le=q'al
 Ali hither-come.PFV-PRET-CVB=MOD
 ‘(Apparently) Ali came.’

(Forker 2017: 9)

Forker states that it is the use of the predicative particle =*q'al* that adds the “mirative flavor”²³ to this utterance, which, in the absence of =*q'al* would presumably only have evidential meaning.

Similarly to *ǰoj*, =*q'al* is used in existential clauses, as the example in (80) demonstrates.

(80) [Describing the making of traditional medicine:]

hešt:i it'in miriq^w-e le-d=q'ale
 these red worm-PL exist-N.PL=MOD
 d-uqen-ne d-ax-an-te hel-t:i=ra
 N.PL-long-ADVZ N.PL-go-PTCP-ATTR.PL that-PL=ADD

‘There are these red worms, (you know). These long ones that also crawl around ...’ [during the rain. I put them into vodka, ...]

(Forker 2017: 9)

Forker notes that when used in these constructions =*q'al* confirms the existence of the referent under discussion.

²³ I assume that the term ‘mirative’ is used by Forker in DeLancey’s (1997, 2001) understanding of this category as that encoding information which is “new or surprising to the speaker” (1997: 33). I follow Hill (2012) in his skepticism towards identifying ‘mirativity’ as a linguistic or cross-linguistic category, since morphemes which have been described as expressing this category usually have several other functions. As Hill points out, the full semantics of these morphemes must be considered in detail before any claims for their ‘mirative’ function could be made. For a detailed and full discussion, the reader is referred to Hill (2012).

Forker rejects a number of analyses of *=q'al* which focus on the pragmatic effects the use of this particle produces, rather than providing a unified analysis of it. By unified analysis Forker means identification of a functional core component of *=q'al*, which would explicate and unify all of its distinct but related readings. Since *=q'al* resembles Russian *že*, which is typically used for Russian translations of *=q'al*, Forker first applies McCoy's (2003) analysis of *že* as a contrast marker (mentioned in the preceding section) to *=q'al*. She concludes, however, that this analysis is not suitable for *=q'al*, as it does not provide a unified explanation of all its uses. Forker observes, that *=q'al* can only reinforce contrast in contexts where it is already present, but it does not itself convey contrast.

Forker mentions Feldman's (2001) analysis for *že* – presented in the previous section – but does not follow it entirely, stating her preference for Ariel's (1988, 1998) “comparable but far more articulated account” (2017: 11) for the Hebrew particle *harey*. According to Ariel's (1988, 1998) Relevance Theory based approach, then, *harey* is a procedural marker indicating accessible, given information, which already has representation in the hearer's mind. Accessibility/givenness is treated as a conventional implicature, which is a non-truth-conditional inference attached by convention to a particular item, rather than derived from superordinate pragmatic principles like the Gricean maxims, for example.

Forker applies this analysis to *=q'al*, and states that “the conventional implicature *=q'al* conveys is accessibility or givenness” (2017: 12). The particle signals that the information carried by the utterance in which it appears already has a mental representation in the hearer's mind, or that the speaker expects the hearer to have that mental representation. Forker posits that it is the interaction of the particle's proposed core function with certain contexts that results in the multitude of additional specific meanings, such as contrast as in (78), request for confirmation as in (77), request for information, and surprise as in (79).

The status of *=q'al* as a procedural rather than conceptual item is also addressed by Forker. Recall the distinction that is made between conceptual and procedural elements, as briefly outlined in the preceding sub-section. On the basis of this distinction, Forker analyses *=q'al* as a procedural rather than a conceptual item. She

observes that the particle is not consciously accessible, easily interpretable or paraphrasable by native speakers.

Instead of denoting some concept, being a true procedural item *=q'al* instructs the hearer to process the proposition in which it appears as accessible/given or “mutually manifest” in Sperber and Wilson’s (1995: 38-46) terms. The speaker uses the particle to instruct the hearer to process the information from the utterance as given, or as accessible from the common ground. This essentially imposes a procedural constraint, which can be narrowed down even further in certain contexts, where the particle has more specific readings, as demonstrated above for a number of specific contexts.

To sum up, in this section, the Sanzhi Dargwa particle *=q'al* and Forker’s analysis of it were presented. Forker identifies *=q'al* as a modal and predicative particle, which “conveys the same meanings in its occurrence as a pure modal particle and when it serves the additional predicative function” (2017: 16). Following Ariel’s (1988, 1998) Relevance Theory based approach to the Hebrew particle *harey*, Forker proposes that the core meaning of *=q'al* is signalling given, accessible information. Since it carries no truth-conditional meaning, and is used by the speaker to instruct the hearer on how to process the information from utterances in which it appears, it is placed in the category of procedural rather than conceptual items.

It is clear that *=q'al* is very similar to *ǰoj*, as both items are used in analogous contexts and functions, and produce comparable pragmatic effects. Forker’s Relevance Theory analysis accounts for the pragmatic effects the use of *=q'al* creates in various contexts, and neatly identifies the particle as a procedural item. However, I am reluctant to follow this analysis for *ǰoj*, since it does not sufficiently explicate all its uses. This primarily concerns the use of *ǰoj* as a copula in cleft construction, where the contrastively focused constituent is followed by *ǰoj* which, in turn, is followed by a relative clause. The proposed Relevance Theory analysis would not cover: a) the syntactic role *ǰoj* plays in these constructions; and b) the appearance of *ǰoj* between a correctively focused constituent and the pragmatically ‘given’ cleft/relative clause.

6.6.3. Tundra Yukaghir Particle $m\partial(r)=$

In this sub-section I consider Matić and Nikolaeva's (2014) analysis of the Tundra Yukaghir particle $m\partial(r)=$. Although this particle does not bear as clear a resemblance to $\check{g}oj$ as some of the previously discussed items from Russian and Sanzhi Dargwa, the semantic analysis proposed for it by Matić and Nikolaeva can be successfully applied to $\check{g}oj$, which is demonstrated in sub-Section 6.6.4.

In their paper on the realis mood, focus and existential semantics in Tundra Yukaghir (TY), Matić and Nikolaeva examine the realis/irrealis dichotomy in Tundra Yukaghir (isolate, north-eastern Siberia). They argue that "the TY realis is a hybrid category, composed of elements from different domains (information structure, lexical semantics, and quantification), but unified by a higher-level notion of existential closure of events in Davidson's (1967) sense" (2014: 204). It is demonstrated that realis is not a category of TY grammar, however, it is a relevant notion in this language. According to the authors, a realis clause in TY has the following minimal structure:

(81) Realis clause: X + V

where X is: (i) a non-verbal focus element, or
(ii) a verb modifier, or
(iii) the verbal particle $m\partial(r)=$

(Matić and Nikolaeva 2014: 206)

At first glance the items in (i)-(iii) in (81) above do not share any features, that is, they do not form a natural class, which could explain their participation in the process of forming a realis clause. I omit the detailed description of the first two classes and concentrate on the particle $m\partial(r)=$, as it is most relevant to the present work. Suffice it to say that the first and second class of items are not compatible with the verbal particle $m\partial(r)=$ as the occurrence of $m\partial(r)=$ in a sentence with a non-verbal focus or a verb modifier results in ungrammaticality; in other words the first two elements and the last element of class X are in complementary distribution.

An example containing $m\partial(r)=$ is presented in (82) below. The particle is glossed as EXST for existential quantifier.

- (82) lawjə moj-l-γa mə=pugeč
 water hold-1/2-DS EXST=be.warm.N.INTR.3SG
 ‘When I touch the water, it’s warm.’

(Matić and Nikolaeva 2014: 207)

Matić and Nikolaeva note that there is not consensus in the literature on the function of $m\partial(r)=$, and it has been analysed as a positive polarity indicator (Krejnovič 1958), a declarative marker (Fortescue 1996; Kurilov 2006), an affirmative prefix which signals narrow focus on the verb (Maslova 2003), and an indicator that the verb is within the focus domain (Matić and Nikolaeva 2008). Matić and Nikolaeva state that unlike all the previously proposed analyses analysing $m\partial(r)=$ as an existential quantifier exhaustively describes its meaning.

In order to identify the realis clauses, Matić and Nikolaeva make use of the terms Event Time (ET) and Topic Time (TT), the former being the time in which the event specified by the proposition takes place, and the latter - the interval of time “for which the particular utterance makes an assertion” (Matić and Nikolaeva, 2014: 222, citing Klein, 1994: 37). According to the authors, then, only the propositions that encode events for which ET is contained in TT ($ET \leq TT$) can be considered realis in TY. The role of the particle $m\partial(r)=$, according to this approach, is in existential quantification, that is, “it explicitly marks that the denoted event exists in the actual / preferred world and that the temporal location of the entire event is within the TT” (Matić and Nikolaeva 2014: 223).

Matić and Nikolaeva also explore in detail the connection between the first two elements in class X (as given in (81)), and demonstrate that a sentence containing focus or a verb modifier receives the realis interpretation by default and there is no need for an overt existential modifier $m\partial(r)=$ in these cases. This also leads to the understanding of the mutual exclusion of the first two categories of class X and $m\partial(r)=$. As all these elements in essence perform the same function - existential bounding of the event variable - once one of these elements is present in a proposition

there is no need for another one to be used. Matic' and Nikolaeva conclude that “the particle $m\partial(r)=$ is a last resort solution to overtly close the event variable” and that “[I]t is employed in those realis sentences where the alternative means for existentially closing the variable fail” (2014: 226).

Although we might not be dealing with the realis/irrealis distinction as far as the use of *ǰoj* in Kazakh is concerned, applying Matic' and Nikolaeva's analysis of $m\partial(r)=$ as an existential operator to *ǰoj* opens up a promising avenue of investigation. This is explored further in the following section.

6.6.4. *ǰoj* as an Existential Operator

In this sub-section I apply Matic' and Nikolaeva's analysis of $m\partial(r)=$ to *ǰoj*, and view both the particle and the copula *ǰoj* as carriers of existential semantics. It is demonstrated that this approach provides for a unified analysis of *ǰoj* in all its uses, as well as for the pragmatic effects resulting from the use of *ǰoj* in different contexts.

Let us first consider the existential operator semantics for the post-predicative use of *ǰoj* as a particle. As was shown in Sections 6.4 and 6.5, the particle *ǰoj* can optionally appear after different types of predicates, and produce a variety of pragmatic effects. The most common pragmatic effect is the effect of ‘givenness’ or ‘shared information’, as shown, for instance, in the example in (24), repeated below in (83).

(83) [The opening line of a folk-tale about camels.]

Tüje	qazırdıñ	özinde	ädemi	ǰoj.
camel	now	itself	beautiful	ǰoj

‘Camels nowadays are beautiful, aren't they. / As is well known, camels are beautiful at present time. [But a long time ago... (and the tale continues to say that camels used to be more beautiful and how they lost that beauty)].’

(ACKL)

The particle *ǰoj* here follows a predicate expressed by an adjective, and can be removed from the sentence in which it appears without affecting its grammaticality, meaning that it is used here for pragmatic rather than syntactic reasons. As the English approximation shows, the addition of *ǰoj* creates the pragmatic effect of ‘given’ or ‘shared’ information. That is to say, that the hearer or, in this case, the reader, is not simply informed that camels are beautiful, which would have been the case for the same sentence without *ǰoj*.

The addition of *ǰoj* instructs the reader to either: a) re-activate the information expressed in the proposition preceding it; or b) accept the information expressed in the proposition preceding it as ‘given’. As mentioned in the previous sections, the information does not have to have been physically shared by the interlocutors, it may be a piece of general knowledge accessible, or at least considered to be accessible, to all interlocutors belonging to a certain social or cultural group.

The proposed analysis of *ǰoj* as an existential operator allows for a simple and elegant explanation of this pragmatic effect. By adding *ǰoj* after a proposition *p*, the speaker indicates that *p* exists in the CG. The notion of CG is not restricted to shared information between a particular speaker and a particular hearer; it covers a wider CG, which the speaker or writer considers to be accessible for her hearers or readers. The sentence in (83) can roughly be paraphrased as ‘it exists that camels nowadays are beautiful’ or ‘camels nowadays are beautiful exists’ – *ǰoj* scopes over the complete proposition ‘camels nowadays are beautiful’, and confirms its existence. Since the proposition over which it scopes is already a fully-formed grammatical independent clause, *ǰoj* does not participate in its syntactic structure and can be easily removed or omitted.

Let us consider how the proposed analysis of *ǰoj* works for a sentence like that in (21), repeated below in (84).

- (84) [After finding out that one of the female employees who is not married has a son, one manager says to another:]
 Āke-si bol-u kerek qoj?!
 father-POSS.3 be-INF must ǰoj

‘There’s got to be a father, hasn’t there/right?!’

In this example *ğoj* appears after a modal verb *kerek* ‘must’ in its epistemic function, and creates the effect of intensification of that epistemic modality – what Straughn (2011) refers to as the ‘expressive’ or ‘emotive’ effect. Additionally, as indicated by the question mark, this utterance is pronounced with interrogative intonation, which, in combination with *ğoj* creates the effect of a rhetorical tag question. Notably, it is *ğoj* that is pronounced with a rising intonation, indicating interrogation – the existence of the preceding proposition in the CG is under question. This leads to the interpretation along the lines of: ‘does it exist that there must be a father?’, which translates into the English approximation given in (84).

The same effect is observed in (25), repeated below in (85).

- (85) Sen bugün universitet-ke bar-a-siñ ğoj?
2SG today university-DAT go-FUT/PRS-2SG ğoj
‘You are going to university today, aren’t you?’
‘You are going to university today, right?’

The proposition *you are going to university* is followed by an interrogatively pronounced *ğoj*, which questions the existence of this proposition in the CG. This questioning of the existence of the proposition in the CG, as opposed to questioning of the content of the proposition, creates the pragmatic effect of ‘previously shared information’ or ‘given information’. How the predicate is expressed in the proposition the existence of which is questioned by the interrogatively pronounced *ğoj* does not appear to be of importance here, as long as the proposition and the sentence are grammatically complete and independent.

Interestingly, a parallel item – *innit* – can be found in the colloquial London English. It is the heavily truncated version of ‘is it not’, used to form tag questions as shown in (86) and (87) below.

- (86) This dress looks good, innit?

(87) She will come to the party tonight, innit?

In Standard English, the tag question is formed by using an auxiliary verb (*be*, *do*, or *have*) in the same tense as the predicate of the clause followed by the tag, but in contrasting polarity; if the verb of the main clause displays agreement with the third person singular subject, so does the tag question. We see all these rules violated by the use of the tag *innit* in (86) and (87): in the former, the negated form auxiliary *do* in the third person singular would be used in Standard English, and in the latter, the negated form of *will* would appear.

Additionally, the form of the pronoun in the tag question in Standard English depends on its antecedent in the main clause. This, however, is not the case when *innit* is used, as the example in (87) demonstrates – the pronoun *she* would have been used in the Standard English version of the tag question there.

It appears that *innit* is a shortcut of sorts which signals a tag-type question whatever the form of the predicate in the main clause. This is very similar to what we have observed for *ğoj*, which doesn't follow any agreement patterns. Interestingly, the proposed existential semantics for *ğoj* matches the existential semantic of *innit*, which originated from the existential verb *to be*. Instead of forming a canonical tag question, *innit* questions whether the proposition expressed by the main clause applies or 'exists'. The sentences from (86) and (87) can be roughly paraphrased as 'is it the case that this dress looks good' and 'is it the case that she will come to the party tonight' respectively. This cursory observation requires further research, which is out of the scope of this work.

Returning to the use of the particle *ğoj*, let us examine one more example where *ğoj* follows a correctively focused predicate or verum focus, as shown in (50), repeated in (88) below.

(88) Speaker A: Qajrat keşe düken-ge bar-ğan žoq.
Kairat yesterday shop-DAT go-PTCP NEG
'Kairat did not go to the shop yesterday.'

Speaker B: Ol bar-di ěoj!
 3SG go-PST(3) ěoj
 ‘Of course he went! / He did too!’

As in all the examples considered in this section so far, *ěoj* follows the predicate of a complete main clause. In this case, the predicate expresses verum focus (described in Section 2.1 of Chapter 2), that is to say, its polarity is in opposition with the polarity of the predicate in the preceding sentence. Verum focus is expressed by the main stress falling on the verb itself, and the addition of *ěoj* creates the effects of intensification, correction, and givenness. These effects can be easily explained through the existential semantics of *ěoj*: once again it scopes over the proposition it follows, and confirms that this proposition exists in the CG. In other words, the proposition ‘he DID go’ or ‘he WENT’ followed by *ěoj* results in the reading: ‘it is the case that he DID go’ or ‘that he WENT exists’, which yields the intensification of verum focus.

McCoy states that using *ěe* adds the following meaning to an utterance: “You are wrong! And more than that, you are capable of arriving at the correct conclusion yourself, but nevertheless you are sticking to the wrong conclusion” (2003: 125). Some uses of *ěoj* might be interpreted in this way as well, however, I believe that this is also the effect produced by the existential nature of *ěoj*. If the speaker has explicitly shared some information with the hearer, and the hearer asks a question regarding the same piece of information again, the speaker is highly likely to add *ěoj* to her response to signal that this proposition already exists in the CG. This would then trigger further implications such as the pragmatics of reproach described by McCoy.

However, as we have seen from the examples above, the speaker might add *ěoj* after a proposition that has not been previously explicitly shared between the interlocutors. This, then, instructs the hearer to treat the information as present in the CG. No reproach is intended or perceived in this case, but the effect of previously shared or given information remains, since *ěoj* places the proposition into the CG.

The proposed analysis is also applicable to the copula *ěoj*, which is not unexpected since copulas often have existential semantics. As was shown in 6.4.2, copula *ěoj* participates in the cleft construction where it follows a

correctively/contrastively focused argument or adjunct, and is followed by a topical present or past participial verb form. The example from (61) illustrating this construction is repeated in (89) below.

(89) Speaker A: Keše Bolat düken-ge bar-dï.
 yesterday Bolat shop-DAT go-PST(3)
 ‘Bolat went to the shop yesterday.’

Speaker B: Arman ğoj bar-ğan!
 Arman ğoj go-PTCP
 ‘(Of course) it was Arman who went!’

It seems that the existential nature of *ğoj* in these examples is self-evident and uncontroversial. The interaction of the contrastively/correctively focused element and the existential semantics of *ğoj* result in the perceived intensification of the correction/contrast, which, in turn, is interpreted as emotional or expressive speech. The existential semantics of *ğoj* also yield the pragmatic effect of givenness or shared information, as previously discussed in this chapter for the particle *ğoj*.

As has been noted throughout this chapter, the pragmatic effects arising from the use of *ğoj* are context-dependent, and vary subtly from one context to the next – this confirms, that ‘givenness’ or ‘shared information’ cannot be the underlying meaning of *ğoj*, since they themselves are highly context-dependent. In proposing an unchanging existential semantic meaning for *ğoj* whose interaction with different contexts produces different pragmatic effects, we overcome the issue of equating the cause and the effect, as appears to have been the case in the analyses of *ğoj* and similar items.

One of the main questions that naturally arises is why *ğoj*, and not another existential element available in Kazakh (e.g. *bar* ‘there is’ or ‘there exists’ and *bol-* ‘be’) is used as a copula in this cleft construction. This question requires an in-depth diachronic and synchronic investigation, which is outside of the scope of this work. However, as a preliminary hypothesis, I can propose that the answer to this question

lies in the typology of copular clauses (see e.g. Higgins (1979), Declerck (1988), Mikkelsen (2005, 2011)).

Following Higgins (1979), four types of copular sentences or constructions have traditionally been identified in the literature: predicational, specificational, identificational, and equative. Den Dikken and O’Neill (2017) observe that the distribution of copular elements in languages with multiple-copula systems is often determined by types of copular construction. This is observed in many languages with multiple copulas, as for example, Geist (2007) reports for Russian, Pustet (2003) for Lakota, Gibson (2012) for Rangi, Michaelis et al. (2013) for Saramaccan, or Hedberg and Schneider-Zioga (2015) report for Kinande. It is therefore not unusual that a different copular element is used in the Kazakh cleft construction, which is distinct from other copular constructions in this language.

While this analysis is comparable with some of the Relevance Theory based analyses, its advantage is in the ability to be formally represented in syntactic/semantic theories, as the following section demonstrates.

6.7. Dynamic Syntax Analysis of *ǰoj*

In this section I present a sketch of a formal Dynamic Syntax analysis of both types of *ǰoj* identified in the preceding sections – the particle and the copula. The previous section demonstrated that analysing *ǰoj* as an existential operator provides for a convincing explication of the multiple pragmatic effects that result from its use as either a particle or a copula. This unified analysis also lends itself nicely to being formalised within the framework of Dynamic Syntax. In what follows I propose a single Dynamic Syntax lexical entry for both uses of *ǰoj*, which not only reflects the existential nature of the item, but also helps account for the difference in their syntactic distribution pattern.

Before considering the lexical entry for *ǰoj* and its application in the semantic trees, let us recall some of the basic facts about the syntactic behaviour of both types of *ǰoj*. The particle *ǰoj* typically follows the predicate, which can be verbal or non-verbal. The particle is optional, and its presence is determined by the pragmatic rather

than syntactic factors. It can be removed without affecting grammaticality of the sentence. The particle links the proposition expressed in the utterance it follows to some wider context by asserting the existence of this proposition in the wider common ground.

The copula *ğoj*, on the other hand, is not optional, and its omission leads to ungrammaticality. It participates in the cleft construction, where it follows a contrastively/correctively focused argument or adjunct expressed by case-marked or unmarked nouns. It is obligatorily followed by a relative clause headed by a participle, and containing presupposed information. It can be said then, that this copula links new information carried by the correctively focused element with the context in relation to which this new information must be processed.

Notably, despite the differences in the syntactic behaviour, both the particle and the copula *ğoj* perform some sort of a ‘linking’ function, which is in line with the existential semantics analysis proposed for these items in the preceding sections. The proposed Dynamic Syntax lexical entry for *ğoj* presented in the following sub-sections reflects these semantic and syntactic properties of the particle and the copula.

6.7.1. The Dynamics of Kazakh Relative Clauses

Before we consider the lexical entry for *ğoj*, we must take a look at the relative clauses in Kazakh, since the copula *ğoj* has a direct relationship with them in the cleft constructions. Recall three main features characteristic of relative clauses in Kazakh outlined in Section 1.1 of Chapter 1: a) relative clauses precede the head noun; b) there are no relative pronouns; and c) relative clauses are non-finite, and contain *-GAN* or *-AtIn* participles, as in (90) and (91).

- (90) Maqpal süj-etin Marat kel-di.
 Maqpal love-PTCP Marat come-PST(3)
 ‘Marat, whom Maqpal loves, arrived.’

- (91) Maqpal süj-gen Marat kel-di.

Maqpal love-PTCP Marat come-PST(3)
 ‘Marat, with whom Maqpal has fallen in love, arrived.’

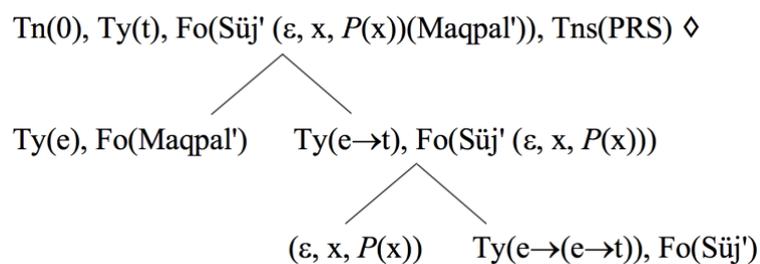
The relative clause in (90) is constructed with the present participial form of the verb *süju* ‘to love’, while that in (91) contains the past participial form of the same verb. This difference is reflected in the English approximations of the clauses. Let us consider how a semantic tree would develop for a clause like that in (90).

I present a simplified, sketch-like account, since the main purpose I pursue here is not to re-iterate what has already been shown in the literature, but to utilise the DS tools to formally represent the effect *ğoj* has on the sentences in which it appears as either a particle or a copula.

Recall from Section 2.2 of Chapter 2, that relative clauses are analysed in DS as a Linking relation between two trees: the tree containing the head noun, and that of the relative clause (Kempson et al. 2001; Cann et al. 2005). In the head-initial language, like English, for example, this means that the head noun is parsed first, and then a Linked structure for the relative clause is built from it. This plays out differently in head-final languages, like Kazakh or Japanese, for example.

Following Kempson and Kurosawa (2009), and Seraku (2013) in their DS analysis of Japanese relative clauses, I develop the DS representation of the parsing process for Kazakh relative clauses on the example from (90). The requirement for a $Ty(t)$ is set out by the Axiom, and the propositional structure of the relative clause *Maqpal süj-etin* is developed underneath it, as shown in (92). Since the object argument of the verb *süj-* ‘love’ is missing, it is represented as the epsilon term ($\epsilon, x, P(x)$), P being an abstract predicate (Kempson and Kurosawa 2009: 65).

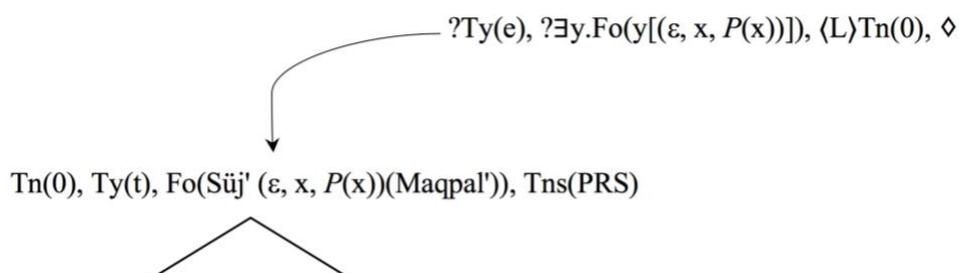
(92) Parsing *Maqpal süj-etin...*



Notably, there is a tense decoration at the root node of this tree, despite it being a non-finite relative clause. Recall (Chapter 1, Section 1.1), that Kazakh non-finite forms can carry tense features, and that finiteness in this language is contributed by the subject agreement suffixes.

The next step is the construction of the Link transition initiated by the general action Link Introduction, which builds an inverse Link (L^{-1}) relation from the current node (the root node) of the relative clause tree to a new node of the main tree with the requirement for $Ty(e)$. Another requirement decorating this new $?Ty(e)$ node is the requirement for this node to be decorated with a term y containing $(\varepsilon, x, P(x))$ as a sub-term. This requirement guarantees a shared term in the paired trees, which corresponds to the ‘gap’ in the relative clause tree.

(93) Link Introduction



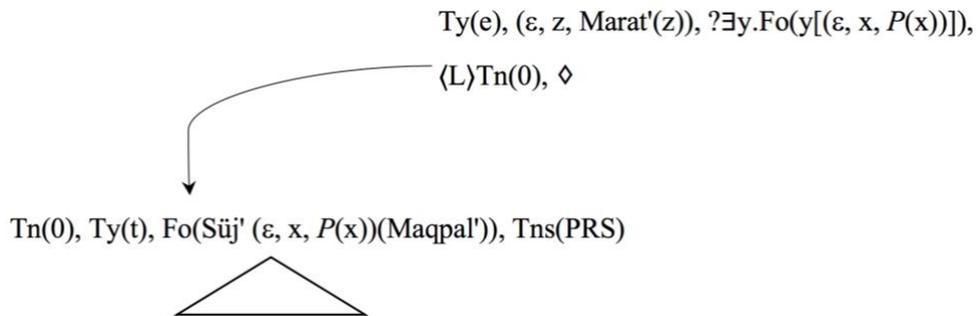
The formalism of the Link Introduction action is given in (94). The initial IF-statement ensures that the process is not initiated anywhere but at a type t node. Should this condition be satisfied, another IF-statement checks that some node underneath is decorated with a term α – in our example instantiated as $(\varepsilon, x, P(x))$. In case this IF-statement is also satisfied, an inverse Link relation is initiated by the parser, and subsequently decorated with $?Ty(e)$, and $?\exists y.Fo(y[\alpha])$.

(94) Link Introduction

IF	Ty(t)	
THEN	IF	$\langle D \rangle(\alpha)$
	THEN	make $\langle L^{-1} \rangle$, go $\langle L^{-1} \rangle$; put(?Ty(e), $?\exists y.Fo(y[\alpha])$)
	ELSE	Abort
ELSE	Abort	

The next word in the sentence under consideration is *Marat* – the head noun of the relative clause. It is parsed onto the newly constructed node with the requirement for Ty(e). This results in the tree from (93) being developed into the tree in (95) below.

(95) Parsing *Maqpal süj-etin Marat...*



Following this, the process of Link Evaluation (Cann et al. 2005) takes place. This means that the current Ty(e) node is enriched with the information from the Linked tree. The content of the relative clause is utilised by the parser, and the epsilon term $(\epsilon, x, P(x))$ is replaced by the variable z , which makes the variable z restricted by the information of the head noun and the relative clause.

Seraku labels this process Link Enrichment (Seraku 2013).²⁴ He provides a lexical entry for Link Enrichment as shown in (96) below. The term in the first IF-

²⁴ Link Evaluation is the original term appearing in Cann et al (2005), however, Seraku (2013) makes a point of labelling this process Link Enrichment in order to differentiate this operation from Quantifier Evaluation. I do not take a hard stance on this question, therefore both labels are presented for the convenience of the reader.

statement is the content of a head, and ψ in the second IF-statement is the content of a relative clause containing the term $(\varepsilon, x, P(x))$ as one of its arguments. If both the IF-statements are satisfied, then the parser takes the content of the relative clause ψ , and compiles it up into the term $(\varepsilon, y, \varphi(y))$ as an additional restrictor through replacing the term $(\varepsilon, x, P(x))$ within ψ with the variable y .

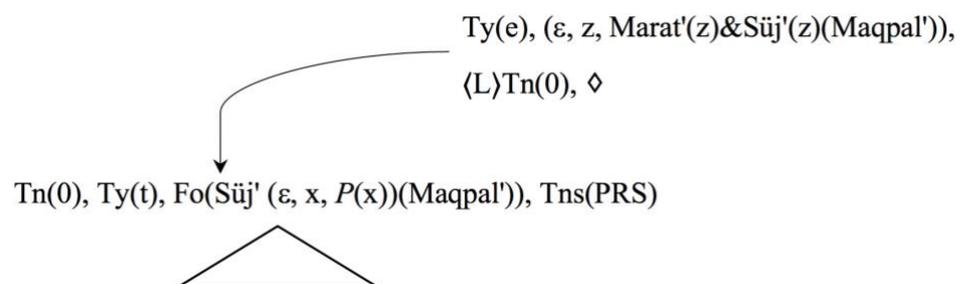
(96) Link Enrichment

IF	$(\varepsilon, y, \varphi(y))$	
THEN	IF	$\langle L \rangle (\psi[(\varepsilon, x, P(x))])$
	THEN	$\text{put}(\varepsilon, y, \varphi(y)) \& \psi[y/(\varepsilon, x, P(x))]$
	ELSE	Abort
ELSE	Abort	

(Seraku 2013: 97)

After Link Evaluation/Enrichment takes place, we have a term simultaneously representing both: an individual loved by *Maqpal*, and *Marat*. The result of the Link Evaluation/Enrichment process is presented in (97) below.

(97) Link Enrichment



Seraku points out that in a Japanese string the nominative case particle *ga* follows the head noun, as exemplified by *ishi-ga* in (98) below.

- (98) Sheeta-ga motsu ishi-ga hika-tta.
 Sheeta-NOM hold stone-NOM shine-PST
 ‘A stone which Sheeta holds shone.’

This, however, is problematic for the tree we have built so far, since, according to the standard DS understanding of nominative case presented for Korean in Chapter 2, and repeated in (99) below for Japanese (Seraku’s example (3.59), p.97), the current node, onto which the head noun is parsed must be a locally unfixed node of type *e*.

- (99) Lexical entry for *-ga* (nominative)

IF	Fo(α), Ty(<i>e</i>)
THEN	IF $\langle \uparrow_0 \uparrow_i^* \rangle$ (Tn(<i>a</i>), ?Ty(<i>t</i>))
	THEN put(? $\langle \uparrow_0 \rangle$ Ty(<i>t</i>))
	ELSE Abort
ELSE	Abort

This issue is also applicable to our Kazakh example, in which an IP boundary fixes a non-case-marked nominal, initially parsed onto an unfixed node, as the subject of a clause in which it appears (see Section 3.5 of Chapter 3).

As per Seraku’s observation, however, when the second nominal *Marat* is parsed in our sentence, instead of finding itself on an unfixed node being ready to be fixed as the subject of the clause, it is on a node which is fixed in relation to the tree Linked to it. Seraku proposes to resolve this issue by developing the nominative case lexical entry in such a way that it allows for a bottom-up construal of partial structure from a node that has no dominating nodes above it, until it reaches a node with ?Ty(*t*). A simplified version of Seraku’s revised lexical entry is shown in (100) (Seraku’s example (3.66), p.100).

- (100) Revised Lexical entry for *-ga* (nominative)

IF		Fo(α), Ty(e)
THEN	IF	$\langle \uparrow_0 \uparrow^* \rangle$ (Tn(a), ?Ty(t))
	THEN	put(? $\langle \uparrow_0 \rangle$ Ty(t))
	ELSE	IF $\langle \uparrow_0 \rangle \perp$
		THEN make($\langle \uparrow_0 \rangle$), (go($\langle \uparrow_0 \rangle$)); put(?Ty(t))
		ELSE Abort
ELSE		Abort

The first and second IF-statements in this lexical entry are the same as in the previous entry for nominative case shown in (99). The third IF-statement, found within the first ELSE-statement introduces the bottom-up construal process: it checks that the parser is at a node where there is nothing else above it. If this applies, the actions in the THEN-statement build an argument node, go up that node and add a requirement for a Ty(t) there.

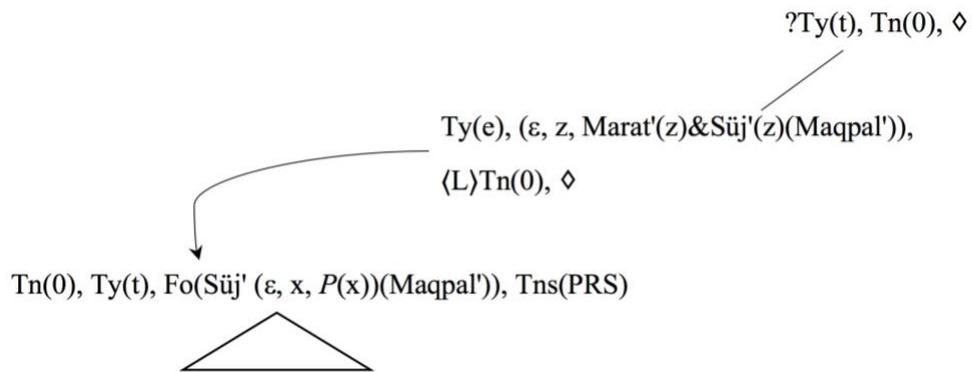
The lexical entry for an IP boundary originally presented in sub-Section 2.2.8 of Chapter 2 can be expanded in the same way:

(101) Revised Lexical entry for Kazakh IP boundary

IF		Fo(α), Ty(e)
THEN	IF	$\langle \uparrow_0 \uparrow^* \rangle$ (Tn(a), ?Ty(t))
	THEN	put(? $\langle \uparrow_0 \rangle$ Ty(t))
	ELSE	IF $\langle \uparrow_0 \rangle \perp$
		THEN make($\langle \uparrow_0 \rangle$), go($\langle \uparrow_0 \rangle$); put(?Ty(t))
		ELSE Abort
ELSE		Abort

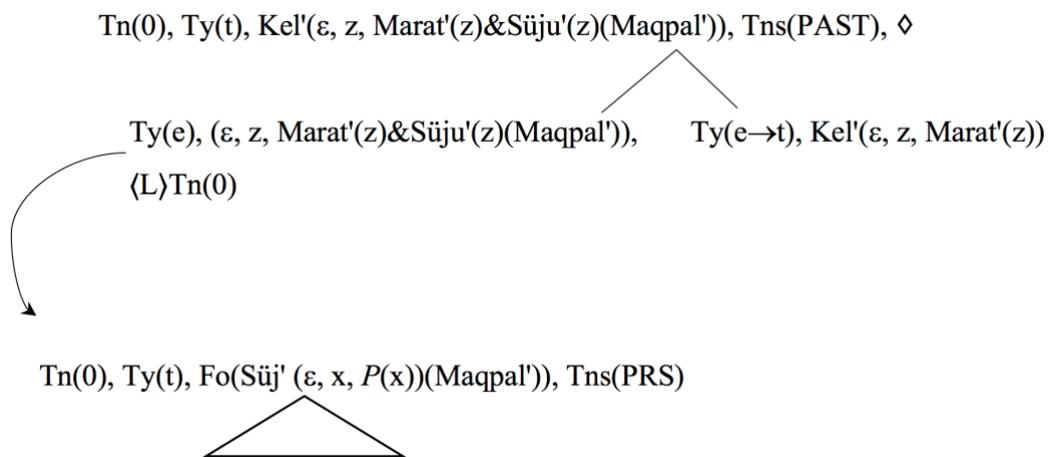
Following this approach, we can continue constructing from the tree shown in (97). The tree in (102) reflects the update from the parsing of the unmarked proper name *Marat*.

(102) Parsing *Maqpal süj-etin Marat %...*



The final step in our parse is the verb *kel-* ‘come’, which fleshes out the main tree. The process of Elimination cleans up the tree. The root node of the final tree given in (103) represents the propositional content of the whole sentence including the relative clause.

(103) Parsing *Maqpal süj-etin Marat kel-di*



The final tree in (103) essentially contains two clauses: one containing the proposition about *Maqpal* loving *Marat*, and the other – the proposition about *Marat* arriving. This is reflected in the root node of the final tree. The relative clause *Maqpal süjetin* is connected with the main clause via a Link relation, which ensures that one of the

formulas is shared by both clauses. This encapsulates the idea that one of the trees is constructed in the context of the other.

Notably, the final tree for the English version of this sentence would look exactly the same, which attests the semantic nature of the DS trees. The important cross-linguistic differences are captured within the tree-growth process, however, which would develop differently for the English version of the sentence, where the relative clause containing a relative pronoun would follow its head rather than precede it.

6.7.2. The DS Analysis of *ǰoj*-clefts

Now that we have seen how the parsing process of a Kazakh sentence with a relative clause typically unfolds, we can turn to *ǰoj*-cleft sentences, as in the example in (104) below, repeated from one of the preceding sections of this chapter.

- (104) [Speaker A: Keše Bolat düken-ge bar-dï.
yesterday Bolat shop-DAT go-PST(3)
'Bolat went to the shop yesterday.']
- Speaker B: Arman ǰoj bar-ǰan!
Arman ǰoj go-PTCP
'(Of course) it was Arman who went!'

Recall the important observations that have been made about these cleft constructions: a) the correctively focused element is immediately followed by *ǰoj*; b) only participial verb forms can be used in the post-*ǰoj* part of the sentence; and c) the post-*ǰoj* part of the sentence contains the backgrounded, presupposed information, usually repeated from the preceding utterance.

Recall also, that the proposed overall meaning of *ǰoj* is that of an existential operator. In this sketch DS analysis I take a shortcut and represent the content of *ǰoj* as Fo(Exist'), in much the same way as Cann et al. (2005) simplify their representation

of the English *be* to Fo(**BE**). The crucial difference between the English *be* and the Kazakh *ǵoj* is in that the former projects an underspecified one-place predicate whose content is established by context, and the latter builds a Linked structure and takes some or all of the content of the original tree as an argument of an existential clause.

In other words, the English *be* is a meta-variable of type *t*, whereas the Kazakh *ǵoj* behaves more like a full predicate which has a special requirement to be parsed on a Linked structure. A more detailed examination of formal semantics of *ǵoj*, especially in comparison with similar elements from other languages is out of scope of this work, but would provide grounds for a fascinating typological investigation.

Based on these observations, it can be said that *ǵoj* links the new, corrective information which precedes it with the presupposed context which follows it. Speaker B's sentence can be roughly paraphrased as: *Arman exists such that he is the one who went*. The role of *ǵoj*, then, is to structure the sentence in such a way that the relative clause creates context for the existential sentence. In other words, *ǵoj* triggers the construction of a Linked structure, as is reflected in the lexical entry presented in (108) below.

The difference between the construction of a canonical sentence with a relative clause, and a cleft-sentence, is that, in the latter, the parsing of a relative clause is 'interrupted' by the main existential clause constructed by *ǵoj*. This interruption contributes to the pragmatic effect of corrective focusing of one of the terms of the relative clause. Thus, the final tree in (110) is structurally identical to that in (103), however, the steps whereby these trees are built up are different, which reflects the differences in their syntactic structure, and in the diverse pragmatic effects associated with these constructions.

Another important observation must be made before the parsing process for our *ǵoj*-cleft from (104) is presented – it concerns prosody. Recall, that an IP boundary plays an important role in identifying an unmarked nominal as the subject of a sentence in which it appears, canonically, sentence-initially. Conversely, the lack of an IP boundary leaves a non-case-marked nominal unfixed.

In *ǵoj*-cleft sentences *ǵoj* appears to form a prosodic phrase with the nominals that it follows – that is to say, no IP boundary is observed between a nominal and *ǵoj*,

as demonstrated by Figures 3 and 4 which represent the sentence-initial elements from (105) and (106) (repeated from (104)) respectively.

- (105) Arman dükenge bardî.
 ‘Arman went to the shop.’
- (106) Arman ğoj barğan.
 ‘It was Arman who went.’

Figure 3. Sentence-initial NP with an IP boundary

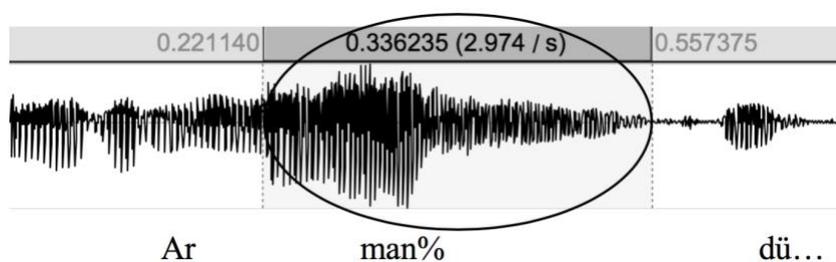
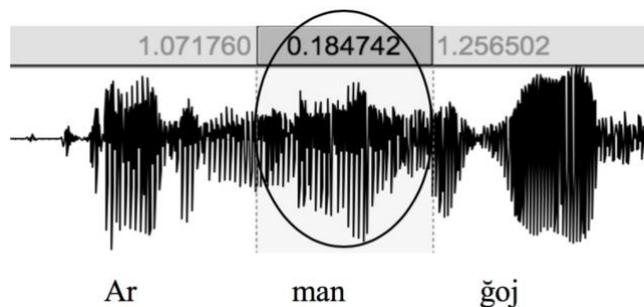


Figure 4. Sentence-initial NP – no IP boundary

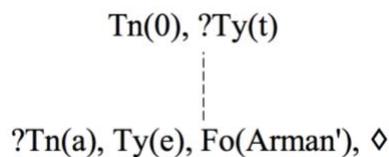


Comparing the length of the final syllable of sentence initial *Arman* from Figure 3 and Figure 4 – highlighted and circled – we can see a significant difference in the length of that syllable: it is 0.3362 milliseconds long in the former, and 0.1847 milliseconds in the latter.

This final syllable lengthening in the example where *Arman* is not followed by *ǰoj* indicates an IP boundary, which allows this sentence initial non-case-marked constituent to be parsed as the subject of the sentence, as was demonstrated in Chapter 3. Conversely, no lengthening of the final syllable of *Arman* in (106) – meaning no IP boundary – together with lack of case marking indicate that this constituent would remain unfixed until the verb is parsed (like unmarked direct objects do, as shown in Chapter 4). This is not an issue, however, as is shown below. We are now ready to consider the steps of the parsing process for our sentence from (104)/(106).

The tree in (107) shows the first step in this parsing process. As discussed above, the sentence-initial constituent here is not case marked, nor is it identified as the subject by an IP boundary. It therefore remains unfixed for the time being:

(107) Parsing *Arman*...



ǰoj appears next in this sentence, and triggers the construction of an inversed Linked structure as per the lexical entry in (108). Note, that the parsing of *goj* starts from the node of *Ty(e)* to ensure that the Linked structure is built in relation to the nominal immediately followed by *goj* and not some other nominal that might already be present in the tree. In this sense, *goj* behaves as a clitic or a case marker.

(108) Lexical entry for copula *ǰoj* – to be revised

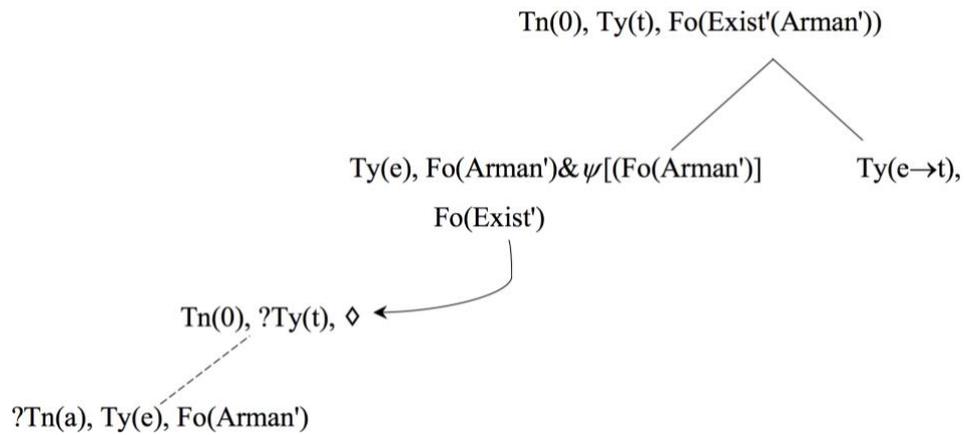
IF	Ty(e), Fo(α)
THEN	go($\langle \uparrow_0 \uparrow_1^* \rangle$, ?Ty(t)); make($\langle L^{-1} \rangle$), go($\langle L^{-1} \rangle$); put(Ty(e), Fo(α)& $\langle L \rangle \psi$ [Fo(α)]); make($\langle \uparrow_0 \rangle$), go($\langle \uparrow_0 \rangle$); put(?Ty(t)); make($\langle \downarrow_1 \rangle$), go($\langle \downarrow_1 \rangle$); put(Ty(e \rightarrow t), Fo(Exist')); go($\langle \uparrow_1 \rangle$); go($\langle L \rangle$)
ELSE	Abort

The IF statement checks that *g*oj is at a node with a type *e* formula decoration on it. The parser is then instructed to go up to the root node, which might be immediately dominating the current node – $\langle \uparrow_0 \rangle$; or be found one argument node and several functor nodes above it – $\langle \uparrow_0 \uparrow_1^* \rangle$.

The construction of an inverted Link is then initiated from the root node, as per the ELSE statement. Once the Link is constructed, a node is built and decorated with Ty(e) and Fo(α); this essentially copies Fo(α) across from the partially built structure into the new Linked structure. Additionally, the content of the clause down the Link (ψ) needs to be copied across to this node as well – this is achieved through the Link Enrichment process. This cannot happen at this point, since the relative clause has not appeared in full yet, however, this information will be compiled up during the final Elimination processes, as variables must be instantiated by the end of parse.

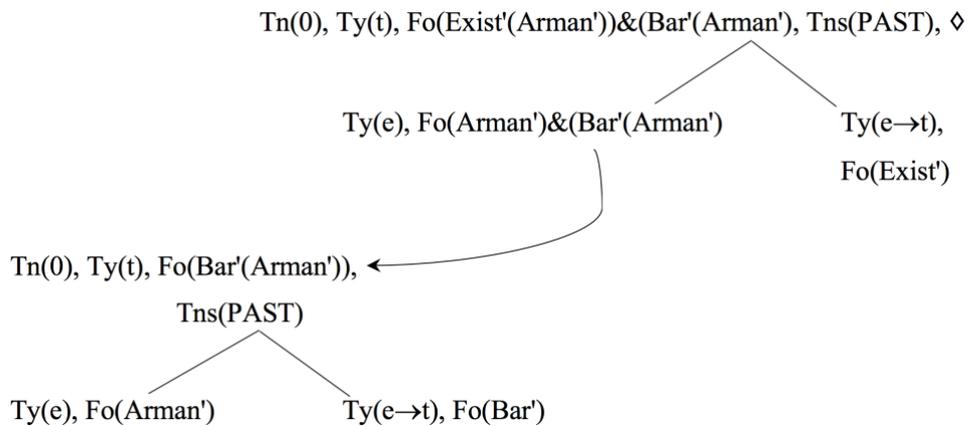
The new node is then fixed as the subject of the new tree under construction, and the parser moves up to its root node, from which a functor node is subsequently built. The functor node is decorated with Fo(Exist') reflecting the semantic content of *g*oj. Having built this tree containing an existential proposition, the parser goes back down the Link as per the last line of the THEN statement, and ends up at the root node of the original incomplete tree. At the end of this parse we are left with the tree in (109).

(109) Parsing *Arman g*oj...



The lexical verb in participial form can now be parsed. It projects the argument structure, and the previously unfixed node merges with the subject node. All the information is compiled up in the Elimination processes (for the relative clause tree, and for the main clause) with the resulting final tree given in (110).

(110) Parsing *Arman goj bar-gan* + Elimination (twice)



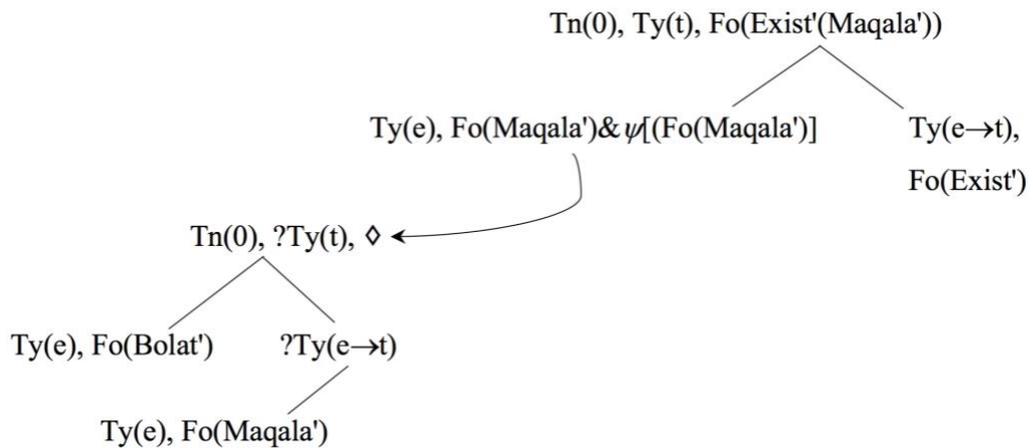
The parsing process develops in the same way regardless of which term is correctively focused and precedes *goj* in these constructions. If a case-marked direct object is followed by *goj*, as in Speaker B's utterance in (111), then the initial tree would contain a partially constructed tree with a subject and a direct object, as shown in (112) below.

(111) Speaker A: Bolat mīna kitap-ti žaz-dī.
 Bolat this book-ACC write-PST(3)
 ‘Bolat wrote this book.’

Speaker B: Ol mīna maqala-ni ġoj žaz-ğan.
 3SG this article-ACC ġoj write-PTCP
 ‘It was this article that he wrote.’

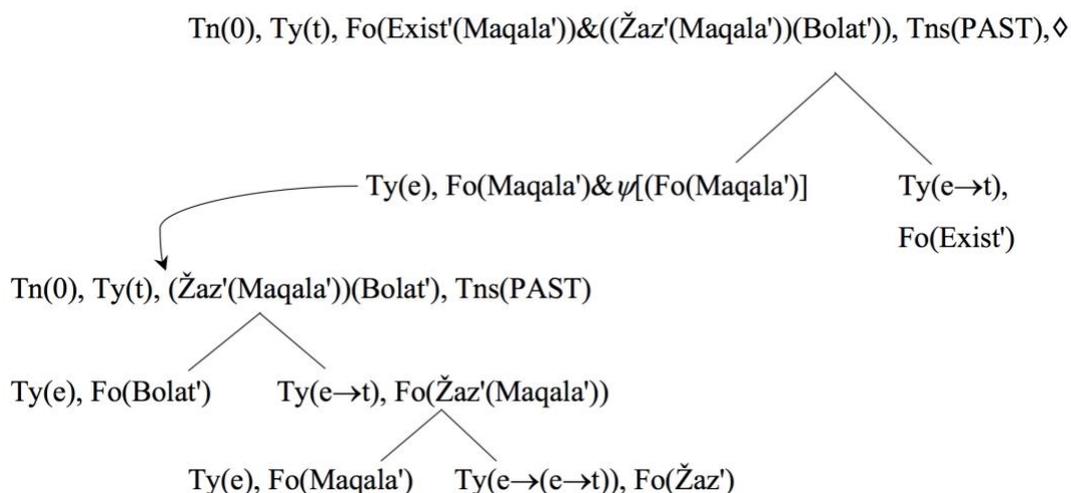
Notably, unlike in the previous parsing process, both nominals appearing before *ģoj* in this example can be parsed onto fixed nodes thanks to being case-marked.

(112) Parsing *Ol maqala-ni ġoj...*



The parsing of the participle *žazğan* completes the tree, and final tree after all the clean-up and finalising processes have taken places.

(113) Parsing *Ol maqala-ni ġoj žazğan* – final tree



To sum up, it was demonstrated that when *ğoj* appears post-nominally in cleft sentences, it initiates the construction of an inverse Linked tree structure, which takes the nominal immediately preceding *ğoj* as the subject of the existential predicate. This new tree functions as the main clause context for which is provided by the relative clause under construction. Since only participial verb forms can appear in relative clauses in Kazakh, no other verb forms are grammatical in these cleft constructions. At the same time, *ğoj* creates the main clause to which the relative clause can be Linked, thus avoiding ungrammaticality of a stand-alone relative clause.

The ‘interruption’ that *ğoj* causes in the construction of a clause and the ‘extraction’ of a nominal with its subsequent utilisation as the subject of an existential main clause contribute to the pragmatic effects associated with: a) cleft constructions, and b) the use of *ğoj*. The speaker uses *ğoj* to assert the existence of the ‘clefted’ nominal, which results in both the corrective focus and shared knowledge pragmatic effects.

Since this was only intended as a sketch account which would help shed some light on the nature of *ğoj*, many technical questions remain unanswered. For example, the question of finiteness/non-finiteness, as well as of tense and aspect contributions of different Kazakh verb forms, amongst others. These are out of the scope of this work, and must be left for further research. Let us now move to the post-predicative use of *ğoj*.

6.7.3. The DS Approach to the Post-verbal Particle *ĝoj*

In this section I present a sketch of a Dynamic Syntax analysis to the particle *ĝoj*. Recall, that this particle appears post-predicatively, and produces the pragmatic effect of ‘given’ or ‘shared information’. Unlike the post-nominal, copular *ĝoj*, the particle *ĝoj* is optional, and does not affect the grammaticality of a clause which it follows. In DS terms I take it to mean that *ĝoj* does not make a direct contribution to the root node annotation of the clause, but rather asserts the existence of the proposition carried by the root node, in much the same way as we so copular *ĝoj* do in the previous section.

The difference in the behaviour of the copula and the particle is reflected in the lexical entries. The lexical entry for the particle *ĝoj* is given in (114) below.

(114) Lexical entry for particle *ĝoj* – to be revised

IF	Ty(t), Fo(α)
THEN	make($\langle L^{-1} \rangle$), go($\langle L^{-1} \rangle$); put(Ty(t), Fo(α)); make($\langle \uparrow_0 \rangle$), go($\langle \uparrow_0 \rangle$); put(?Ty(t)); make($\langle \downarrow_1 \rangle$), go($\langle \downarrow_1 \rangle$); put((Ty(t \rightarrow t), Fo(Exist'))); go($\langle \uparrow_1 \rangle$); go($\langle L \rangle$)
ELSE	Abort

We can note the difference in the IF statement of the copula *ĝoj* in (108) and the particle *ĝoj* above in that the latter must be parsed at the complete Ty(t) node. This condition ensures that the particle appears after the predicate has been parsed and no outstanding requirements are present in the tree. Identically to the copula, however, the particle also builds an inverse Link to a leaf node of a new tree; the information from the complete Ty(t) node is subsequently copied onto that leaf node.

I follow Cann et al. (2005: 195) in the assumption that a formula of Ty(t) can decorate a subject node, which is posited in the analysis proposed by the authors for the expletive *it* constructions in English. The rest of the lexical entry proceeds to build

a new root node with the requirement for a $Ty(t)$ and the functor node which is decorated with the $Ty(t \rightarrow t)$ and $Fo(Exist')$.

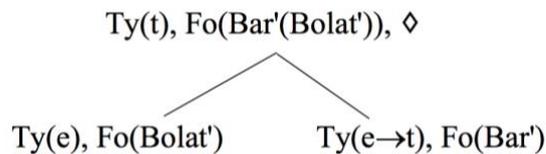
Let us now consider how the parsing process for a sentence with the particle *ǰoj* would develop on the example of Speaker B's utterance from (115) below.

(115) [Speaker A: Keše Bolat düken-ge bar-ma-dī.
yesterday Bolat shop-DAT go-NEG-PST(3)
'Bolat didn't go to the shop yesterday.']

Speaker B: Ol bar-dī ǰoj!
3SG go- PST(3) ǰoj
'He DID (too) go!' / 'Of course he went!'

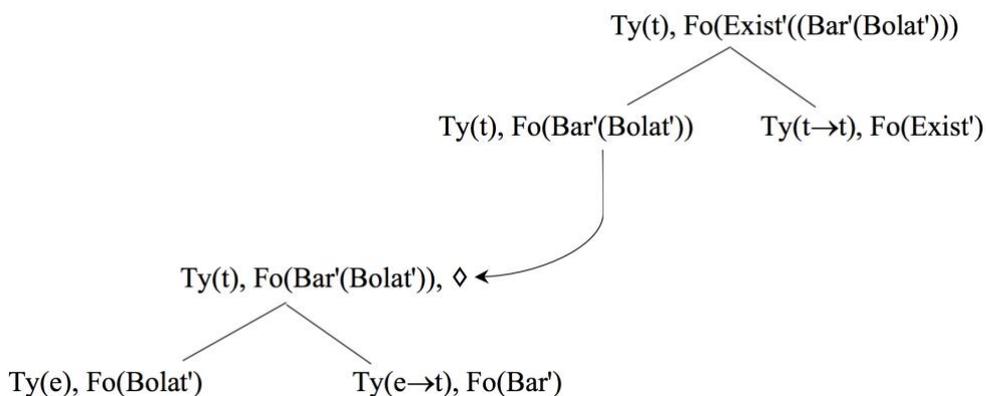
The tree in (116) illustrates the complete tree built from the parse of the string *Ol bardī*. Note that the pronominal *ol* got instantiated with $Fo(Bolat')$, to which it anaphorically refers.

(116) Parsing *Ol bar-dī* (in the context of *Bolat dükenge barmadī*, as above)



The tree in (117) is the update of the tree from (117) after the particle *ǰoj* has been parsed. The tree constructed by the original string *ol bardī* remains as is, while the actions from the lexical entry for *ǰoj* build the inverted Link relation and the existential clause. A copy of the formula of $Ty(t)$ from the initial root node decorates the argument node of the Linked existential tree thereby asserting the existence of the original proposition.

(117) Pasring *Ol bar-di ğoj*



Note that in accordance with the lexical entry from (114) the pointer returns back to the root node of the initial tree. This allows for the parsing of any post-predicative elements, like in the example in (20), repeated below in (118), where the direct object expressed by a pronoun appears after the predicate and the particle *ğoj*.

(118) [One friend utters to another as they hug after not having seen each other for a long time.]

Fusun, äbden	sağın-dī-q	qoj	seni.
Fusun very.much	miss-PST-1.PL	ğoj	2SG.ACC

‘Fusun, we did miss you a lot.’

(MBT)

This sentence-final, post-verbal direct object does not satisfy any outstanding requirements – all the requirements in this tree have already been satisfied. Its parsing simply builds the same direct object node, which is already present and complete in the structure, having been projected by the lexical actions of the predicate and updated from the context. This construction of the node that is already present in the tree yields the pragmatic effect of afterthought, but does not interfere with the normal tree construction and completion processes preceding it. Once any post-verbal elements (if there are any) are parsed, the final elimination processes take place, and all the

information gets compiled up to the root node of the existential clause, where the pointer also ends up.

The process of tree growth and the final tree provide a good representation of both the syntactic distribution and the semantic contribution of the particle *ǰoj*. In regards to the syntactic distribution, it is shown that the particle is optional since it does not satisfy any outstanding requirements in the tree constructed by the string which it follows. The trigger condition in the IF statement ensures that the particle follows the predicate.

As far as the semantic contribution is concerned, the existential clause constructed by the lexical actions of *ǰoj* reflects the existential semantics of *ǰoj*, and allows for a convincing account of the pragmatic effects associated with *ǰoj*, discussed in preceding sections. The final tree essentially renders the reading of ‘that Bolat went exists’ or ‘it exists / it is the case that Bolat went’, which creates the pragmatic effect of the proposition carried by the main clause as being shared or given.

The proposed analysis allows for an elegant account of the use of the particle *ǰoj* in interrogative sentences, like in the example in (25), repeated as (119) below for convenience.

- (119) Sen bügin universitet-ke bar-a-siñ ǰoj?
 2SG today university-DAT go-FUT/PRS-2SG ǰoj
 ‘You are going to university today, aren’t you?’
 ‘You are going to university today, right?’

As mentioned in the preceding sections, in the interrogative sentences with *ǰoj*, the interrogative prosody is realised on the particle, which is pronounced with a rising intonation. Contrary to some statements in the literature, I posit that it is this interrogative intonation that makes a sentence containing the particle *ǰoj* interrogative, not the particle itself – the lexical entry in (114) reflects this since it makes no reference to interrogation.

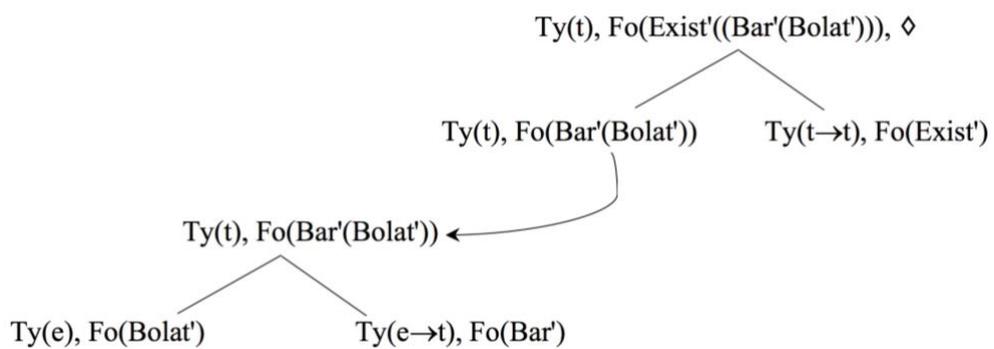
Let us return to our string *Ol bardı ǰoj* from above, which can also be pronounced with an interrogative intonation, as shown in (120) below.

(120) [Friends discussing Bolat's offer to go to England for studies.]

Ol bar-dī ğoj?
 3SG go-PST(3) ğoj
 'He went, didn't he?'
 'He did go, didn't he?'

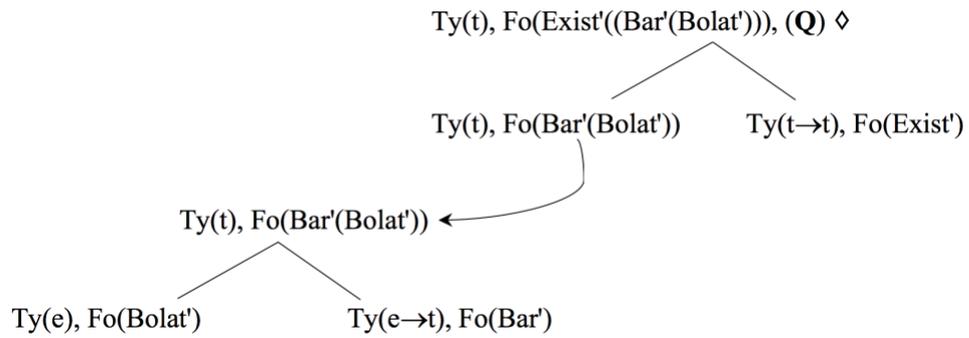
The tree in (121) shows the final tree for the string *Ol bardī ğoj* – all the tree construction and clean up processes have taken place, and the pointer is at the root node of the existential clause. This is where we see the interaction between the original proposition, the existential semantics of *ğoj*, and the interrogative intonation.

(121) Parsing *Ol bar-dī ğoj*



The tree in (122) has an additional decoration – **(Q)** – at its root node; this decoration is just a simplified representation of interrogation, which is not intended to be taken as a serious account of interrogation, but suffices for the illustrative use in our sketch.

(122) Parsing *Ol bar-dī ğoj?*



The addition of interrogation to the root node of the existential clause built by the lexical actions of *ǰoj* reflects the semantic and pragmatic effects created by *ǰoj* pronounced with interrogative prosody. These types of questions can be rephrased as ‘Is it / Isn’t it the case that...?’, or, literally, ‘Does it / doesn’t it exist that...’ – and this is just the meaning arising from the tree in (122). In other words, the question is not to the ‘main’, meaningful proposition, but to the existential clause constructed by *ǰoj*, as reflected in (122).

6.7.4. Unified Lexical Entry for *ǰoj* and Conclusions

I conclude this DS section with a unified lexical entry for the copula and the particle *ǰoj*, which is constructed by bringing together the lexical entries presented in (108) and (114). The possibility of constructing one lexical entry that can reflect the diverse syntactic behaviour and distribution of an item presents another reason for the use of the DS formalisms as the theoretical underpinning for the analysis of *ǰoj*.

Combining the lexical entries from (108) and (114) results in the lexical entry presented in (123).

(123) Lexical Entry for copula and particle *ǰoj* – revised

```

IF          Ty(e), Fo( $\alpha$ )
THEN
  go( $\langle \uparrow_0 \uparrow_1^* \rangle$  ?Ty(t)); make( $\langle L^{-1} \rangle$ ), go( $\langle L^{-1} \rangle$ );
  put(Ty(e), Fo( $\alpha$ ) &  $\langle L \rangle \psi$ [Fo( $\alpha$ )]);
  make( $\langle \uparrow_0 \rangle$ ), go( $\langle \uparrow_0 \rangle$ ); put(?Ty(t));
  make( $\langle \downarrow_1 \rangle$ , go( $\langle \downarrow_1 \rangle$ ); put(Ty(e  $\rightarrow$  t), Fo(Exist'));
  go( $\langle \uparrow_1 \rangle$ ); go( $\langle L \rangle$ )
ELSE
  IF          Ty(t), Fo( $\alpha$ )
  THEN
    make( $\langle L^{-1} \rangle$ ), go( $\langle L^{-1} \rangle$ );
    put(Ty(t), Fo( $\alpha$ ));
    make( $\langle \uparrow_0 \rangle$ ), go( $\langle \uparrow_0 \rangle$ ); put(?Ty(t));
    make( $\langle \downarrow_1 \rangle$ ), go( $\langle \downarrow_1 \rangle$ ); put((Ty(t  $\rightarrow$  t), Fo(Exist')));
    go( $\langle \uparrow_1 \rangle$ ); go( $\langle L \rangle$ )
  ELSE
    Abort

```

As mentioned in Section 6.7.3, the main differences between the lexical entry of the copula and the particle *ğoj* lie in the trigger conditions (the IF-statements), and in the type of the formula that gets copied across a Link onto a leaf of the new existential tree constructed by *ğoj*. This accurately accounts simultaneously for the differences in the syntactic distribution of these items, and for the similarity of the pragmatic effect of ‘givenness’ or ‘shared information’ that they create.

It must be noted, however, that this lexical entry can and should be enriched with another algorithm for *ğoj* – the imperative form of *qojju* –, as discussed in Section 6.3; this is left for further research, since it requires a detailed analysis of how complex/compound Kazakh verb forms are parsed, which is out of the scope of this work. It can be predicted, however, that the lexical entry for the imperative *ğoj* would have a different trigger condition, which would allow to identify its distinct function, diverse from the copula and the particle.

To sum up, this section has provided a sketch of a formal Dynamic Syntax analysis for the copula and the particle *ğoj*. It has been demonstrated that despite displaying disparate syntactic characteristics, both the copula and the particle *ğoj* initiate the construction of a Linked structure from the main propositional tree to a new existential tree. In the case of the copula, a Ty(e) formula gets copied across the

Link onto the subject node of the existential clause, whereas the particle uses the copy of the formula for the whole clause – Ty(t) – as an argument for an existential clause it constructs.

It has been shown that a single DS lexical entry for *ǰoj* can be constructed since the trigger conditions for these items are not the same. The copula *ǰoj* always follows a nominal element of Ty(e), and appears in an incomplete clause; whereas the particle follows a complete Ty(t) formula – which also accounts for its optionality.

6.8. Conclusions

This chapter has presented previously unseen or discussed data on *ǰoj*. It has been demonstrated that contrary to previous claims of *ǰoj* being an optional particle of some kind, word form *ǰoj* can function as the imperative form of the verb *qoju*, a copula, and a particle. Focusing on the latter two functions, this chapter shows that, as a copula, *ǰoj* participates in cleft constructions which have not been mentioned in the literature to date. When functioning as a copula *ǰoj* is not optional, and cannot be omitted without disrupting the grammaticality of the clause. The particle *ǰoj*, on the other hand, is optional and can be omitted without affecting the grammaticality of the clause in which it appears.

Since there are no developed analyses of *ǰoj* in the literature, I considered the analyses for two similarly functioning items in this chapter: the Russian (Slavic) *že*, and the Sanzhi Dargwa (Nakh-Daghestanian) *=q'al*. Both analyses propose a pragmatic notion as the foundation for unified analyses of the multi-functional nature of those items. I propose to follow a different path, and apply Matic and Nikolaeva's (2014) analysis of the Tundra Yukaghir particle *mə(r)=* as an existential operator to both the copula and the particle *ǰoj*. Not only does this analysis successfully account for the pragmatic effects produced by the use of *ǰoj*, but it also allows for a formal representation of *ǰoj*, to which the final section of this chapter are dedicated.

In that section, I provided a sketch of a formal Dynamic Syntax analysis for the copula and the particle *ǰoj*. The existential nature of both items is reflected in their respective lexical entries, which are subsequently combined into one. While the lexical

actions contained in the copula and the particle *ǰoj* are almost identical, some important differences explicate the divergence in their syntactic behaviour and distribution. This is one of the advantages of using the DS formalisms for analysing these items. Another advantage is the ability to formally represent the difference in the development of the parsing process for strings containing different types of *ǰoj*.

7. Main Results, Contribution, and Further Research Questions

This chapter summarises and concludes this thesis and proceeds as follows: section 7.1 provides the summary of the main results of this thesis; in section 7.2 I outline the contributions these results make to several areas of linguistics research; and section 7.3 contains some questions for further research.

7.1. Summary of the Main Results

The objective of this thesis has been dual: to provide the first detailed, synchronic, data-rich description of information structural characteristics of the Kazakh language; and to formally model some of the Kazakh information structural phenomena using the tools of the syntactic theoretical framework of Dynamic Syntax (DS) (Kempson et al. 2001; Cann et al. 2005). The first two chapters of this thesis provide the foundation for the subsequent four core chapters by introducing the necessary grammatical and theoretical background. The core chapters – Chapters 3-6 – are dedicated to the relation between word order and IS, DOM, topic markers *še*, *bolsa*, and *degen*, and the particle *ğoj* respectively. Each of the four core chapters focuses first on describing an information structural phenomenon, and then presenting a DS analysis of it.

In regard to the word order-IS relation in Kazakh, this thesis has demonstrated that scrambled word orders in this language are information structurally/pragmatically motivated. This puts Kazakh into the class of discourse-prominent, rather than discourse-configurational (as defined by É. Kiss (1995)) languages. As noted by Surányi (2016), in contrast to discourse-configurationality, discourse-prominence does not assume a systematic and restricted association of a particular syntactic position with the notions of topic or focus, but simply means that a sentential word order of a language is best described with respect to information-structural categories.

And indeed, it has been shown that in Kazakh, a sentence-initial element, for instance, is more likely to have a particular information structural status, namely, to be topical rather than be of a particular syntactic or grammatical category. The same

applies to constituents appearing in the IPV position: they are typically focal, but can belong to diverse syntactic or grammatical categories.

On the other hand, however, it has also been demonstrated that the neutral and most of the scrambled word orders in Kazakh can have more than one information structural configuration, especially if the notions of topic and focus are expanded to include their contrastive counterparts; the latter are typically distinguished via prosodic means. The inclusion of the notion of contrastive focus into the repertoire of descriptive information structural terminology also leads to the replacement of the term ‘focus position’ with the term ‘focus field’, which comprises the whole of the preverbal area up to and including the verb.

The DS sketches of diverse word orders and information structural configurations have further revealed that all information structural effects observed in a Kazakh sentence are to do with the way the linguistic strings are parsed incrementally within context, rather than with separate layers of grammar, specific syntactic positions, information structural features, or movement. That is to say, that words in a linguistic string appear in a certain order so as to create the information structural effects of topicality or focusing as appropriate within a given context.

The main outcomes from the description and analysis of the word order-IS relation in Kazakh can be summarised as follows. Firstly, it has been observed that different information structural configurations can be expressed either via ordering of constituents within a sentence, or with the help of prosody; the latter is especially pertinent of contrastive items.

Secondly, it has been shown that the key information structural notions of topic and focus must be expanded to include their contrastive counterparts to achieve a comprehensive description of information structural configurations available for neutral and scrambled Kazakh sentences.

Thirdly, these key information structural notions should only be treated as descriptive categories that allow to consistently outline various information structural configurations for diverse or one and the same word orders. These notions should not be viewed as primary grammatical categories, since they are context-dependent information structural / pragmatic effects.

And finally, no additional devices are necessary for explaining the information structural effects of topic, focus and contrast, as they naturally arise in the process of incremental contextualised parsing of linguistic strings, which can be formalised through of DS semantic trees.

This thesis has also demonstrated that Kazakh has special items that ensure that a constituent is parsed as a (contrastive) topic; these items are the topic markers *še*, *bolsa*, and *degen*. It has been shown that the first two of these topic markers follow and indicate contrastive topics, and the third one can follow contrastive or non-contrastive topics. Several interesting ongoing grammaticalisation processes have been uncovered for all three topic markers (Chapter 5). The DS formalisms have been successfully used to explicate the employment of conditional verb forms (*še* and *bolsa*) in the topic marking function.

Kazakh DOM has been considered in this thesis due to its direct relation with the matters of word order variation, and pragmatic statuses of marked and unmarked direct objects; several main conclusions have been reached. Firstly, it has been shown that marked and unmarked direct objects are linked with the notion of specificity (as outlined in Section 2.1 of Chapter 2) – marked direct objects are perceived as having specific referents (which may be definite or indefinite), and unmarked direct objects are considered to have non-specific referents.

Secondly, the DS analysis of DOM has demonstrated that the difference in specificity between marked and unmarked objects is not encoded in the accusative case markers directly (as has been claimed for Turkish, a language related to Kazakh) but is rather the pragmatic effect arising from the difference in how marked and unmarked objects are parsed: onto fixed and unfixed nodes respectively. This analysis elegantly explicates Kazakh DOM and helps to avoid assigning an additional pragmatic meaning to the accusative case.

Thirdly and finally, the DS analysis of DOM has been extended to explicate the special nature of the IPV position to which unmarked direct objects are confined. It has been shown that unmarked direct objects must appear in that position because they are guaranteed to be fixed in the next step of the parsing process – when the verb is parsed; additionally, this positional restriction prevents two unfixed nodes from

appearing in a tree which would prevent the parsing process from proceeding. This approach to the IPV position does away with the need to assign any assumed syntactic features to it or indeed to constituents appearing in it.

The final core chapter of this thesis examines the particle *ğoj* which has been analysed as a particle (of different types) indicating emotivity, emphasis, or givenness; similar items in other languages have been analysed as indicators of givenness / shared information or contrast. The following main conclusions have been drawn on the basis of the description and analysis of *ğoj* presented in this thesis. Firstly, it has been demonstrated that there are two syntactically diverse variants of *ğoj*: a copula and a particle. The former is used in the construction of cleft sentences, while the latter follows complete propositions and is not directly involved in their syntax.

Secondly, it has been shown that both types of *ğoj* add the same pragmatic effects of givenness or shared information to sentences in which they appear. It has been proposed that these pragmatic effects arise from the existential semantics that both types of *ğoj* carry.

Thirdly, the DS formalisms have been used to demonstrate that both types of *ğoj* initiate almost identical sequences of lexical actions, but at different levels – the copula *ğoj* acts ‘within’ a sentence, and the particle *ğoj* ‘over’ it. This difference has been formally represented through the difference in the trigger conditions of the copula and the particle. It has also allowed for a unified DS lexical entry for *ğoj* to be constructed.

7.2 Contribution to the Field

As the previous section demonstrates, this thesis covers several diverse phenomena which are unified by being relevant to Kazakh IS. On the one hand, this diversity is a welcome and necessary quality in a thesis like this, since it has allowed for a broad coverage of previously undescribed and unanalysed IS-related matters. On the other hand, however, it makes it somewhat challenging to frame the main contributions of this thesis within narrow theoretical discussions. However, below, I situate the

material covered, analyses presented, and conclusions drawn in this thesis within some wider linguistic fields.

Firstly, this thesis has started to fill a considerable gap in the research on the Kazakh language and more specifically, on its IS. It provides the first detailed description of the relation between word order and IS in Kazakh, which represents a significant move away from the previous restricted and oversimplified understanding of information structural characteristics of a Kazakh sentence, by demonstrating that a linguistic string can encode more than one information structural configuration. It has also become evident that the number of information structural configurations that can be identified for one and the same string very much depends on the number of information structural notions, means of their identification, and contexts one takes into consideration.

These observations not only advance our understanding of Kazakh word order-IS relation, but also provide for a basis for further investigation of this relation in other Turkic languages. Together with the DS approach to the matter adopted in this thesis these observations feed into a much larger discussions on discourse-configurationality and discourse-prominence (cf. Halliday 1967, Li and Thomsson 1976, Vilkuna 1989, É. Kiss 1995, Wedgwood 2003, Neeleman and van de Koot 2008, Surányi 2016, amongst many others), which are far from being settled. Additionally, this contributes to the wider areas of typological research on IS, where Turkic languages do not appear to be very well represented judging, for instance, by their absence from the *Oxford Handbook of Information Structure* (2016, edited by Féry and Ishinara).

Moreover, these observations, together with the conclusions on the context-dependent nature of the key information structural notions outlined in the previous section, contribute to a broader theoretical discussion on the notions of grammatical and information structural categories (cf. Newmeyer 2007, Rijkhoff 2009, Haspelmath 2010, amongst many others for discussions on grammatical categories). As far as the information structural categories are concerned, the data and the analyses presented in this thesis support Matić and Wedgwood's (2013) view against considering such information structural category as focus, for example, a linguistic category (see also

Matić and Nikolaeva 2018 for a discussion). This view is also intrinsic to the theoretical framework of DS adopted in this thesis.

Secondly, this thesis has presented the first attempt at formally analysing Kazakh DOM. While previously fairly well described, this phenomenon in Kazakh has not been formally analysed, unlike its Turkish counterpart, for which diverse analyses had been proposed. The proposed analysis for Kazakh DOM is novel not only by virtue of its existence, but also due it being developed through the tools and formalisms of DS.

The proposed DS analysis of Kazakh DOM breaks away from the traditional analyses (proposed for Turkish DOM) which ultimately either assign some additional meanings to the accusative case marker, or assume a difference in the syntactic positions of marked and unmarked direct objects (see e.g. Kornfilt 1997, Gerdts 1998, Kuribayashi 1989, Mithun 1984, Erguvanlı 1984, Aydemir 2004, Orgun and Inkelas 2004, amongst others).

Instead, it has been demonstrated that the differences in the pragmatics and the syntactic behaviour of marked and unmarked objects can be straightforwardly explicated through the basic DS concepts of fixed and unfixed nodes, and their applications to the parsing of strings with differently marked direct objects. The proposed DS approach to Kazakh DOM can be extended to Turkish DOM, DOM in other Turkic languages, and, potentially, in non-related languages.

Thirdly, the descriptions of Kazakh topic markers *še*, *bolsa*, and *degen* provided in this thesis make a significant contribution to the fields of Kazakh and Turkic studies, since no detailed accounts of these items had been published previously. These descriptions also contribute to the broader fields of information structural studies and typological studies, as not much has been written about Turkic topic markers to date.

Fourthly, the detailed description and analysis of the particle *ğoj* that have been developed in Chapter 6 of this thesis are the first of their kinds for this item thanks to the great number of contextualised natural and elicited examples that have been collected. This makes an important contribution to the body of research on Kazakh

and other Turkic languages in which similar items exist but have been ignored in linguistic literature.

Great care has been taken in this thesis to tease apart the pragmatic effects from the semantic meaning of *ǰoj*. Instead of taking the pragmatic effects of givenness / shared information yielded by *ǰoj* to be its meaning, it has been suggested that these effects arise from the interaction of *ǰoj*'s existential semantics with the proposition in which it appears. This contributes to many ongoing discussions about the meaning of similar items in different languages (e.g. Feldman 2001 and McCoy 2003 for Russian; Ariel 1988, 1998 for Hebrew; Werner 1991 and Karagjosova 2001, 2006 for German; Forker 2017 for Sanzhi Dargwa; amongst many others).

Additionally, these data-rich description and data-led analysis of *ǰoj* can prove useful for a larger debate on classification of such items as modal particles, discourse particles/markers, or something else (e.g. Degand et al., 2013; Diewald, 2013; Bayer and Struckmeier, 2017, amongst many others).

A separate note is in order regarding the contribution this thesis makes to DS. It must be said that all central notions of DS remained unchanged, and have been used in the same form as originally defined in Kempson et al. (2001) and Cann et al. (2005). However, the DS analyses proposed here for several phenomena add to the growing body of DS research, and demonstrate this syntactic framework's capability to deal with much-debated-over matters in an elegant and straightforward way.

7.3. Questions for Further Research

As mentioned earlier, this thesis has covered several connected yet diverse topics, which prevents us from identifying a single direction for further research. Many questions for further synchronic and diachronic research arise out of the findings of this thesis – some have been pointed out throughout, and in this section I focus on three potential future research avenues.

The first major line of enquiry that can develop from and at the same time enrich the account of word order-IS relation in Kazakh is to do with the relation between prosody and IS. A detailed empirical examination of various prosodic

contours, perhaps in combination with IP boundaries or pauses, and their connection with such information structural notions as focus, topic, and contrast (cf. Hedberg and Sosa 2008 for English, Downing et al. 2004 for Chichewa, Sudhoff 2010 for German, amongst many others) in Kazakh and other lesser-researched Turkic languages would not only contribute to Turkic linguistics, but also allow for further cross-linguistic comparisons and observations.

The findings of this prosodic research would enable subsequent formalisation of prosodic signals in DS, along the same lines as Kiaer's (2007, 2014) lexical entry for an IP boundary in Korean, and the lexical entry for an IP boundary in Kazakh presented in this thesis. Including all prosodic signals into the DS representation of parsing would provide for a comprehensive formal account of incremental language processing, and the relation between prosody and information structural effects.

Another interesting channel for further research would be a diachronic investigation into the origins and paths of development and grammaticalisation of Kazakh topic markers. Some useful insights into this question could be gained from comparative synchronic and diachronic studies of related Turkic languages such as Turkish, Uzbek, Kyrgyz, and others. As has been shown in this thesis, many grammaticalisation processes to do with topic markers are currently ongoing in Kazakh.

One of these processes – grammaticalisation of *bolsa* as a topic marker – appears to have completed in Turkmen, but is still ongoing in Kazakh and Uyghur. A comparative analysis of the development of this topic marker in these and other related languages would provide valuable insights not only into the general understanding of the origins and evolution of topic markers, but also into grammaticalisation in general by teasing out the linguistic and other factors that might explicate the difference in the completion rate of the same grammaticalisation process.

Further research is certainly required for the pragmatic marker *ǵoj* and its equivalents in other Turkic languages, since hardly any research has been carried out on these items. Diachronic and comparative study of *ǵoj* were out of the scope of this thesis, but are essential to the full understanding of its origins and development

paradigm. Application of the analysis for *ǰoj* proposed in this thesis to other similar items in unrelated languages – the Russian *že* and *ved'*, for instance – would

ǰoj-cleft constructions also warrant further attention, since, despite being recognised by the speakers as grammatical constructions in Kazakh, their use is in decline, possibly under the influence of the Russian language over the many decades of its dominance in Kazakhstan. Further research can reveal whether this is also the case for other Turkic languages spoken the former USSR republics.

These are just several of many possible developments for further research that can be based on findings presented in this thesis, since so little research has been carried out on Kazakh in general, and its information structure and related phenomena in particular.

Bibliography

- Abbott, B., 2000. Presuppositions as Nonassertions. *Journal of Pragmatics*, 32, 1419-1437.
- Abels, K. & Muriungi, P., 2008. The Focus Marker in Kîtharaka: Syntax and Semantics. *Lingua* 118, 687-731.
- Abish, A., 2014. *Modality in Kazakh as Spoken in China*. Ph.D. Uppsala Universitet.
- Ahmed, T., 2006. Spatial, Temporal and Structural Uses of Urdu *ko*. In: Butt, M. & King, T.H. (eds.), *Proceedings of the LFG 06 Conference*. Stanford: SCLI Publications, 1-13.
- Aralbaev, Ž.A., 1988. *Qazaq Fonetikasi bojînša Etjudter* [Notes on the Kazakh Phonetics]. Almatî: Ğilim.
- Ariel, M., 1988. Retrieving Propositions from Context: Why and how. *Journal of Pragmatics* 12, 567-600.
- Ariel, M., 1998. Discourse Markers and Form-function Correlations. In: Jucker, A.H. & Ziv, Y. (eds.), *Discourse markers: Descriptions and theory*. Amsterdam: Benjamins, 223-259.
- Aronson, H., 1967. The Grammatical Categories of the Indicative in the Contemporary Bulgarian Literary Language. In: *To Honor Roman Jakobson*. The Hague: Mouton, 82-98.
- Arregi, K., 2016. Focus Projection Theories. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 185-202.
- Aydemir, Y., 2004. Are Turkish Preverbal Bare Nouns Syntactic Arguments? *Linguistic Inquiry* 35, 465-474.
- Bailey, L., 2010. Sentential Word Order and the Syntax of Question Particles. *Newcastle Working Papers in Linguistics* 16, 23-43.
- Bailey, L., 2013. Sentential Word Order and the Syntax of Question Particles. *Proceedings of International Conference on Final Particles, University of Rouen, 27-28 May 2010*. Amsterdam: John Benjamins Publishing Company, 408-425.

- Balakaev, M.B., 1959. *Sovremennij Kazakhskij Jazik. Sintaksis*. [Modern Kazakh Language]. Alma-Ata: Izdatel'stvo Akademii.
- Balakaev, M.B., Baskakov, N.A., & Kenesbaev, S. K., 1962. *Sovremennij Kazakhskij Jazik* [Contemporary Kazakh Language]. Alma-Ata: Izdatel'stvo Akademii Nauk Kazakhskoi SSR. [The Kazakh SSR Academy of Science Publishing].
- Baskakov, N.A., 1966. *Sopostavitel'naja Grammatika Russkogo i Kazahskogo Jazykov. Morgologija*. [Comparative grammar of the Russian and Kazakh languages. Morphology.]. Alma-Ata: Izdatel'stvo Akademii Nauk Kazakhskoi SSR. [The Kazakh SSR Academy of Science Publishing].
- Bauer, B.L.M., 2009. Word Order. In: Baldi, P. & Cuzzolin, P. (eds.), *New Perspectives on Historical Latin Syntax. 1. Syntax of the Sentence*. Berlin: Mouton de Gruyter, 241-316.
- Bayer, J., & Struckmeier, V., 2017. *Discourse Particles: Formal approaches to their syntax and semantics*. Berlin; Boston: De Gruyter.
- Beckman, M. & Pierrehumbert, J., 1986. Intonational Structure in Japanese and English, *Phonology Yearbook III*, 15-70.
- Bisang, W., 2007. Categories that Make Finiteness: discreteness from a functional perspective and some of its repercussions. In: Nikolaeva I. (ed.), *Finiteness: Theoretical and empirical foundations*. Oxford ; New York: Oxford University Press, 115-137.
- Bizakov, S., 2014. *Kazakhskij Jazik: Prosto o Sloznom* [The Kazakh Language: Simply about the Difficult]. Almaty: Eltanim.
- Blackburn, P. & Meyer-Viol, W., 1994. Linguistics, Logic and Finite Trees. *Bulletin of Interest Group of Pure and Applied Logics* 2, 2-39.
- Blakemore, D., 2002. *Relevance and Linguistic Meaning: The semantics and pragmatics of discourse markers*. Cambridge, UK: Cambridge University Press.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott International*, 5:9/10, 341-345.
- Bolinger, D., 1961. Contrastive Accent and Contrastive Stress. *Language* 37, 83-96.

- Bol'shoj Kazahsko-Russkij Slovar'* [Big Kazak-Russian Dictionary], 1998. Almaty: Slovar'.
- Bossong, G., 1991. Differential object marking in Romance and beyond. In: Wanner, D., & Kibbee, D. (eds.), *New Analyses in Romance Linguistics, Selected Papers from the XVIII Linguistic Symposium on Romance Languages 1988*. Amsterdam: Benjamins, 143–170.
- Bouzouita, M., 2008. At the syntax-pragmatics interface. In Cooper, R. & Kempson, R. (eds.), *Language in Flux*. London: College Publications, 221-264.
- Büring, D., 1997. *The Meaning of Topic and Focus: The 59th Bridge Street Accent*. London: Routledge.
- Büring, D., 2006. Focus Projection and Default Prominence. In: Molnár, V. & Winkler, S. (eds.), *The Architecture of Focus* (Studies in Generative Grammar 82). Berlin/New York: Mouton De Gruyter, 321-346.
- Butt, M., 1993. Object Specificity and Agreement in Hindi/Urdu. *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, 1, 89-103.
- Butt, M. & Ahmed, T., 2010. The Redevelopment of Indo-Aryan Case Systems from a lexical semantic perspective. *Morphology*, 21(3/4), 545–572.
- Butt, M. & King, T.H., 1996. Structural Topic and Focus without Movement. In: Butt, M. & King T.H. (eds.), *On-line Proceedings of the LFG96 Conference*.
- Butt, M. & King, T. H., 2000. Null elements in discourse structure. In: Subbarao, K. V. (ed.), *Papers from the NULLS Seminar*. Delhi: Motilal Banarasidas.
- Cann, R., Kempson, R. & Marten, L., 2005. *The Dynamics of Language: an Introduction*. Amsterdam: Elsevier.
- Cann, R. & Miljan, M., 2012. Differential Case-marking: Syntactic descriptions and pragmatic explanations. *Linguistic Review*, 29(4), 585-605.
- Caron, B., 2006. Condition, Topic and Focus in African Languages: Why conditionals are not topics. *ZAS Papers in Linguistics* 46, 69 – 82.
- Carpenter, B., 1998. *Type-Logical Semantics*. Cambridge, MA: MIT Press.
- Chafe, W., 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In: Li & Thompson (eds.), 25-56.

- Chao, Y.R., 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chappell, H., 1996. Inalienability and the personal domain in Mandarin Chinese discourse. In: H. Chappell & W. McGregor (eds).1996. *The grammar of inalienability. A typological perspective on body part terms and the part-whole relation*. Berlin: Mouton de Gruyter, 465-527.
- Chatzikyriakidis, S., 2010. *Clitics in four dialects of Modern Greek: A dynamic Account*. Ph.D. King's College London.
- Chelliah, S.L. & De Reuse, W.J., 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.
- Chomsky, N., 1971. Deep structure, surface structure, and semantic interpretation. In: Steinberg, D. & Jakobovits, L. (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge, UK: Cambridge University Press, 183–216.
- Chomsky, N., 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N., 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Dalrymple, M. & Nikolaeva, I., 2011. *Objects and Information Structure*. Cambridge, UK: Cambridge University Press.
- Davidson, D., 1967. The logical form of action sentences. In: Rescher, N. (ed.), *The Logic of Decision and Action*. Pittsburgh : University of Pittsburgh, 81-95.
- Declerck, R., 1988. *Studies on Copular Sentences, Clefts, and Pseudo-Clefts*. Dordrecht: Foris.
- Degand, L., Cornillie, B., & Pietrandrea, P., 2013. Modal particles and discourse markers: Two sides of the same coin? In: Degand, L., Cornillie, B., & Pietrandrea, P. (eds.), *Discourse markers and modal particles: Categorization and description*. Amsterdam: John Benjamins Publishing Company, 1-18.
- De Hoop, H. & Malchukov, A., 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117, 1636-1656.
- De Hoop, H., 1992. *Case Configuration and Noun Phrase Interpretation*. Ph.D. University of Groningen.

- DeLancey, S., 1997. Mirativity: The Grammatical Marking of Unexpected Information. *Linguistic Typology* 1, 33-52.
- DeLancey, S., 2001. The Mirative and Evidentiality. *Journal of Pragmatics* 33, 369-382.
- Demirci, K., 2006. *Kazakh Verbal Structures and Descriptive Verbs*. Maryland: Dunwoody Press.
- Den Dikken, M. & O'Neill, T., 2017. Copular Constructions in Syntax. *Oxford Research Encyclopedia of Linguistics*, [online]. DOI: 10.1093/acrefore/9780199384655.013.137
- Diesing, M., 1992. Bare Plural Subjects and the Derivation of Logical Representations. *Linguistic Inquiry*, 23(3), 353-380.
- Diewald, G., 2013. "Same same but different" – Modal particles, discourse markers and the art (and purpose) of categorization. In: Degand, L., Cornillie, B., & Pietrandrea, P. (eds.), *Discourse markers and modal particles: Categorization and description*. Amsterdam: John Benjamins Publishing Company, 19-45.
- Dmitriev, N. K., 1962. Varvarizmy v bashkirskoj rechi [Barbarisms in Bashkir Speech]. In: Dmitriev, N. K. (ed.), *Stroj tjurkskih jazykov* [Structure of Turkic Languages]. Moscow: Izdatel'stvo Vostochnoj Literatury, 433-464.
- Downing, L.J., Mtenje, A. & Pompino-Marschall, B., 2004. Prosody and Information Structure in Chichewa. *ZAS Papers in Linguistics* 37, 167-186.
- Dryer, M. S., 1992. The Greenbergian Word Order Correlations. *Language* 68, 81-138.
- Dryer, M. S., 1997. On the Six-Way Word Order Typology. *Studies in Language* 21, 69-103.
- Dryer, M.S., 2007. Word Order. In: Shopen, T. (ed.), *Clause Structure, Language Typology and Syntactic Description*, vol. I. Cambridge, UK: Cambridge University Press, 61-131.
- Dryer, M. S., 2013. Determining Dominant Word Order. In: Dryer, M. S. & Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/s6>, Accessed on 2018-07-01.)

- Du Bois, J., 1985. Competing motivations. In: Haiman, J. (ed.), *Iconicity in Syntax*. Amsterdam: John Benjamins Publishing Company, 343-365.
- É. Kiss, K., 1995. *Discourse Configurational Languages*. Oxford: Oxford University Press.
- Enç, M., 1991. The semantics of specificity. *Linguistic Inquiry*, 22,1-25.
- Erdal, M., 2002. Relativisation in Krymchak. In: Nurettin, D. & Fikret, T. (eds.), *Scholarly Depth and Accuracy. A Festschrift to Lars Johanson*. Ankara: Grafiker.
- Erdocia, K., Laka, I. & Rodriguez-Fornells, A., 2012. Processing Verb Medial Word Orders in a Verb Final Language. In: Lamers, M. & de Swart, P. (eds.), *Case, Word Order and Prominence: interacting cues in language production and comprehension*. Dordrecht, Heidelberg, London and New York: Springer.
- Erguvanlı, E.E., 1984. *The function of word order in Turkish grammar*. Berkeley, Los Angeles, London: University of California Press.
- Erkü, F., 1983. *Discourse Pragmatics and word order in Modern Turkish*. Ph.D. University of Minnesota.
- Farkas, D., 1997. Evaluation indices and scope. In: Szabolcsi, A. (ed.), *Ways of Scope Taking*. Dordrecht, Boston, London: Kluwer, 183-211.
- Feldman, A., 2001. Discourse markers – Accessing ‘hearer-old’ information: The case of Russian *že*. *The LACUS forum* 27, 187-201.
- Féry, C. & Samek-Lodovici, V., 2006. Focus Projection and Prosodic Prominence in Nested Foci. *Language*, 82(1), pp.131-150.
- Féry, C. & Ishihara, S., 2016. Introduction. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 1-15.
- Filimonova, E., 2005. The noun phrase hierarchy and relational marking: problems and counterevidence. *Linguistic Typology* 9(1): 77–113.
- Fodor, J.A., 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Fodor, J.A., 2002. Psycholinguistics cannot escape prosody. *Proceedings of the SPEECH PROSODY 2002 Conference*, Aix-en-Provence, France.

- Forker, D., 2017. *What Relevance Theory Can Teach us about Modal Particles: A lesson from Sanzhi Dargwa*. (Submitted for publication in March 2017).
- Fortescue, M., 1996. Grammaticalized focus in Yukagir: is it really grammaticalized and is it really focus? In: Engberg-Pedersen, E., et al. (eds.), *Content, Expression and Structure: Studies in Danish Functional Grammar*. Amsterdam: Benjamins, 17-38.
- Friedman, V., 1978. On the Semantic and Morphological Influence of Turkish on Balkan Slavic. In: Farkas, D. et al. (eds.), *Papers from the Fourteenth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 108-118.
- Geist, L., 2007. Predication and equation in copular sentences. Russian vs. English. In: Comorovski, I. & von Heusinger, K. (eds.), *Existence. Syntax and Semantics*. Dordrecht: Springer, 79–105.
- Gerdts, D., 1998. Incorporation. In: Spencer, A. & Zwicky, A.M. (eds.), *The Handbook of Morphology*. Oxford: Blackwell, 84-100.
- Gibson, H.C., 2012. *Auxiliary Placement in Rangî: A Dynamic Syntax Perspective*. Ph.D. School of Oriental and African Studies, University of London.
- Godjevac 2000. *Intonation, Word Order, and Focus Projection in Serbo-Croatian*. Ph.D. Ohio State University.
- Göksel, A. & Özsoy, A. S., 2000. Is there a focus position in Turkish? In Göksel, A. & Kerslake, C. (eds.), *Studies on Turkish and Turkic Languages; Proceedings of the Ninth International Conference on Turkish Linguistics*. Wiesbaden: Harrassowitz, 219-228.
- Göksel, A. & Kerslake, C., 2005. *Turkish, A Comprehensive Grammar*. London: Routledge.
- Göksel, A., 1998. Linearity, Focus and the Postverbal Position in Turkish. In: Johanson, L., & Csató. É. Á. & Locke, V. & Menz, A. (eds.), *The Mainz Meeting*. Wiesbaden: Otto Harrassowitz Verlag.
- Gorelova, L., 2005. Typology of Information Structures in the Altaic Languages. In: Rybakov, R. B., & Boikova, E. V. (eds.), *Kinship in the Altaic world*:

- Proceedings of the 48th Permanent International Altaistic Conference, Moscow 10-15 July, 2005.* Wiesbaden: Harrassowitz, 149-172.
- Greenberg, J.H., 1966 [1963]. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In: Greenberg, J.H. (ed.), *Universals of Language*. Cambridge, MA: MIT Press, 73-113
- Gregoromichelaki, E., 2006, *Conditionals in Dynamic Syntax*. Ph.D. King's College London.
- Gregoromichelaki, E., 2010, Conditionals in Dynamic Syntax. In: Kempson, R., Gregoromichelaki, E. & Howes, C. (eds.), *The Dynamics of Lexical Interfaces*. Stanford: CSLI Publications, 231-271.
- Gundel, J., 1988, Universals of topic-comment structure. In: Hammond M., Moravczik, E. & Wirth, J. (eds.), *Studies in Syntactic Typology*, Amsterdam: John Benjamins Publishing Company, 209-239.
- Gussenhoven, C., 2007. Types of Focus in English. In: Lee, C., Gordon, M. & Büring, D. (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*. Heidelberg: Springer, 83-100.
- Gutzmann, D., 2012. Verum – Fokus – Verum-Fokus?. In: Blühdorn, H. & Lohstein, H. (eds.), *Wahrheit – Fokus – Negation*. Hamburg: Buske, 67–103.
- Gutzmann, D. & Castroviejo Miró, E., 2011. The Dimensions of Verum. *Empirical Issues in Formal Syntax and Semantics*, 8, 143-165.
- Haiman, J., 1978. Conditionals are Topics. *Language*, 54, 564-589.
- Hale, K., 1983. Warlpiri and the Grammar of Non-configurational Languages. *Natural Language & Linguistic Theory*, 1, 5–47.
- Halliday, M., 1967. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, 3, 199-244.
- Harrison, K.D., 2005. A Tuvan hero tale. *Journal of the American Oriental Society*, 125, 1–30.
- Hartmann, K. & Zimmermann, M., 2012. Focus Marking in Bura: Semantic Uniformity Matches Syntactic Heterogeneity. *Natural Language & Linguistic Theory*, 30(4), 1061-1108.

- Hartmann, K. & Veenstra, T., 2013. Introduction. In: Hartmann, K. & Veenstra, T. (eds.), *Cleft structures*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-32.
- Haspelmath, M., 1995. The converb as a cross-linguistically valid category. In: Haspelmath, M. & König, E. (eds.), *Converbs in cross-linguistic perspective*. Berlin: Mouton de Gruyter, 1-55.
- Haspelmath, M., 2010. Comparative Concepts and Descriptive Categories in Crosslinguistic Studies. *Language*, 86(3), 663-687.
- Hawkins, J.A., 1983. *Word Order Universals*. New York: Academic Press.
- Hawkins, J.A., 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hedberg, N., 2008. *Approaching Clefts: Syntax, Semantics, Pragmatics*. Paper presented at the Workshop on Clefts, Zentrum für Allgemeine Sprachwissenschaft (ZAS). Berlin, Germany, Nov. 28-29, 2008.
- Hedberg, N., 2013. Multiple Focus and Cleft Sentences. In: Hartmann, K. & Veenstra, T. (eds.), *Cleft structures*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 227-250.
- Hedberg, N. & Schneider-Zioga, P., 2015. Predication, specification and information structure in Kinande. Poster presented at *Generative Linguistics in the Old World (GLOW 38)*, April 15-18, 2015. Available at: http://www.sfu.ca/~hedberg/GLOWPoster2015_NH_PSZ_NH%20copy.pdf [Accessed 20 March 2018].
- Hedberg N., & Sosa J.M., 2008. The Prosody of Topic and Focus in Spontaneous English Dialogue. In: Lee C., Gordon M. & Büring D. (eds.), *Topic and Focus. Studies in Linguistics and Philosophy*, vol 82. Dordrecht: Springer, 101-120.
- Higgins, F., 1979. *The pseudo-cleft construction in English*. New York: Garland.
- Hill, N.W., 2012. "Mirativity" does not exist: *hdug* in "Lhasa" Tibetan and other suspects. *Linguistic Typology*, 16, 389-433.
- Höhle, T., 1988. Vorwort und Nachwort zu Verumfokus. *Sprache und Pragmatik* 5, 1-7.

- Höhle, T., 1992. Über Verum-Fokus im Deutschen. In: Jacobs, J. (ed.), *Informationsstruktur und Grammatik*. Opladen: Westdeutscher Verlag, 112-141.
- Hopper, P.J. & Traugott, E.C., 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Horn, L.R., 1997. Presupposition and implicature. In: Lappin, S. (ed.), *The handbook of contemporary semantic theory*. Oxford: Blackwell, 299-320.
- Huddleston, R.D. & Pullum, G.K., 2002. *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.
- Isaev, M.I., 1970. *O Sovetskom Opyte Razreshenija Jazykovoï Problemy v Mnogonacionalnom Gosudarstve* [About Soviet Experience of Solving the Language Issue in a Multilingual State]. In: Nikolskij, L. B. (ed.), *Problemy Izuchenija Jazykovoï Situaicii i Jazykovoï Vorpos v Stranah Azii i Severnoi Afriki* [Issues in Studying Language Situation and Question of Language in countries of Asia and North Africa]. Moscow: Nauka.
- Ippolito, M., 2013. *Subjunctive Conditionals: A Linguistic Analysis*. Cambridge, MA: MIT Press.
- Jackendoff, R., 1972. *Semantic Structures*. Cambridge, MA: MIT Press.
- Jayaseelan, K., 2008. Topic, focus and adverb positions in clause structure. *Nanzan Linguistics*, 4, 43-68.
- Jun, S., 2003. Prosodic Phrasing and Attachment Preferences, *Journal of Psycholinguistic Research*, 32(2), 219-249.
- Kabak, B. & Schiering, R., 2004. A corpus study on the distribution of Turkish *ise* and its clitic form. *Turkic Languages*, 8, 232-244.
- Kajdarov, A.V., 1997. Kazahskij Jazyk [The Kazakh Language]. In: Tenishev, E.R. (ed.), *Jazyki Mira. Tjurkskije Jazyki* [Languages of the World. Turkic Languages]. Moscow: Indrik, 242-253.
- Kang, S. & Speer, S. R., 2005. Effects of prosodic phrasing on syntactic disambiguation. *Studia Linguistica*, 59 (2-3), 244-258.
- Kaplan, R. & Zaenen, A., 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In: Baltin, M. & Kroch, A. (eds.), *Alternative*

- Conceptions of Phrase Structure*. Chicago: University of Chicago Press, 17-42.
- Karagjosova, E., 2001. Towards a comprehensive meaning of the German *doch*. In: Striegnitz, K. (ed.), *Proceedings of the Sixth ESSLLI Student Session*, August 2001. Helsinki, 131-141.
- Karagjosova, E., 2006. The German response particle *doch* as a case of contrastive focus. In Gyuris, B., Kálmán, L., Piñon, C. & Varasdi, K. (eds.). *Proceedings of the Ninth symposium on logic and language*. Besenyötelek, Hungary, 90-98.
- Kažbulatova, L.S., 2009. *Osnovy Grammatiki Kazahskogo Jazyka* [Foundations of Kazakh Grammar]. Almaty: LEM Publishing.
- Kempson, R., Cann, R., & Kiaer, J., 2006. Topic, Focus and Structural Dynamics of Language. In: Molnar, V. & Winkler, S. (eds.), *The Architecture of Focus*. Berlin; New York: Mouton de Gruyter, 59-82.
- Kempson, R. & Chatzikiyriakidis, S., 2009. The PCC as a Restriction on Tree-growth. *Formal vs. Processing Explanations of Syntactic Phenomena conference*, York.
- Kempson, R. & Kurosawa, A., 2009. At the syntax-pragmatics interface: Japanese relativeclause construal. In: Hoshi, H. (ed.), *Language, Mind and Brain: Perspectives from Linguistics and Cognitive Neuroscience*. Tokyo: Kuroshio, 47-84.
- Kempson, R. & Meyer-Viol, W. & Gabbay, D., 2001. *Dynamic Syntax: The Flow of Language Understanding*. Oxford: Blackwell Publishers.
- Kennelly, S.D., 1997. The presentational focus position of nonspecific objects in Turkish. In: İmer, K. & Uzun, N.E. (Eds.), *Proceedings of the VIIIth International Conference on Turkish Linguistics*. Ankara: Ankara University Press, 25-36.
- Kennelly, S.D, 1999. The Syntax of the P-focus position in Turkish. In: Rebuschi, G. & Tuller, L. (eds.), *Grammar of Focus*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 179-211.
- Kiaer, J., 2007. *Processing and Interfaces in Syntactic Theory: the case of Korean*. Ph.D. King's College London.
- Kiaer, J., 2014. *Pragmatic Syntax*. London: Bloomsbury Academic.

- Kim, H., 2004. *The Use of Multiple Sources of Information in Korean Sentence Processing*. Ph.D., Stanford.
- King, T.H., 1997. Focus Domains and Information Structure. In: Butt, M. & King, T.H. (eds.), *Online Proceedings of the LFG97 Conference*. Available at: <https://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG2-1997/lfg97king.pdf> [Accessed 10 January 2018].
- Kirchner, M., 1998. Kazakh and Karakalpak. In: Johanson, L. & Csató, É. (eds.), *The Turkic languages*. London and New York: Routledge, 318-332.
- Klein, W., 1994. *Time in Language*. London and New York: Routledge.
- Kohler, K.J., Peters, B. & Wesener, T., 2005. Phonetic Exponents of Disfluency in German Spontaneous Speech. In: Kohler, K.J., Kleber, F. & Peters, B. (eds.), *Prosodic Structures in German Spontaneous Speech*, 35(1). Kiel: IPDS, 185-201.
- Kornfilt, J., 1997. *Turkish*. London: Routledge.
- Kratzer, A., 1981. The Notional Category of Modality. In: Eikmeyer, H. & Rieser, H. (eds.), *Words, Worlds, and Context. New Approaches in Word Semantics*. Berlin: de Gruyter, 38-74.
- Krejnovič, E.A., 1958. *Jukagirskij Jazyk*. Moscow and Leningrad: Nauka.
- Krifka, M., 2007. Basic notions of information structure. In: Féry, C., Fanselow, G. & Krifka, M. (eds.), *Working Papers of the sfb632, Interdisciplinary Studies on Information Structure*, vol. 6. Universitätsverlag Potsdam, 13-36.
- Kučerová, I., 2007. *The Syntax of Givenness*. Ph.D. MIT.
- Kučerová, I., 2012. Grammatical marking of givenness. *Natural Language Semantics*, 20(1). 1-30.
- Kural, M., 1992. *Properties of scrambling in Turkish*. Ms., UCLA.
- Kuribayashi, Y., 1989. Accusative Marking and noun-verb constructions in Turkish. *Journal of the Linguistic Society of Japan*, 95, 94-119.
- Kurilov, G.N., 2006. *Sovremennyj Jukagirskij Jazyk*. Jakutsk: Ministerstvo obrazovanija RS(Ja)/IPMNS.

- Kuroda, S-Y., 1988. Whether We Agree or Not: A Comparative Syntax of English and Japanese. In: Poser, W. (ed.), *Papers from the Second International Workshop on Japanese Syntax*. Stanford, CA: CSLI. pp.103-143.
- Ladd, R., 1996. *Intonational Phonology*. Cambridge, UK: Cambridge University Press.
- Lambrecht, K., 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge, UK: Cambridge University Press.
- Lamers, M.J.A. & de Swart, P., 2012. The Interaction of Case, Word Order and Prominence: Language Production and Comprehension in a Cross-linguistic Perspective. In: Lamers, M. & de Swart, P. (eds.), *Case, Word Order and Prominence: interacting cues in language production and comprehension*. Dordrecht, Heidelberg, London and New York: Springer, 1-16.
- Lehmann, W.P., 1973. A Structural principle of language and its implications. *Language* 49, 42-66.
- Lewis, G., 1967. *Turkish Grammar*. Oxford: Clarendon.
- Li, C.N. & Thompson, S.A., 1976. Subject and Topic: A New Typology of Language. In: Li, C.N. (ed.), *Subject and Topic*. New York: Academic Press, 457-490.
- Lin, H., 1981. *Essentials of modern Chinese grammar*. Boston: Cheng and Tsui.
- Lohnstein, H., 2016. Verum Focus. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 290-313.
- Lucas, C., 2014. Indefinites and negative concord in Maltese: towards a dynamic account. In: Borg, A., Caruana, S. & Vella, A. (eds.), *Perspectives on Maltese Linguistics*. Berlin: Akademie Verlag, 225-248.
- Luraghi, S., 1995. The pragmatics of verb initial sentences in some ancient Indo-European languages. In: Downing, P.A. & Noonan, M. (eds.), *Word Order in Discourse*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 355-386.
- Lyons, C., 1999. *Definiteness*. Cambridge, UK: Cambridge University Press.
- Mahanjan, A., 1990. *The A/A-Bar Distinction and Movement Theory*. Ph.D. MIT.

- Malchukov, A.L., 2004. Towards a Semantic Typology of Adversative and Contrast Marking. *Journal of Semantics*, 21, 177-198.
- Marten, L., 2002. *At the Syntax-Pragmatics Interface: Verbal Underspecification and Concept Formation in Dynamic Syntax*. Oxford: Oxford University Press.
- Maslova, E., 2003. *Tundra Yukaghir*. München: LINCOM Europa.
- Mathew, R., 2013. Recursion of FocP in Malayalam. In: Hartmann, K. & Veenstra, T. (eds.), *Cleft Structures*. Amsterdam: John Benjamins Publishing Company, 251-268.
- Matić, D. & Nikolaeva, I., 2008. Predicate focus and the particle *mə(r)=* in Tundra Yukaghir. Paper Delivered at Predicate Focus Workshop. University of Potsdam.
- Matić, D. & Nikolaeva, I., 2014. Realis mood, focus, and existential closure in Tundra Yukaghir. *Lingua*, 150, 202-231. doi:10.1016/j.lingua.2014.07.016.
- Matić, D. & Nikolaeva, I., 2018. From polarity focus to salient polarity: From things to processes. In: *Polarity focus*. Amsterdam: John Benjamins. (Linguistics Today) (Forthcoming).
- Matić, D. & Pakendorf, B., 2013. Non-canonical SAY in Siberia: Areal and Genealogical Patterns. *Studies in Language*, 37(2), 356-412.
- Matić, D. & Wedgwood, D., 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(1), 127–163.
- McCoy, S., 2003. Connecting Information Structure and Discourse Structure through “Kontrast”: The Case of Colloquial Russian Particles *-TO*, *ŽE*, and *VED'*. *Journal of Logic, Language and Information*, 12, 319-335.
- Mikkelsen, L., 2005. *Copular Clauses: Specification, Predication and Equation*. Amsterdam: John Benjamins Publishing Company.
- Mikkelsen, L., 2011. Copular clauses. In: Maienborn, C., von Stechow, K., & Portner, P., (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2. Berlin: Mouton de Gruyter, 1805-1829.
- Michaelis, S. M., Maurer, P., Haspelmath, M., & Huber, M., 2013. *The survey of pidgin and creole languages*, vol. 1. Oxford: Oxford University Press.

- Mithun, M., 1992. Is Basic Word Order Universal? In: Payne, D.L. (ed.), *Pragmatics of Word Order Flexibility*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 15-61.
- Mohanan, T., 1994. *Argument structure in Hindi*. Stanford: CSLI Publications.
- Molnár, V., 2006. On different kinds of contrast. In: Molnár, V. & Winkler, S. (eds.), *Architecture of Focus*. München: Walter de Gruyter, 197-233.
- Morrill, G., 1994. *Type-Logical Grammar*. Dordrecht: Kluwer.
- Muhamedowa, R., 2011. From Participle to Topic Particle: Kazakh *degen* in Comparison with its Functional Equivalents in Turkish. *Turkic languages*, 15, 259-274.
- Muhamedowa, R., 2016. *Kazakh: A Comprehensive Grammar*. London and New York: Routledge.
- Mycock, L., 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of 'Wh'-Questions*. Ph.D. University of Manchester.
- Neeleman, A. & van de Koot, H., 2008. Dutch Scrambling and the Nature of Discourse Templates. *Journal of Comparative Germanic Linguistics*, 11, 137-189.
- Neeleman, A. & van de Koot, H., 2016. Word Order and Information Structure. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 383-401.
- Newmeyer, F., 2003. "Basic Word Order" in Formal and Functional Linguistics and the Typological Status of "Canonical" Sentence Types. In: Willems, D. & Defrancq, B. & Coleman, T. & Noël, D. (eds.), *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison*. Basingstoke: Palgrave, 69-88.
- Newmeyer, F. 2007. Linguistic Typology Requires Crosslinguistic Formal Categories. *Linguistic Typology*, 11(1), 133-157.
- Nikolaeva, I., 2001. Secondary topic as a relation in information structure. *Linguistics*, 39 (1), 1-49.
- Nikolaeva, I., 2005. Review of Elena Maslova "A grammar of Kolyma Yukaghir". *Linguistic Typology*, (9) 2, 299 -325.

- Nikolaeva, I., 2007. Introduction. In: Nikolaeva I. (ed.), *Finiteness: Theoretical and empirical foundations*. Oxford and New York: Oxford University Press, 1-22.
- Nikolaeva, I., 2012. Unpacking finiteness. In: Corbett, G., Brown, D. & Chumakina, M., (eds.), *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Orgun, C.O. & Inkelas, S., 2004. Turkish direct objects: incorporation or word order. In: *CLS 40-1: The Main Session Proceedings*. Chicago: The Chicago Linguistic Society, 265-278.
- Ótrott-Kovács, E., 2015. *The Syntax of Non-finite Clauses in Kazakh*. Ph.D. University of Szeged.
- Öztürk, B., 2005. *Case, Referentiality and Phrase Structure*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Payne, D.L., 1992. Introduction. In: Payne, D.L. (ed.), *Pragmatics of Word Order Flexibility*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 1-14.
- Pierrehumbert, J., 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. MIT.
- Pinto, M., 1997. Definite Subjects and Inversion. In: Dekker, P., van der Does, J. & de Hoop, H. (eds.), *Proceedings of the Workshop on Definites*, 113-124.
- Predolac, E., 2017. The Subjunctive and Indicative Moods in Turkish. *IULC Working Papers: Papers in Turkish and Turkic Linguistics*, 18(2) [online]. Available at: <https://www.indiana.edu/~iulcwp/wp/issue/view/Turkish%20and%20Turki%20Linguistics> [Accessed 13 June 2018].
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C., 1991. The use of prosody in syntactic disambiguation, *Journal of the Acoustical Society of America*, 9, 2956-2970.
- Primus, B., 2012. Animacy, generalized semantic roles, and differential object marking. In: Lamers, M. & de Swart, P. (eds.), *Case, word order, and prominence. Interacting Cues in Language Production and Comprehension*. Dordrecht: Springer, 65-90.

- Prince, E. F., 1981. Toward a taxonomy of given-new information. In: Cole, P. (ed.), *Radical Pragmatics*. New York and London: Academic Press.
- Pustet, R., 2003. *Copulas*. Oxford: Oxford University Press.
- Radetzky, P., 2002. *The Functions and Evolution of Topic and Focus Markers*. Ph.D. University of California, Berkeley.
- Ramchand, G., 1993. *Aspect and argument structure in modern Scottish Gaelic*. Ph.D. Stanford University, Stanford, CA.
- Rando, E., 1980. Intonation in Discourse. In: Waugh, L.R., van Schooneveld, C.H. & Bolinger, D. (eds.), *The Melody of Language*. Baltimore: University Park Press, 243-278.
- Reese, B. & Asher, N., 2006. Prosody and the Interpretation of Tag Questions. In: Puig Waldmüller, E. (ed.), *Proceedings of Sinn und Bedeutung 11*. Barcelona: Universitat Pompeu Fabra, 448-462.
- Reinhart, T., 1982. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27, 53-94.
- Repp, S., 2010. Defining ‘contrast’ as an information-structural notion in grammar. *Lingua*, 120(6), 1333-1345.
- Rijkhoff, J., 2009. On the (Un)suitability of Semantic Categories. *Linguistic Typology*, 13(1), 95-104.
- Rivarlor, A. de, [1784] 1929. *Discours de l’universalité de la langue française*. Paris: Delagrave.
- Rizzi, L., 1997. The fine structure of the left periphery. In: Haegeman, L. (ed.), *Elements of grammar: A handbook of generative syntax*. Dordrecht: Kluwer, 281-337.
- Rochemont, M., 1998. Phonological focus and structural focus. In: Culicover, P. & McNally, L. (eds.), *The Limits of Syntax. Syntax and Semantics*, 29, 337-363.
- Rochemont, M., 2016. Givenness. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 41-63.
- Romanenko E.I., 2011. *Kazakhskaja Grammaticka dlja Russkojazychnih* [Kazakh Grammar for Russian Speakers]. Almaty: A-level.

- Romanenko E.I., 2015. *Sovremennij Kazakhskij* [Modern Kazakh]. Almaty: Kazbooka.
- Romanova, A. V. & Myreeva A. N., 1964. *Dialektologičeskie materialy po govoram evenkov Jakutskoj ASSR*, 1–3. Moscow/Leningrad: AN SSSR.
- Romero, M. & Han, C., 2004. On negative y/n-questions. *Linguistics and Philosophy*, 27, 609-658
- Ross, J.R., 1967. *Constraints of Variables in Syntax*. Ph.D. MIT.
- Rozenal, D.A. & Telenkova, M.A., 1985. *Slovar' trudnostei russkogo jazyka*. Moskow: Russkii Jazik.
- Saint-Dizier, P., 2006. Introduction to the Syntax and Semantics of Prepositions. In: Saint-Dizier, P. (ed.), *Syntax and Semantics of Prepositions*. Dordrecht: Springer, 1-23.
- Saito, M., 1985. *Some Asymmetries in Japanese and Their Theoretical Implications*. Ph.D. MIT.
- Schafer, A.J., 1997. *Prosodic Parsing: The Role of Prosody in Sentence Comprehension*. Ph.D. University of Massachusetts, Amherst.
- Schafer, A.J. & Jun, S., 2002. Effects of Accentual Phrasing on Adjective Interpretation in Koran. In: Nakayama, M. (ed.), *East Asian Language Processing*. Stanford: CSLI, 223-255.
- Schauber, E., 1978. *The syntax and semantics of questions in Navajo*. New York: Garland Publishing.
- Schneider-Zioga, P., 2015. The Linker in Kinande Re-examined. In: Kramer, R., Zsiga, E.C. & Tlale Boyer, O. (eds.), *Selected Proceedings of the 44th Annual Conference on African Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 264-276.
- Schwabe, K. & Winkler, S., 2007. *On Information Structure, Meaning and Form: Generalizations across languages*. Amsterdam: John Benjamins Publishing Company.
- Seiler, W., 1985. *Imonda, a Papuan Language*. Canberra: Australian National University Printing Service.

- Selkirk, E., 1995. Sentence Prosody: Intonation, Stress, and Phrasing. In: Goldsmith, J. A. (ed.), *The Handbook of Phonological Theory*. Cambridge, UK: Blackwell, 550–569.
- Senft, G., 2009. Elicitation. In: Senft, G., Östman, J.-O. & Verschueren, J. (eds.), *Culture and language use*. Amsterdam: John Benjamins Publishing Company, 210-225.
- Seraku, T., 2013. *Clefts, Relatives, and Language Dynamics: The Case of Japanese*. Ph.D. St. Catherine's College, University of Oxford.
- Shamina, L.A., 2001. *Polipredikativnye Sintetičeskie Predloženiya v Tuvinskom Jazyke* [Polypredicative Synthetic Sentences in the Tuvinian Language]. Novosibirsk: Sibirskij Hronograf.
- Siewerska, A., 1988. *Word Order Rules*. London: Routledge.
- Simons, M., 2006. Foundational Issues in Presupposition. *Philosophy Compass*, 1(4), 357-372.
- Simons, G. F. & Fennig, C. D., (eds.), 2018. *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas: SIL International. [online] Available at: <http://www.ethnologue.com>
- Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C. & Krifka, M., 2006. *Questionnaire on information structure (ISIS Vol. 4)*. Potsdam: Universitätsverlag. [online] Available at: <https://www.sfb632.uni-potsdam.de/images/isis/isis04.pdf>
- Smailov, A.A., 2011. *Results of the 2009 National population census of the Republic of Kazakhstan* [analytical report]. Astana: Agency of Statistics of the Republic of Kazakhstan.
- Soames, S., 1989. Presupposition. In: Gabbay, D. & Guenther, F. (eds.), *Handbook of Philosophical Logic*. Dordrecht: Springer, 553-616.
- Song, J.J., 2001. *Linguistic Typology: Morphology and Syntax*. Harlow: Longman.
- Song, J.J., 2009. Word order patterns and principles: an overview. *Language and Linguistics Compass*, 3(5), 1328-1341.
- Speas, M., 1990. *Phrase Structure in Natural Language*. Dordrecht: Kluwer Academic Publishers.

- Speer, S.R., Kjelgaard, M.M. & Dobroth, K.M., 1996. The influence of Prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25 (2), 249-271.
- Sperber, D. & Wilson, D., 1986/1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Stalnaker, R., 1974. Pragmatic Presuppositions. In: Munitz, M.K. & Unger, P. (eds.), *Semantics and Philosophy*. New York: New York University Press.
- Stalnaker, R., 1978. Assertion. In: Cole, P. (ed.), *Pragmatics*. New York: New York Academic Press, 315-332.
- Stalnaker, R., 2002. Common Ground. *Linguistics and Philosophy*, 25, 701-721.
- Steedman, M., 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.
- Stommel, H., 2011. *Verum-Fokus im Deutschen*. Marburg: Tectum.
- Straughn, C., 2011. *Evidentiality in Uzbek and Kazakh*. Ph.D. The University of Chicago.
- Sudhoff, S., 2010. *Focus Particles in German: Syntax, prosody, and information structure*. Amsterdam: John Benjamins Publishing Company.
- Surányi, B., 2016. Discourse-configurationality. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 422-440.
- Takano, Y., 1998. Object Shift and Scrambling. *Natural Language and Linguistic Theory*, 16, 817-889.
- Titov, E., 2012. *Information Structure of Argument Order Alternations*. Ph.D. UCL.
- Torrence, H., 2013. The Morphosyntax of Wolof Clefts. In: Hartmann, K. & Veenstra, T. (eds.), *Cleft structures*. Amsterdam: John Benjamins Publishing Company, 187-223.
- Truckenbrodt, H., 2015. Intonation phrases and speech acts. In: Kluck, M., Ott, D. & de Vrieds, M. (eds.), *Parenthesis and Ellipsis. Cross-linguistic and Theoretical Perspectives*. Berlin: Walter de Gruyter, 301-349.
- Tuite, K., Agha, A. & Graczyk, R., 1985. Agentivity, transitivity, and the question of active typology. *Chicago Linguistics Society*, 21. pp.252–270.

- Valjajeva, T., 2018. *Kazakhskij Jazik* [Kazakh Language]. [online] Available at: <http://kaz-tili.kz>
- Vallduví, E., 1992. *The Informational Component*. New York: Garland.
- Vallduví, E. & Engdahl, E., 1996. The Linguistic Realization of Information Packaging. *Linguistics*, 34(3), 459-520.
- Vallduví, E. & Vilkkuna, M., 1998. On Rheme and Kontrast. *Syntax and Semantics 29* (The Limits of Syntax), 79-108.
- van Geenhoven, V., 1998. *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic*. Stanford, CA: CSLI Publications.
- Vasilyeva, A. N., 1972. *Particles in Colloquial Russian: Manual for English-Speaking Students of Russian*. Moscow: Progress Publishers.
- Vilkkuna, M., 1989. *Free Word Order in Finnish*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- von Stechow, K., 2002. Specificity and Definiteness in Sentence and Discourse Structure. *Journal of Semantics*, 19, pp.245-274.
- Watters, J.R., 1979. Focus in Aghem: A Study of its Formal Correlates and Typology. In: Hyman, L. (ed.), *Aghem Grammatical Structure*. Los Angeles: University of Southern California, 137-197.
- Wedgwood, D., 2003. *Predication and Information Structure: A Dynamic Account of Hungarian Preverbal Syntax*. Ph.D. University of Edinburgh.
- Weil, H., [1848] 1978. *The Order of Words in the Ancient Languages Compared with That of the Modern Languages*. Revised edition of the translation by Charles Super (1887). Amsterdam: Benjamins.
- Werner, A., 1991. Discourse Particles in German: How does their illocutive force come about? In: Werner, A. (ed.) *Discourse Particles in German*. Amsterdam: John Benjamins Publishing Company.
- Wilson, D., 2011. The Conceptual-procedural Distinction: Past, present and future. In: Escandell-Vidal, V., Leonetti, M. & Ahern, A. (eds.), *Procedural meaning: Problems and Perspectives*. Bingley: Emerald, 3-11.

- Witzlack-Makarevich, A. & Seržant, I.A., 2018. Differential Argument Marking: Patterns of Variation. In: Seržant, I.A. & Witzlack-Makarevich, A. (eds.), *Diachrony of Differential Argument Marking* (Studies in Diversity Linguistics 19). Berlin: Language Science Press, 1-40.
- Woolford, E., 2000. Object Agreement in Palauan: Specificity, Humanness, Economy and Optimality. In: Ileana, P., Phillips, V. & Travis, L. (eds.), *Formal Issues in Austronesian Linguistics*. Dordrecht: Springer, 215-245.
- Yu Cho, Y-M., 1990. Syntax and Phrasing in Korean. In: Inkelas, S. & Zec, D. (eds.), *The Phonology-Syntax Connection*. Chicago and London: The University of Chicago Press, 47-62.
- Zimmerman, M., 2016. Predicate Focus. In: Féry, C. & Ishihara, S. (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 314-335.
- Zimmerman, M. & Féry, C. (eds.), 2009. *Information Structure: Theoretical, Typological, and Experimental Perspectives*. Oxford: Oxford University Press.
- Zimmerman, M. & Hole, D., 2008. Predicate focus, verum focus, verb focus. Similarities and difference. Talk at the *Postdam-London IS Meeting*. 12.12.2008.
- Zubizarreta, M.L., 1998. *Prosody, Focus and Word Order*. Cambridge, MA: MIT Press.