Pronouns in Kumyk Discourse: A Cognitive Perspective

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Linda Anne Humnick

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jeanette Gundel, Adviser

March 2009

Acknowledgements

Just as this research began with my initial field work among the Kumyk people of Dagestan, Russia, I would like to begin by expressing my gratitude for the generous hospitality and linguistic contributions of my Kumyk friends and colleagues. First, I would like to thank several Kumyk scholars who contributed to my research. The late Zhangishi Khangishiev was my first patient tutor in the intricacies of Kumyk grammar and morphology. The late Farida Muratchaeva inspired me with her scholarship on the Kumyk language and encouraged me in my own research. Ruhaniyat Musaeva, a dear friend, began as my tutor in spoken Kumyk, assisted in the collection of data, and generously gave me many opportunities to share the beauty of her family and culture. Finally, Agaragim Sultanmuradov has been invaluable as a colleague in scholarship, language consultant, contributor to the text database, and as the administrator of the questionnaire on site in Dagestan. I value his friendship, the patience and insight with which he has answered all my questions over the past 5 years, and his model of scholarship in Kumyk language research.

I am grateful to the members of my examining committee: Nancy Stenson, with whom I share a passion for field research, Hooi Ling Soh, who sets a wonderful example in striving for excellence in research, and Michael Kac, whose creative and analytical thinking have been an inspiration. I also appreciate the contribution of Bruce Downing, who served on my committee at an earlier stage and shares my enthusiasm for the Turkic world. I am deeply in debt to my adviser, Jeanette Gundel, who not only has shared her insight and expertise in research and taught me much about the process of writing, but has also opened her home and her friendship to me over these years, mentoring me in a number of areas of life and scholarship.

I would not have been able to complete a work of this scope without the encouragement of friends. Though I cannot begin to name all the friends who have encouraged me with their prayers and acts of kindness in these busy years, I would at least like to mention two fellow graduate students who have supported me along the way.

In particular, I thank Michiko Buchanan, whose encouragement and friendship reminded me that the bulbs lying dormant in the winter will eventually sprout and bloom. A special privilege during my graduate study has been having my friend Amel Khalfaoui as a fellow researcher, and a secondary coder on some of my texts. More importantly, we have shared many areas of friendship beyond the Givenness Hierarchy—including dreams for our families and careers, and even an appreciation of good food.

Finally, I acknowledge that this work has only been possible through the support and sacrifices of my family. I owe much appreciation to my loving grandparents, Iva Dean Winkler and the late Gordon Winkler for teaching me the value of education and for making provisions for my college and post-graduate education even before I was old enough to go to college. I especially want to thank my grandmother, for being one of my biggest fans while I was in graduate school. I want to acknowledge my deep appreciation for my parents, Bob and Barbara Dunnigan, for their love, prayers, words of encouragement, and unconditional support. They have always encouraged me to do my best, but have also reminded me to "take time to smell the roses." I deeply appreciate the love and support of my husband, Paul, and my children, Abraham and Katy, and I thank them for their patience with me as I invested time and energy in this research. On the practical side, I benefitted greatly from Paul's technical computer support and willingness to help with many aspects of our family life, and I am also grateful to have Paul as a colleague in field research. His travels during the last five years were an invaluable assistance in the process of gathering data, and his mutual interest and willingness to listen to my ideas have been a great encouragement to me. With Katy and Abraham, many hugs and moments of fun have kept me going during the long hours of research and writing. For the three of you who have walked most closely with me through this process, your love has made all the difference. Finally, I give my thanks and praise to my Heavenly Father, who created me with this interest in language, who provides for all my needs, and who has guided my foot along this path.

Dedication

To the glory of the Creator of language and the honor of the Kumyk people

Abstract

This dissertation investigates pronominal forms of referring expressions in Kumyk, a Turkic language spoken primarily in the Dagestan region of Russia. The Kumyk language has six third person pronominals, including null arguments, demonstratives, and reflexives. Morphologically, each of these forms is unmarked for gender or animacy. This work provides an explanatory account of the distribution and interpretation of different pronominal forms in Kumyk primarily in terms of what these forms communicate about the status of their referents in the minds of the speech participants, specifically claiming that different pronominal forms signal differences in the cognitive status of their referents, following the Givenness Hierarchy model of Gundel, Hedberg, and Zacharski (1993). The analysis is based primarily on data from a corpus of oral and written Kumyk texts with supporting evidence from grammaticality judgments of constructed examples in questionnaires. According to the analysis, null arguments and reflexives signal the status, 'in focus', while demonstratives signal the status, 'activated'. Particular attention is given to the role of scalar implicatures which arise from the unidirectional entailment of statuses on the Givenness Hierarchy and the fact that the demonstrative *sho*, which signals activation, has a particular association with this implicature. A unique contribution of the analysis is the evidence for the fact that *sho* not only gives rise to a scalar implicature in contexts where two referents have different maximal cognitive status(e.g. one in focus versus one at most activated), but also in contexts where two referents have the same maximal cognitive status, a fact which leads to the conclusion that this form specializes in indicating the less salient of two or more entities. The study also provides evidence that the demonstrative *bu* specializes in indicating the more prominent of two or more entities that are at least activated. Finally, in addition to the role of pronominals in signaling cognitive status and communicating the relative prominence of multiple referents, the study explores contextual effects such as imposed salience, point of view, empathy, or contrastive focus that are associated with particular forms.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The goal of this dissertation is to provide an explanatory account of the distribution and interpretation of different pronominal forms in Kumyk discourse primarily in terms of what these forms communicate about the status of their referents in the minds of the speech participants. The basis for this analysis is evidence from a text corpus, with supporting data from grammaticality judgments and constructed examples in questionnaires.[1]

This study follows the Givenness Hierarchy framework of Gundel, Hedberg, and Zacharski 1993, who claim that pronouns and other types of referring forms signal the cognitive status of their referents in the mind of the addressee as part of their lexical meaning. Cross-linguistic studies within this model show that pronominals refer to entities that the speaker expects to be represented in the working memory of the addressee (activated) and a subset of pronominals are restricted to reference to entities that are at the center of the addressee's attention (in focus) (Gundel et al. 1993, Gundel, Bassene, Gordon, Humnick and Khalfaoui 2007). This study likewise finds that the use of Kumyk pronominals correlates with activated referents and that two types of forms, nulls and the pronominal reflexive *o'ziu* are restricted to referents in focus. My research also provides evidence to support the Givenness Hierarchy model's claim that the

[1] A significant part of this research is carried out under NSF grant #0519890 to Jeanette Gundel.

distinction between the set of forms that signal that a referent is activated in the mind of the addressee and the set that signals that a referent is in focus is the basis for scalar implicatures that play a role in referent disambiguation through the indication of relative salience.

In addition to communicating information about the cognitive status of a referent, the choice of pronominal may communicate other information about a referent, such as the speaker's empathy towards the entity or the fact that one entity is being contrasted with another. In Kumyk, differences in these types of features occur among forms that signal the same cognitive status. A significant aspect of this study, moreover, is the evidence that, among the forms that signal 'activated', two forms appear to have a specialized function of indicating how the speaker wishes the hearer to interpret the relative salience of the entity in comparison to other entities in the context—with one often indicating the most prominent of a set of entities, while the other indicates the least prominent.

While the primary significance of this research is to support the cross-linguistic validity of theories which correlate the use of restricted forms with attention state, the dissertation will also be a significant contribution to the general knowledge of Kumyk, a language about which relatively little has been written in general and even less in English. The dissertation provides an overview of pronominal forms in the language, including aspects of morphology, and a characterization of the use of these forms in discourse. In particular, the discussion of null arguments and pronominal reflexives provides information of wider interest to syntactic theory and linguistic typology.

1.0. Kumyk language facts

The Kumyk language is a member of the West Kipchak group of the Kipchak (or Western) branch of Turkic languages. The majority of the approximately 450,000 speakers of this language live in the Republic of Dagestan in the Russian Federation.[2] Kumyk is one of approximately 40 different languages spoken in this region, which lies north of Azerbaijan between the Caucasus Mountains and the Caspian Sea. Smaller groups of Kumyk speakers are also reported in other parts of the Caucasus region, Turkey, and Kazakhstan (Gordon 2005).

The majority of Kumyk speakers are bilingual in Russian and Kumyk. While Russian is both the national language and the primary language of education for the region, Kumyk is a vital and well-developed language within Dagestan. Kumyk has been a literary language for more than a century. The first orthography, based on Arabic script, was used until 1928, and then replaced by a Latin-based script. Since 1938, Kumyk has used a Cyrillic orthography.

Kumyk is currently a language of education, but the degree of use in education varies from community to community according to the composition of the local population. In predominately Kumyk communities, Kumyk is a language of instruction at the primary level, while in secondary schools and in higher education, the language is taught only as a subject. In multi-ethnic communities, Russian is the language of

---

[2] The population estimate reflects information from the 2002 Russian census. Previous information reports approximately 280,000 speakers (cf. Gordon 2005). A map showing the geographical distribution of Kumyk speakers in relation to the distribution of speakers of other languages in Dagestan appears as Appendix 1.

instruction, and Kumyk literature is offered as a subject only in areas with a significant percentage of Kumyk speakers. For this reason, while Kumyk speakers who live in predominantly Kumyk communities tend to be literate in Kumyk as well as Russian, the majority of Kumyk speakers in multi-lingual communities such as Makhachkala, the capital of Dagestan, use Kumyk only as an oral language.

Within the Dagestan region, Kumyk is used in both the media and the arts, as is evident from the existence of Kumyk theatre, journals, newspaper, radio programs and TV broadcasts. Many local scholars affiliated with institutions of higher education and research institutions such as the Dagestan branch of the Russian Academy of Sciences are active in language-related research.

Very little documentation of the Kumyk language exists in English, but there are a significant number of descriptive works in Russian and one notable comprehensive work in Kumyk. Dimitriev 1940 provides the earliest comprehensive grammar of the Kumyk language in Russian. Khangishiev 1995 provides a more recent overview of the morphology and grammar written in the Kumyk language.[3] In addition to these comprehensive works, a number of articles and theses have been written in Russian on various aspects of the language. Among these are descriptions of the verbal system (Djanmavov 1967, Hangishiev 1985, Gadzhiakhmedov 1987, Sultanmuradov 1997 inter alia), morphophonology (Ol'mesov 1994), historical and derivational morphology (Gadzhiakhmedov 1998, Kadyradzhiev 1999, Abdurakmanova 2005), and complex

[3] An unpublished grammar overview in English based on these works was compiled by Clinton 1997.

clause constructions (Abdulaeva 1973).[4] Of particular relevance to this work is a brief description of demonstratives by Muratchaeva (2001).

To this point, little attention has been given to the comprehensive description of the use of referring expressions in Kumyk discourse, though a brief overview is offered in Humnick 2002 and smaller-scale analyses of certain aspects of the use of pronouns and null arguments are offered in Humnick 2005, 2007 and 2008.

1.1. Kumyk pronouns

The Kumyk language has a variety of third person pronominal forms. These include null arguments, the overt pro-forms, *bu*, *o*, *sho*, and *shu*, and the inflected 3rd person reflexive, *o'ziu*. Morphologically, each of these forms is unmarked for gender or animacy. In grammatical descriptions, *o* is described as a personal pronoun, with *o'ziu* as its reflexive counterpart, while *bu*, *o*, *sho*, and *shu* are described as demonstratives (Dimitriev 1940, Hangishiev 1995). In spite of the fact that these proforms are associated with different labels such as 'reflexive', 'demonstrative', or 'personal pronoun', to a large extent they share the same syntactic distribution, as illustrated in examples (1) - (4). Examples (1) - (3) exemplify the alternation between a null argument, the overt pronoun *o,* and the inflected reflexive *o'ziu* in the second sentence of each example. Example (4) illustrates an alternation between *o* and *bu* in the two parallel sentences of the example.[5]

---

[4] For a more complete description of linguistic works on the Kumyk language, see Gjulmagomedov, A. G., I. X. Abdullaev, M.Sh. Khalilov, and F. A. Muratchaeva, eds. 1998. Bibliographiia po dagestanskomu iazykaznaniiu. Makhachkala: Izdatelsvo Instituta Iazyka, Literatury I iskusstva D.N.Ts. R.A.N.

[5] The vernacular text has been transliterated according to the ALA-LC 1997 standards. The following abbreviations are used in glossing: ABL=Ablative, ACC=Accusative, COND=Conditional, CON=Conjunction,

(1) *Salimat erine ash ete. Song, Ø savut-saba zhuĭe.*
Salimat husband.3POSS.DAT food make.PR later dishes wash.PR
'Salimat fixes her husband a meal. Later (s)he washes the dishes.'

(2) *Salimat erine ash ete. Song, o savut-saba zhuĭe.*
Salimat husband.3POSS.DAT food make.PR later 3.2 dishes wash.PR
'Salimat fixes her husband a meal. Later (s)he washes the dishes.'

(3) *Salimat k"yzyny yashy iuretdi.*
Salimat daughter.3POSS.ACC well teach.PST

Xali o'ziu savut-saba zhumaĭ; k'zyny da zhuĭe.
Now self.3POSS dishes wash.NEG.PR daughter.3POSS EMPH wash.PR
'Salimat taught her daughter well. Now (s)he herself doesn't wash the dishes; her daughter does.'

(4) *O savut-saba zhuĭmady. Bu zhuĭdy.*
3.2 dishes wash.NEG.PST 3.1 wash.PST
'That one didn't wash the dishes. This one did.'

The existence of such a diversity of proforms sharing the same syntactic distribution but lacking grammatical distinctions of gender or animacy strongly suggests that different forms are used to signal other characteristics of the referent.

COP=Copula, DAT=Dative, EMPH=Emphatic, FUT=Future, GEN=Genitive, GER=Gerund/Converb, INF=Infinitive/Verbal Noun, LOC=Locative, MOD=Modifier, NOM=Nominalizer, PASS=Passive, PL=Plural, POSS=Possessive, PR=Present, PRT=Participal, PST=Past, RFLX=Reflexive (verbal), S=Singular, SPEC=Speculative. 3.1, 3.2, 3.3 and 3.4 indicate the four distinct third person pronouns, *bu*, *o*, *sho*, and *shu*, respectively.

A number of linguists suggest that the explanation of the use of different pronominal forms lies primarily within the realm of pragmatics and information structure. For example, the use of null arguments in contrast to overt pronouns is often associated with the topicality or salience of the referent (Givón 1983, Ariel 1988, Gundel, Hedberg, and Zacharski 1993, inter alia). In the literature on reflexives, which predominantly focuses on the syntactic properties of coreference relations, a number of studies associate certain uses of such forms with non-syntactic properties of the referents such as discourse prominence, contrast, or empathy (Kuno 1987, Baker 1995, Zribi-Hertz 1995). For demonstratives, differences in distribution are associated with differences in the referent's location in relation to the speaker/hearer (Diessel 1999, Dixon 2003, inter alia) or different communication functions related to establishing joint attention (Özyürek and Kita 2001 (cited in Enfield 2003), Enfield 2003, Diessel 2006). Finally, some approaches correlate the use of all three types of proforms, in addition to other forms of referring expressions, with differences in the attention state of the addressee in relation to the referent (Chafe 1976, Ariel 1988, Gundel et al. 1993, inter alia). These approaches, which are not necessarily in contradiction to the other types of approaches mentioned here, differ in the sense that they address the distribution of proforms as part of a unified explanation of the distribution of referring forms in language and the cognitive processes which determine referential choice.

Reflecting such unified approaches, this study proposes that the characteristics of referents which are signaled by different pro-forms in Kumyk relate to how the referent is represented in the mind of the hearer. More specifically, I claim that different

pronominal forms signal differences in the cognitive status of their referents, following the Givenness Hierarchy model of Gundel et al. (1993). Proposing six cognitive statuses which represent different levels in the hearer's memory and attention state, Gundel et al. claim that different forms of referring expressions can be used to signal different cognitive statuses. In other words the choice between forms in a given language, such as the choice between a null form versus the pronoun *o* in Kumyk, represents a difference in how the speaker assumes the referent of the form is represented in the hearer's mind. The signals that the hearer receives via the form of referring expression, in turn, guide the hearer in the process of reference resolution by constraining reference to the set of possible referents with a particular status. Particularly in the case of reduced forms with minimal descriptive or conceptual content, such constraints significantly reduce the processing effort required for the hearer to make the association between a form and its referent.

1.2. Organization of the study

In Chapter 2 of this dissertation, I provide a comprehensive discussion of theories related to the distribution of referring forms in discourse and the particular features of the cognitive model of Gundel et al. 1993. The second part of the chapter describes the methodology used to analyze referring forms within this particular framework, in particular, the coding protocol used to determine the cognitive status of the referent of a referring expression. Chapter 3 defines the scope of the term 'pronominal' for the Kumyk language and provides general morphological data for the six distinct forms considered in this research. Chapters 4 presents an analysis of pronoun tokens and their

referents within a corpus study, in particular, the coding of the cognitive status of the referents and the hypothesized correlations between a particular form and a particular cognitive status within the Givenness Hierarchy. This chapter also describes distinct featues of each form, including specific contextual effects associated with their use. Chapter 5 further explores the hypothesized form-status correlations on the basis of grammaticality judgments and other types of questionnaire data. In this chapter I also explore the way in which different forms are used to indicate the relative prominence of entities, both through scalar implicatures and other ranking effects. Finally, Chapter 6 summarizes the conclusions and theoretical implications of the integrated analysis.

Chapter 2: Theoretical Framework

2.0. What is a cognitive model?

Associating the form of a referring expression with characteristics of a referent is basic to the act of reference. While content nouns such as *tree* may be said to be associated with referents which have tree-like qualities, the difference between the expression *a tree* and the expression *the tree*—and, even more so, the difference between expressions such as *it* or *that*—are not related to the inherent qualities of the tree-entity, but to qualities associated with the representation of that entity in the speech context, in particular what the speaker believes about the way in which the referent of a particular expression is represented in the mind of the addressee at a given moment (Chafe 1976, Prince 1981, Ariel 1988, Gundel et al. 1993, Lambrecht 1994, inter alia).

In order to illustrate the relationship between the form of a referring expression and how an entity is represented in the addressee's mind, consider the following example. Suppose that Anne calls Bob on the telephone and Anne utters the speech represented in (5).

(5) Anne (to Bob): Remember the woman we met at the Christmas party?
She came to see me yesterday.

The first utterance brings a representation of a particular woman to the center of Bob's attention. In the second utterance Anne uses *she* to refer to the same woman. A number of potential expressions could have referred to the same individual—for example, *the woman with the red hair, that woman, a woman*, among others. The fact that Anne uses

she indicates that she expects the woman from the party to be at the center of Bob's attention (due to the fact that she is the entity brought into focus by the previous utterance). Bob, in turn, can easily interpret this form even though it has very limited semantic content because he associates it with the woman who is at the center of his attention.

Further evidence that a speaker communicates something about a referent to the hearer via the choice of referring expression can be demonstrated in the following scenario. Suppose that Anne communicates the utterance in (6) to Bob.

(6) Anne (to Bob): Remember the woman we met at the Christmas party?
A woman came to see me yesterday…

By using the form *a woman* Anne communicates to Bob that she believes the referent is not at the center of his attention because it is someone who is not even identifiable by Bob by virtue of being in his long or short term memory or on the basis of the descriptive content encoded in the phrase itself.[6] Since Anne just referred to a woman from the Christmas party (who is now at the center of Bob's attention), Bob assumes that Anne

---

[6] Within the Givenness Hierarchy model proposed by Gundel et al. 1993, this is due to the scalar implicatures inherent in a hierarchical model of the use of referring forms interpreted in light of Grice's Maxim of Quantity, which are explained more fully in section 2.2. The implicature that the referent is not more than type identifiable, arises from the fact that the determiner *a* is interpreted as signaling the status 'type identifiable'. Since the speaker does not use a stronger form that indicates the addressee should be able to identify the referent—that is, *the woman*, *that woman*, or *she*, as shown in (i) below—the use of *a woman* implicates that the referent is not uniquely identifiable by Bob. When the stronger form is used, on the other hand, as in example (i) below, the hearer would have to search memory to find a uniquely identifiable referent, and the woman from the Christmas party would satisfy that search with minimial processing effort.

(i) Anne (to Bob): Remember the woman we met at the Christmas party?
The woman/that woman/she came to see me yesterday.

must be talking about a woman who is different from the woman from the Christmas party.

Reduced forms such as the pronoun *she* have extremely limited information content, yet they are effective in communication because they are constrained to use with highly salient referents. From the exchange between Anne and Bob, we see that associations between the form of a referring expression and salience of potential referents allow Bob to identify the referents of the various referring expressions in Anne's speech. The next section presents theoretical material which describes the constraints on referring expressions and their effects in communication more precisely.

2.1. Givenness, topic continuity, centering and accessibility

In a landmark article, Chafe 1976 is one of the first to talk about the information status of referents of noun phrases. Citing statuses like given-new, focus of contrast, definite-indefinite, subject, topic, and point of view, Chafe prepares the way for theories which link the use of distinct forms of referring expressions with properties of the referent in the minds of the interlocutors. Subsequent studies focus on describing these properties in terms of givenness/familiarity (Prince 1981), topic continuity (Givón 1983), accessibility (Ariel 1988, 1990), center of attention (Grosz, Joshi, and Weinstein 1983, 1995, Walker, Joshi, and Prince 1998), or cognitive status (Gundel et al. 1993). While the Givenness Hierarchy (henceforth GH) model of Gundel et al. provides the most comprehensive framework for describing the cognitive properties of referents of referring expressions, it is useful to look at the key contributions of earlier works and the

differences between the GH model and other frameworks which associate forms of referring expressions with mental properties.

Prince 1981 introduces a comprehensive view of "givenness", unifying various types within a single taxonomy and introducing distinctions such as "new", "inferrable", and "evoked". This work recognizes two types of givenness which are directly related to the way in which a referent is represented in the mind of the addressee: 1) givenness which is a result of the fact that a referent is assumed to be in the consciousness of the addressee because it has either been introduced in the discourse or is present in the physical context of the utterance; and 2) givenness which results from shared cultural knowledge, which is crucial to the concept that a referent is "inferrable", which means that it can be logically inferred from other discourse entities in the context (1981:236).[7] A third type of givenness described by Prince relates to properties such as parallelism, predictability, and recoverability. According to Prince, this type of givenness refers to the assumption that "the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence" (1981:226). Though the role of this type of givenness is not addressed specifically by the GH model, my analysis indicates that syntactic parallelism plays a role in the distribution of reduced forms of referring expressions in Kumyk.

---

[7] In the utterance, 'I went to the house, but the door was locked,' the entity represented by *the door* is inferred from the existence of the house entity (activated in the previous clause) based on cultural or encyclopedic knowledge about the structure of a house.

Givón (1983) establishes correlations between various degrees of topic continuity in discourse and a hierarchy of forms of referring expressions based on the amount of coding material. Working within the domain of functional linguistics, Givón claims that different "topics", or discourse participants, are associated with different degrees of accessibility for processing based on their predictability or "continuity" (1983:7). Forms of referring expressions, in turn, correlate with the degree of continuity of the topic. For example, in many languages, a null form is associated with a continuing topic, while an overt pronoun signals a change in topic. Givón proposes a cross-linguistic generalization that a hierarchy of forms based on phonological size (zero anaphora > unstressed/bound pronouns or grammatical agreement, stressed/independent pronouns, full NP's) correlates with the degree of topic continuity (1983:18). To a large extent, the ordering of forms in the hierarchy is reflected in the correlations suggested by the GH model; however, the latter model is a hierarchy of cognitive states, not of forms. Though Givón's claims account for aspects of the distribution of forms in discourse, he makes no claims about the use of referring expressions for entities which are not accessible via the linguistic context (that is, via cultural knowledge or the physical context). Givón's predictions, moreover, represent probable correlations, in contrast to the absolute predictions of the GH model that result from the structure of unidirectional entailment.

Building on Prince's familiarity scale and Givón's hierarchy of forms, Ariel (1988, 1990) proposes that the distribution of different types of referring expressions can be accounted for by the degree of accessibility of their referents in terms of a hierarchy of high, mid, and low accessibility levels. According to Ariel, a referent is accessible via

three hierarchically ordered means (from lowest to highest accessibility): "general knowledge, physical surroundings and previous linguistic material" (1988:68). At each of these levels of accessibility, different forms represent different rankings on a scale of low to high accessibility. For example, within the level of accessibility via general knowledge, a form such as *Joan Smith the president* represents lower accessibility than the form *Joan*, and within the level of accessibility via the linguistic context, a reflexive anaphor such as *herself* represents higher accessibility than the pronoun *she* (1988:81). Ariel's theory is comprehensive in the sense that it claims to account, not only for a variety of forms of expressions, but also for the distinction between pronominal and anaphor (in the binding theory sense); however, she does not suggest a methodology by which one can test form-status correlations in cross-linguistic corpus studies. Also, like Givón, her predictions are probabalistic and do not distinguish those cases where pronouns are completely disallowed from cases where a pronoun or full NP is equally appropriate.

Centering Theory (Grosz et al., 1983, 1995, Walker, Joshi, and Prince 1998) is another approach which claims that the relative salience of discourse entities places constraints on the form of referring expression which may be used. While the Givenness Hierarchy is a comprehensive model of form-status correlations and how they derive from both discourse and aspects of the extra-linguistic context (including encyclopedic knowledge, as for inferrables), Centering Theory focuses on the relationship between the form of referring expression and the coherence of discourse, describing focus of attention at points in a discourse and how changes in attention state take place (Grosz et al.1995:4).

This theory describes attention state as a product of the discourse alone, primarily relating it to previous segments, but also suggesting that higher-level discourse units play a role. One of its primary goals is to predict via a "centering algorithm" how the concept of focus can be used to identify the referents of anaphoric expressions such as pronouns, ellipsis, and definite noun phrases (Grosz et al.1995:5). As its name suggests, Centering Theory concentrates on identifying the focus or center of attention—that is, the most salient from among the set of referents of referring forms represented at a particular point in the discourse. From this set, the theory predicts the highest-ranking referent primarily on the basis of syntactic role, but also considering properties such as pronominalization and surface position (word order). As Mulkern 2003 points out, this conceptualization reduces salience to the level of the local utterance and does not allow a means of considering factors such as frequency of mention outside of the local utterance, overall importance of the topic to the discourse as a whole, participants in the extra-linguistic context, or higher level entities evoked by the discourse (such as propositions) (37). Considering that the center of focus is the emphasis of Centering Theory, it is not surprising that one of its largest contributions has been in the description of the use of reduced forms of referring expressions such as null arguments and pronouns. Of particular relevance to the present study is the description of Turkish by Turan (1998:140), which describes one aspect of the distribution of null arguments according to the following Rule of Centering: "The C_b (U_n) is the highest ranked entity of the C_f(U_n-1); if any entity is realized with a null subject, it is the C_b." While such a rule only

addresses a specific subset of null arguments, Turan's study provides evidence for specific factors which contribute to the salience of referents in Turkish (cf. section 2.3).

2.2. The Givenness Hierarchy model

While Ariel describes statistical correlation between categories of forms and levels of accessibility of the referent, Gundel et al. assume that the form that a speaker chooses from among multiple possible categories signals the speaker's assumptions about the referent's memory or attention status in the hearer's mind—that is, its 'cognitive status' (1993: 274-5). A significant difference between Ariel's model and the GH model is that the former addresses degrees of accessibility while the latter primarily addresses manner of accessibility (Gundel, Ntelitheos and Kowalsky 2007).[8] This model proposes six cognitive statuses which are assumed to be part of the conventional meaning of lexical items like pronouns and determiners: in focus, activated, familiar, uniquely identifiable, referential, and type identifiable. For each of these statuses, the theory proposes specific minimal criteria by which an entity may be associated with that status (see section 2.3).

Unlike in Ariel's work, in the Givenness Hierarchy model, the set of cognitive statuses is implicationally related and hierarchically ordered from most restricted (in

[8] One should note, however, that, while the difference between the cognitive status of familiar and uniquely identifiable primarily lies in how one accesses the referent, the difference between in focus and activated is, indeed, one of both manner and degree of accessibility.

focus) to least restricted (type identifiable), as shown in Figure 1, along with the corresponding referring expressions in English (1993:274-5).[9]

Status	In Focus >	Activated >	Familiar >	Uniquely > Identifiable	Referential >	Type > Identifiable
Form	*it,* unstressed *(s)he*	*this, that,* stressed *(S)HE, this* N	*that* N	*the* N	indefinite *this* N	*a* N

Figure 1: *The Givenness Hierarchy*

Correlating a type of referring expression with a cognitive status in the hierarchy means that its referents must always have at least that status. For example, in English, the pronoun *it* is analyzed as signalling 'in focus' and may not be used to represent a referent which is at most activated. The unidirectional entailment relationship between the statuses means that any form that is in focus is also activated, familiar (and so forth). In other words, the statuses are not mutually exclusive.

The entailment of statuses in the hierarchy also predicts, as Gundel 2003 explains, "a one to many mapping between statuses and forms in language use, since forms are underspecified for higher statuses, rather than excluding them" (134). This means that a form that signals one status may be used to represent referents of a higher status. For example, in English, the pronoun *that* can refer to an entity that is at most activated or an

[9] Note that pronouns referring to speech act participants are not generally included in discussions of theory about the distribution of types of referring forms because of the special contextual feature that the referents are always implicit in the speech act context.

entity that is in focus. It also follows that a referent that is at most activated can be represented by any one of the following forms: *this*, *that*, a stressed pronoun, *this* N, *that* N, *the* N, (indefinite) *this* N or *a* N (but not *it*). One of the implications of this claim is that the use of a particular form is constrained to referents which have a particular status, but within these constraints the use of one form over another is not predictable by cognitive status alone. The choice of one form over another is assumed to be governed by pragmatic principles such as Grice's Maxim of Quantity and Sperber and Wilson's Relevance principle (Grice 1975, Sperber and Wilson 1986/1995, 2004). Factors which may affect this choice include the number of competing referents, higher-level discourse structure, the relative salience of participants, and the desired contextual effects. Examples (7) and (8) illustrate how these contextual factors may influence the choice of referring expression.

(7) a. John has a very energetic Golden Retriever named Sam.
b. Every afternoon he and Sam go for a walk.
c. Usually after an hour, John is ready to come home, but the dog isn't.

In this example, according to the GH model, the referents of both *the dog* and *John* are in focus at the beginning of utterance (c); however, neither entity is referred to with a pronoun. Imagine the same utterance with pronouns, as shown in (c_2). As Relevance Theory predicts, speakers do not use an utterance of this form because the ambiguity of the pronouns increases the processing effort without adequate cognitive effects (Sperber and Wilson 2004:3-8).

c_2. Usually after an hour, he is ready to come home, but he isn't.

Ariel 1990 describes an interesting type of contextual effect created by the use of a lower status form—in this case, a definite NP—in reference to an entity with high activation status (in focus status, according to the GH model). Her example is reproduced here as example (8).

(8) After I'd prepared *our evening meal*, I cooked *our evening meal* and washed up after *our evening meal*. I didn't eat much of *our evening meal* because I was then sick of the sight of it.
(D. Harpwood, *The Diary of a Happy Housewife*, as cited in Ariel 1990:200).

In this case, the speaker's use of the definite NP is not related to the need to avoid ambiguity, but, as Ariel describes it, related to "the wish to invoke in the addressee the same feeling of repugnance with 'the evening meal'"(200).

Another unique claim of the GH model is that the form-status correlation represented in the hierarchy is part of the conventional lexical meaning of a form, and that the use of these forms creates conversational implicatures—more specifically, scalar implicatures—according to Grice's Quantity Maxim ("Make your contribution as informative as required/Do not make your contribution more informative than is required" (1975:162)). According to this principle, the use of a form which signals that a referent is activated could create the implicature that the referent is not in focus in the same way that *some* implicates *not all* (Gundel et al. 1993:294-299). As Gundel (2003:136) explains this means that, for example in English, the demonstrative pronoun *that* is "typically used only when conditions for using the more restrictive (hence more informative) pronoun, *it*, are not met". This point is illustrated in Gundel et al. 1993 and

Gundel 2003 through the contrast in the following examples, represented here as (9) and (10). According to Gundel et al., at the time the pronoun is uttered, the kitchen is in focus, but the closet is at most activated; hence the hearer associates *it* in example (9) with the kitchen, but *that* in example (10) with the closet.

(9) Anyway, going back from the kitchen then is a little hallway leading to a window. Across from the kitchen is a big walk-through closet and next to *it*…

(10) Anyway, going back from the kitchen then is a little hallway leading to a window. Across from the kitchen is a big walk-through closet and next to *that*…
(Gundel et al. 1993:298, ex. 56; Gundel 2003:136, ex. 24 - 25, italics mine)

To summarize the effects of the implicational nature of the Givenness Hierarchy, the forms in the hierarchy signal that a referent is at least a particular status, and, at the same time, often implicate that the referent is at most that status (i.e. not a higher status). These scalar implicatures create a ranking effect which can play a role in disambiguating referents, as in the examples above. According to Gundel et al. 1993 and Gundel and Mulkern 1998:26, the implicature created by the use of a demonstrative also explains why these forms are associated with focus shift. In addition to the disambiguating function, the ranking effect created by scalar implicatures has other discourse effects. For example, Mulkern 2003 notes that if two entities have the same cognitive status at some point in the discourse, the choice of referring form can indicate that the speaker is imposing a higher level of salience for one entity in relation to another (25). As explained by Mulkern, this "imposed salience" contrasts with the "inherent salience of the entities represented by their actual cognitive status.

A remaining question related to pragmatic principles guiding the use of forms to signal cognitive status is whether or not certain uses can represent "flouting"—or blatantly not following their cognitive status correlations in a sense parallel to the flouting of maxims described by Grice (1975). While this idea is not specifically expressed in the literature related to the GH model, it could explain the exceptional case of the use of demonstrative pronouns in cataphoric references. A demonstrative pronoun should not normally be used in reference to an entity which is not even activated, but is occasionally used in such a way in English, for example, as in (11).

(11) Wait until you hear this: Bill just proposed to Mary!

Such a "misuse" of the activated form may create extra processing effort, as the hearer must anticipate the continuation of the utterance in order to resolve the reference, but it seems to produce an added effect of emphasis via such anticipation.

2.3. Coding procedure for the GH model

The claims of the Givenness Hierarchy are based on cross-linguistic studies of referring forms in natural discourse as well as constructed examples tested with speakers. In both cases, the analysis is based on a process of coding the referent of a referring expression in order to determine the cognitive status of that referent. In order to determine this status, the Givenness Hierarchy model uses a coding protocol which describes sufficient conditions for assigning a particular cognitive status (Gundel et al. 1993, Gundel 2004). The criteria in the coding protocol refer to features such as the following: the syntactic and linear position in which a referent was previously

mentioned, whether or not a referent is a higher-level topic in the discourse, whether a referent is an event or state introduced directly by the content of a clause, or a speech act or higher order entity introduced more indirectly, whether or not the introduction of an entity is accompanied by gesture or eye gaze, and whether a unique representation of an entity can be created via the linguistic form by which it is introduced.

The most relevant cognitive statuses in a study of pronominals are 'in focus' and 'activated', as it is highly unlikely for a reduced form of referring expression to be used with a referent whose status is lower than activated. According to Gundel et al., a referent is in focus if it is "at the center" of an individual's attention (2003:284). Specific criteria for coding this status are listed in Figure 2.

A referent is IN FOCUS if it meets at least one of the following criteria:
1. It is mentioned in subject position in the immediately preceding sentence/clause. 2. It is mentioned earlier in the same sentence. 3. It is mentioned in syntactic focus position of the immediately preceding clause (postcopular position of a cleft or existential sentence in English). 4. It is a higher-level topic that is part of the interpretation of the preceding clause (whether it is overtly mentioned there or not). 5. It is mentioned in the two immediately preceding clauses. 6. It is the event denoted by the immediately preceding sentence.

Figure 2: *Coding protocol for in focus status (Gundel 2004)*

The status 'activated' describes entities which are represented in working memory, as opposed to long-term memory (Gundel et al. 2003:284). While any of the criteria mentioned in Figure 2 are sufficient for coding a referent as activated (since the status 'in focus' entails the status 'activated'), Figure 3 shows additional sufficient criteria for activated status according to Gundel 2004.

A referent is ACTIVATED if it meets at least one of the following criteria:
1. It is mentioned in one of the immediately preceding two sentences. 2. It is something in the immediate spatio-temporal context that is activated by means of a simultaneous gesture or eye gaze. 3. It is a proposition, fact, or speech act associated with the eventuality (event or state) denoted by the immediately preceding sentence(s).

Figure 3: *Coding protocol for activated status (Gundel 2004)*

2.3.1. Some coding choices in this study

While the criteria of the coding protocol have been tested cross-linguistically, the theoretical model does not exclude the possibility of certain language-specific adjustments to the criteria which are in character with their purpose. In this study, there are at least two areas worth noting where I have interpreted the criteria in ways which are, perhaps, language specific.

First, the criteria which refer to 'sentence' and 'clause' necessitate a language-specific definition of the terms. In terms of surface syntactic features, I define 'sentence' as any clause with a verb which includes both tense/aspect and number markers (regardless of punctuation), while I define 'clause' as a co-subordinate or subordinate structure with a non-finite verb (one which lacks tense/aspect or number features). Based on discussion with Jeanette Gundel about current revisions to the coding protocol under consideration, I interpret criterion 5 as applying to finite clauses (hence 'sentences').[10]

[10] Current discussion of the revisions of the coding protocol suggests that revising criterion 5 to replace 'clauses' with 'sentences' may actually be more accurate cross-linguistically. To illustrate the difference, consider the example in (i).

Secondly, I interpret the category of 'subject' as including non-subject arguments which are the experiencers of psychological verbs. This decision is based on evidence from Turkish in Turan (1998:153) that the experiencer object, particularly in the case of a psychological verb, is of a similar level of salience to grammatical subjects.[11] It should be noted that current revisions of the coding protocol by Gundel replace specific mention as a grammatical subject with the more general criterion, mention in a position of syntactic prominence. A number of other questions concerning the application of the coding which arise in the analysis, such as the coding of plural and composite entities or the treatment of contextual features such as pictures and video representation of referents, are discussed in Chapter 4.

(i) I talked to my mother early this morning. After I went to work, I decided to call her again. But **she** wasn't at home.

In thtis example, if criterion 5 applies to all types of clauses—including subordinate and non-finite clauses, then the referent of the pronoun *she* (bold) in the last sentence would not be automatically coded as in focus, as the referent is mentioned in the previous main clause by the form *her*, but not in the subordinate clause which precedes it (that is, *after I went to work*). If criterion 5 is formulated to apply to sentences, then the referent of *she* in the last sentence would be coded as in focus because it is mentioned by the form *her* in the previous sentence and by the expression *my mother* in the sentence before that. The application of criterion 1 also merits some elaboration given the discussion of the clause-sentence distinction. For the sake of clarification, consider the following example:

1. **Bob** drove the car because **Anne** asked **him** to.
2. **She** wanted **him** to have some practice driving on the highway.

When coding the referent of *she* in sentence 2, note that Anne is mentioned as the subject of the previous clause, but not as the subject of the main clause of the previous sentence. The criterion is intended to apply in a way that includes these cases as in focus. Moreover, the referent of *him* in the second sentence is Bob. Bob is mentioned in the previous clause, but not as the subject; however, *Bob* is the subject of the previous sentence. The criterion should be applied in a way such that this case is also in focus.

[11] An example of an experiencer object in Kumyk is illustrated in the expression *paĭkhammarny esine gele,* which translates literally, 'it came to the prophet's mind (that…)'.

As an example of how the criteria of the coding protocol are applied in Kumyk in order to determine the cognitive status of the referent of a referring expression, consider the text excerpt in (12), which consists of the first three sentences (a-c) of a narrative text. In this illustration, the referring expressions selected for coding are underlined, while other mentions of the same referents (which are relevant to the application of the coding criteria) are marked in bold type. Clause boundaries are marked with brackets.

(12) a. *[**Bir paĭkhammary bir k"yzy**,* *bir esheki,* *bir*
one prophet.GEN one girl.3.POSS one donkey.3.POSS one

mishiki, *bir* *de* *iti* *bolg"an.]*
cat.3.POSS one EMPH dog.3.POSS be.3S.PST
'A prophet had a daughter, a donkey, a cat and a dog.'

b. *[**K"yzny*** *erge* *berme* *iaraĭg"an* *vak"tisi*
girl.ACC husband.DAT give.INF possible.PST.PRT time.3.POSS

gelip] *[**bug"ar*** *gelechi* *gele.]*
come.GER 3.1.DAT suitor come.3S.PR
'When it comes time for the daughter to be given in marriage, a suitor comes to her.'

c. *[**K"yzny atasy*** *da* *razi* *bolup,]*
girl.GEN father.3.POSS and agreed be.GER

*[**Ø*** *k"yzyn* *berezhek* *bola.]*
girl.3.POSS.ACC give.FUT be.3S.PR
'The girl's father agreeing, (he) will give the girl.'

The referent of the expression *k"yzny* 'his daughter' in the first clause of (b) is the daughter of the prophet. The same referent is mentioned in the previous sentence/clause

in (a), but not as the unique grammatical subject; hence no criterion for in focus is satisfied.[12] The referent of the pronoun *bug"ar* in (b) is also the daughter. In this case, the referent is evaluated as in focus by criterion 2, "mentioned earlier in the same sentence." Criterion 5, "mentioned in the two immediately preceding clauses" would also apply to this case. In (c), *k"yzny atasy* "girl's father" refers to the prophet, who is mentioned in (a), but not in either clause of the previous sentence in (b); thus the referent fails to satisfy a criterion for in focus. The prophet is also the referent for the null argument in (c). Since the prophet is mentioned earlier in the same sentence, this referent is in focus by Criterion 2. Criterion 1 would also apply, since the referent is mentioned in subject position in the previous clause.

2.3.2. Coding clausally-introduced and higher order entities

The types of referents for any category of referring expression may include not only entities such as people, places, or things introduced by nominals, but also entities introduced by clauses, such as events, facts, and propositions. For the example in (13), possible referents introduced in the initial sentence include Ansar, his mother, the apple, the event of Ansar giving an apple to his mother, or the fact that Ansar gave an apple to his mother. Sentences (a) – (e) illustrate possible continuations of the text with each of these referents.

---

[12] The referent of *k"yzny* 'his daughter' in (b) is mentioned previously as part of the complex subject, *bir paĭkhammary bir k"yzy, bir esheki, bir mishiki, bir de iti* 'a prophet's one daughter, one donkey, one cat, and one dog'. Though the complex entity set would be brought into focus via mention in subject position, an individual entity would not necessarily be brought into focus by the mention of the set in subject position; thus the referent of *k"yzny* is coded as 'activated'.

(13) Ansar gave his mother an apple.

a. He is a nice boy.
b. She was pleased.
c. It was not ripe enough.
d. It happened this morning.
e. That surprised everyone.

In (a), (b), and (c), the personal pronouns refer to entities introduced by the nominals in (13), *Ansar*, *his mother*, and *an apple*, respectively. In (d) and (e), on the other hand, the referents of *it* and *that* are entities introduced by the whole clause. The referent of *it* is an event, while the referent of *that* is a fact.

The identification and coding of entities introduced by clauses (and verb phrases) has been established in Gundel's coding protocol based on corpus evidence from English and Norwegian (Gundel, Borthen and Fretheim 1999, Hegarty, Gundel, and Borthen 2001, Gundel, Hegarty and Borthen 2003, Hedberg, Gundel and Zacharski 2007) and supporting claims about the ranking of abstract entities in discourse (Asher 1993, Cornish 1999).[13] The most complete documentation of the distribution of referring forms in reference to clausally-introduced entities in English appears in Gundel et al. 2003. Based on previous evidence that the pronoun *it* requires its referent to be in focus while the pronoun *that* only requires activation (Gundel et al. 1993), the authors propose that clausally-introduced entities which may subsequently be referred to by the pronoun *it* are in focus, while those that cannot are at most activated. They argue that events are

[13] Cornish (1999:47-48), bases his distinction on Lyons' (1977) 3-tiered classification of entity order which is similar to the distinctions made by Asher.

immediately brought into focus by an utterance, since a clause directly denotes the event it describes, and the event can thus be brought into focus simply by virtue of processing the clause. Facts and propositions associated with a clause, on the other hand, are typically at most activated (cf. (13)). In making this claim, they also refer to the work of Asher (1993) who suggests a ranking of abstract entities in terms of "world immanence", with events having greater world immanence due to their properties of individuation and spatio-temperal features.[14] Subsequent work of Gundel et al. (2005, 2007) describes a major difference between higher order entities in terms of a distinction between a direct versus indirect relationship between a referring form and its antecedent trigger.[15] According to Gundel et al., events, activities, or states are directly introduced by clauses or verb phrases, while facts, situations, propositions typically must be inferred (or in the case of speech acts are activated but rarely brought into focus) and, therefore, involve further processing (2005:356, 2007:31).[16]

A significant concern in research methodology is the proper identification of higher order entity types. Hegarty (2007), following Vendler (1967) proposes that these types can usually be identified based on the selectional properties of verbs (171-173). For example, the complements of verbs such as *believe* and *think* are usually

---

[14] Fraurud 1996 also supports the role of individuation in entity categories, though this is not specifically linked to distinctions between clausally-introduced entities.

[15] Note that the concept of 'indirect anaphor' was first discussed by Erkü and Gundel 1987 but the application to entities introduced by clauses is discussed in later works.

[16] Speech acts performed by virtue of an utterance are always at most activated, since the focus of attention will be on the content of the speech act, not the fact that it has been performed (cf. Gundel, Hegarty, and Borthen 2003 for more detail). Cornish (1999) mentions that Dik (1997), like Gundel et al. considers speech acts to be even less accessible than facts, situations and propositions.

propositions, according to Hegarty, and facts are complements of verbs like *notice* and *point out* in English. Following similar logic, a speech act, would be recognizable by virtue of the fact that the referring form is the complement of a speaking verb, as in *John said **that***.

Events and states are considered to have spatio-temporal features and often occur with verbs of perception (e.g. *see hear*, *feel*), or with verbs meaning "happen," "repeat" or "do" (Asher 1993, Gundel et al. 2003, Hegarty 2007). Hegarty also clarifies the distinction between an event and a situation. While an event has specific temporal boundaries, a situation is a composite of an event and its consequences and, thus, has an initial temporal point which coincides with that of the event, but an undetermined end point (2007:171-172). References to situations are often associated with verbs or adjectives which express emotion, such as in the expressions, *that embarrassed me*.

Activities also fall within the category of entities not introduced by nominals, but they are considered within the Givenness Hierarchy literature to be introduced directly by a verb phrase rather than a full clause (Hedberg et al. 2007).[17] Like events, activities have spatio-temporal features, but they do not include reference to the agent/experiencer. In reference to the example in (13), a sentence continuation which includes a pronominal reference to an activity could be, *Did you see him do **it**?—* where *it* refers to the activity of (Ansar) giving an apple to his mother. One issue that arises within this study is reference to an activity type rather than a token activity with specific spatio-temporal

[17] Note that this description of activities versus events differs from literature on the semantic classification of events, which considers activities to be one of four sub-categories of events (Vendler 1967, inter alia).

features. As with reference to individuals versus entity types, reference to specific activities activates, but does not necessarily bring into focus the activity type. Subsequent reference to the type can therefore be made only with a form that does not require in focus status. For example, in the utterance *that is not allowed* in response to the example in (13), *that* refers not to a specific instance of giving an apple but to an abstracted type of activity that does not involve a specific agent,recipient, time, or place. The fact that *it is not allowed* sounds odd in this context indicates that the type is not in focus; in other words, *it*, which is restricted to entities in focus, cannot be used with an activity type that is at most activated.

In the analysis of clausally-introduced entities and other entities not introduced by nominals, one factor relevant to the discussion is the degree of nominalization of the clause/verb. The verb to noun scale is a continuum, which includes various degrees of nominalization in Kumyk, such as various types of non-finite suffixes (participle/verbal noun), nominalizing suffixes, and case suffixes. In this study entities introduced by a non-finite clause with some degree of nominalization are considered to be more easily accessible than those introduced by a finite clause, even though this is not specified in the coding protocol.

2.4. The GH model and other pragmatic factors in referential choice

Both the theoretical framework and the coding procedures of the GH model provide an excellent means of explaining the distribution and interpretation of referring expressions in discourse. Following this framework, the focus of the dissertation is to describe the degree to which choices between reduced forms of referring expressions can

be described in terms of the contrast between two levels of cognitive status: 1) in focus and 2) activated. I assume that the contrasts represented by the association between form and cognitive status are both inherent to referential choice and also used to constrain the set of referents for the hearer.

Beyond the GH framework, given the fact that there are six reduced forms in Kumyk, it is likely that the differences in form represent functional differences in addition to the two-way distinction in the cognitive status of referents. Based on the 1993 research, it is not unusual to have more than one form signaling a particular cognitive status, even when these forms are grammatically interchangeable in terms of gender or animacy. All five languages researched in Gundel et al. 1993 have two or more interchangeable forms representing the status, 'activated'. Three of the four languages which use null pronouns to represent entities in focus also have an overt pronoun which signals 'in focus'. A secondary interest of this dissertation is to consider distinctions in the use of pro-forms whose referents require the same cognitive status. This secondary task could be approached from two perspectives: 1) considering further distinctions in the upper end of the GH hierarchy—that is, statuses which lie between the distinction between activated and in focus, and 2) considering "reference-specific" functions of the multiple referring forms which represent the same level of cognitive status.[18] The first

[18] According to Brown-Schmidt et al. 2005, the "reference-specific" framework implies that "the mapping between referential forms and referents is mediated by multiple factors, with different referential forms being more sensitive to some factors than others." They further state that this type of framework assumes that a theory of reference must specify the set of constraints and their relative weights for each referring form.

possibility is not incompatible with the GH model, though it would require revisions in the structure of the hierarchy; however, I focus on the latter perspective which considers some specialized functions of some referring forms. Some recent works view evidence for reference-specific functions as a challenge to what they view as the "unidimensional salience approach" of the GH model (Kaiser and Trueswell 2004, Kaiser 2005, Brown-Schmidt, Byron and Tanenhaus 2005). My view, however, is that such functions can be identified without contradicting claims related to cognitive status.

Referential differences suggested by the literature that are discussed in this study include differences in reference to singular versus composite entities, as suggested in the work of Brown-Schmidt et al. (2005), differences in reference to directly-introduced versus indirectly-introduced entities (Cornish 1999, Gundel et al. 2007), or forms specialized for marking contrast (Mulkern 2003, Kaiser and Trueswell 2004). Another possibility suggested by the literature on demonstratives and reflexives is that forms might be specialized for marking speaker point of view or empathy (Kuno 1987, Cole, Hermon, and Huang 2001).[19]

Finally, a possible source of multiple forms representing the same cognitive status is the tension between eliding redundant forms and the need to represent highly salient entities with forms which bear phonological stress. A marked or stress-bearing form might be necessary in an emphatic context, such as context which contrasts two entities in the physical context. A marked form (versus a null form) would also be necessary in a

[19] Note that others, such as Kibrik 1999 and Grüning and Kibrik 2005 include point of view as a criterion that contributes to salience, not as an independent factor.

context in which the speaker desires to impose salience or in focus cognitive status for a referent which does not inherently have that status (Mulkern 2003).

2.5. Methodology in this study

Following the pattern of previous cross-linguistic study using the Givenness Hierarchy model, this analysis is based on the distribution of referring forms in a corpus study supplemented by an analysis of constructed examples through questionnaires and grammaticality judgments. In this case, the greater emphasis is given to corpus study.

2.5.1. Description of corpus

The text corpus for this study consists of twelve Kumyk texts, totaling approximately 100 interlinearized pages. Of these texts, four are transcriptions of oral elicitations, and eight are written texts. In terms of percentage of the total number of pages, 27.3 % of the corpus consists of oral material and 72.7 % is written. The written texts include one unpublished essay, three newspaper articles, two excerpts from a published novel and two complete texts from published collections of anecdotes and essays. One oral text is a description of a short video, one is a narrative of a wordless picture book, and the remaining two are free accounts of cultural or historical events. The texts in the corpus include a variety of genres, including narrative, hortatory, and expository discourse. Most of the texts represent contemporary style, while one text dates from the early 20th Century. Appendix 2 lists all the texts in the corpus by name along with information about length, mode, genre, and style for each text. Appendix 3 displays a sample text with coded forms.

In this corpus, the referents of 193 tokens of the overt pronouns *bu*, *o*, *shu*, *sho*, and *o'ziu* and the referents of 229 null arguments are analyzed and coded for cognitive status.[20] The distribution of each of these forms within the corpus is shown in Table 1. Note that for nulls and reflexives, the analysis is limited to those forms which represent third person entities. The analysis also excludes pronouns which occur within quoted speech in narrative texts, due to the difficulty of determining the cognitive status of the referent.[21]

Form	Number of Tokens
Ø	229
o'ziu	36
bu	33
o	90
sho	33
shu	1

Table 1: *Distributiton of pronouns in text corpus*

2.6. Description of questionnaire and grammaticality judgments

While corpus studies provide an excellent means of forming and testing correlations between referring forms and cognitive status, it is also important to test

[20] See Chapter 3, section 3.2 for my explanation of how null pronouns are defined in this study.
[21] While there are cases in which there is not enough material to code the referent of a pronoun found within a quotation, recent research by Watters and Gundel (2007), shows that for the English pronoun *it*, there is enough material in the majority of cases to code the referents of pronouns found within quotations.

whether or not grammaticality judgments of constructed examples indicate the same constraints on the use of referring forms. The challenge of testing a proposed correlation/constraint lies in the importance of providing adequate textual or extralinguistic context in which the cognitive status of a referent is evident as represented by the criteria of the coding protocol.[22] While preliminary research tests the substitution of forms within a naturally occurring text, this section primarily describes the use of a questionnaire to test the acceptability of pronoun use in constructed sentence pairs.

The questionnaire also provides an opportunity to elicit data related to a number of other issues of interest to this study: correlations between pronoun choice and resolving ambiguity in reference resolution, ranking effects in sentences with multiple pronouns, and correlation of pronouns with animate versus inanimate entities. Some samples are specifically designed to provide evidence for constraints in referring to certain types of clausally-introduced entities. In the sections below, more information will be provided about the methodology of the questionnaire in relation to these goals.

The questionnaire, consisting of 5 parts, was administered in written form to 5 respondents in Dagestan, Russia.[23] The respondents represent both males and females

---

[22] I initially experimented with testing predictions by substituting certain forms within the texts of my natural text corpus. For example, in a given text, I would substitute a form which is restricted to use with entities in focus (that is, *bu*, *o'ziu* or *Ø)* for a form in the text that refers to an entity which is at most activated by the criteria of the coding protocol of the GH model. I would then ask a Kumyk speaker to read the text and mark any forms that appear to be awkward or unnatural. In such a procedure, it is also important to ask the speakers to note which entity they think the form refers to. The scope of things I could test with this method, however,was somewhat limited.

[23] The difficulty of access to the region by foreigners did not allow me to administer the questionnaires in person, but the written questionnaires were administered by a local proxy who had been trained to administer them.

between the ages of 19 and 64, from several different dialectal regions. In order to eliminate some potential bias related to the order of presentation of examples for judgment (see Schütze 1996), I use two different versions of the questionnaire which primarily differ in the presentation order of the grammaticality judgments. (There are also a couple of examples which only appear in one version or the other, which allowed me to have a greater sampling of material without making the questionnaire too long.) Version A of the questionnaire is displayed as Appendix 4.

In the first part of the questionnaire, respondents are given 8 sentences, each of which is followed by a demonstrative pronoun in parentheses (no nulls or reflexives). For each sentence, the respondent is asked to write a one-sentence continuation using a particular pronoun. An example of this type of exercise is shown in (14).

(14) Given: *Muallim studentge kitap berdi. (bu)*
teacher student.DAT book give.3
'The teacher gave the student a book.'

Response: *Bu shossahat okhuma bashlady.*
3.1 immediately read.INF begin.3PST
'(S)he started to read right away.'

The methodology for this section is based on Kaiser 2005:272, whose goal is to test predictions that the choice of referring form influences the respondent to select a particular referent from the preceding sentence. While my primary goal in this section is to elicit a natural use of the given pronoun, this section also intends to provide evidence that contrasts in form guide anaphora resolution and help resolve ambiguity. Kaiser and Trueswell 2004 suggest that a sentence completion method can provide evidence that the

choice of proform (from among a presumed set of forms) allows the hearer to identify the referent of that form (from among a set of multiple referents). For example, provided the Kumyk context in (14), the subject is asked to continue the text using a specific pronominal form (*bu*). The choice of form is assumed to influence the subject's choice of topic of the continuing text from among possible referents introduced in the original sentence. In response to (14), most respondents use *bu* to refer to the student, not to the book or a higher-order entity like an event or fact. This type of exercise is particularly interesting in Kumyk, where there is no morphological distinction between animate and inanimate entities, and, thus, cognitive status could be the only property that would determine the possibility of reference with a particular form. (See also the description of section 4 of the questionnaire for further discussion of how the choice of pronoun guides the hearer in selecting from among multiple possible referents.)

In part 2 of the questionnaire, respondents are given 17 pairs of sentences and asked to fill in a blank in the second sentence with one of the possible pronominal forms listed: *bu, o, sho, shu, and o'ziu* (first 9 pairs); *bu, o, sho, shu, o'ziu* + Ø (next 8 pairs). The respondent is also given the choice that, if no pronoun is possible, (s)he should indicate the type of form which could be used to fill in the blank. Unlike part 1, the blanks are situated in sentential contexts biased for a particular referent, and the response elicits a form that is possible for that referent. For example, in (15), the context is constructed so that the respondent will choose a pronoun that refers the referent of *kitap* 'book' in the first sentence.

(15) Given: *Muallim studentge kitap berdi.* __ *Tangcholpan edi.*
teacher student.DAT book give.3PST Tangcholpan be.3PST
'The teacher gave a book to the student. __ was Tangcholpan.'

Response: *Ol Tangcholpan edi.*
3.2 Tangcholpan be.3PST
'It was Tangcholpan.'

The responses in this section indicate which forms can be used with certain types of referents. Since the respondent chooses only one form to fill in the blank, the responses also give an indication of preferences for pronoun use. The lack of use of a particular form with a certain type of referent, however, does not provide direct evidence of constraints on the use of a particular form, since that form may be a possible but not a preferred choice. For most of these sentences, the acceptability of various forms is tested separately in the grammaticality judgment portion

Part 3 (13 pairs of sentences) is structured in the same way as part 2, except that, in all but 3 cases, the second sentence of the pair has more than one blank, and respondents are also asked to identify the referents of the pronouns they used to fill in the blanks (4 cases have 3 blanks, and 6 cases, 2 blanks). The identification of referents is important in examples where there is more than one possible referent, as in (16), where the pronoun in the third blank could refer to either Patimat or the teacher.[24]

[24] I had to exclude data from one questionnaire in this section because the respondent did not identify the referents of the pronouns.

(16) A. *Muallim Patimatg"a kitap berdi.*
teacher Patimat.DAT book give.3PST
'The teacher gave a book to Patimat.'

B. …1… …2… *këp ushatdy.* *Song,* …3… *u'ĭge getdi.*
much please.3PST later home.DAT go.3PST
'It pleased her. Later she went home.'

Response: 1 = *O* (Patimat); 2 = *shonu* (book); 3 = *Ø* (Patimat)

Section 3 not only provides information about which types of pronouns can be used for which types of referents, but also illustrates tendencies in the use of pronouns in sentences with more than one pronominal form. Some analysis of a "ranking affect" is possible from this data.

The purpose of section 4 of the questionnaire is to test the respondents' intuitions about whether the use of a particular pronominal form signals something to the hearer about the choice of referent in contexts with multiple possible referents. This section is constructed specifically to test a previous hypothesis that *o* and *sho*, though signaling the same cognitive status, vary in the degree or frequency with which they give rise to a ranking effect via scalar implicatures.[25] For this part, pairs of sentences were given, as in (17) (A) and (B).

(17) A. *Bolat atyna minip barag"anda ĭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

[25] The form *bu* was not included in this section, as it was hypothesized that the "ranking effect for this pronoun in relation to *o* or *sho* is due to the fact that it signals 'in focus' while the other two signal 'activated'.

B_1. *Og"ar* *bolg"an* *zat* *boldumu?*
3.2.DAT happen.3PST.PRT thing happen.3PST
'Did anything happen to PRO?'

Response: *og"ar* = ___

B_2. *Shog"ar* *bolg"an* *zat* *boldumu?*
3.3.DAT happen.3PST.PRT thing happen.3PST
'Bolat fell off while riding his horse. Did anything happen to PRO?'

Response: *shog"ar* = ___

The second sentence of the pair allows for ambiguity in pronominal reference. For example, in (17) (B_1), the pronoun *og"ar* in the second sentence could refer either to Bolat or to the horse. The respondent is asked to identify the referent of the underlined pronoun. For each sentence pair in this section, there are two versions, illustrated here as (17) (B_1) and (B_2), which differ only in whether *o* or *sho* is used in the second sentence. The two nearly identical pairs are not given side-by-side, but, rather, interspersed with other sentences of a similar type.

Finally, section 5 provides grammaticality judgments of 41- 42 sentence pairs containing pronominal forms in the second sentence, as illustrated in (18). Each respondent is asked to judge the second sentence in the pair according to the following scale: 1=sounds great; 2=sounds good; 3=possible, but people don't usually say it that way; 4=no Kumyk person would say it that way.[26]

[26] I decided to exclude one set of data in which the respondent only rated one sentence as 3 and none as 4. (from Babayurt, not Makhachkala)

(18) *Muallim studentge kitap berdi. Sho Tangcholpan edi.*
teacher student.DAT book give.3PST 3.3 Tangcholpan be.3PST
'The teacher gave a book to the student. It was Tangcholpan.'

Response: 1 2 3 4 (circle one)

In order to provide more types of sentences, I vary a subset of sentence pairs for 2 of the questionnaires, which means that for that subset of the examples (16 sentences), only 2 respondents provide data, while the other examples have data from 4 respondents.

In this section, negative judgments for certain forms which represent referents coded as activated are interpreted as evidence that these forms are restricted to referents which are in focus. Likewise, positive judgments of pronouns used to refer to entities considered to be at most activated are interpreted as evidence that the forms are not constrained to referents in focus.[27] One of the caveats for this type of methodology is that the restrictions on the use of different forms of referring expression might represent 'soft constraints' as opposed to 'hard constraints'. While the cognitive status constraints might be strong enough to guide the production of referring forms (and hence explain the distribution of forms in a corpus), they might not be strong enough to produce judgments of unacceptability. The data from this section also provide evidence for other possible constraints on the use of pronominals, such as whether certain forms can be used with indirectly constructed referents or are restricted to referents with particular animacy

[27] Since the coding criteria are sufficient but not necessary criteria, some referents that do not meet any of the criteria for being in focus may actually be in focus. Decisions about which types of referents are considered to be at most activated are discussed further in Chapter 5 on the basis of parameters suggested by the questionnaire results.

features, or whether there are restrictions which correlate with previous mention in subject position.

To summarize the parts of the questionnaire, Sections 1-3 provide evidence for the set of pronouns which are possible in certain contexts, but do not provide clear evidence for which pronouns are not possible, as the absence of a particular pronoun in these contexts may be due to chance rather than due to restrictions. Section 3 provides additional evidence about tendencies for the distribution of forms and their referents in sentences with multiple pronominals. Section 4 primarily provides evidence for differences between *o* and *sho* in relation to scalar implicatures. Finally, Section 5 is intended to provide evidence on the cognitive status constraints and other possible constraints on the use of pronouns.

Chapter 3: Kumyk Pronouns and Their Properties

3.0. General comments

Within the range of possible forms of referring expressions, this study focuses on reduced forms, or those which have minimal semantic content.[28] In the literature, these forms are known as indexicals, deictic forms, pro-forms or anaphors. Within this class, various distinctions have been made in the literature. Hankamer and Sag (1976) distinguish between surface anaphors and deep anaphors, the former representing reduced forms which require a syntactic antecedent, and the latter representing those with a pragmatically controlled antecedent. Later syntactic literature formalizes a distinction between pronominals and anaphors in terms of binding conditions (Chomsky 1980 1982, 1986, 1995). Anaphors represent the class of pro-forms which are syntactically bound within their local domain (clause), which means these forms have a c-commanding antecedent within that domain. Pronouns represent forms which are locally free—or forms which cannot have an antecedent in the local domain but can have an antecedent in the sentence but outside the local domain or in the discourse context (via previous mention or mutual knowledge). The distinction between anaphor and pronoun as defined by Hankamer and Sag and Chomsky, among others, suggests that the antecedent for an

[28] In this study, as is the case in most comparitive studies of referring expressions, the focus is on third person pronouns, as the referents of these forms can be compared to the referents of other categories of referring forms, such as noun phrases. Also, first and second person pronouns, which refer to speech act participants, are of limited interest, since their referents are always in focus in the speech context.

anaphor is determined through a syntactic process, while the antecedent for a pronoun is determined semantically or pragmatically.

The work in accessibility or cognitive status neither relies on nor refutes the distinction between anaphor and pronoun. In one respect, it is arguable, as Ariel 1990 suggests, that apparent syntactic restrictions are based on inherent accessibility. In other words, since an anaphor's antecedent is constrained to the local domain, the antecedent is always highly accessible. In the same way, within the Givenness Hierarchy model, any form which fits the syntactic definition of anaphor would normally be in focus. In spite of the potential for cognitive models to account for the distribution of anaphors, the current study focuses on describing the distribution of forms which cannot be described solely in terms of binding conditions and control—that is, pronominal (as opposed to anaphoric—in the binding sense) uses of null arguments and reflexives, and overt non-reflexive pronouns. Sections 3.2 and 3.3 provide evidence that some types of null arguments and reflexives, which are anaphors in many other languages are pronominals in Kumyk and, therefore, merit consideration in this study.

Among overt pronominal forms, several categorical distinctions are made (primarily in relation to the use of demonstratives), and it is important to clarify the role of such distinctions in this study. Halliday and Hasan (1976:57-76) make a distinction between exophoric uses and endophoric uses. Exophoric uses, more commonly called deictic uses, represent reference to entities in the physical context of the speech situation,

while endophoric uses include all other uses.[29] Though the GH framework applies equally to both endophoric and exophoric uses, most of my data comes from endophoric uses. Some works further subdivide endophoric uses into two or more categories based on whether they refer to discourse participants or to higher level entities such as propositions or facts (Diessel 1999:93). This study, however, does not make any further distinctions among endophoric uses, applying the framework to all types of referents, whether individuals, objects, events, propositions, or other types of entities.

Following the GH framework, but in contrast to much of the literature on demonstratives, differences between multiple demonstrative forms are not described in terms of distance features such as proximal, medial, or distal, but in terms of differences in cognitive status. It is worth noting that recent literature outside the GH framework also prefers to describe demonstratives in terms of their communicative or cognitive function rather than in terms of distance features (Diessel 2006:12). These approaches propose that a primary function of demonstratives, whether endophoric or exophoric, is to establish joint focus of attention (Özyürek and Kita 2001 (as cited in Enfield 2003), Küntay and Özyürek 2002, Enfield 2003, Levinson 2004, Diessel 2006).

This chapter describes six third person pronominals in Kumyk, providing details of form, general aspects of use, and a summary of how these forms are described in the literature on Kumyk, as well as some comments on Turkish, a closely-related and more widely analyzed Turkic language. The description of four overt non-reflexive forms also

[29] This study does not address the issue of whether one type of usage—exophoric or endophoric—is more primitive.

raises the question of whether a distinction should be made between demonstratives and personal pronouns in Kumyk, and the sections on reflexives and null arguments provide justification for including the inflected reflexive and a subset of null arguments in Kumyk among pronominals.

3.1. Personal pronouns or demonstratives

The Kumyk language has four third person pronouns, occurring in their root forms as *bu, o, shu* and *sho*.[30] Morphologically unmarked for gender or animacy, these four forms refer both exophorically and endophorically to various types of animate and inanimate entities, including both human or non-human individuals, objects, abstract concepts, and clausally-introduced entities such as events, propositions, and facts.

In his early description of Kumyk grammar, Dimitriev (1940) classifies *o* as a neuter personal pronoun, while *bu, o, sho,* and *shu* are classified as demonstrative pronouns.[31] There is, however, no syntactic basis for making a distinction between personal pronouns and demonstratives in Kumyk, either in morphological form or in syntactic distribution. Moreover, since many would expect the term 'demonstrative' to be associated with distance features (which is incongruent with the framework I am using), this study chooses to use the more general term 'pronoun' in reference to these four non-reflexive forms and to study them comparatively as one class. It is useful, nevertheless, to be aware of how linguists have described the differences in meaning

---

[30] *Ol* also occurs as a dialectal variation of *o*.
[31] The root or nominative pronominal forms in Kumyk are identical to the demonstrative determiners *bu*, *o*, *shu* and *sho*, as in the expression, *sho adam* 'that man'.

between the four demonstratives. Dimitriev (1940:79) and, likewise, Muratchaeva (2001) describe the differences in meaning in terms of distance and "known-ness" as follows: *bu* represents near and known, *shu* represents near and not known, *o* represents distant, and *sho* represents the medial distance (further than *shu*, nearer than *o*).[32] The distinction between *shu* and *sho* is of particular interest, as the form *shu* is rare in discourse, and a number of other Turkic languages exhibit a 3-way, rather than a 4-way demonstrative system. Turkish, for example, encodes a 3-way distinction between the forms *bu*, *şu*, and *o*. Lewis (1967) describes *şu* as a medial demonstrative, while others claim that the same form encodes a referent that the addressee is not attending to (Özyürek and Kita 2001, as cited in Enfield 2003:109). While it is important to be aware of how Turkic pronouns are described in the literature, this study suggests, rather, that the difference in meaning between the four forms in Kumyk can be described more effectively with reference to the cognitive status of the referents they represent.

An overview of the morphology of third person pronouns provides information about various case forms and the types of entities to which they refer. The basic forms *bu, o, shu* and *sho* represent the nominative singular. As is the case with other nouns in Kumyk, plural pronouns are formed by the addition of the suffix *–lar* to the basic form, as illustrated in the first column of Table 2.[33]

---

[32] Neither Dimitriev (1940) nor Muratchaeva (2001) clearly states whether *o* or *sho* represent entities which are known or unknown to the hearer; however I assume that both forms represent entities which are known.

[33] In some cases, the singular form can be used to represent a plural entity, as with the adverbial/adjective forms such as *sholaĭ* (not *sholarlaĭ*).

Pronominal case forms other than nominative reflect the combination of one of the following case suffixes with the singular or plural form: genitive *(-nI*), dative *(-gA*), accusative *(-nI*), locative (-*dA*), ablative (-*dAn*), and possessive ((*sIi*), as shown in (19).[34] The full declension of singular and plural pronouns is displayed in Table 2.

(19) *o* + *lar* + *nI* = *olany*
root + plural + accusative

	NOMINATIVE	GENITIVE	DATIVE	ACCUSATIVE	LOCATIVE	ABLATIVE	POSSESSIVE
SING.	*bu*	*munu*	*bug"ar*	*munu*	*munda*	*mundan*	*busu*[*]
	shu	*shunu*	*shug"ar*	*shunu*	*shunda*	*shundan*	*shusu*
	o	*onu*	*og"ar*	*onu*	*onda*	*ondan*	*osu*
	sho	*shonu*	*shog"ar*	*shonu*	*shonda*	*shondan*	-
PLURAL	*bular*	*bulany*	*bulag"a*	*bulany*	*bularda*	*bulardan*	-
	shular	*shulany*	*shulag"a*	*shulany*	*shularda*	*shulardan*	-
	olar	*olany*	*olag"a*	*olany*	*olarda*	*olardan*	-
	sholar	*sholany*	*sholag"a*	*sholany*	*sholarda*	*sholardan*	-

Table 2: *Morphological declensions of pronominal forms*

(*The possessive case is only documented with the form *bu* in this data. The other two forms are listed by Dimitriev 1940:79.)

[34] Capital letters in suffix forms, such as -*mA* represent vowel harmony alternations represented in the orthography.

Pronouns plus the locative (*-dA*) or ablative (*-dAn*) suffixes are used in the sense of the English locational deictics *here* or *there* in referring to places, as illustrated in (20), where *onda* refers to the market.

(20) *Murat bazarg"a gete.* ***Onda*** *Ø alma aldy.*
Murat market.DAT go.3S.PR there apple buy.3S.PST
'Murat went to the market. (He) bought apples there.'

Class-changing suffixes such as *–laĭ* (modifier) combine with pronouns, as in the adverbial *bulaĭ* shown in (21). Semantically, these forms are roughly equivalent to the forms *thus* or *so* or *like that* in English, and they primarily refer to entities introduced by clauses or verb phrases. Interestingly enough, *-laĭ* adverbials may also be nominalized, as in the form *olaĭlar* 'ones like that' in (22).

(21) *Kuranda* ***bulaĭ*** *aĭtila:*
Koran.LOC 3S.MOD say.PASS.PR

"katingishige atasyny ag"lusunu kollari
oman.DAT father.3POSS.GEN brother.3POSS.GEN hand.PL.3POSS

tiĭse gunakh tugul bashg"a adamlar tiĭse gunakh."
touch.3COND sin NEG other man.PL touch.3COND sin
'In the Koran it is said like this: "If a woman's father's or brother's hands touch her, it is not a sin; if other men touch (her), it is sin.'

(22) *Bibliotekag"a bardym – cheber kitaplar da okhumaĭ eken.*
library.DAT go.1S.PST – literature book.PL EMPH read.INF COP

Klassda ***olaĭlar*** *dag"yda bolg"an.*
lass.LOC 3.2. MOD.PL more be.PST.PRT
'I went to the library—to read literature. In the class there were others like that.'
[Marian.045]

Pronominal forms in various cases occur with a number of postpositions, also labeled subordinating conjunctions or logical connectives, as in *o saialy* 'because of 3S'/'therefore'/'according to' or *shog"ar gere*, literally, 'seeing that', but conveying the meaning 'therefore'/'because of'. A list of connectives incorporating pronouns is displayed in Figure 4.

Nominative form	**Pronouns(+Case)+Postposition**
bu	*munu uchun* 'for 3S'
shu	*shug"ar gëre* 'according to3S' *shundan song* 'after 3S' *shuunu bulan birge* 'together with 3S'
o	*onu uchun* 'for 3S' *o saialy* 'because of 3S' *o sebepli* 'according to 3S' *og"ar gëre* 'according to 3S'
sho	*sho(nu) saialy* 'because of 3S' *sho sebepli* 'according to3S' *sho sebepden* 'because of 3S' *shonu uchun* 'for 3S' *shondan song* 'after 3S' *shondan taba* 'through that' *shog"ar gëre* 'according to 3S' *shog"ar de karamayly* 'not consisdering 3S'

Figure 4: *Pronouns in logical connectives*

The pronouns within these expressions most often refer to events, facts, propositions or other higher order entities, as illustrated in (23), where *o saialy* refers to the fact that the mother-in-law wants to invite people to a feast.

(23) *Yash bolnitsadan chykg"an song bir zhumany ichinde kaĭnana*
child hospital.ABL exit.PST.PRT after one week.GEN inside.of mother.in.law

takhana etip adamlany chakyrma baĭram etme suĭe.
takana make.GER man.3PL.ACC invite.INF feast do.INF want.PR

O saialy *bu chakyryvlar ĭiberip kyzny adamlaryn*
3.2 according.to 3.1 invitation.PL send.GER girl.GEN man.PL.3POSS.ACC

o'ziunu adamlaryn chakyra.
self–3POSS.GEN man.PL.3 POSS.ACC invite.PR
'After the child leaves the hospital, within one week, the mother-in-law wants to hold a feast, making taxana and inviting people. Because of this, she sends invitations and invites daughter(in-law)'s people, (self's) own people.'
[Birth1.008-9]

In a few cases, it is difficult to determine if derivational forms of *bu, o, shu* and *sho* are truly referential pronouns. One set of examples involves the form *shonchaky* 'that much'/'so much', in which the pronoun root combines with the morpheme *-cha/-chaky*, which carries the meaning of 'amount', as shown in (24).

(24) *sho + chak"y = shonchak"y*
root + amount 'that much'/'so much'

In some cases *shonchak"y* appears to be referential, referring to a specific amount mentioned in the discourse context (25), while in other cases the same form functions as an intensifier, as in (26).

(25) *Muratny on rubil' bar. Mennike shonchak"y ak"cha ëk".*
Murat.GEN ten ruble exist.3S my 3.3.much money exist.NEG
'Murat has 10 rubles. I don't have that much money.'

(26) *Nege_tiugiul de shonchaky uzak" zaman ishlegen*
because EMPH 3.3.much distant time work.PST.PRT

ishibizni *natizhasyn* *gërmek*
work.1PL.POSS.GEN result.3POSS.ACC see.INF

ullu *siuiunch* *ullu ish.*
great joyful great work
'Since the result of our work had taken such a long time to see, it was a great joy, a big deal.' [NT.012-.013]

Non-referential cases in which *shonchak"y* serves as an intensifier are not considered in this study.

3.2. Null pronouns

Within binding theory (Chomsky 1980 1982, 1986, 1995) a distinction is made between syntactically controlled null arguments (anaphors) and null arguments which are not syntactically controlled (pronominals); this study focuses on the latter. The study of pronominal null arguments is more interesting to cognitive approaches because, unlike anaphors, which are obligatory, pronominal nulls alternate with other forms of referring expressions. Much attention has been given to explaining elided subjects or 'pro-drop' in terms of syntactic parameters such as recoverability (see Huang 2000 for a comprehensive overview); however, it is widely accepted that, cross-linguistically, the occurrence of pronominal null arguments versus fuller forms in their various argument positions depends on non-syntactic parameters. Non-syntactic parameters proposed in the literature include topic continuity (Givón 1983), centering (Grosz et al., 1983, 1995, Walker, Joshi, and Prince 1998), accessibility (Ariel 1988, 1990), and cognitive status (Gundel et al. 1993), with the approach of this study focusing on the last of these. Most of the literature which analyzes null arguments based on non-syntactic parameters such as

accessibility or cognitive status applies more broadly to all types of null arguments (not just subjects) and is assumed by many to apply equally to both anaphors and pronouns (in the binding sense). Theories which associate the form of referring expression with attention state, mental accessibility, or the cognitive status of the referent propose that null arguments are associated with the highest level of accessibility or the most restricted level of cognitive status – that is, located at the center of the hearer's attention.[35] Several of these theories also associate null forms with the concepts of continuity, parallelism or predictability (Prince 1981, Givón 1983), a factor which appears to play a role in the distribution of null arguments in Kumyk, as discussed in Chapter 4).

In Kumyk, null pronouns occur in a variety of syntactic contexts: null subjects of finite clauses, as illustrated in (27), null subjects of adverbial clauses as in (28) and non-subject arguments such as objects or possessors, as illustrated in (29).[36]

(27) *Murat bazarg"a gete. Onda* ***Ø*** *alma aldy.*
Murat market.DAT go.3S.PR there apple buy.3S.PST
'Murat went to the market. (He) bought apples there.'

(28) *Murat bazarg"a gete. Onda* ***Ø*** *alma alyp,*
Murat market.DAT go.3S.PR there apple buy.GER

o bizge geldi.
3S 1PL.DAT come.3S.PST
'Murat went to the market. Having bought apples there, he came to us.'

---

[35] It should be noted that this type of characterization of null arguments applies only to referential cases, and not to expletive or generic uses of null arguments.
[36] Humnick 2006 provides a more complete description of syntactically restricted (anaphoric) versus unrestricted (pronominal) null arguments in Kumyk.

(29) *K"yzny atasy$_i$ da razi bolup,*
girl.GEN father.3.POSS and agree be.GER

*Ø$_i$ [subject] **Ø**$_i$ [possessor] k"yzyn$_j$ berezhek bola.*
girl.3.POSS.ACC give.FUT be.PR
'The girl's father, agreeing, will give (his) daughter.'

In example (27) the null argument is the subject of the second matrix clause. As is the case in many pro-drop languages, the finite verb *aldy* 'bought' is morphologically marked for subject agreement in person and number. In this case, the morpheme *–dy* indicates third person singular. In example (28) the null form occurs in an adverbial clause as the subject of the gerund, *alyp*.[37] In this type of complex clause construction, the null subject could occur either with the gerund or with the matrix verb (as in the first sentence of (29)), or with both. Though, in most cases, it would be infelicitous to have overt subjects in both clauses it is possible to have an overt subject in both clauses if one is a reflexive.[38]

The justification for the category of null possessive pronouns illustrated in (29) is based on the syntax of possessive constructions in Kumyk. In a possessive construction

---

[37] Non-finite verbs with the suffix *–Ip* are also described as 'converbs' in Turkic literature (see also Haspelmath 1995).

[38] While the general use of a reflexive pronoun with a coreferent overt subject appears to be infelicitous, cases which are clearly emphatic usages are acceptable. Example (i) occurs in the extended text corpus. Since the reflexive here is marked with a 1st person suffix, there is no question that it is coreferential with the matrix subject, *men* "I".

(i) *O'zium de bilmeǐ, men bir k"yzny betine tiklenip turg"anman.*
self.1S EMPH know.NEG.GER 1S one girl's face.at stare.GER remain.PST.1S
'Not even realizing it myself, I kept staring at one girl's face.'

in which all elements are overt, the NP representing the possessor exhibits a genitive suffix, while the possessee exhibits a possessive suffix which agrees in person and number with the possessor.[39] The structure of the overt possessive construction is illustrated in (30), where (a) is a first person possessive and (b) is a third person possessive.

(30) a. *meni* *ash-ym*
1S.GEN food-1S.POSS
'my food'

b. *mishik-ni* *ash-y*
cat-3.GEN food-3.POSS
'the cat's food'

In contrast to (30), in which the third person possessor is overt, in example (29) above, the possessor is realized as a null argument, while the possessee is realized as an overt form, *k"yz-y-n* 'his daughter-3POSS-ACC' that exhibits agreement in person and number with the possessor (the father) via the suffix, *-y*.

This study excludes two types of syntactically controlled null arguments: 1) null subjects in non-finite complement clauses that are syntactically restricted to coreference with a particular argument of the matrix clause by the category of the verbal suffix and 2) null subjects of adnominal clauses (Humnick 2008). For example, in a complement clause that ends with the suffix *–mA* or *-mAg"A*, as shown in (31), the null subject must

[39] The third person singular and third person plural forms, however, are indistinguishable.

be interpreted as coreferent with the matrix subject, *k"atyn* 'woman'.[40] (The non-finite clause is enclosed in brackets, with its subject in bold, and the matrix subject is underlined.)

(31) <u>*K"atyn*$_i$</u> *[baĭramg"a* ***Ø$_i$*** *adamlany* *chak"yrma]* *siue.*
woman feast.DAT man.3PL.ACC invite.INF want.3S.PR
'The woman wants to invite people to the feast.'

As shown in (32) (a) and (b), in this type of non-finite clause, it is ungrammatical for the subject to occur as an overt argument, whether coreferent with the matrix subject or not.

(32) a. *<u>*K"atyn*$_i$</u> ***[gelin$_j$*** *adamlany* *baĭramg"a* *chak"yrma]* *siue.*
woman bride man.3PL.ACC feast.DAT invite.INF want.3S.PR
'The woman wants the bride to invite people to the feast.'

b. * <u>*K"atyn*$_i$</u> ***[o'ziu$_i$*** *adamlany* *baĭramg"a* *chak"yrma]* *siue.*
woman self man.3PL.ACC feast.DAT invite.INF want.3S.PR
'The woman wants herself to invite people to the feast.'

The second type of null arguments excluded in this study are those which occur as arguments within adnominal participial clauses with the suffixes *–Ag"An* (present) or *–g"An* (past)—that is the equivalent of the relative clause. Such clauses, as illustrated in (33) (a) and (b), can modify any argument of the matrix clause, and the null argument of the participial clause is always interpreted as coreferent with the head.

---

[40] There are also semantic types of verbs taking the same suffix in which the null argument is controlled by the object, as with the case of the phrasal verb like *ixtiiar bere* 'gives the right to'.

(33) a. *[Ø$_i$ ash etegen] k"atyn$_i$ gele.*
food make.PR.PRT woman come.3S.PR
'The woman who makes the food coming.'

b. *[K"atyn **Ø**$_i$ etegen] ash$_i$ iakhshy.*
woman make.PR.PRT food good
'The food that the woman makes is good.'

In (33) (a), the participial clause (in brackets) modifies *k"atyn* 'woman', and the null subject of the non-finite verb *etegen* 'making' must be interpreted as coreferent with *k"atyny*. Similarly, in (33) (b), the null argument which is the object of the adnominal must be interpreted as coreferent with *ash* 'food'. For the types of null arguments in both (31) and (33), the referent is syntactically determined, and the null argument cannot alternate with other forms of referring expressions, thus such examples have comparatively little interest to this study.

3.3. Reflexives

3.3.1. General properties

Kumyk exhibits two types of nominal reflexives (in addition to verbal reflexivization): the uninflected reflexive *o'z* and the inflected form of the same root morpheme. The uninflected form, which is not included in this study, serves primarily as a modifier and does not necessarily imply specific reference, as in example (34).

(34) *Anam **o'z** k"olu bulan ash etgen.*
mother.1S.POSS self hand.3POSS with bread make.3PST
'My mother made homemade bread.' (Lit. 'My mother made bread by own hand.')

The inflected reflexive is composed of the noun root *o'z* 'self' followed by a possessive suffix which, in a genitive construction, indicates agreement in person and number with the possessor. Example (35) illustrates the parallel structure between a first person singular genitive construction (a) and the inflected first person singular reflexive (b).

(35) a. *meni gëz - ium*
1S.GEN eye – 1S.POSS
'my eye'

b. ~~*meni*~~ *o'z - ium*
1S.GEN self – 1S.(POSS)[41]
'myself'

Reflexive pronouns, like other Kumyk pronouns, do not indicate animacy or gender, but are marked for person and number, as shown in Table 3. [42] (Although 1st and 2nd person forms are given here for reference purposes, the focus of this research is 3rd person forms.) Like all nouns, reflexives exhibit case suffixes agreeing with their argument function. Table 4 shows the case declension for 3rd person singular and plural forms.

---

[41] In the remainder of this paper I have dropped the 'POSS' portion of the gloss and have chosen to represent the glosses for reflexives simply as 'self.1S, self.2S, self.3S, self.1P, self.2P, self.3P'.

[42] When the referent of a reflexive is third person plural, often a singular reflexive form is used. Unlike third person singular forms, third person plural reflexive forms do not exhibit the possessive suffix, but are declined for case like other plural nouns.

[44] Cole et al. describe three types of long-distance reflexives: 1) "long-distance bound anaphors", 2) "forms which are used as reflexives locally and as pronominals non-locally", and 3) "forms that are "primarily" bound anaphor reflexives, but which can be used non-locally in specific syntactic and discourse contexts" (2001:xviii). For languages like Kumyk, the fact that the same form behaves locally like an anaphor and non-locally like a pronominal is still somewhat problematic. For similar data on the Malay reflexive *diri* + pronoun, Cole and Hermon 2005 offer the explanation that the features 'anaphor' or 'pronominal' are not

Singular	1st person	*o'zium*
	2nd person	*o'ziung*
	3rd person	*o'ziu*
Plural	1st person	*o'ziubiuz*
	2nd person	*o'ziugiuz*
	3rd person	*o'zler*

Table 3: *The inflected reflexive*

	Nominative	**Genitive**	**Dative**	**Accusative**	**Locative**	**Ablative**
Singular	*o'ziu*	*o'ziuniu*	*o'ziune*	*o'ziun*	*o'ziunde*	*o'ziunden*
Plural	*o'zler*	*o'zleni*	*o'zlege*	*o'zleni*	*o'zlerde*	*o'zlerden*

Table 4: *Case declensions of 3rd person reflexives*

3.3.2. Syntactic properties

Kumyk reflexives do not have the prototypical properties of a form that is strictly an anaphor in the binding theory sense, that is, a form whose antecedent is a c-commanding element found within the same local domain, or a pronominal, a form which cannot have a local c-commanding antecedent. The ambiguous interpretation of the sentence illustrated in (36) shows that the antecedent of a reflexive may be either a c-

commanding element within the local domain (Murat) or a syntactic element outside the local domain (Ali).

(36) *Ali*$_i$ *Murat*$_j$ ***o'ziun***$_{i/j}$ *siuegenine* *inanaď.*
Ali Murat self.3S.ACC like.PR.PRT.3POSS.DAT believe.PR
'Ali$_i$ believes that Murat$_j$ likes him$_i$ /himself$_j$.'

The evidence of long distance antecedents raises the question of whether the reflexive in Kumyk is more like a long-distance bound anaphor or a pronominal.[44] Cole, Hermon, and Huang 2001 demonstrate that long-distance reflexives in a variety of typologically diverse languages vary in their resemblance to pronouns or bound anaphors. For example, based on evidence provided by Kornfilt (2001) in the same volume, Cole et al. categorize the Turkish inflected reflexive *kendi* as a pronominal in its use beyond the local domain (2001:xx). Their categorization cites the following types of evidence that a long-distance reflexive is a pronominal rather than an anaphor: 1) lack of restriction to certain types of complex clause structures, 2) lack of restriction to subject antecedents, 3) the possibility of use with extra-sentential antecedents and 4) the lack of restriction to logophoric contexts (Pica 1987, Cole et al. 2001). [45]

An analysis of Kumyk reflexives in complex clause structures demonstrates that long-distance reflexivization is found in all types of clause combinations (Humnick

[45] Cole et al. 2001 also point out that bound anaphors require sloppy readings under VP ellipsis, while pronouns allow for both strict and sloppy readings under VP ellipsis; however,Kornfilt's analysis of Turkish does not provide data on this aspect of inflected reflexives.

2007). Examples (37) – (38) illustrate two of these types: complement clause constructions and relative clauses.

(37) *Iug'an*$_i$ *[[Pavel*$_j$ ***o'ziun***$_{i/j}$ *siuegenine]*
John Paul self.3S.ACC like.PR.PRT.3POSS.DAT

men inanag"anymny] *bile.*
1S believe.PR.PRT.1S.POSS.ACC know.3S.PR
'John$_i$ knows that I believe that Paul$_j$ likes himself$_j$/him$_i$.'

(38) *Ibrag'im*$_i$ ***[o'ziu***$_i$ *(Ø*$_j$*)* *tizip* *getgen]* *zatlag"a*$_j$ *k"arap*
Abraham self.3S build.GER leave.PST.PRT thing.PL.DAT look.GER

bir zatlar aĭtdi.
one thing.PL say.3S.PST
'Abraham$_i$, looking at the things$_j$ which (he) himself$_i$ built, said some things.'

As illustrated in (37) reflexives within nominal clause complements can have either the local clause subject or the matrix clause subject as an antecedent, and there is no blocking effect from intermediate subjects.[46] In this example, *o'ziun*, the object of the verb, *siuegenine* 'like' can refer either to *Pavel*, the local subject or to *Iug'an*, the matrix subject, and the intervening subject *men* 'I' does not block the long-distance reference. Example (38) demonstrates a reflexive occurring as an argument of a relative clause. In this case, the reflexive *o'ziu* is coreferent with the matrix subject, *Ibrag'im*. In Kumyk relative clauses, the head is represented as a gap (in other words, not overtly represented).

[46] In some languages, such as Chinese, the scope of long-distance reflexivization cannot cross a clause in which the subject has different person and number features from the reflexive, a phenomenon which is known as a 'blocking effect' (Cole et al. 2001:xiv, Huang 2000:98).

Reflexives are not (usually) coreferent with the head, but are coreferent with the matrix subject.

Though the majority of reflexives have subject antecedents, there is evidence that long-distance reflexivization allows for non-subject antecedents in some cases, as demonstrated in (39).

(39) ***O'ziu***$_{i/j}$ *biblioteka gelse,* *Iug'an*$_i$ *Pavela*$_j$ *gërezhek.*
self library.DAT come.COND John Paul.DAT see.FUT
'If self comes to the library, John will see Paul.'

In this example, the reflexive occurs in the non-finite conditional clause, and it can be coreferent either with the subject of the matrix clause (*Iug'an*) or a non-subject argument of the matrix clause (*Pavela*). In other words, this example can mean either "If *Paul* comes to the library, John will see Paul" or "If *John* comes to the library John will see Paul".

Long-distance reflexivization also extends to extra-sentential antecedents, as shown in (40), where the reflexive *o'ziu*, the subject of the second sentence, is understood to be coreferent with *muallim* 'teacher' in the previous sentence.[47]

(40) *Muallim studentge kitap berdi.*
Teacher student.DAT book give.PST.

[47] This example is from section 3 of the questionnaire responses. Both *o'ziu* and *onu* are forms used to fill-in the blank in two blanks in this example.

Song ***o'ziu*** *onu* *eksameni* *bulan* *k"utlady.*
Later self.3S 3S.ACC exam.ACC with congratulate.PST
'The teacher gave the student a book. Then he(*o'ziu* = teacher) congratulated him(*onu* = student) with the exam.'

The evidence that long-distance reflexivization in Kumyk is not restricted to certain clause types and allows for non-subject and extra-sentential antecedents supports the claim that the Kumyk inflected reflexive is a pronominal whose distribution cannot be characterized on a purely syntactic basis (that is, on the basis of locality and antecedent type constraints). It is appropriate, therefore, to consider the distribution of this form in terms of non-syntactic parameters such as the cognitive status of the referent.

3.4. Summary

To summarize, the primary interest of this study is the endophoric use of pronominal forms which are not constrained to an antecedent based purely on syntactic constraints such as binding conditions. This class includes five overt forms, *bu*, *shu*, *o*, *sho*, and *o'ziu* and a subset of null arguments. This chapter provides evidence that the differences in the distribution of these five forms cannot be described solely in terms of syntax, and categories such as 'demonstrative' or perhaps even 'reflexive' (to the extent that it presumes an anaphoric distribution) are not very meaningful in describing the use of these forms. The purpose of this study is primarily to describe the use of these forms in terms of a correlation with the cognitive properties of their referents.

Chapter 4: Corpus Study

4.0. General comments

The goal of the corpus analysis is to compare the six pronominal forms in Kumyk on the basis of features of their referents—most importantly how the referent in each use of these forms corresponds to one of the cognitive statuses on the Givenness Hierarchy. Table 5 displays the numerical results of this coding of referents according to the highest status for which a referent could be coded.[48] For example, for the form *bu*, 30 tokens have a referent that is coded as in focus, while one referent is coded as at most activated. Since the status 'in focus' entails that a referent is also activated, all 31 coded forms have referents that are at least activated, but only one has a referent that is coded as at most activated. This table also lists 12 forms for which the cognitive status of the referent remains undetermined. These are primarily cases which fall within the scope of the study, but for which there is a particular question related to the application of the coding protocol. Table 6 displays the distribution of tokens of a given form in terms of percentages of the total number of referents coded for that form. In other words, for the form *bu*, 96.8% of the 31 coded referents are in focus. Note that, as no referent of a pronominal is coded as a status lower than familiar in this study, only the three relevant statuses, in focus, activated, and familiar, are represented in both of these tables.

---

[48] Since the statuses of the Givenness Hierarchy are in a relationship of unidirectional entailment, if a referent is coded as in focus, the referent is also activated, familiar, and so forth. In the table the referent of a given token is only listed under the highest status for which it is coded and not the other statuses which it entails.

Form	Number of occurrences	Referent in focus	Referent activated	Referent familiar	Referent undetermined
Ø	229	213	8		8
o'ziu	36	35	-		1
bu	33	30	1		2
o	90	64	24	1	1
sho	33	4	29		
shu	1		1		

Table 5: *Coding Results for Pronominal Forms in the Text Corpus*

Form	Total coded referents	% in focus	% activated	% familiar
Ø	221	96.4%	3.6%	
o'ziu	35	100%		
bu	31	96.8%	3.2%	
o	89	71.9 %	27.0 %	1.1 %
sho	33	12.1%	87.9%	
shu	1	0 %	100%	
TOTAL	410	84.4%	15.4	0.2%

Table 6: *Percentage of coded referents with a specific cognitive status*

According to the Givenness Hierarchy model, if the referents of a given form are distributed among more than one status on the hierarchy, the form is said to explicitly encode the lowest status, i.e. to signal that the referent of the form is at least that status. For example, if the referents of a form are distributed between the statuses 'in focus' and 'activated', as for the pronoun *sho* in table 5, where 29 referents are coded as activated and 4 referents are coded as in focus, the form is said to signal that the referent is activated. In some cases, however, when the number of referents corresponding to the lowest status is less than 5% of the coded forms, these forms may be considered

exceptional or miscoded, and the form is analyzed as correlating with the higher status. For example, for the form *o*, the total number of coded forms in the corpus is 89, and the referents are distributed as follows: 64 (71.9 %) in focus, 24 (27.0 %) activated, and 1 (1.1 %) familiar. Since the number of familiar referents is less than 5.0 % the form *o* is analyzed as signaling that the referent is activated. According to these principles, the corpus data results in the hypothesized form-status correlations: nulls, *o'ziu* and *bu* signal 'in focus', while *o*, *sho*, and *shu* signal 'activated' (see Figure 5). Note that, since there is only one token of the pronoun *shu*, the hypothesized form-status correlation is at best preliminary.

FOC	**ACT**
Ø, o'ziu, bu	***o, sho, shu**** *(* insufficient data)*

Figure 5: *Form Status Correlations*

The remainder of this chapter discusses the distribution of each form primarily in terms of the cognitive status of the referent. My analysis also notes correlations between particular forms and other characteristics of the referent, including animacy features (human, non-human, inanimate), previous mention as a subject (a subset of the coding criteria for in focus) and whether a referent is introduced previously by means of a noun phrase or a whole clause. Within the discussion of each form, I provide samples as well as a discussion of exceptional or undetermined cases, cases of particular interest, difficult

coding decisions, and unique features of the distribution, including contextual effects associated with the form.

4.1. Nulls in the corpus

Null arguments are the most frequent form of pronominal occurring in this corpus, with a total of 229 tokens. Of this total, eight tokens are cases of backwards anaphora which are classified as undetermined. The majority of coded referents of null arguments are in focus, with only eight referents coded as activated. Since the number of activated referents is such a small portion of the total number of referents (3.6 %), I posit that these are exceptions and that null arguments signal that their referents are in focus.[49]

According to the corpus data, the use of null arguments is neither restricted to reference to previous subjects nor restricted in terms of animacy features. While the use of null arguments with clausally-introduced entities is rare, there is one example in the corpus which is discussed in section 4.1.5. [50]

4.1.1. Coding the referents of null arguments

In example (41), a narrative text excerpt from the first chapter of a novel, I illustrate a number of different cases of the coding of null arguments.[51]

---

[49] The term 'exception' here could include both cases which are miscoded—that is, cases for which there is no coding criterion from the coding protocol which applies, but the referent is actually in focus—or speech performance errors. The numerical percentage of 'exceptions' in my research is similar to results found in Gundel et al. 1993, who, for Chinese, posit that null forms signal 'in focus', though 1 out of 25 forms is coded as activated (291).

[50] There are also two ambiguous cases which possibly refer to a clausally-introduced entity, but which I have interpreted as referring to a person.

[51] In order to facilitate the discussion of a larger unit of text, I present only the vernacular text line, in which null arguments are identified with the symbol Ø, accompanied by the free translation.

(41) Excerpt from “I Am Guilty, Marian.”

a. *Bizin iurtny iag”asynda **u’ch ak” terek** bar edi.*

b. *Ullular **olag”a** “iurt bulan bir iashly” dep aĭtag”an edi.*

c. *Sho gertimi iada ialg”anmy, bilmeĭmen.*

d. *Amma **sho terekler** iurtumny iakhshysyna da iamana da shag’at.*

e. *Elni k”aĭg”yly giunlerinde **olar**, bukhari bërkler ĭimik, **Ø** bashlaryn sabur kiuĭde o’rge gëterip, **Ø** maslag”at etgenler,*

f. *Shatly giunlerinde busa, **Ø** butak”laryn, **Ø** iaprak”laryn oĭnatyp **Ø** k”uvanchg”a-k”uvanh k”oshg”anlar.*

g. *Ullu tamasha tiugiulmiu…*

h. *K’aravulchular ĭimik, **olar** u’cheviu de iurtnu u’ch iag”ynda$_i$ o’sgen (Ø$_i$ dërtiunchiu, giunbatar ,iag”ynda busa o’zen ag”a edi).*

i. ***Ø** K”ybladag”ysy k”ara giunleni këp gërgen,*

j. *Ø k”ara giĭgen k”atyndaĭ këp kerenler mungaĭg”an,*

k. *Ø Gëziash ornuna chyk” tëkgen.*

l. *Nege tiugiul de ol k”aburlag”a barag”an sok”mak”ny u’stiunde o’sgen edi.*

m. *Iurtda bolg”an chak”y o’liu onu tiubiunden o’tgen.*

n. *Sho zatlag”a Ø talchyg”yp busa iaraĭ, onu bashy tez k”urudu.*

o. *[**Ø** giuntybyshdag”ysy] bash elleni aldynda tok”tap, [Ø] elge ik”lyk” etgen;*

p. *[Ø] k”abu ellerden k”uvang”an (“Bash el balag’ geltirer, k”abu el k”ap tolturar”).*

a. ‘At the edge of our village were three white trees.
b. The elders said that they were the same age as our village.
c. Whether this is true or false, I don’t know.
d. But those trees were witnesses to the good and bad in my village.
e. In the mournful days of wind they, like sheepskin coats, pointed (their) heads up supportively, (they) gave advice.
f. In the happy days (their) branches, (their) leaves played, (they) added joy to joy.
g. It’s not a big surprise…

h. Like guardians, they grew on the three sides of the village (on the fourth, the west side, flowed a river).
i. The southern one (of them) saw many black days,
j. like a woman clothed in black, (it) mourned many times,
k. instead of tears, (it) cried dew.
l. That's because it grew on the path to the graveyard.
m. Whoever died in the village passed by it.
n. Maybe (it) worried over those things, and its top (lit. head) dried up quickly.
o. The eastern one (of them) rested in front of the headwinds, (it) made shade for the people;
p. (it) rejoiced in the west winds ("The head wind will bring disaster, the west wind will fill the sack").'

In the first section of text, sentences (a) – (i), there are six null arguments for which the referent is a set of three trees which are important to a village (all mentions of this referent, whether overt or null forms, are in bold). The first null argument which refers to the trees occurs in sentence (e) as the possessor in the possessive phrase, *Ø bashlaryn* 'their heads'. This null is coreferent with the overt pronoun *olar*, the subject of the previous clause, and, since the referent is mentioned overtly earlier in the same sentence, it is in focus according to criterion 2 of the coding protocol. The second null argument in (e) is the subject of the second clause and is also in focus by criterion 2. Sentence (f) contains three null arguments, two of which are part of the possessive phrases, *Ø butak"laryn* 'their branches' and *Ø iaprak"laryn* 'their leaves'. For the first null argument in (f), the fact that the referent (the three trees) is mentioned as the subject of the previous sentence satisfies the first criterion the status, 'in focus'. For the next two cases, the referent is in focus due to its previous mention in the same sentence. In sentence (i) the null argument is part of the possessive phrase, *Ø k"ybladag"ysy* 'the

southern one of them', and its referent is coded as in focus according to criterion 1 (mentioned as subject of previous sentence).

In sentences (j) – (n), the discourse topic is the southern tree, which is first mentioned as the subject of sentence (i), and all the null arguments refer to this entity. (All references to the southern tree are underlined.) In (j) and (k) each of the null arguments may be coded as in focus because each of their referents is mentioned as the subject of the previous sentence. In (n), the referent of the null argument is mentioned in the previous sentence, but not as the subject; however, since it is also mentioned in the sentence before that, it is in focus by criterion 5.

The coding of the first null argument in sentence (o), which refers to the group of trees, is of particular interest. While sentences (j) – (n) clearly mention one of the trees, there is no overt mention of the group of trees in this local section. However, this plural entity has been mentioned overtly as a set at the beginning of the text as a higher-level topic for this unit of discourse. Each mention of an individual member of the set—that is each mention of an individual tree—is enough to maintain the plural entity as a higher level topic which is "part of the interpretation of the preceding clause (whether it is overtly mentioned there or not)" (criterion 4) and, thus, it is in focus. The second null argument in (o) and the one null argument in (p) both refer to the eastern tree, which is introduced as the subject of sentence (o). (Each of these references to the eastern tree is marked in brackets). The referent of each of these null forms can be coded as being in focus straightforwardly according to criteria 2 and 1, respectively.

4.1.2. Discussion of undetermined cases

All eight cases in this study where the cognitive status of the referent of a null argument remains undetermined are cases of backwards anaphora, or cases in which a null or overt pronominal within a non-finite clause linearly precedes an overt mention of the same referent in the main clause. These are special cases where the referent is not resolved until the whole sentence is processed, and they occur only under restricted syntactic conditions. Cases of backwards anaphora are not considered to be counterexamples to the claim that null arguments require referents in focus, but merely cases in which the referent cannot be coded by the current formulation of the criteria.

Unlike English, where a subordinate clause may precede or follow the main clause, in Kumyk, a subordinate or non-finite clause always precedes the main clause. In a complex clause construction, coreferent subjects can only occur once as an overt argument—either in the main clause or the non-finite clause, but not in both. Previous research on null arguments in Kumyk (Humnick 2008) indicates that, when a null subject occurs in a non-finite adverbial clause, its referent is usually already in focus from previous discourse. The eight undetermined cases in this corpus study represent only cases of null arguments in a non-finite adverbial clause for which the referent cannot be coded as in focus because it cannot be resolved at the point where the null is encountered (as opposed to the majority of null arguments of non-finite clauses for which the referent is in focus due to mention in the previous sentence(s)). Five of these occur in non-finite clauses ending with an *–Ip* converb/gerund, and three occur in non-finite clauses ending with the gerund, *-gEndE*. The text excerpt in (42) illustrates a case of backwards anaphora involving a *–gende* clause in the final sentence of the text. The non-finite

clause, *Meni gjorgende* ‘when (they) saw me’ contains a null argument that is coreferent with the expression *doskany aldy bulan ereturg”an k”yzlar* ‘the girls standing in front of the blackboard’.

(42) *Bir giun akhsham darslardan song, vozhatyĭ k”yz*
One day evening lesson.PL.ABL after lead girl

o’ziuniu klassyn k”aldyrg”an.
self.3POSS.GEN class.3POSS stay.CAUS.PST.PRT
‘One day after the afternoon lessons, the lead girl asked her class to stay late.’

Men de, baiag”y tergeme girdim.
1S and well watch.INF come.in.1S.PST
‘And I, well, came to watch.’

Okhuvchular baĭramg”a ĭyrlar u’ĭrene bolg”an eken.
student.PL holiday.DAT song.PL learn.PR be.PST COP
‘The students had been learning songs for a celebration.’

Ø** Meni gërgende, **doskany aldy bulan ereturg”an k”yzlar
1S.ACC see.PST.PRT.LOC board.GEN front with stand.PST.PRT girl.PL

uialdy.
become.embarrassed.PST
‘When (they) saw me, the girls standing at the blackboard became embarrassed.’
[Marian.026-.029]

Considered to be a special type of cataphora, cases of backwards anaphora require more effort to process and may create particular contextual effects. One possible effect, known as *in medias res*, is an artificially created impression at the beginning of a narrative text that a referent/participant is already at the center of attention. In other cases, as in (42), backwards anaphora is a result of ordering constituents according to principles of coherence. In the last sentence of this example, either the locative clause *meni gërgende* ‘when (they) saw me’ or the overt subject, *doskany aldy bulan ereturg”an*

k"yzlar 'the girls standing at the blackboard', could occur in sentence-initial position; however, the placement of the object, *meni*, at the beginning of the sentence provides greater coherence with the previous event-line sentence, *Men de, baiag"y tergeme girdim* 'And I, well, came to watch.'

4.1.3. Discussion of activated cases

Of the eight cases in the data in which a null argument is associated with a referent coded as at most activated, four represent cases in which the entities are very comparable to referents in focus, but do not fit any particular criterion of the coding protocol, while four appear to be exceptions licensed by a form of parallelism. The four cases in which the referent is comparable to in focus include two referents mentioned as syntactically prominent non-subject elements of the previous sentence and two cases of a plural referent for which at least one member of the plural entity is in focus at the time of mention.

Example (43) displays one case of a null argument for which the referent is a syntactically prominent non-subject argument of the previous sentence, but not one that is specifically mentioned in the coding protocol as bringing a referent into focus.

(43) *Khoziaĭstvoda* ë*lbashchy* *alyshyng"an.*
farm.LOC leader change.PST
'The head of the farm has been replaced.'

Iangy *directoru* *Mag"ammat* *Khamavovnu* *tezden* *tanyĭman.*
new director.ACC Magamet Khamavov.ACC from.long.ago know.PR.1S
'I have known the new director, Magamet Khamavov a long time.'

Ø *G'arakatchy* *ëldash.*
expert friend
'He's an expert.' [Agr.053-055]

In this case, the null argument in the third sentence refers to Magamet Khamavov, mentioned in the second sentence as the direct object, *iangy directoru Mag"ammat Khamavovnu* 'the new director Magamet Khamavov'. The object occurs in sentence initial position and most likely in topic position—in other words, in a place of syntactic prominence equivalent to that of the subject in other types of sentence constructions. It is, thus, understandable that in recent revisions of the coding protocol the criteria are revised to generalize the criterion, "mentioned as the subject" or "mentioned in syntactic focus position" to "mentioned in a position of syntactic prominence" (Gundel 2006).

The criterion of a referent being mentioned in a position of syntactic prominence is also relevant to analysis of the first of two null arguments displayed in the second sentence of (44).

(44) *Ullu alim iaryk"landyryvchu Abusupiyan Akayevni$_i$ eki k"yzy$_j$*
great scholar holy.one Absupian Akaev two girl.3poss

bolg"an.
be.3pst
'The great scholar, the enlightened one, Absupian Akaev, had two daughters.'

Sho eki de k"yzny u'ĭlenmege zamany etishgende
3.3 two EMPH girl.GEN get.married.INF time.3POSS arrive.PST.PRT.LOC

Ø$_i$ *Ø*$_j$ *u'ĭlendirgen.*
get.married.CAUS.3PST
'When the time came for the two girls to get married he gave them in marriage.'
[Editorial.025-.026]

The referents of these two nulls are mentioned as part of a possessive phrase which occurs as the grammatical subject of the first sentence: *ullu alim yaryk"landyryvchu*

Abusupiyan Akayevni eki k"yzy 'the great scholar, the enlightened one, Absupian Akaev's two daughters'. The daughters are mentioned again at the beginning of the second sentence. Based on the coding criteria, the daughters can be coded as in focus based on criteria 1 or 2, but Absupian Akaev is coded as activated because this referent is not mentioned as the syntactic subject in the first sentence; however, it is in a prominent position as the possessor in the possessive noun phrase in subject position.

Another type of case which is very similar to that of a referent in focus is the case of a plural entity for which all members are activated and at least one member of the set is in focus at the time of mention. There are two cases of the use of null arguments with such entities, one of which is illustrated in example (45).

(45) *Bularda geche de Ø$_i$ (prophet) k"almaǐ mishigine Ø$_i$ bara.*
3.1.PL.LOC night EMPH stay.NEG.PR cat.3.POSS.DAT go.PR
'Not staying with them for the night, (he) went to the cat.'

Hosh-besh ***Ø*** *etip sorashyp ashg"a Ø olturalar.*
greeting do.GER greet.GER food.DAT sit.3PL.PR
'Greeting each other, (they) sit down to eat.' [Prophet.0026-27]

Of the four null arguments in this text, the form of interest to this discussion is the one first mentioned in the second sentence (represented in bold). The two null arguments in the first sentence refer to the prophet, who is in focus based on the previous context not shown here. Another important referent which is mentioned in the first sentence is the cat (that is, the cat who turned into a girl earlier in the narrative), mentioned overtly by the form *mishigine* 'to the cat'. The two null arguments in the second sentence both have a plural referent consisting of at least the prophet and the cat-girl, and perhaps other

members of the household (e.g. the cat-girl's husband). According to the methodology of this study, the first mention of a plural referent is the first time all the members of the set are mentioned as part of the same noun phrase, so the individual mentions of the prophet and the cat-girl in the previous sentence are not enough to bring the plural referent into focus at the time of mention of the first null argument of the second sentence (however, the second null argument of the same sentence is in focus by criterion 2). This exceptional use of the null form with an at most activated referent appears to be acceptable due to the fact that at least one of the members of the plural set is in focus at the time of mention.

In four cases in which null arguments are used in reference to at most activated entities, their use appears to be licensed by a form of parallelism in the syntactic structure or event structure. All of these (represented in bold) occur in the excerpt from the text called "The Prophet", which is displayed in (46).

(46) *Excerpt from "The Prophet"*

a. *Erten* *Ø$_i$ (prophet)* *turg"anda* *tilegenler$_j$* *gelip*
morning get.up.PST.PRT.LOC requester.PL come.GER

Ø$_j$ *k"yzny* *elteler.*
girl.ACC lead.3PL.PR
'In the morning, when [the prophet] gets up, the requesters come and take away the girl.'

b. *K"yz* *getip* *gechesinde* *paĭkhammarny$_i$* *u'ĭunde*
girl leave.GER night.3POSS.LOC prophet.GEN home.3POSS.LOC

o'ziuniu$_i$ *k"yzyndan* *da* *isbaĭy* *k"yz* *bola.*
self.3POSS girl.3POSS.ABL and slender girl be.PR
'After the girl leaves, that evening, in the prohet's home, besides his own daughter there appears a slender girl.'

c. *O - Ø$_i$ u'ĭdegi mishigi bolg"an.*
3.2S - home.LOC.MOD cat.3POSS be.PST
'She had been (his) house cat.'

d. *Erten Ø$_k$ gelip geleshegenler$_k$ onu da elte.*[52]
morning come.GER ask.for.in.marriage.PR.PRT.PL 3.2.ACC EMPH lead.PR
'Coming in the morning, the ones asking [for her] in marriage lead her [away], too.'

e. *U'chiunchiu geche Ø$_i$ iti k"yz bola,*
third night dog.3POSS girl become.PR
'The third night (his) dog becomes a girl,…'

f. *erteninde Ø$_l$ gelip o k"yzny da*
morning.3POSS.LOC come. GER 3.2.DET girl.ACC EMPH

Ø$_l$ alyp gete.
take.GER leave.PR
'…in the morning [they]come and take that girl, too.'

g. *Aradan bir - eki giun getip bir geche*
interval.ABL one - two day pass.GER one night

Ø$_i$ esheki k"yz bola.
donkey.3POSS girl become.PR
'After a couple of days pass, one night (his) donkey becomes a girl.'

h. *Erteninde onu da Ø$_m$ elteler.*
morning.3POSS.LOC 3.2S.ACC EMPH lead.3PL.PR
'In the morning, [they]take her, too.' [Prophet.0009-15]

In this excerpt, three sentences contain possessive phrases in which the possessor is elided (underlined in the text): *Ø$_i$ u'ĭdegi mishigi* '(his) house cat', *Ø$_i$ iti* '(his) dog' and *Ø$_i$ esheki* '(his) donkey'. In each case the possessor is the prophet, the main

[52] The null argument in the first clause of this sentence is a case of backwards anaphora and is, thus, listed among the undetermined cases of coding.

participant in the narrative. The null argument representing the possessor in (c) is coded as in focus, as the prophet is mentioned in both sentence (a) and (b) – though not as the subject of (b) – thus criterion 5 applies. In the case of both *Ø$_i$ iti* 'his dog' in (e) and *Ø$_i$ esheki* 'his donkey' (g), the referent of the null arguments, that is, the prophet, is not mentioned in the previous sentence; therefore it is not coded as in focus, but as activated. In the latter two cases I propose that the parallelism in argument structure between the three cases is enough to render the use of a null argument acceptable even though the referent is coded as activated.[53]

Two other null arguments used with activated entities in the same text involve a slightly different form of parallelism. These cases, which occur in sentences (f) and (h), are particularly interesting in the sense that they involved parallelism at the level of discourse structure. In this text there are four groups of groom's representatives (suitors) that come to take away each of the prophet's four daughters, three of which were previously household animals. In (a), the representatives of the first groom are mentioned via the overt form *tilegenler$_j$* in the first clause and a null argument in the second clause, the referent of which is in focus by criterion 2. The representatives of the second groom are mentioned in (d) via a null argument in the first clause and the form, *geleshegenler$_k$* in the second clause. In this case the null argument is considered a case of backwards anaphora (listed as undetermined). By the time the third groom's representatives are mentioned in (f), null forms are used in both the first and second

[53] Alternatively, it could be argued that the referent (the prophet) is in focus because it is a higher-level topic that is part of the interpretation of the whole unit of text represented in the example.

clause, even though the referent of these forms was not mentioned in the preceding sentence (e). Similarly, in the single clause of (h), a null argument is used in reference to the fourth groom's representatives, though no representatives are mentioned in the previous sentence. In these cases, the acceptability of the null forms in (f) and (h) appears to be justified by a type of parallelism I call script repetition, referring to the repetition of a series of events in the narrative which involves parallel sentences with the same argument structure (Humnick 2008:57). The evidence from this example suggests that "script repetition" gives rise to a particular pattern and the expectation that this pattern will be continued. I believe that degree of expectedness or predictability is a factor which contributes to the inherent salience or cognitive status of an entity (cf. Prince's "parallelism principle" (1981:228)).

4.1.4. Higher-level topics

One of the more subjective areas of the coding process is the application of criterion 4 for coding a referent in focus: "It is a higher-level topic that is part of the interpretation of the preceding clause (whether it is overtly mentioned there or not)." While I apply this criterion conservatively in the analysis, it is worthwhile to present a couple of cases where this criterion clearly applies to the coding of a null referent.

In the first of these cases, the use of null forms is associated with a shift in narrative point of view in the Aihalay text, presented in (47). (Lines (a) – (d) of this text are presented in the form of an English translation, while line (e), where the null argument occurs, is presented as glossed text.)

(47) a. ‘For several weeks Aihalay wore her slippers on her days off when there were meetings, gatherings in the club, or when she visited someone.

b. Lo and behold, at that time, the sole started coming off the slippers she wore to work.

c. It was said that the boot maker was going around drinking like evil; to put it briefly, there was no way out except either to stay home and not to go to work or to go to work barefoot.

d. That day there were especially a lot of meetings—the meeting of the clean up commission, the meeting of the town soviet.’

e. *Bulany barysyn_da uzak" zaman sëĭlediler erishdiler,*
3.1.PL.GEN every.one long time say.3PL.PST argue.3PL.PST

u'ĭge busa ***Ø*** *geche sag'at birde kaĭtmag"a tiushdiu.*
home.DAT however night hour one.LOC return.INF descend.3PST
‘At every one of them (people) talked a long time, and argued, and she ended up coming home at one o'clock at night.’ [Aihalay.07-.011]

In lines (a) – (b), the narrative point of view is that of a third-person omniscient narrator. Lines (c) – (e) represent a shift to narration from the point of view of Aihalay’s thoughts, expressing her evaluation of the situation and her only options: “there was no way out except either to stay home and not to go to work or to go to work barefoot (line (c)).” Since Aihalay’s thoughts are being represented, it follows that she is implicitly represented in this line of text, even though she is not overtly mentioned. Moreover, the fact that Aihalay is not mentioned overtly has the dramatic effect of allowing the hearer (reader) to empathize—to put herself in Aihalay’s place. In line (d), the meetings discussed are the meetings Aihalay attended, and the way the meetings are described represents her evaluation. Since Aihalay is part of the interpretation of both sentence (c) and sentence (d), she is arguably a ‘higher-level topic, and the null argument in (e) can be coded as in focus.

Another case in which a referent of a null argument is coded on the basis of being a higher level topic in the previous context occurs in the Agriculture text in a description of how one farm tries to negotiate a deal to use the land of another farm (48).

(48) ‘When we knew that our neighbors' land was empty, we wanted to sow crops for common (profit)," said Soltanyangiyurt GPU's agricultural expert, Abusayit Gamzatov. Where is that? (idiom) [**Soltanyangiyurt's leaders** didn't take it close (didn’t like it). Maybe **they** didn't accept the conditions. How could this be? The seeds would have been from us, and the technology, and we would have tilled and even when the wheat was ripe we would have gathered it.’

Alyng"an *tiushiumnu* ***tenge-teng*** ***etme*** *khyialybyz*
be.taken.PST.PRT crop.ACC equally do.INF thought.1PL.POSS

bar *edi.*]
exist 3.PST.AUX

Ø *Ø*[54] *Razi* *bolmadilar…*
(they) (conditions) agree be.NEG.3PL.PST
‘We planned to divide the harvest evenly. They didn't agree to it...’
[Agr.032-039]

The higher level topic in the unit of text marked with brackets is the Soltanyangiyurt’s leaders’ response to the conditions of the deal. Soltanyangiyurt’s leaders are overtly mentioned several times in the previous context, but not in the sentence preceding the null argument (in bold). In this sentence, however, the Soltanyangiyurt leaders are implicitly part of the semantics of the verb *tenge-teng etme* ‘divide’ by virtue of being the people to the harvest would be divided with.

[54] The set of conditions are mentioned overtly as a plural nominal earlier in this excerpt and the specific details of the conditions are mentioned overtly in the two sentences prior to the occurrence of this null argument; thus it is in focus by criterion 5.

4.1.5. Clausally-introduced entities

Following the assumption of the GH model that events or states denoted by a preceding sentence are brought into focus, one would expect that null arguments could be used to refer to events or states. In the corpus data, however, there is only one clear case in which a null argument refers to event or state introduced by a clause, which is shown in (49).

(49) - *Baliki Ø shartlaryn k"abul_etmegendir.*
- maybe (they) condition.3PL.ACC accept.NEG.PR.PRT.SPEC
'Maybe they didn't accept the conditions.'

- ***Ø*** *Nechik bola?*
- how be.PR
'How could this be?' [Agr.035-.036]

In this example, the null subject of the second sentence refers to a potential state introduced by the previous sentence. Note that the verb is marked with the suffix *–dir*, which is a particle expressing conjecture or speculation (Clinton ms:28).

The lack of other examples of null reference to clausally-introduced entities suggests that nulls may actually be restricted to reference to nominally-introduced entities and that example (49) is an exceptional case. Informal testing suggests that certain exclamations or expressions of propositional attitude such as *nechik bola* 'how could this be' and *anlamaĭman* 'I don't understand' and *bilmedim* 'I didn't know' are acceptable

without overt objects, whether the referent is an event, state, or fact.[55] Such cases may actually be conventionalized expressions that could be considered to be non-referential.

4.2. Reflexives in the corpus

All 35 coded referents of the reflexive pronoun *o'ziu*, are in focus. While the coding of referents of *o'ziu* is straightforward in most cases, one case remains undetermined—a case of backwards anaphora which is discussed in section 4.2.1. The remainder of this section presents data about the referents of reflexives which focuses primarily on the linear distance and argument position of previous mention, as such data is useful in supporting my analysis of *o'ziu* as a pronominal.

As for other features of the distribution of referents, while most referents of *o'ziu* are human, there is one case in which this form is used with a non-human referent (cf. example (57), section 4.2.2.2). *O'ziu* is not, however, used with entities introduced directly or indirectly by full clauses, such as events and facts.

4.2.1. Discussion of undetermined cases

The one referent of the reflexive *o'ziu* for which the coding remains undetermined is a case of backwards anaphora involving a reflexive used within a relative clause, as illustrated in (50).

(50) *Milletleni g'alklany arasynda bulaĭ aĭtyvlar bar:*
nations peoples among 3.1.MOD sayings exist.PR

---

[55] Note that a fact denoted by a previous sentence is assumed to be at most activated by the GH model, hence one would not expect reference with a form which signals 'in focus' to be acceptable.

*"**o'ziuniu**$_i$ milletin g'alkyn siuiup*
self.GEN nation.3POSS.ACC people.3POSS.ACC love.GER

bazharmaǐg"an adamlar$_i$ o'zge milletleni de
know.how.NEG.PRT man.PL other nation.PL.ACC EMPH

g'alklany da siuiup bazharmas".
people.PL.ACC EMPH love.GER know.how.NEG.FUT
'Among the nations and peoples there is a saying like this: "Men who don't know how to love their own nation and people will not know how to love other nations and people.' [Editorial.07]

In this example, the reflexive is coreferent with the matrix subject *adamlar* 'men', but linearly precedes it. (And, since there is no other mention in the previous context, there is no other criterion by which the referent can be in focus.) As in the case of nulls, backwards anaphora with reflexives is restricted to cases in which the form linearly precedes a mention of the same referent in matrix subject position. Though such referents cannot be coded as in focus under the current formulation of the coding protocol, a formulation of the coding protocol which defines the criteria in terms of syntactic hierarchy rather than linear order would result in the coding of these forms as in focus. On the other hand, the fact that such cases are exceptions to generalizations about linear order may result in added contextual effects and, thus, it would be more insightful not to code them conventionally in the same way as other referents in focus.

Another interesting case in the corpus looks like a case of backwards anaphora involving a local possessive reflexive, but the referent of the reflexive is actually coded as in focus by virtue of being a higher level topic (criterion 4). As shown in (51), *o'zleni* (plural, genitive) is coreferent with *khoziaǐstvo* 'the (collective) farm'. The mention of the possessive reflexive linearly precedes the mention of the farm in the subject position

of the clause; however, it is highly likely that the referent of the reflexive is already in focus at the time of that mention, as it is coreferent with the pronominal form *munda* 'here' in the previous sentence, and the farm is the higher-level topic of a larger section of the text from which this excerpt is taken

(51) *Munda ashlyk" tarlavlany maĭdanyn*
3.1.LOC crops field.PL.GEN area.3POSS.ACC

dag"y_da gengleshdirme umut ete.
even.more widen.INF hope do.3PR
'Here, it is hoped that the area of the grain fields will be widened even more.'

O'zleni *besh kombaĭny bulan khoziaĭstvo tiushium*
self.3PL.GEN five combine with farm crop

k"aĭtaryvnu chalt wa k"urumlu kiuĭde o'tgere.
harvest.ACC quickly and organized state.LOC carry.out.3PR
'With their five combines the farm carried out the harvest in a quick and organized manner.' [Agr.016-017]

4.2.2. Referents of reflexives and 'antecedent' types

Though an analysis within the GH model focuses on the cognitive features of the referent of a form, many referents are coded on the basis of how the syntactic features of previous mentions of the referent—features such as argument position and linear distance—contribute to cognitive status. The evaluation of features of previous mention is, to a large degree, parallel to the way in which syntacticians categorize reflexives in terms of their antecedents. In using the term *antecedent* here, I am using a conventional shortcut to refer to the linguistic form which constitutes a mention of the same referent as the form being coded. In this analysis, I divide the tokens of reflexives into two main categories: those which have an antecedent in the same simple clause (18 tokens) versus

those which have a long-distance antecedent outside of the simple clause (17 tokens).[56] Section 4.2.2.1 discusses reflexives with same clause antecedents. Most of these are possessives, but there are also a few cases of non-possessive reflexives with local clause antecedents which are discussed separately. Among cases where a reflexive has an antecedent in the local clause, there is only one case of a form that has a non-subject antecedent (and no other antecedent in the previous clause or sentence). Section 4.2.2.2 presents reflexives with long-distance antecedents. Long-distance reflexivization, by definition, involves forms that do not have a subject antecedent within the local clause. In this corpus most of the long-distance antecedents in this corpus are subjects of the matrix clause (or, in one case, an adjacent non-matrix clause), though there are two cases in which reflexives have a long-distance non-subject antecedent. This section also discusses two cases of long-distance reflexivization that are extra-sentential. Though infrequent, these cases strongly support of the categorization of *o'ziu* as a pronominal rather than an anaphor.

4.2.2.1. Same clause antecedent

The most common use of the form *o'ziu* in the corpus is as a local possessive reflexive, where it is coreferent with the local clause subject (whether expressed as an overt or implicit argument). Nearly half (15 out of 35 coded cases) of the tokens of *o'ziu* are local possessives, most of which are in focus by criterion 2. The example in (52)

---

[56] By using the distinction between antecedents within the same clause as opposed to those outside the clause I follow the type of criteria in the coding protocol (e.g. linear and hierarchical distance) without capturing all of the distinctions between local and non-local domains which are represented within binding theory.

illustrates a local possessive reflexive where *o'zleni* (plural, genitive) has the same referent as *iurtda ing ariv giĭinegen katynlar* 'the most beautifully dressed women in the village,' mentioned previously in the same clause as the subject.

(52) *Iurtda* *ing* *ariv* *giĭinegen* *katynlar*$_i$
village.LOC very pretty be.dressed.PR.PRT wife.PL

o'zleni$_i$ *machilerin* *Aĭkhalaĭg"a* *bichdire;…*
self.3PL.GEN slipper. PL.3.POSS.ACC Aihalay.DAT cut.out.CAUS.PR
'The most beautifully dressed women in the village have their slippers made by Aihalay;…' [Aihalay.003]

There is one case, shown in (53), in which a possessive reflexive has a non-subject antecedent in the same clause (in brackets). Though this case is different in terms of syntactic relations, the coding of the referent of *o'ziuniu*—that is, the prophet—is straightforward, as the prophet is overtly mentioned earlier in the sentence, and, thus, is in focus by criterion 2.

(53) *K"yz* *getip* *[gechesinde* *paĭkhammarny*$_i$ *u'ĭunde*
daughter.3poss leave.GER that.evening prophet.GEN home.3POSS.LOC

o'ziuniu$_i$ *k"yzyndan* *da* *isbaĭy* *k"yz* *bola.]*
self.3S.GEN girl.3POSS.ABL EMPH slender girl be.PR
'After the girl/daughter$_i$ leaves, that evening, in the prophet's$_j$ home, besides his$_j$ own daughter there appears a slender girl$_k$.' [Prophet.0010]

Among reflexives which have an antecedent in the local clause, there are only two that are not possessives. This infrequency is due, in part, to the fact that reflexive patients are most often expressed by a verbal morpheme (which renders the overt mention of the patient unnecessary) rather than a pronoun. One example of a non-

possessive reflexive with a local clause antecedent is illustrated in the second sentence of (54).

(54) *Oramda* *bary_da* *g'alk* *o'ziune* *k"arap* *kiulegendir*
street.LOC every people self.3S.DAT look.GEN laugh.PR.PRT.COP

dep *esine* *gelip* *Aĭkhalaĭ*$_i$ *ĭylamag"a*
COMP mind.3POSS.DAT come.GER Aihalay cry.INF

az *k"alyp* *takhtamekni* *u'stiunde* *olturdu.*
little stay.GER bench.ACC on sit.3S.PST
'Aihalay$_i$ sat down on the bench on the verge of tears, thinking that all the people on the street would look at her$_i$ and laugh.'

Tek *birden_eki* ***o'ziuniu***$_i$ *aldinda* *tashlanyp*
but suddenly self.3S.GEN in.front.of be.thrown.GER

turag"an *machiĭni* *Ø*$_i$ *gëriup* *k"aldi.*
stay.PR.PRT slipper.ACC see.GER stay.3S.PST
'But all of a sudden she saw in front of her a slipper that had been thrown down.' [Aihalay.028-.029]

The referent of the reflexive *o'ziuniu* is Aihalay, who is mentioned overtly in the first sentence of the example, but is also mentioned in the second sentence as a null subject argument.

4.2.2.2. Long-distance antecedent

Of the 17 tokens of *o'ziu* with a long-distance antecedent, the most common type consists of the 7 cases of reflexives which occur in a relative clause and have a matrix clause subject antecedent. In Kumyk, relative clauses consist of a head preceded by a clausal modifier in which the head is represented by a gap or null. For example, in (38), the reflexive occurs in the relative clause *o'ziu tizip getgen* 'which self built', which modifies the head, *zatlag"a* 'things'.

(55) *Ibrag'im*$_i$ ***[o'ziu***$_i$ (*Ø*$_j$) *tizip* *getgen]* *zatlag"a*$_j$ *k"arap*
Abraham self.3S build.GER leave.PST.PRT thing.PL.DAT look.GER

bir *zatlar* *aĭtdi.*
one thing.PL say.3S.PST
'Abraham$_i$, looking at the things$_j$ which (he) himself$_i$ built, said some things.'
[Video1.014] (Example repeated from Chapter 3)

In this example, *o'ziu* refers to the referent of *Ibrag'im* 'Abraham', the matrix clause subject.

In addition to relative clause reflexives with subject antecedents, there are two cases of relative clause reflexives with non-subject antecedents, as illustrated in (56) and (57).[57] Example (56) illustrates a case in which a reflexive within a relative clause has a referent that is mentioned, not as the matrix subject, but, still, as the most prominent argument of the matrix clause.

(56) *[Adamlag"a*$_i$ *[o'ziu*$_i$ (*Ø*$_j$) *siuiegen]* *adamyndan*$_j$
man.PL.DAT self.3S love.PR.PRT man.3.POSS.ACC.ABL

aĭrylmag"a] *bek* *k"yĭyn;*
be.separated.from.INF very difficult
'It is hard for people to be separated from someone they love;…' [Obituary.006]

[57]This is the only type of long-distance non-subject antecedent illustrated in the corpus; however, there are other possible types of long-distance, non-subject antecedents. For example, in the following object-controlled non-finite clause, a reflexive in the non-finite clause can be coreferent with either the object or the subject of the matrix clause.

Murat$_i$ *Arsenge*$_j$ ***[o'ziuniu***$_{i/j}$ *topurag"y* *satmag"a]* *ikhtiĭar* *bergen.*
Murat Arsen.DAT self.3S.GEN land.3POSS sell.INF right give.3S.PST
'Murat$_i$ gave Arsen$_j$ the right to sell his$_{i/j}$ land.'

In this example, *o'ziu* has the same referent as *adamlag"a* 'people', which occurs in the dative case in the matrix clause *adamlag"a adamyndan aĭrylmag"a bek k"yĭyn…* 'It is hard for people to be separated from a man…'[58]

Example (57) is an unusual case in which a reflexive in a noun phrase has as its antecedent a noun phrase embedded within the same noun phrase.

(57) *Adabiiat – inche saniiatny bir tarmag"y.*
literature fine arts.GEN one area.3POSS
'Literature is one area of the fine arts.'

[[Inche saniiatny g'ar tarmag"ny$_i$ *busa [(Ø*$_j$*)* ***o'ziune***$_i$
fine arts.GEN every area.3POSS.GEN however self.3S.DAT

(58) *khas_bolg"an] k"uraly*$_j$*]] bola.*
belong.PST.PRT material.3POSS exist.3S.PR
'Each area of the fine arts (however) has materials which belong to it.' (Lit. 'Fine arts'each area's [(they) belong to it(area)] materials exist.') [Mother Tongue.05]

In (57) the relative clause in which *o'ziune* occurs modifies *k"uraly* 'material' the head of the subject noun phrase. Interestingly, the relative clause (single brackets) occurs in the middle of the extended noun phrase (double brackets) that serves as the syntactic subject. The reflexive within the clausal modifier co-refers not with the subject noun phrase, but with the head of the possessive phrase *inche saniiatny g'ar tarmag"ny* 'each area of the fine arts' that is the genitive complement (possessor) of *o'ziune khas bolg"an k"uraly* 'materials which belong to it'.

---

[58] In Kumyk plural agreement is sometimes omitted in third person; thus in this case it is acceptable for the third person singular form *o'ziu* to be used for the referent of *adamlaxa*, a plural noun.

While reflexives which serve as relative clause arguments are the most common type of long-distance reflexive, there are only two occurrences of long-distance possessive reflexives, one of which is shown in (59).

(59) *[Karimulla$_i$ iashadan sholaǐ tez geter] dep*
Karimulla life.ABL 3.3.ADV early go.3S.FUT COMP

ia ***o'ziuniu***$_i$ *iada bizin onu*$_i$ *iuvuk"laryny*
nor self. 3S.GEN or our 3.2.ACC near.one.PL.3.POSS.GEN

bir de esibizge gelmeǐ edi.
no.one memory.1PL.DAT come.NEG.PR 3S.PST.AUX
'Neither he nor we, his dear ones, none of us thought that Karimulla would leave this life so early.' [Obituary.011]

Interestingly, the antecedent of this reflexive (which occurs in the matrix clause) is the subject of the complement clause which linearly precedes the reflexive rather than the subject of the matrix clause.

The second long-distance possessive reflexive, shown in (60), occurs within semi-direct speech and is coreferent with the subject of the matrix clause. This type of case, which involves logophoricity, will be discussed further in section 4.2.3.

(60) *Ø Ashamag"a tok"tag"an zamanda Aǐkhalaǐ*$_i$ *bary_da*
eat.INF stop.PST.PRT time.LOC Aihalay all

ishleǐgenleni aǐlanyp olardan "ëlda ***o'ziuniu***$_i$
work.PR.PRT.PL.ACC bother.GER 3.2.PL.ABL road.LOC self.3POSS.GEN

machiǐin tapmag"anmyken" dep sorady.
slipper.3POSS.ACC find.NEG.PST.INTER.being COMP ask.3.PST
'When it was time to stop to eat, Aihalay bothered all the workers, asking them, "have you found my slipper on the road?".' [Aihalay.021]

Another type of long-distance reflexive I categorize as an emphatic reflexive. This type of reflexive occurs as the subject of one clause and is coreferent with the subject (either overt or implicit) of another clause in a complex clause construction. Since these reflexives alternate with null arguments in giving a coreferent reading in this type of context, the use of the overt form is interpreted as having an emphatic effect. Note that in the same context, any other type of overt pronoun would be interpreted as not coreferent with an overt subject of the adjoining clause (cf. example (66) for a case in which an emphatic reflexive is coreferent with an overt subject in an adjoining clause).

(61) *Geche* *Ø*$_i$ *(prophet)* *iatg"anda* *o'ziu*$_i$ *etgen* *ters*
night lie.down.PST.PRT.LOC self..3S do.PST.PRT wrong

ishge ***o'ziu***$_i$ *g'ëkiuniup* *tobag"a* *Ø*$_i$ *tiushe* *turup*
work.DAT self.3S regret.GER repentance.DAT descend.PR stay. GER

Ø$_i$ *birdag"y* *namaz* *k"yla.*
one.more prayer do. PR
'[That] night when [he] lay down,[he]regrets the wrong thing that he had done, falls into repentance, [and] prays one more prayer.' [Prophet.0006]

In example (61), from the narrative about the prophet, the prophet is in focus from the previous context and is the referent of all the nulls and reflexives. Note that the first occurrence of the reflexive is not emphatic, but is part of a relative clause and is needed because *ters ish* 'difficult work' is not the subject in the clausal modifier (thus the subject within the modifying clause must be expressed overtly). In this example, the antecedent of the second reflexive is the null subject of the sentence-initial non-finite clause.

In two cases, emphatic reflexives have extra-sentential subject antecedents (where sentence is assumed to be equivalent to finite clause and independent of punctuation). In

the example in (62), the reflexive *o'ziu* occurs in the second finite clause and is coreferent with *o*, the subject of the finite clause represented in the first line. (Both forms refer to Aihalay, the major participant of the text.)

(62) *Geche* o_i *iaman tiushler gëriup chyk"g"an;*
night 3.2 bad dream.PL see.GER exit.3S.PST

nechik tiushler ekenin chi ***o'ziu***$_i$ *de bilmeǐ;*
what.kind dream.PL COP.ACC EMPH self.3S EMPH know.NEG.3S.PR

nege_tiugiul u'iuiu tiubiunden o'tiup ishge
because home.3POSS.GEN from.under pass.by.GER work.DAT

barag"an arbalany k"avg"alary onu$_i$ *birden – eki uiatg"an…*
go.PR.PRT carriage.PL.GEN noise.PL.3POSS 3.2.ACC suddenly wake.3S.PST
'That night she$_i$ had bad dreams; how bad, (she) herself$_i$ didn't even know; that's because the noise of carriages going to work passing by underneath (her) room woke her$_i$ suddenly…' [Aihalay.017]

The reflexive in (62) is emphatic, both in the sense that a null argument would have been sufficient in this context and in the sense that the form *o'ziu* is followed by the emphatic particle, *de*.

In (63) the reflexive *o'zler* occurs in the final sentence of the example, and the referent is the narrator's town-mates who are in charge of a particular farm.

(63) (Translation of preceding context):

'Truly, my town-mates didn't agree. You don't want to say it. But when you tell about the beginning, how can you not report the end? The main reason, why should we give (our) land to outsiders? In other words, let the grass grow there, but we don't give it to anyone else.' [Agr.041-.045]

Ø$_i$ *Anglashylmaǐ.*
be.understand.PASS.NEG.PR
'They can't be understood.'

O'zler$_i$ *de k"ollamaǐ,* $Ø_i$ *bazharag"anlag"a da bermeǐ.*
self.3PL EMPH use.NEG.PR know.how.PR.PRT.PL.DAT EMPH give.NEG.PR
'They are not using it themselves, and they don't give it to those who know how.'
[Agr.046-.047]

In this case the referent is in focus, and the emphatic use of the reflexive form allows for the contrastive focus between the people who are not using the land and those who would like to use it. Such pragmatic or contextual effects will be discussed further in the next section.

4.2.3. Contextual effects of reflexives

In a number of syntactic studies, the long-distance use of reflexives is associated with properties such as contrast, emphatic usage, and logophoricity (Baker 1995, Zribi-Hertz 1995, Cole et al. 2001). In classifications of long-distance reflexives, moreover, Cole et al. point out that long-distance bound anaphors are often restricted to specific types of discourse contexts, such as logophoric constructions (2001:xix-xx). While I propose that the Kumyk form *o'ziu* is a pronominal which is restricted to referents in focus but not restricted to certain types of discourse contexts, the corpus analysis does provide evidence that discourse properties or contextual effects such as logophoricity and imposed salience are often associated with this form. On the other hand, there are cases, such as the long-distance use of reflexives in relative clauses, where reflexives merely represent obligatory arguments and are free of added contextual effects. In this analysis, I propose that in any case where a reflexive can alternate with a null argument (both syntactically and because the referent is in focus) the use of a reflexive has added contextual effects. First I discuss examples where the contrast between the reflexive

form *o'ziu* and other possible forms (e.g. nulls or demonstrative pronouns) emphasizes a contrast between particular entities referred to in a text. Reflexives can also be used emphatically in cases which express some type of unexpected element. Finally, previous research proposes a correlation between the long-distance use of reflexives and logophoric contexts, where logophoricity can be defined as the property of referring expression which signifies coreference with the speaker/thinker—for example, in a complex clause involving a verb of speaking or cognitive process (Cole et al. 2001).[59] In Kumyk, though not all long-distance reflexives are associated with logophoric contexts, corpus evidence suggests that logophoric constructions require the use of reflexive pronouns.

In the corpus data there are several notable cases where a reflexive possessive is repeatedly used in a context where a null possessive pronoun would be adequate. Note that, since the person and number of the possessor is recoverable from the possessive suffix on the head noun in a possessive phrase, null possessives are adequate in many cases. (In fact, there are about four times as many null possessives as reflexive possessives in the corpus.) In a number of such cases, the form *o'ziu* is used to express a contrast between two or more entities. For example, in the Aihalay text, Aihalay, a woman who makes a living by embroidering slippers, has two pairs of slippers: her everyday pair, which she wears to work, and a new intricately embroidered pair which

---

[59] Cole et al. 2001 give credit to Hagège 1974 for coining the term *logophoric*.

she only wears on special occasions. Once both pairs of slippers have been introduced in the text, continuing references to the everyday pair do not contain overt possessive reflexives, but a number of references to the holiday slippers contain the possessive marker. This contrast can be seen in example (64).

(64) *Maĭny birinde Aĭkhalaĭ bir giĭim machiĭge material bulan savg"atlang"an. Aĭkhalaĭ sho machiĭleni bek aryv etip, tiurliu omuzlar va k"yzyl sak"tiiandan besh miuiushliu iuldyz da tutup tikgen edi. Machiĭler deseng machiĭler bolg"anmu dag"y, k"arap turmag"a bir zatlar, o'zliugiunden iuriup getezhek. Bir-eki zhumalar Aĭkhalaĭ* ***o'ziuniu machiĭlerin*** *vyxodnoĭ giunlerde sobraniialar bolg"an geziklerde, klubdag"y zhĭynlag"a iada k"onak"laĭ barag"anda giĭip iuriudiu. Muna sho vak"tide* ***onu ishge giĭegen machiĭlerini*** *ultany aĭrylyp k"aldy.*

'On the first of May Aihalay rewarded herself with material to make slippers. Aihalay made these slippers very beautifully, embroidering them with various designs and sewing on a red leather five-pointed star. If you can say it about slippers, it's as if they could walk by themselves. For several weeks Aihalay wore **her own (*o'ziu*) slippers** on her days off when there were meetings, gatherings in the club, or when she visited someone. Lo and behold, at that time, the sole started coming off **her (*onu*) slippers that she wore to work**.' [Aihalay.004-8]

In many of the contexts of the Aihalay text in which reflexive possessives are used, a null argument would have been unambiguously interpreted, as in example (65).

(65) *Ø Iukhlamag"a iatag"anda Aĭkhalaĭ oĭlashdy*
sleep.INF lie.down.PR.PRT.LOC Aihalay think.3PST

wa erten ishge ***o'ziuniu/Ø*** *baĭram machileri*
and morning work.DAT self.3S.GEN holiday slipper.PL.3POSS

bulan barazhak"g"a Ø tok"tashdi.
with go.FUT.DAT decide.3PST
'When she lay down to sleep, Aihalay thought, and she decided that she would go [to work] in the morning with her (very own) special occasion slippers.'
[Aihalay.016]

In such cases, the use of the reflexive with the special slippers in contrast with the lack of use for the everyday slippers, however, produces the effect of conveying the importance of the special slippers to Aihalay and/or to the narrative as a whole. One could also claim that the use of this form, not only conveys the importance of the referent, but also requires the hearer/reader to view this referent with empathy.

In the narrative called "The Prophet's Daughters", the prophet has one biological daughter and three household pets which become daughters at a later point in the story. The reflexive is usually used in reference to the biological daughter to emphasize the contrast between his real daughter and the ones that were formerly household animals. The use of the reflexive in possessive expressions referring to his own daughter is notable, as overt possessives are rarely used in conjunction with kinship terms.

Emphatic subject reflexives are also used for contrast and to indicate unexpectedness. In example (63) in the previous section, the reflexive subject is used to contrast the owners of one farm who are not using their land with the owners of the neighboring farm who have asked to use the first farm's land. The contrastive focus in this utterance conveys the narrator's feelings of surprise and lack of understanding of the first farm owners' decision not to lease their land.

In example (66), a reflexive in the expression *o'zium de bilmeǐ* 'not knowing myself' is used to convey the unexpectedness that the speaker would not know what he himself is doing.

(66) ***O'zium***$_i$ *de* *bilmeǐ,* *men*$_i$ *bir* *k"yzny*
self.1S EMPH know.NEG.3S.PR 1S one girl's

betine *tiklenip* *turg"anman.*
face.DAT stare.GER remain.1S.PST
'Not even realizing it myself, I kept staring at one girl's face.' [Marian.037]

In another case, found in a newspaper obituary, a reflexive is used to convey the unexpectedness of the deceased's untimely death, as displayed in (67).

(67) *Bu* *ǐylny* *avgust* *aǐyny* *31-nde* *iuvuk"ubuz*
3.1 year.GEN August month.3POSS.GEN 31th near.one.1PL.POSS

Karimulla *Dag'irovg"a*$_i$ *45 ǐyl* *bitme* *gerek* *edi.*
Karimulla Dagirovga 45 year finish.INF must 3.PST.AUX
'On August 31st of this year, our dear friend Karimulla Dagirovkha should have been 45 years old.'

Tek *ol*$_i$ ***o'ziuniu***$_i$ *sho* *chak"yna* *etishmedi.*
but 3.2 self. 3S.GEN 3.3 amount.DAT arrive.INF.3PST
'But he didn't reach this age (of his).' [Obituary.001-.002]

The form *o'ziuniu*, which refers to Karimulla Dagirov, is used as a genitive modifier of *sho chak"yna* 'this age'. Since, grammatically, *sho chak"yna* 'this age' would have been sufficient, the addition of the reflexive modifier is interpreted as having an emphatic function—probably associated with the shock of his untimely death.

Finally, in many languages, a particular set of pronouns is used for logophoric reference which may be distinct from other types of third person reference. For example,

in Kumyk, in a sentence which translates, 'Zuchra thought that everyone would laugh at her,' as shown in (68) (based on the corpus example shown in (54)), the use of the form *o'ziu* in (a) is interpreted as coreferent with Zuchra, while the use of the form *og"ar* in (b) is interpreted as not coreferent with Zuchra.

(68) a. *Zukhrag"a*$_i$ *bary_da_g'alk* ***o'ziune***$_i$ *k"arap* *kiulegendir*
Zuchra everyone self.3S.DAT look.GEN laugh.PR.PRT.COP

dep *esine* *geldi.*
COMP mind.3POSS.DAT come.PST
'Zuchra thought that everyone would laugh at her (Zuchra).'

b. *Zukhrag"a*$_i$ *bary_da_g'alk* ***og"ar***$_{j/*i}$ *k"arap* *kiulegendir*
Zuchra everyone 3S.DAT look.GEN laugh.PR.PRT.SPEC

dep *esine* *geldi.*
COMP mind.3POSS.DAT come.PST
'Zuchra thought that everyone would laugh at her (not Zuchra).'

The corpus data includes two cases of the logophoric use of reflexives similar to example (68), one of which occurs in an indirect speech clause, and the other of which occurs within semi-direct speech (cf. examples (54) and (60) respectively). In these cases, the choice of a reflexive as opposed to an overt non-reflexive pronoun seems to be related to coreferentiality restrictions specific to logophoric contexts and, therefore, does not have contextual effects beyond that of providing the information needed to disambiguate the referent.

4.3. The pronoun bu in the corpus

The corpus analysis of *bu* is based on 33 tokens, 31 of which have referents which could be coded for cognitive status. The two undetermined cases are both cases of

cataphoric reference. Of the 31 coded referents, 30 are coded as in focus, with the remaining referent being coded as at most activated. The one referent coded as activated also happens to be the only coded referent which is introduced by a verb phrase (cf. section 4.3.4). With only one activated referent, representing 3.2 % of its referents, the form *bu* is analyzed as signaling in focus.

Some other noteworthy characteristics of the corpus distribution of *bu* are 1) that its occurrence is somewhat higher in oral versus written texts (14 cases in 27 pages of oral text versus 19 cases in 72 pages of written text , 2) that the majority of referents of *bu* are humans, 3) that the majority but not an overwhelming percentage of referents of *bu* are previously mentioned as syntactic subjects, and 4) that the use of *bu* with clausally-introduced entities is rare and limited to cases of cataphoric reference. The frequency of *bu* in oral versus written texts is an interesting phenomenon, but one which does not fall within the direct scope of this study. The frequency of human versus non-human referents of *bu* most likely reflects the frequency of humans versus other types of entities as participants in texts in general rather than any type of restriction or tendency related to the use of the specific form, *bu*. In support of this conclusion, the list of non-human referents of *bu* includes a variety of types of inanimate entities: villages, farm, tree, meetings, time, speech, slander, hypocrisy, thus this factor is not discussed further in this study. The data on *bu* shows that a third of referents were not previously mentioned as subjects, confirming that subjecthood of the antecedent is not the primary factor affecting the distribution of this form. Of the four generalizations mentioned above, the last seems to be the most significant. The form *bu* refers only three times to an entity not

introduced by a nominal. One of these cases is an activity introduced by a verb phrase that is part of a clausal argument. The other two are cases of cataphoric reference to speech such as the expression, *bulaĭ aĭtyvlar* 'this kind of saying' with the particular feature that the speech itself is introduced by a noun phrase. The apparent restrictiveness of *bu* in relation to entities not introduced by nominals merits further discussion.

4.3.1. Coding the referents of *bu*

The coding of the majority of referents of *bu* is straightforward, and the primary purpose of this section is to illustrate the coding and to provide examples of animate and inanimate referents. In (69), an excerpt from the beginning of an oral narrative, the referent of *munu* (accusative/genitive form of *bu*) is a human named Layla.

(69) ***Laĭla degen kumuk k"yz*** *ulan tapdi.*
Layla say.PST.PRT Kumyk girl son give.birth.3PST
'A Kumyk girl called Layla gave birth to a son.'

Bir zhuma getip ***o*** *bolnitsadan chyk"dy.*
one week leave.GER 3.2 hospital.ABL exit.3PST
'One week having passed, she left the hospital.'

Bolnitsany aldina ***munu*** *k"arshylama*
hospital.GEN before.2POSS.DAT 3.1.ACC meet.INF

munu *absunlary k"aĭynk"yzlary geldi.*
3.1S.GEN sister.in.law.PL.3POSS sister.in.law. PL.3POSS come.3PST
'Her sisters-in-law (brother's wives) and sisters-in-law(husband's sisters) came to meet her in front of the hospital.' [Birth1.001-3]

Previous references to this participant (in bold) occur in the first and second sentences. At the time of the first occurrence of *munu* in the third sentence, the referent is coded as in focus (either by criterion 1, as it was mentioned as the subject of the previous sentence or by criterion 5, as it is mentioned in the two preceding clauses). At the time of the

second use of *munu*, the referent is clearly still in focus by criteria 1 and 5 or also by criterion 2 (mentioned earlier in the same sentence).

Example (70) illustrates reference to an inanimate entity—in this case, a character trait.

(70) *Adamny aldynda bir tiurliu, Ø artynda bir*
man.GEN front.3.POSS.LOC one various back.3.POSS.LOC one

tiurliu sëĭleĭgen gishige - munafik" dep aĭtyla.
various say.PR.PRT person.DAT faith. betrayer say.PRT be.called.PR
'A person who says one thing in front of a person and another behind is said to be a betrayer of faith.'

Munafik"leni g'ech_kimni iag"ynda k"yĭmaty ikhtibary
faith .betrayer.PL noone on.the.part.of worth.3POSS trust.3POSS

bolmas.
be.NEG.FUT
'Betrayers of faith don't have worth, trust in anyone's eyes.'

G'ak"ylly balalar ***bulaĭ*** *iaman k"ylyk"dan bek*
wise young.ones.PL 3.1.MOD bad trait.ABL much

sak"lanmag"a tiĭishli.
keep.INF suitable
'Wise children deserve to be protected from a bad character trait like this.'
[Character.01-.03]

In this example, the character trait is introduced in the first sentence by the description that is given the label, *munafik"*, then *munafik"* is repeated in the second sentence. By the time of its mention in the third sentence, therefore, the referent is in focus by criterion 5.

4.3.2. Discussion of undetermined/cataphora cases

The two uncoded referents of *bu* are cases of cataphora. Like backwards anaphora, cataphoric references are cases in which the referent of a pronoun cannot be resolved because there is no mention of the referent in the context (linearly) preceding the occurrence of the pronoun, but, rather, the referent of the pronoun is overtly mentioned in the context following the pronoun. Cataphora is less specific than backwards anaphora, which is restricted to particular syntactic contexts (cf. section 4.1.2). In (71) the expression *bulaǐ aǐtyvlar* 'this kind of saying' refers to the content of the quoted speech which follows.

(71) *Milletleni g'alklany arasynda* ***bulaǐ*** *aǐtyvlar bar:*
nations peoples among 3.1.MOD sayings exist.PR

"o'ziuniu milletin g'alkyn siuiup
self.GEN nation.3POSS.ACC people.3POSS.ACC love.GER

bazharmaǐg"an adamlar o'zge milletleni de
know.how.NEG.PRT man.PL other nation.PL.ACC EMPH

g'alklany da siuiup bazharmas".
people.PL.ACC EMPH love.GER know.how.NEG.FUT

'Among the nations and peoples there is a saying like this: "Men who don't know how to love their own nation and people will not know how to love other nations and people.' [Editorial.07 (also shown as ex. (50)]

It is interesting that both cases of cataphora involve expressions which introduce speech with a noun phrase that refers to a speech act (*bulaǐ aǐtylar* 'sayings like this' and *bulaǐ zhawap* 'an answer like this'). The contextual use of cataphora, particularly to introduce speech is described by Levinsohn as a means of 'highlighting' or imposing prominence

(2008:82). Interestingly, in this corpus, *bu* is the only overt pronoun that is used cataphorically.

4.3.3. Discussion of examples that were difficult to code

In one particular text, several tokens of the pronoun *bu* are difficult to code because of contextual factors related to the method of elicitation. The elicitation of this text involved a process of showing the speaker a series of pictures in a wordless picture book. The speaker narrated the story while looking at each picture in turn, thus the referents were within the eye gaze of the speaker and hearer whenever they were portrayed in that particular picture (or page of pictures). While the coding criteria for an activated referent refer to eye gaze via criterion 2, "It is something in the immediate spatio-temporal context that is activated by means of a simultaneous gesture or eye gaze," this contextual factor is only mentioned as a means of activating a referent, and the role of eye gaze in continuing reference to an activated participant is not discussed in the coding protocol. According to Jeanette Gundel (personal communication), however, a picture of an activated referent can be considered the equivalent of an overt mention of the referent (or an implicit mention of a higher-level topic), and I therefore choose to code entities as in focus if they appear in more than one contiguous picture. Example (72) illustrates two tokens of *bu* which can be explained in connection with pictures which are interpreted as mentions of a referent.

(72)	*Ø*	*suratlag"a*	*k"arag"anda*	***bu***	*iangy*	*ĭylny*	*aldyndag"y*
		picture.PL.DAT	look.PR.PRT.LOC	3.1	new	year.GEN	in.front.of.MOD
	giunler	*Isany*	*tuvg"an*	*giuniu*	*bulan*	*baĭlavlu*	
	day.PL	Jesus.GEN	be.born.PST.PRT	day.3POSS	with	connected.ADJ	

baĭramlany vak"tisi.
holiday.PL.GEN time.3POSS
'Looking at the pictures, this is the time before New Year's, the time of the holiday connected with Jesus' birth (Christmas).'
[in the picture the woman is walking on a street decorated for Christmas]

Adamlar baĭramlag"a g'azirlik gëriup
man.PL holiday.PL.DAT ready.NOM see.GER

aĭlanag"any gëriune.
go.around.PR.PRT.3POSS appear.3PR
'People appear to be going around as if getting ready for the holidays.'
[in the picture the woman is on the side of the street playing the accordian while people are passing by with Christmas trees and packages]

Song ***bu*** *tiukenge gire.*
then 3.1 store.DAT enter.3PR
'Later this one (she) goes into a store.' [Wordless Book.021-.022]
[in the picture the woman is going into a store]

In this narrative, an old woman is the VIP participant, and in the series of pictures with which this text was elicited, the woman appears in all but one picture (the exception is a landscape picture early in the story—though there are pictures of the woman on the same page). The first token of *bu* in example (72) occurs in the first sentence, and its referent is not the woman, but, rather, the picture itself. Since the picture is overtly mentioned earlier in the sentence, this referent can be coded as in focus without reference to the fact that the picture is within the eye gaze of the speaker and hearer. For the second token of *bu*, however, it is necessary to refer to the role of the picture as a mention of the referent. The second token of *bu* refers to the woman who is the VIP participant. She is not overtly mentioned in the previous three sentences, but she is within the eye gaze of speaker and hearer via the pictures and, therefore, coded as in focus.

4.3.4. An activated entity not introduced by a nominal

The form *bu* only has one instance in the corpus where the referent is coded as at most activated, and this case also happens to be the only case where *bu* refers to an entity introduced previously by a syntactic unit other than a nominal. In the case shown in (73), the referent of *bu* is *ëk" zatny sëĭlemekge* 'saying something that does not exist', an activity that is introduced by a verb phrase within a nominal clause in the dative case.

(73) *G'ibat degen iaman zat bolsa da*
slander say.PST.PTCP bad thing be.COND EMPH

Ø gishini artyndan bar zatny sëĭlemekdir.
person.GEN back.3POSS.ABL exist thing say.INF.COP
'Although slander is bad, it is to say behind someone's back something that exists.'

Ëk" zatny sëĭlemekge – bug'tan dep aĭtyla.
exist.NEG thing.ACC say.INF.DAT – defamation COMP be.said.3PST
'Saying something that does not exist is called defamation.' [Literally, '(The word) *bug'tan* is said to [saying something that does not exist.']

Bu *busa g'ibatdan da iaman.*
this however slander.ABL EMPH bad
'This is worse than slander.' [Character.015-.017]

While activities introduced by verb phrases are treated as in focus, I apply this only in cases where the activity is introduced by the finite verb phrase of the matrix clause or as a subject complement of the matrix clause. Since, in this case, the activity is introduced by a verb phrase embedded in a non-subject nominal clause, I code the entity as activated; however, it could potentially be in focus based on the fact that it is, semantically, the most prominent argument of the clause. Considering that this is the only referent of *bu* not introduced by a noun phrase, it is significant that the verb phrase which introduces the

referent involves a form of nominalization, which is evident from the non-finite verb form and the dative case suffix in *sëĭlemekge* ('say' + participial suffix + dative suffix). This is, perhaps, an indication that the use of *bu* with clausally-introduced entities (or entities introduced by a verb phrase) is restricted to cases in which there is some degree of nominalization.[60]

4.3.5. Contextual effects associated with *bu*

Based on the corpus analysis, a number of interesting contextual effects are potentially associated with the form *bu*. Section 4.3.2 above discusses its cataphoric use as a highlighting device. Other contextual effects associated with *bu* include signaling an established topic at a point of topic switch and use in elaboration.

In many cases in the corpus, the use of *bu* correlates with an established topic, or the thematic referent for a unit of text, and a number of these involve continuing reference to a previous established topic (subject) in cases of topic shift (subject shift). As shown in a number of corpus examples, a referent (in focus) mentioned in subject position (as topic) in one sentence is likely to be mentioned as a null argument if it continues to be mentioned in subject position, but not likely to be mentioned as a null argument if it occurs in another argument position in a sentence in which a new subject (topic) is introduced. In such situations, the pronoun *bu* seems to be the preferred form of

---

[60] In the two uncoded cases of cataphora, one of which is displayed as example (71), *bu* refers to a speech act introduced by a following clause; however, in each of these cases, the speech act itself is first introduced by a nominal-- *bulaĭ aĭtyvlar bar* 'there is a saying like this' and *bulaĭ zhawap bergen* 'gave an answer like this'. This fact can be interpreted as supporting the idea that the use of *bu* with clausally-introduced entities requires some degree of nominalization.

overt reference. In GH model terms, *bu* is often used to refer to an established entity in focus in an utterance in which a new entity is brought into focus.

The text "I Am Guilty Marian" provides a clear example of the contrasting use of *bu* in cases of topic shift as opposed to the use of nulls for continuing topics. In this text, the author writes about three trees which stand guard over the three sides of the village. Each tree, in turn, is the topic of a paragraph which gives further details about the nature of that tree. In the paragraphs describing the first two trees, displayed as example (41) in Section 4.1.1, overt noun phrases are used in initial reference to the trees, while null arguments are used for continuing subject reference. The text excerpt about the third tree, displayed as example (74), likewise, begins with an initial noun phrase reference *tëbendegisi* 'the lower one' (mentioned as subject of the matrix clause), followed by three null subject references in the next three matrix clauses.

(74) *Tëbendegisi busa iurtg"a iakhshylyk" bulan giregenlege k"ychak" iaĭyp bash ekgen,*
'The lower one, however, lowers its head, opening its arms to those entering the village with good intentions;

Ø namart khyial bulan gelegenlege k"ash tiugiul k"opaĭg"an.
'to those entering with evil thoughts, doesn't blink an eye.'

Ø iurtdan getegen ulanlag"a iakhshy ël etgen,
'To the young men leaving the village, wishes good travels;'

Ø[61] *gëriunmeĭgen bolg"uncha alg"ysh aĭtyp, Ø artyndan k"arag"an.*
'until (they) disappear, saying praises, looking after them.'

[61] This is a reference to the young men leaving the village, not the tree.

Saparg”a chyg”ag”anlar ***bug”ar*** *savbol etgen,*
‘Those leaving on a journey said goodbye to it,’

k”aĭtag”anlar Ø salam bergen, sorashg”an.
‘those returning greeted (it).’

Iurtdan getegende iurekge syĭynmaĭg”an k”aĭg”y, ichim tolg”an dert bulan men de ***munu*** *tiubiune geldim.*
‘When I left the village, in my distress, my heart filled with sorrow, I also came to the foot of it.’

In the fifth finite clause, however, a new subject *saparg”a chyg”ag”anlar* ‘departing ones’ is introduced, while the tree continues its status as a referent in focus. In this sentence, the form *bug”ar* refers to the tree. In the following finite clause, another new subject *k”aĭtag”anlar* ‘returning ones’ is introduced, but a null argument is used for the tree because it is mentioned in the same argument position as the previous clause. In the next sentence another new subject is introduced—*men*, referring to the first person narrator—and the tree is again referred to with the genitive form of *bu* (*munu*)—probably because it occurs in a different non-subject argument position. This line is particularly interesting, as the first person narrator of the novel is introduced in connection with his coming to the foot of this tree at the time of his departure from the village. One may assume, therefore, that the third tree has particular salience for the author and the narrative, and that at least the second instance of *bu* (if not both) could also be interpreted as a means of imposing higher relative salience on the third tree compared to the other trees mentioned in the previous context.

Another case where *bu* appears to be associated with topic shift is illustrated in example (69) above. In this case, a referent is mentioned by the overt pronoun *o* in the

subject position of one clause, then by the accusative form of *bu* (*munu*) in a non-subject position in the following clause. In a similar case shown in (75), the topic of the first clause, that is, the girl that is the referent of *k"yzny*, is not the syntactic subject, but is in an important syntactic position as an animate possessor (Poesio and Nissim 2001).

(75) *K"yzny* *erge* *berme* *iaraĭg"an* *vak"tisi* *gelip*
girl.ACC husband.DAT give.INF possible.PST.PRT time.3.POSS come.GER

bug"ar *gelechi* *gele.*
3.1.DAT one.who.requests.engagement come.PR
'When it comes time for the daughter to be given in marriage, a person requesting engagement comes to her.' [Prophet.0002]

In the second clause of this sentence, the subject is *gelechi* 'suitor', and the girl is mentioned by *bug"ar*, a dative pronoun. This example is particularly interesting, as the topic shift occurs between an adverbial clause and matrix clause of the same sentence.

There are also cases where the use of *bu* is associated with a more generalized form of topic shift, as in the various changes of location in "The Prophet's Daughters". Each animal/daughter's home is a local higher level topic, and a shift of location usually takes the form of an utterance in which *bu* is used to refer to the location which has been the center of attention and either another form of demonstrative or a noun phrase is used to refer to the new location. Example (76) illustrates the use of *bu* in a context with a change in location. (The text associated with the first location is presented as an English translation, while the text containing the actual pronominal forms is fully presented.)

(76) '(He) begins by going to the donkey. In the evening, having eaten and taken drink, beginning the conversation about this and that, the prophet asks his son-in-law about his life situation.

‘Son-in-law: “Everything’s okay, but one thing is bad – enough time has passed to teach an animal. Whether you say something to her or not, it’s the same; very peaceful in attitude towards work.” “A donkey will not know how to work,” the prophet thought.’

Geche Ø ***bularda*** *k"onak" da bolup erten*
night 3.1.PL.LOC guest thus be.GER morning

itine Ø bara.
dog.3POSS.DAT go.3PR
‘Staying the night as a guest at their place, in the morning (he) goes to the dog.’

Shunda *da sholaĭ gech bolg”uncha*
3.4.LOC EMPH 3.3.ADV late become.before.GER

iashav g'aldan Ø Ø sorashyp…
life situation.ABL greet. GER
‘At that place, thus (as before), before it got late, (he) asks about the situation of life, …’ [Prophet .0017-23]

In the final part of this text, the location that has been the topic of this section, that is, the donkey-daughter’s home, is referred to with the pronoun *bularda*, while the new location, the dog-daughter’s home is referred to with the pronoun *shunda*.

Like the reflexive, the use of the pronoun *bu* is sometimes associated with emphatic prominence—in particular emphatic prominence that is associated with unexpectedness. In Kumyk, emphatic prominence is often marked by the particle *de/da*. In some cases, such as the one illustrated in example (77), the use of the form *bularda* (plural locative form) combines with the emphatic particle *da* to express unexpectedness or surprise.

(77) *Tiurliu-tiurliu milletler iashaĭg”an shag”arlarda iurtlarda*
different.INT nation.PL live.PR.PRT city.PL.LOC village.PL.LOC

rus til aslu erni tutma bashlag”an, ***sholardag”y***
Russian tongue important place.ACC take.INF begin.PST 3PL.LOC.ADJ

o'siup *gelegen* *naslu* *ana* *tilin* *iarty-iurtu*
grow.up.GER come.PR.PRT generation mother tongue.ACC mish-mash

iada *birdok"da* *bilmeĭgen* *bolup* *bara.*
or not.at.all know.NEG.PST.PRT be.GER go.3PR
'In cities and villages in which many different nations live, the Russian language has begun to take an important place; the generation growing up in **3PL** has begun to know the mother tongue poorly or not at all.'

Iangyz *bir* *milletni* *vakilleri* *iashaĭg"an* *iurtlarda* *da*
only one nation.GEN member.PL.3POSS live.PR.PRT village.PL.LOC EMPH

g'al *onchak"y* *maktardaĭ* *tiugiul,* *nege_ tiugiul* *de* ***bularda*** *da*
state so.much better NEG because EMPH 3.PL.LOC EMPH

rus *tilni* *etegen* *ta"siri* *bashg"a* *az* *tiugiul.*
Russian tongue.GEN do.PR.PRT influence other little not
'The state in villages in which members of only one nation live is not much better, because even in these, the influence of the Russian language is not little.'
[Mother Tongue.012-.013]

In the case of this example, the speaker assumes that the hearer/reader expects Russian to be used as a lingua franca in villages where more than one local language is spoken, but expects minimal use of Russian in villages that are predominantly Kumyk. The use of *bu* with the emphatic particle to refer to villages that are predominantly Kumyk coincides with the expression of a fact that is counter to expectations: that the influence of Russian is significant in villages that are predominately Kumyk.

Finally, another function which seems to be associated primarily with the form *bu* is elaboration. In elaboration, the speaker is giving more detailed information about the topic rather than actually adding information/continuing the narrative. One example is illustrated in (78), where the pronoun *bu* refers to the people gathered at the celebration

and introduces further details about the actual categories of relatives who come to such celebrations.

(78) *Baĭramg"a* *zhyĭylg"an* *g'alk* *këbiusiu* *g'alda*
feast.DAT be.gathered.PST.PRT people most situation.LOC

bu - *k"atyngishiler,* *k"yzny* *kardashlary,* *ulanny* *kardashlary,*
3.1 woman.PL girl.GEN relative.PL.3POSS son.GEN relative.PL.3POSS

bular *bari_da* *iashg"a* *uzak"* *o'miur*
3.1.PL all child.DAT distant lifetime

nasipli *yashav* *ërаĭ.*
fortunate life wish.3PR
'The people gathered for the celebration, usually these are women, the girl's relatives, the boy's relatives, all of these wish the child 'long life', 'happy life'.' [Birth1.018]

As in the case of topic shift, the use of *bu* seems to mark unexpectedness. Rather than the unexpectedness of participant change, the unexpectedness in this case is the fact that the narrative is being interrupted by the addition of more specific information about the topic. Calling such cases 'renewal' of the topic, Levinsohn primarily associates this with the introduction of background material (2008:45).[62]

4.4. The pronoun o in the corpus

O is the most commonly occurring overt pronoun and occurs 90 times in the corpus. The referents of *o* include both animate and inanimate entities, as well as abstract

[62] Though I credit Levinsohn for the original idea, I claim that elaboration is a subset of background information and that the form *bu* is specifically associated with this subset, while other overt pronouns are also associated with the introduction of other types of background information.

entities such as times and facts. As a matter of fact, approximately 25% of referents are inanimate entities.

Out of 90 tokens in the corpus, the referents of all but one of these forms could be coded for cognitive status. The one referent which is undetermined is an indirectly constructed entity and is discussed in section 4.4.2. Of the 89 referents coded, 64 are in focus, 24 are at most activated, and one is coded as at most familiar. Since the number of at most familiar cases is only 1.1 % of the total cases coded, I hypothesize that *o* signals that the referent is activated and the one case in which the referent is coded as familiar (1.1 % of referents coded for this form) is within the proposed range of possible exceptions (< 5.0 %).

4.4.1. Coding referents of *o*

This section provides some sample tokens of *o* in the corpus and the coding of the referents in these examples. The excerpt in (79) contains 3 occurrences of *o* – one in the second sentence and two in the third sentence.

(79) *Ibrag'im girp geldi.*
Abraham enter.GER come.3S.PST
'Abraham came in.'

Onu *k"olunda bir sumka bar.*
3.2.GEN arm .3POSS.LOC one suitcase exist.PR
'There is a suitcase in his hand.'

Ondan ***o*** *bir zatlar boshatdy.*
3.2.ABL 3.2 one thing.PL empty.3S.PST
'From it he emptied some things.' [Video1.003]

In this example, the first token, *onu*, refers to Abraham, who is mentioned as the subject of the previous sentence and, thus, is in focus by criterion 1. Just over one third (34/89) of the referents of *o* in the corpus are in focus by criterion 1. The second token, *ondan*, refers to the suitcase/bag, which is mentioned in a syntactic focus position of the previous sentence—the precopular position of an existential sentence—and is in focus by criterion 3.[63] The third token, *o*, refers to Abraham again. Though this entity is not mentioned in a subject or focus position in the previous sentence, it is mentioned in each of the immediately preceding clauses, and, thus, is in focus by criterion 5.

In a number of cases, including the text portion represented in (80), a referent of *o* is mentioned in the preceding sentence but not in one of the two positions automatically coded as in focus—that is, as the subject or in focus position.

(80) *Ishge erli segiz kilometr ël bar;*
work.DAT until eight kilometer road exist.PR

onda *akhsham giun batg"ynchag"a Ø turmag"a da gerek...*
3.2.LOC evening sun until.setting stay.INF.DAT EMPH necessary
'It was nearly 8 kilometers to work, and it would be necessary to stay there until the sun set, but the slippers were in a state of disrepair.' [Aihalay.014]

If the referent is not mentioned in the pre-previous sentence, then it must be coded as activated (criterion 1). In example (80), the referent of the form *onda* 'at that (place)'is the place of work, which is mentioned for the first time in the previous sentence by *ishge*

[63] Note the word-order difference from the coding protocol (precopular versus postcopular). As is typical for SOV languages, in Kumyk the focal constituent in an existential sentence is preverbal.

'to work', but not as the subject or in focus position, and, thus, its referent is coded as activated.[64]

One interesting case involves a plural referent that is coded as activated. The plural entity representing two children is mentioned overtly as a plural pronoun at one point in the text. This is followed by six sentences in which the children are not mentioned as a plural entity, but one of the two children is mentioned in each of the sentences of this span. As for the cases mentioned in section 4.1.3 on null arguments, I consider that, once the plural referent has been activated, the mention of an individual member of the set maintains the plural entity as at least activated, if not in focus; however, this aspect of plurals is not specified in the coding protocol.

4.4.2. Discussion of noteworthy cases

Of the 90 referents of *o*, one is noteworthy because it is coded as at least familiar but does not meet any criteria for being activated from the coding protocol, and one is undetermined because it is only indirectly constructed from an entity overtly mentioned in the corpus. The text excerpt shown in (81) illustrates the undetermined case and the case coded as familiar (lines (i) and (j), respectively).

(81) a. *Sho giunden song men onu es etegen boldum.*
b. *Song okhuvun tergep de k"aradym:*
c. *bary zatdan dërt-besh k"yĭmat bulan iuriuĭ,*

---

[64] The fact that *ish* combines with the dative/directional morpheme *–ge* constitutes an overt refererence to a place. There is some degree of uncertainty, however, about whether or not a time or place associated with an event is activated or brought into focus by the mention of an event.

d. *tek netesen, ruscha iazyvdan osal edi.*

e. *Bibliotekg"a bardym –*

f. *cheber kitaplar da okhumaĭ eken.*

g. *<u>Klassda</u> olaĭlar dag"y da bolg"an.*

h. *Sho saĭaly da, vozhatyĭ bulan da oĭlashyp, <u>spor</u> etme tok"tashdyk"…*

i. *Biz erten, **olar** busa tiushden song okhuĭ edi.*

j. *Sho saĭaly da men chyg"ag"anda **ol** gelip tiupde tabula edi.*

k. *Esimde bar: Mar'ian koridorny bag"anasyna da taianyp turag"an edi.*

a. 'From that day forward, I remembered her.
b. Later on, I looked at my grades
c. everything was going about 4-5,
d. but, what to do….Russian language was looking bad.
e. I went to the library -
f. – it was to read literature.
g. In the class, there were others like that.
h. For this reason, I talked to the leader and we decided to form a group…
i. We studied in the morning, and they studied in the afternoon.
j. Because of that, when I was leaving, she was coming in.
k. In my mind, Marian stood out among the many in the corridor.'
[Marian.040-.049]

The entity for which coding is undetermined is related to two specific entities mentioned overtly in lines (g) and (h) of the text: *klassda*, referring to the narrator's class and *sbor*, referring to a group of students having trouble with Russian writing/literature. The referent of *olar* in line (i), however, refers neither to the whole class nor to the remedial group, but to the students in the class who were not part of the remedial group. This entity is a plural entity which must be constructed by subtracting certain members of the set of individuals in the class. Though such an entity is not specifically mentioned in the text, it is acceptable to refer to it with a pronoun. Since the class as a whole has already been mentioned, the reconstruction of the group seems to function in a similar way to an

activated entity, allowing for pronominal reference. Note that, this is somewhat similar to a case in which a plural entity is indirectly introduced via the mention of two or more individuals, which I also consider to be an activated entity. Though the coding of indirectly introduced entities of this type is not specifically mentioned in the coding protocol, I would consider this entity to be activated.

The one referent of *o* coded as familiar occurs within the same text in reference to Marian, a girl in the class whom the narrator is particularly interested in. (As a matter of fact, the whole introductory section of text is about the moment when he first saw Marian.) The text portion in (81) begins with the words, *sho giunden song men onu es etegen boldum* 'from that day forward I remembered her'. A pronoun is in line (j), even though Marian has not been mentioned overtly in the previous eight finite clauses. This seems to be an exceptional use of a form that signals 'activated' which is likely used for a particular literary effect—that is, to portray the cognitive prominence that Marian has for the narrator.

4.4.3. Reference to clausally-introduced entities

The pronoun *o* is used five times in the corpus in reference to an entity introduced by a clause. In four cases, *o* refers to a fact or proposition introduced by the preceding clause(s), which is coded as at most activated. Two of these cases (one fact and one proposition) are illustrated in (82) and (83).

(82) *Iash bolnitsadan chyk"g"an song bir zhumany ichinde kaĭnana*
child hospital.ABL exit.PST.PRT after one week.GEN inside.of mother.in.law

takhana etip adamlany chak"yrma bayram etme suĭe.
takana make.GER man.3PL.ACC invite.INF feast do.INF want.PR

O *saĭaly* *bu* *chak"yryvlar* *ĭiberip* *k"yzny*
3.2S according.to 3.1S invitation.PL send.GER girl.GEN

adamlaryn *o'ziuniu* *adamlaryn* *chak"yra.*
man.PL.3POSS.ACC self.3POSS.GEN man.PL.3 POSS.ACC invite.PR
'After the child leaves the hospital, within one week, the mother-in-law wants to hold a feast, making taxana and inviting people. Because of this, she sends invitations and invites daughter(in-law)'s people, (self's) own people.'
[Birth.008]

(83) *Gelezhekde* *Karimulla* *iuvug"ubuznu* *ullu* *ishleni*
come.FUT.LOC Karimulla near.one.1PL.POSS.GEN great work.3PL.POSS

bashyn *tutup* *gërmege Ø* *umut* *ete* *edik.*
beginning.3POSS.ACC hold.GER see.INF hope do.PR 1P.PST.AUX
'In the time to come, we hope that our dear Karimulla's great works will be seen from the beginning (leading the way?).'

Ol savlaĭ g'alkny *bashyn* *tutma* *da*
3S whole people.GEN head. 3POSS.ACC hold.INF EMPH

bazharazhak" *edi.*
know.how.FUT 3.PST.AUX
'He will lead all the people.'

Og"ar *bir* *de* *sheklik* *etmeĭbiz.*
3.2.DAT one EMPH doubt do.NEG.3P.PR
'We don't doubt it.' [Obituary.015-.017]

In (82), *o*, which occurs in the second sentence, refers to the fact that the mother-in-law wants to invite people to a celebration. In (83), *og"ar*, the dative form of the pronoun is an argument of the verb *sheklik etmeĭbiz* 'we do not doubt'. Because doubting relates to truth conditions, the complement of the verb is considered to be a proposition—the proposition that Karimulla will lead all the people (in the afterlife) in good works.

In one case, the form *o* is used to refer to a fact introduced by multiple sentences, which is also coded as activated. In the following example, presented in translation, the pronoun *o* occurs in the expression *ondan k"aĭry da* 'besides this'.

(84) 'In villages where representatives of only one nation live, the situation is not so praiseworthy, because there also the influence of the Russian language is no less. In village schools children study the mother tongue only in the beginning grades; in upper grades, however, all courses are carried out in Russian; mother tongue language and literature are only taught as separate subjects. The number of study hours given to these lessons is even sometimes reduced. Besides this (***ondan k"aĭry da***), in the present age, just like everywhere, in villages the influence of television and other instruments/forms of technology on the people, especially the younger generation, is powerful.' [Mother Tongue.012-015]

In this example, the fact to which *ondan* refers is considered to be the fact (inferred from four consecutive finite clauses) that Russian has a big influence in education because mother tongue education is limited to the early grades and Russian is used in the upper grades for all subjects except mother tongue language and literature, which can sometimes have reduced hours.

The last case in which a referent of *o* is introduced by a clause is a case in which *o* refers to a time associated with a state that has been mentioned in the previous sentence, as in the last clause of example (85).[65]

(85)	*Ish*	*iak"dan*	*alg"anda*	*da*	*Karimulla*	*bazharyvly*
	work	side.ABL	take.PST.PRT.LOC	and	Karimulla	successful

[65] While an event or state is considered to be introduced directly by a clause, the time of the event/state may be introduced indirectly and could be considered to be at most activated.

ulan edi.
son 3.PST.AUX
'With regard to work, Karimulla was a successful man.'

Ol on ĭyldan da artyk" Babaiurt raĭon sudnu
3S ten year.ABL and more Babayurt region court.GEN

predsedateli bolup, song busa Ø Mag'achk"ala
chairman.3POSS be.GER later however Makhachkala

shag'arny Lenin raĭon sud'iasu bolup ishledi;
city.GEN Lenin region judge.3POSS be.GER work.3PST

***ondan** alda Ø segiz ĭyllar prokuraturada zhavaplu*
3.2.ABL beforehand eight year.PL prosecutor.LOC answer.MOD

k"ulluk"larda ishledi.
service.PL.LOC work.3PST
'For more than ten years he served as chairman of the Babayurt court; later he served as a judge for Makhachkala's Lenin region; before that he had worked for eight years in a responsible position in the public prosecutor's office.'
[Obituary.012-.013]

It is noteworthy that *o* is used much less frequently than *sho* for entities introduced indirectly by clauses; however, based on the corpus, it is difficult to determine any differences in the distribution of these two forms with clausally-introduced entities. The fact that *o* is not used at all in reference to an event or state in the corpus is considered to be due to the infrequency of such types of references in discourse rather than a restriction, since *o* can be used with clausally-introduced entities that have a lower cognitive status.

4.5. The pronoun sho in the corpus

The distribution of *sho* in relation to features other than cognitive status, as summarized in Table 6, merits some discussion. This study identifies 33 cases of

pronominal *sho* within the text corpus. In all these cases, the referents are coded as either activated or in focus, with the latter category much less frequent. Only four tokens are references to entities which are coded as in focus, and it is noteworthy that only one of these refers to an entity mentioned as the subject of the previous sentence (one of the coding criteria for in focus).

Among the referents of *sho*, 23 cases are entities introduced by a clause or verb phrase, while ten cases are references to entities introduced by a nominal. Of the latter category, only two refer to humans and both of these references are to groups, not individuals. In these cases, the groups are 'Kumyk readers' and 'minority peoples' (cf. (88)). Though no referents of *sho* are human individuals, *sho* does refer to inanimate individuals in the corpus. Of the non-human referents introduced by nominals, four are individual entities and four are plural entities.[66] Since the distribution of *sho* in terms of entity type is unique, the discussion of coding is divided according to entities introduced by nominals, whether in focus or activated (section 4.5.1) versus entities introduced by clauses and VPs, all of which are activated (section 4.5.2).

[66] Under plural entities, I include one case where a singular form of *sho* refers to an amount which can be construed as a composite entity constructed indirectly from a plural: 667 hectares. (see Byron).

Human singular	Human plural	Animate (non-human) singular	Animate (non-human) plural	Inanimate singular	Inanimate plural-	Introduced by clause or VP
-	2	-	-	4	4	23

Table 6: *Types of referents of sho*

4.5.1. Referents of *sho* introduced by nominals

Of the ten referents of *sho* introduced by nominals, four are in focus and six are at most activated. Examples (86) and (87) illustrate two straightforward cases in which the referents are in focus. In (86) *shonu* refers to a newspaper article which is introduced in the previous sentence as the subject of the passive verb *jerleshdirilgen* 'is located'— that is, the extended noun phrase (marked by brackets) of which *makalasy* is the head.

(86) *Ëldash gazetni bu ĭyl 19-nchu Ianvarda chyk"g"an nomerinde*
friend newpaper.GEN 3.1 year 19th January.LOC go.out.PST.PRT edition.LOC

[K"yzyljurt shag'ardan iazag"an muxbir Izmulla G'adzhievi
Khasilyurt city.ABL write.PR.PRT correspondent Izmulla Gadjiyev's

"g'alk"ybyzny aburun ër_etmeĭik" degen mak"alasy]
people.1PL.GEN regard.3POSS.ACC lose.NEG.1PL.IMP say.PST.PRT article.3POSS

ërleshdirilgen.
be.found.3PST
'In the edition of "The Friend" newspaper published on January 19th of this year, is found an article called, "Let us not lose regard for our people" by the correspondent, Izmullah Gadjiev, who writes from Khasilyurt.'

Men ***shonu*** *tergevliu kiuide okhudum.*
1S 3.3.ACC careful way.LOC read.1S.PST
'I read this carefully.' [Editorial.01-02]

Since the newspaper article is mentioned in subject position in the previous sentence, it is in focus by criterion 1. This is the only case in the corpus in which *sho* refers to an entity introduced as the syntactic subject.

In (87) the form of interest is the pronoun *sho* which occurs in the last line of the example (not the determiner *sho* in the third line).

(87) *Bu* *kaĭtyp* *Ø* *ibadatkhanag"a* *gire.*
3.1 return.GER church.DAT enter.PR
'On the way back, she goes into the church.'

Ø *Ibadatkhanag"a* *girip* *gelegende*
church.DAT enter.GER come.PR.PRT.LOC

sho[67] *onu* *k"olundan* *akchany*
3.3 3.2.GEN hand.3POSS.ABL money.ACC

alyp *getgen* *iash* *mototsikly* *bulang"y* -
take.GER leave.PST.PRT child motorcycle.3POSS with.MOD -

shonu *ichinden* *chyg"yp* *gele.*
3.3.GEN inside.3PSS.ABL go.out.GER come.PR
'As she is going into the church, the kid who took the money out of her hand - the one with the motorcycle - is coming from inside (the church).'
[Wordless Book.029-30]

In this example *sho* refers to a place: a church. The church is mentioned in two consecutive sentences by the form *ibadatkhanag"a*, which brings it into focus even though it is not mentioned in a prominent syntactic position in the previous sentence.

[67] *Sho* is analyzed as a determiner in this phrase, which translates as 'that kid who took the money from her hand'.

In two notable cases in which *sho* refers to an entity in focus, the form is used within a noun phrase to refer to an entity mentioned within the same noun phrase. One of these is illustrated in (88), where s*holany* refers to the referent of *az g'alklany* 'minority peoples' in the expression, *sholany arasinda kumuklany da* 'among these are Kumyks'.

(88) *Artdag"y* *vak"tilerde* *sanav* *iak"dan* *az* *g'alklany*
previous.LOC.MOD time.PL.LOC number side.ABL minority people.PL.GEN

*(**sholany*** *arasynda* *Kumuklany* *da)* *tilleri*
(3.3.PL.GEN midst.3POSS.LOC Kumyk.PL.GEN EMPH language.3PL.POSS

avur *k"ysmatg"a* *taryp* *tura.*
heavy fate.DAT undergo.GER stay.PR
'In recent times, in terms of numbers, minority languages (among them Kumyk) have been subject to a "heavy" fate.' [Mother Tongue.09]

In this case *sho* is used in an expression which gives an example of one member of a set. Though this is similar to elaboration, I make a distinction here between exemplification and elaboration.

Examples (89) and (90) illustrate two of the six activated entities introduced by nominals. In (89) the referent of *shonda* is the noun phrase *o'zleni u'ĭunde Linda iashlaryny g'ak"ynda alg"an bir gichirek kino ĭimik zatni* 'something like a small movie that Linda had taken in their home of the kids'.

(89) *Paul* *mag"a* *kompiuterde* *o'zleni* *u'ĭunde* *Linda*
Paul 1S.DAT computer.LOC self.PL.GEN home.3POSS.LOC Linda

iashlaryny *g'ak"ynda* *alg"an* *bir* *gichchirek* *kino*
child.PL.3POSS.ACC about take.PST.PRT one tiny movie

ĭimik zatni gërsetdi.
like thing.ACC show.3PST
'On his computer, Paul showed something like a small movie that Linda had taken in their home of/about the kids.'

Shonda *Paulnu da Lindany da avletleri Ibrag'imni*
3.3.LOC Paul.GEN and Linda.GEN and child.PL.3POSS Abraham.GEN

de Katiany da g'ak"ynda gërsetilgen.
and Katy.GEN and about be.shown.3PST
'There(in the video), (something) was shown about Paul and Linda's children, Abraham and Katy.' [Video2.001-.002]

In this case *shonda* refers to the movie, mentioned in the previous sentence. In (90) the referent of *sho* is the amount represented by *667 hectares*. In this case, the amount is a composite singular entity which is indirectly constructed from the plural noun phrase *667 hectares*.

(90) *Bu ĭyl munda 667 gektarda giuzliuklege k"ulluk" etildi.*
3.1 year 3.1. LOC 667 hectare.LOC autumn.PL.DAT service be.done.3PST
'This year 667 hectares of autumn ones (wheat) were serviced here.'

Sho *da getgen ĭyl chachylg"anyndan*
3.3 EMPH leave.PST.PRT year be.scattered.PST.PRT.3POSS.ABL

ese 150 gektarg"a artyk" bola.
than 150 hectare.DAT more be.PR
'That is 150 hectares more than what was planted last year.'
[Agr.014-.015]

In another case, shown in (91) a location—in this case a store—is mentioned directly in a text, then not mentioned directly in the next two sentences. However, the

event denoted by the second to last sentence indirectly introduces the location of the event, and, therefore, the store is at least activated. [68]

(91) *Song bu tiukenge gire.*
Then 3.1 store.DAT enter.PR
'Later this one (she) goes into a store.'

Ø Tiukenchi bulan sëĭleĭ.
storekeeper with speak.PR
'(She) talks to the shopkeeper.'

Ø Tiukenchige bir zatlar…[69]
store.keeper.DAT one thing.PL -

ak"chalar berip, Ø Ø arg"anyn sata.
money.PL give.GER accordion.3POSS.ACC sell.PR
'To the shopkeeper something…gives money…(she) sells (him) her accordion.'

Ø Satyp Ø ak"chany da
sell.GER money.ACC and

alyp, Ø ***shondan*** *chyg"yp bara.*
take.GER 3.3.ABL go.out.GER go.PR
'Having sold it, taking the money, she goes out (of there).'
[Wordless Book.023-.026]

4.5.2. Reference to entities not introduced by nominals

Of all the pronominal forms in Kumyk, *sho* is the form most often used with entities introduced by clauses and verb phrases. As a matter of fact, 23 out of 33 or 69.7% of its referents are entities introduced directly or indirectly by clauses/VPs—as

[68] The mention of *tiukenchi* 'storekeeper' two times could also be considered to be an indirect mention of the store, in which case one might code the referent of *shondan* as in focus; however, I choose a more conservative interpretation of the coding here.
[69] This is interpreted as a repair, where the respondent says bir zatlar 'some things', then realizes that this is a sale and money is probably being exchanged; however, the speaker does not correct the text to say that the shopkeeper rather than the woman is giving the money.

compared to 5 out of 90 (5.6%) for *o*, 1 out of 31 (3.2%) for *bu* (restricted) and 1 out of 221 (.4 %) for nulls (restricted). Interestingly, the percentage of Kumyk referents not introduced by nominals for *sho* is similar to percentages reported for the English demonstratives *this* and *that* in previous research: 72 % in Gundel, Hedberg, and Zacharski 2002/2005 and 82.1 % in Hedberg, Gundel, and Zacharski 2007:33)[70].

The majority of clausally-introduced referents of *sho* are facts or propositions, as illustrated in (92) and (93). In example (92) *sho* occurs in the expression *sho saialy* 'therefore, because of that' and refers to the fact mentioned in the previous clause: the fact that Aihalay's squad was in a competition with Umazhat's squad.

(92) *Aĭkhalĭini zvenosu Umuzhatny zvenosu*
Aihalay.GEN squad.3POSS Umuzhat.GEN squad.3POSS

bulan iaryshg"a chyk"g"an edi.
with competition.DAT exit.PST 3.PST.AUX
'Aihalay's squad was in a competition with Umazhat's squad.'

Sho *saialy da ol erten tangdan turup*
3.3 therefore EMPH 3.2 morning dawn.ABL get.up.GER

Ø k"apusta ornatmak" uchun barmag"a gerek edi.
cabbage plant.INF in.order.to go.INF necessary 3.PST.AUX
'Therefore she had to get up in the morning and go plant cabbage.'
[Aihalay.012-013]

Among corpus examples of particular interest, there is one case in which *sho* refers to a proposition which is the conclusion deduced from a set of premises expressed

[70] The latter figure only includes activated entities not introduced by nominals, as there were not entities in focus non introduced by nominals reported in this corpus.

in a series of propositions. The illustration in (93) shows the sentences which express the premises (in the form of an English translation) followed by the sentence expressing the deduction in which the form *sholaĭ* occurs as part of the expression *sholaĭ bolg"an song* 'it follows that'.

(93) 'Literature is one of the areas of the fine arts. Each area of the arts, however, has materials which belong to it. For example, if the materials of painting are paints, literature's (materials) would be words.'

Sholaĭ *bolg"an song, adabiiat tuvradan-tuvra g'alk"ny tili*
3.3.MOD be.PST.PRT after literature along.with people.GEN language

bulan sëz baĭlyg"y bulan baĭlavlu.
with word richness with connected
'It follows that literature, along with the language of the people, is directly connected with its richness of vocabulary.' [Mother Tongue.04-.07]

Another interesting category of referents not introduced by nominals involves other types of entities referred to by the form *sholaĭ*. Kumyk pronouns marked by the suffix *–laĭ* serve as modifiers of both nouns and verbs, and can refer to entities of various types, including individuals, activities, and states, as well as certain features of individuals, activities and states. In several cases in this corpus in which *sholaĭ* modifies a verb or nominalized verb, the referent is defined as the entity introduced by an adverbial constituent in the previous sentence. For example, in (94) the first token of *sholaĭ* (which modifies the nominalized verb *biĭivniu* 'dancing') refers to an entity introduced by the embedded adverbial clause modifier of *biĭimek* 'dance': *biri-biri bulan k"uchak"lashyp* 'hugging one another'.

(94) *Masala busurman dinde ulan da kyz da biri-biri bulan*
for.example Muslim religion.LOC boy and girl and one.another with

k"uchak"lashyp biĭimek geri_urula.
hugging.each.other dance.INF be.forbidden.PR
'For example, in the Muslim religion it is forbidden for a boy and a girl to dance hugging one another.'

Men o'zium de ***sholaĭ*** *biĭivniu bek ershi gëremen.*
1S self.1S EMPH 3.3.MOD dance.NOM.ACC very ugly consider.1PR
'I, myself, consider it very ugly (in poor taste) to dance like that.'

Ulan da k"yz da birge biĭiĭgende, ulan k"olun zaman-zaman
boy and girl and together dance.PR.PRT.LOC boy hand.3POSS time.to.time

k"yzny boĭnuna tiĭdirip biĭiĭgen gezikler de k"arshylasha,
girl.GEN neck touching dance.PR.PRT occasion.PL EMPH come.across.PR

sholaĭ *ulanny men lap osal k"ylyk"ly ulan dep g'isap_etemen.*
such boy 1S very weak character.MOD boy COMP consider.1PR
'One comes across occasions when a boy and girl dance together, that from time to time the boy's hand touches the girl's neck; I consider that kind of boy to be a boy of very weak character.' [Editorial.021-.023]

As Hedberg et al. (2007:34) suggest that an entity introduced by a clausal unit other than the matrix clause should not be coded as in focus, I assume that the entity introduced by the adverbial clause is not in focus.

The second token of *sholaĭ* in (94), which occurs in the phrase, *sholaĭ ulan* 'a boy like that', refers to the boy introduced in the participial phrase, *ulan k"olun zaman-zaman k"yzny boĭnuna tiĭdirip biĭiĭgen* 'a boy dancing with (his) hand from time to time touching the girl's neck' which modifies the noun *gezikler* 'times'. For this case, the referent is considered to be 'a boy who dances with his hand touching the girl's neck', an individual entity which is indirectly constructed from the interpretation of the participial phrase (rather than an event, state, fact, or proposition).

One type of clausally-introduced entity which is less frequently mentioned in the literature is a situation. According to Hegarty (2007), a situation is a complex entity which involves events and states together with their consequences (2007:171). The example shown in (95) has a pronominal form which refers to a situation.

(95) *Bu* *ĭyl* *da* *khoziaĭstvosunu* *avlak"chylary* *1 ming*
3.1 year EMPH farm.GEN field.worker.PL.3POSS 1 thousand

tonnag"a *iuvuk"* *urluk"* *satmag"a* *khyĭal* *ete.*
ton.DAT near seed sell.INF thought do.3PR
'This year farm workers plan to sell nearly one thousand tons of seed.'

Demek ***sholluk"da*** *anadash* *khoziaĭstvosunu* *ekonomikasyn*
say.INF 3.3.means mother.tongue farm.3POSS.GEN economy.3POSS.ACC

dag"y_da *bekleshdirme* *umutlu.*
even.more strengthen.INF hope.MOD
'That is, by such means, it is hoped that the local/mother tongue farm economy will be strengthened even more.' [Agriculture.021-022]

The form *sholluk"da* in this example is a derivational form of *sho* which contains the nominal affix *-luk"* (primarily used to form abstract nouns) and is lexically defined as 'by this means'. The referent in this case is the situation that would result from the farm workers selling a thousand tons of seed. Following Hedberg, Gundel, and Zacharski (2007:35), I code this referent as activated.

In the discussion of clausally-introduced entities, it is worth noting that the corpus data contains no clear cases in which *sho* refers to an event or state directly introduced by a previous sentence, a fact which is most likely related to the natural infrequency of this referent type rather than a restriction for a particular form.

4.6. The pronoun shu in the corpus

The pronoun *shu* is rare in the corpus and occurs only three times, two of which are excluded from the study due to the fact that they occur within direct speech.[71] The one example analyzed, which is displayed in (96), has a referent coded as at most activated.

(96) *Geche Ø bularda k"onak" da bolup erten*
night 3.1.PL.LOC guest EMPH be.GER morning

itine Ø bara.
dog.3POSS.DAT go.3PR
'Staying the night as a guest at their place, in the morning [he] goes to the dog.'

***Shunda** da sholaĭ gech bolg"uncha*
3.4.LOC EMPH 3.3.MOD late become.before.GER

yashav g'aldan Ø Ø sorashyp
life situation.ABL greet. GER

gieviu de bug"ar kant ete…
son.in.law.3POSS EMPH 3.1.DAT complaint do.PR
'At that place, thus (as before), before it got late, (he) asks about the situation of life, and his son-in-law answers him…' [Prophet .0022-23 (cf. (76))]

In this example the referent of *shunda* (literally, 'at that') is understood as the dog. One can think of this colloquially as "at the dog's place". Since the referent of the pronoun is mentioned previously by the form *itine* 'to the dog', a non-subject argument in the previous sentence, it is coded as activated. Though it is not reasonable to make a

---

[71] The examples which occur within direct speech occur in one particular speech portion, displayed as sentences .0041- .0046 in Appendix 3.

[73] In principle, *shu* could be tested in the same way, but the more fundamental question, given the infrequency of this form, is its general acceptability.

hypothesis about what status *shu* signals based on one form, it is more likely that this form signals 'activated' rather than 'in focus'.

An interesting point not directly related to the analysis in this chapter is the way in which the contrast between *bularda* (literally, 'at those') in the first sentence of (96) and *shunda* in the second sentence is used to express the movement of the main participant (the prophet) from one location to the next – that is, from the home of the donkey-daughter and her husband (according to previous context) to the home of the dog-daughter (and her husband). It is interesting to look at how this contrast corresponds to descriptions of the demonstratives in terms of deictic meaning described in terms of distance and known-ness (cf. Chapter 3). According to Dimitriev1940 and Muratchaeva 2001, *bu* represents something that is both near and known, while *shu* represents something that is near and not known (presumably, to the addressee, though this is not specified by Muratchaeva), so this distinction is not defined in terms of distance from the deictic center, but rather by the parameter of known-ness. In this example, the dog cannot be described as unknown, since it is mentioned in the previous sentence. If one thinks of the contrast in terms of cognitive status, however, the difference between the two referents is one between a referent in focus (the donkey-daughter and her husband) and a referent that is activated, but not necessarily in focus (the dog-daughter).

4.7. Summary and interpretation of corpus analysis

4.7.1. Form-status correlations: Where to go from here

The corpus analysis indicates that the overt pronoun *bu*, null pronouns and pronominal reflexives all signal "in focus" status, while the pronouns, *o* and *sho* are both

associated with the status "activated". While the corpus provides insufficient data on the form *shu*, the fact that the one referent coded in this corpus is activated suggests that *shu* is more likely to fall within the group of forms signaling activation.

It is important to keep in mind the need to compare hypotheses about form-status correlations resulting from corpus analysis with hypotheses resulting from the analysis of questionnaire data. More specifically, for two of the forms hypothesized to signal 'in focus', *o'ziu* and null arguments, additional testing should reveal whether the small numbers in the lowest status are exceptions or an indication that the form actually signals the lower status. Also, given the relatively low number of tokens of *bu* (33) in comparison to *o* (91) and nulls (229), the fact that all but one referent is in focus could represent limits in the corpus rather than constraints on the use of the form. (The same principle would apply to *o'ziu*, which has 36 tokens, but is less likely to apply to nulls, with 229 tokens.)

The small number of referents in focus for *sho* – in comparison to the distribution of in focus versus activated entities for *o* – is an interesting aspect of the distribution of pronominals. An important result of form-status correlations is the idea that, following principles of Relevance and Grice's Maxim's of Quantity (cf. Chapter 2, section 3) form-status correlations are used to create scalar implicatures. More specifically, the use of a form which signals that the referent is at least activated may also be used to implicate that the referent is at most activated. (This is enough to explain the distribution of *sho*—and comparable to English data on *this/that* from Gundel et al. 1993.) The data on *sho* are consistent with a strong scalar implicature, 'not in focus', as there are only four in focus

referents; however, since *sho* is not used to refer to individual humans in the corpus, this could be an issue of type of entity rather than cognitive status. The questionnaire tests whether restrictions on *sho* are related to cognitive status issues alone or also to restrictions on types of entities (e.g. that *sho* cannot refer to individual human entities). Also, even though the unidirectional entailment of forms in the GH model predicts that there will not be a form that signals 'not in focus' (just as no language has a form that means 'not all') the small number of referents of *sho* that are in focus (4 tokens) suggests that this could be a possible counterexample and should be investigated further.

4.7.2. Other distributional features

In analyzing the distribution of referents in a given corpus, it becomes evident that certain types of referents are more frequent than others. In particular, in some genres, certain types of referents are more common than others. For example, in narrative texts, individual animate referents (particularly humans) are mentioned more frequently than inanimate entities and plural entities. In expository texts, on the other hand, higher order entities such as facts and propositions are more common than in other types of texts. Any analysis of the distribution of referring forms should take into consideration the underlying (in)frequency of certain types of referents, and further research should involve expanded analysis of infrequent referent types. For example, the fact that a given form refers to animate entities more frequently than inanimate entities is more likely to represent the underlying frequency of entity types rather than any particular correlation between form and entity type. One notable feature of the distribution of referents in this Kumyk corpus is that there is only one clear reference to a clausally-introduced event or

state (which is referred to by a null argument), thus it is difficult to determine if there is a distinction in the distribution of forms between reference to entities introduced directly by clauses (which are assumed to be in focus) and those introduced indirectly (which are assumed to be activated).

With that caveat in mind, the distribution of pronominals in this corpus shows a number of particular features other than those related to the cognitive status of referents. The most obvious distributional fact is simply the relative frequency of forms. Nulls are the most frequent, followed by *o*, *o'ziu*, *bu/sho*, and *shu*. The overall frequency of *o* compared to *bu*, *sho*, and *shu* is interesting, and its relative frequency may be the reason some grammatical descriptions label *o* as a personal pronoun, but not *bu*, *sho* and *shu*. The fact that the corpus includes only one token of *shu* raises a problem for the analysis and suggests that this form may even be disappearing from the lexicon of some speakers. One interesting fact related to frequency is that a higher percentage of tokens of *bu* occur in oral as opposed to written texts, even though the amount of oral versus written material in the corpus is smaller. It could be interesting to determine if this frequency in any way correlates with the use of visual representations of the referent through pictures and video in some oral texts. A more interesting question is whether in planned texts the form *bu* would be more likely to be replaced by nulls or by the pronoun *o*. Answering this question could help to determine why *bu* is more frequent in unplanned oral texts.

My analysis shows several interesting aspects of the distribution of pronouns in relation to entity type—that is both in terms of animacy features and the types of entities introduced by full clauses (events, states, facts, etc.) and VPs (activities). First, nulls and

bu refer to entities that are both animate and inanimate, but for each form, there is only one case of reference to an entity not introduced by a nominal. In the case of *bu*, moreover, the one entity of this type is an activity introduced by a verb phrase exhibiting a high degree of nominalization (cf. 4.3.4). The reflexive form *o'ziu* is even more restricted, referring only to entities introduced by nominals, whether animate or inanimate. The pronoun *o* is similar in distribution to nulls and *bu* in terms of lack of restrictions related to animacy features. There are, however, apparent differences in the degree of restrictiveness with entities not introduced by nominals. While the five referents of *o* not introduced by nominals comprise only a slightly higher percentage of the total number of coded referents than the less frequent form *bu* (5.6 % for *o* versus 3.2 % for *bu*), the linguistic forms which introduce the referents of *o* do not exhibit nominalization, while the one verb phrase which introduces a referent of *bu* exhibits a degree of nominalization. The form *sho* is unique in the sense that the majority of its referents are clausally-introduced entities, and only two of its referents are human. As both of the human referents are groups, it follows that *sho* is not used in reference to individual humans, though it does refer to individual inanimate entities.

A final question which is discussed because it has been raised in a number of studies of demonstratives is whether or not the distribution of any form can be described in terms of restriction to a referent previously mentioned as a subject or a referent not mentioned previously as a subject (Kaiser and Trueswell 2004, Bosch, Katz and Umbach 2007). The corpus analysis indicates that no form is restricted to use with a previous subject, though forms such as nulls and *o'ziu* have a high percentage of referents that are

previously mentioned in subject position. The form *sho*, on the other hand, is used with referents that are not previously mentioned as subjects in all but one case.

4.7.3. Evidence of contextual effects

While the primary effects of the choice of pronominal form discussed in my analysis is the signaling of where/how to search for potential referents in the attention state and the effects of scalar implicatures, my interpretation of the distribution of pronominals also suggests that certain forms are intentionally used by a speaker to create other contextual effects. For example, the data indicate that only nulls and *o'ziu* are associated with backwards anaphora and, among overt non-reflexive pronouns, only *bu* is associated with cataphora. I proposed that some cases of backwards anaphora and cataphora produce contextual effects related to imposed salience, such as *in media* res or highlighting the importance of the information in relation to other information components. In other cases, backwards anaphora appears to be associated with the ordering of information for the purpose of textual coherence. Examples of contextual effects associated with *o'ziu* include the expression of logophoricity, contrast, and imposed relative salience. The form *bu* is associated with continued reference to highly salient participants in contexts which involve some type of contrast such as topic shift or change of location. Finally, certain pronouns appear to be associated with certain types of relationships between units of information. Namely, *bu* is used in contexts involving elaboration, while *sho* is used in contexts which involve exemplification.

4.7.4. Methodological issues

The corpus analysis in this study raises a number of methodological issues that merit further discussion. Four areas worthy of mention are 1) factors that are likely to bring a referent into focus in Kumyk that are not mentioned in the coding protocol, 2) the coding of plural entities, 3) the role of eye gaze and other contextual features in coding for cognitive status, and 4) various issues related to the coding of clausally-introduced entities.

The criteria of the protocol used in coding statuses on the Givenness Hierarchy are minimal criteria, and evidence from the corpus analysis suggests several additional factors that are likely to bring a referent into focus. First of all, the use of null arguments for referents which are mentioned for the first time in the previous sentence but not as the syntactic subject or syntactic focus supports the idea that criteria related to the syntactic prominence of mention in the previous sentence can be more broadly defined—an idea which has already been incorporated in more recent versions of the coding protocol (Gundel 2006). In Kumyk, in particular, there is evidence that the following types of syntactic prominence are likely to bring a referent into focus: topicalization, mention as a possessor in a possessive clause, and, as discussed in Chapter 2, mention as the non-subject agent of a psychological verb.

Another criterion associated with bringing a referent into focus is parallelism. The fact that null arguments are acceptable in cases where there are parallels in argument structure or event structure suggests that the degree of expectedness affects the cognitive status of a referent. In other words, the more expected an activated referent is due to textual patterns, the more likely it is to be brought into focus. In the more unique cases of

parallelism that I call script repetition, parallelism combines with the mention of several different members of the same referent class (e.g. the first suitor, the second suitor, etc.) to create a degree of expectedness that even allows for a null to be used with a referent that would otherwise be coded as at most familiar.

A challenging area of coding in this study is the consistent treatment of plural entities. On one hand, the coding of a plural entity that is mentioned as a plural nominal is just like the coding of any nominally-introduced entity—that is, such an entity is at least activated and can be coded as in focus if it is mentioned in any of the contexts which bring any nominally-introduced entity into focus (e.g. subject or focus position, mention in two consecutive clauses, higher-level topic). An aspect of analysis that is less straightforward (and not specifically mentioned in the 2004 coding protocol), however, is the role of the mention of an individual member of a set in determining cognitive status of the plural entity. In this study, I code a plural entity as at most activated if it has been mentioned once as a plural nominal in a position that does not automatically bring it into focus, and then one member of the set is mentioned again in the following clause(s); however, it is equally likely that such a context would be enough to bring the referent into focus.

The coding of composite entities which are not introduced as plural nominals, but are introduced as distinct mentions of individual entities (or via the subtraction of members from a plural entity, as in one case discussed in section 4.4.2, example (81)), on the other hand, is also not specifically addressed in the coding protocol. Since composites are similar to other indirectly introduced entities in that they are not

introduced by individual nominals, I do not code these entities as in focus, even if all the members of the plural set are in focus at the time of mention of the plural. In another sense, however, composites formed from individual entities introduced by nominals are inherently different than abstract/higher order entities indirectly introduced by clauses. For example, my analysis shows one case in which a null argument, which I claim signals 'in focus', can be used for a plural entity (introduced for the first time) when at least one member of the set is already in focus. These and other questions related to the cognitive status of composite plural entities merit further testing (cf. Brown-Schmidt et al. 2005 for some experimental studies related to reference to composite entities).

Another methodological issue which merits further research is the role of contextual factors such as eye gaze in bringing a referent into focus. While the coding protocol specifically mentions that eye gaze is a sufficient criterion for activating an entity, the criteria do not address the question of whether or not eye gaze should count as a mention of a referent in evaluating the criterion for bringing a referent into focus. The role of contextual features in relation to the activated – in focus distinction is certainly an area of interest to future research.

An area of coding which presents some unique challenges is the coding of entities not introduced by nominals. The identification of higher order entities such as events, states, activities, or situations is less clear than the identification of nominally-introduced entities, and yet the coding of the entity as activated or in focus usually rests entirely on the correct identification of the entity type, since the GH model assumes that events and states are directly introduced by full clauses, while other types of higher order entities are

indirectly introduced by clauses. Moreover, as noted above (cf. 4.5.2) one must also consider the distinction between events (and activities) introduced by a matrix clause and those introduced by embedded clauses. In the latter cases, embedding can prevent an event from being brought into focus by a clause; however, the nominalization which accompanies some types of embedding can also affect referential choice in a different way. Also, the coding protocol does not clearly state how to treat a time or place which may be considered to be introduced indirectly via association with an event—just as a situation would be introduced indirectly in association with an event.

A final issue raised by the coding of clausally-introduced entities in this study is the treatment of entities that are introduced by multiple clauses. While I have listed these cases specifically in my presentation of the data, I choose to treat them in the same way as entities introduced by a single sentence. On the other hand, further analysis is required to determine if such referents share any of the characteristics of plural or composite entities.

Chapter 5: Questionnaire Results

5.0. Goals of the questionnaire

The primary goal of the questionnaire is to test hypotheses about form-status correlations which are based on results of the corpus study. More specifically, following from the fact that only a very small number of referents of null arguments, reflexives, and the overt pronoun *bu* are not in focus, the first hypothesis is that these forms signal that the referent is in focus. If one claims that a form signals 'in focus', it follows that that form cannot be used with a referent that is at most activated. The questionnaire, therefore, tests the claim that the use of these forms requires a referent in focus in order to demonstrate that this is truly a constraint and not just a limitation of the data.

In addition to this first hypothesis, the questionnaire provides a means to test the predictions of the Givenness Hierarchy in relation to *o* and *sho*, the forms predicted to signal 'activated'.[73] According to the GH model, while forms which signal activation may give rise to a scalar implicature that their referents are not in focus, they are not constrained to use with entities that are at most activated, as the unidirectional entailment in the hierarchy, the fact that anything in focus is by definition also activated, means that any form that signals activation can also be used for a form that is in focus. In other words, a form that signals activation is underspecified for any higher status (in focus) rather than specifically excluding it. According to the results of the corpus study, it is clear that *o* is not constrained to referents that are at most activated, as more than half of its referents are in focus. For the form *sho*, on the other hand, only a very small number

of referents are coded as in focus; therefore I use the questionnaire to verify that *sho* is not restricted to referents that are at most activated.

While verifying the absence of constraints on in focus referents for *o* and *sho*, the questionnaire also provides evidence that these forms create scalar implicatures (cf. section 5.8). Corpus data on the distribution of referents in terms of cognitive status provides little information to support or contradict the existence of such implicatures, though one might expect forms signaling activation to be associated predominantly with activated referents. In Gundel, Hedberg, and Zacharski 1993, the majority of pronominal forms that signal 'activated' in four of the five languages investigated have 75-100% of their referents coded as at most activated (1993:291-292).[74] For example, 94.4 % of the referents of the English form *that* and 100 % of the referents for *this* are at most activated. For Russian, 77.8 % of the referents of *eto* 'this' are at most activated. In my corpus study, the distribution of referents of *sho* is similar to that of forms that signal activation in the Gundel et al. study, with 87.9 % of its coded referents having the status 'activated'; but the distribution of *o* differs significantly, as the majority (71.9 %) of its coded referents are in focus. This raises the question of whether both these forms are used to the same degree to implicate 'not in focus'. The questionnaire elicits evidence of scalar implicatures based on their functional effect: the fact that the choice of form can

---

[74] This generalization does not fit the distribution of referents of two of the Japanese pronouns/demonstrative pronouns signalling 'activated': *kare* 'he', for which 100% (4 out of 4) of referents are in focus and *kore* 'this', for which 50% (1 out of 2) referents are in focus. However, for *sore* 'that (medial)', which also signals 'activated', 100 % (1 out of 1) of referents are at most activated (Gundel et al. 1993:291).

be used to disambiguate the intended referent of a pronoun in a context with multiple activated referents. For example, if one referent is in focus and the other at most activated, the use of a form which signals activation (*o* or *sho*) prompts the hearer to associate it with the referent that is not in focus, even though there is no constraint against using the form for a referent in focus. Evidence from the data show that this ranking effect can also be extended to contexts in which the speaker implicates the less prominent of two entities of the same status via the use of *o* or *sho*. This effect does not follow automatically from the Givenness Hierarchy, though it is not inconsistent with it.

Secondary goals of the questionnaire, particularly sentence continuation tests and fill-in-the-blank sections, are 1) to provide further evidence of the natural use of pronouns with referents of a particular cognitive status, 2) to provide evidence for or against constraints (or preferences) related to categories of referents that are relevant but not wholly related to cognitive status—specifically, introduced by a subject versus non-subject, human versus non-human (animacy features), introduced by a nominal versus introduced by a clause, and, finally, 3) to test the general acceptability of *shu* in light of the very limited number of corpus uses.

The first section of this chapter provides general comments about how the questionnaire data is interpreted in relation to the coding guidelines developed within the Givenness Hierarchy model. Sections 5.2 - 5.6 provide general information on referent category correlations for each form and the results of testing in relation to the specific predictions of the Givenness Hierarchy. Evidence for scalar implicatures and ranking

effects is discussed separately in 5.8. The final section of this chapter describes the conclusions based on these results.

5.1. Using the questionnaire to test predictions

The hypothesis that nulls, reflexives, and *bu* are constrained to use with referents in focus is tested by looking at the use of these forms with referents which are likely to be at most activated. If the use of these forms with at most activated referents is unacceptable, this would provide strong evidence that these forms are indeed restricted to referents in focus. The testing of this hypothesis depends on identifying referents which are activated but not likely to be in focus, as determined by the criteria of the coding protocol. Strictly following these criteria, this could mean that a referent is mentioned in one (but not both) of the immediately preceding two sentences, but not mentioned as the subject, higher-level topic, or in the syntactic focus position or that it is "a proposition, fact, or speech act associated with the eventuality (event or state) denoted by the immediately preceding sentence(s)" (Gundel 2004). Within these parameters, any referent introduced as a non-subject argument of the previous sentence would be coded as at least activated, but not in focus. However, one of the underlying premises of the coding protocol is that the criteria are sufficient, but not necessary. In other words, the criteria are intended to capture situations which always bring a referent into focus, yet there are still some types of situations which sometimes, but not always bring a referent into focus. This feature of the coding protocol makes the process of determining that a referent is at most activated an extremely difficult one.

According to Gundel et al.1993, one case where a referent can be coded as at least activated, but may also be in focus in some, but not all contexts is the case of a referent mentioned as a pragmatically prominent non-subject argument (280). In English, for example, there are a number of cases in which the unstressed pronoun *it*, which requires a referent in focus, is used to refer to an entity introduced as the direct object of the previous sentence. Thus, the referent of *a new laptop* in (97) could be brought into focus, which is supported by the fact that it is referred to with *it* in the second sentence.

(97) *Paul bought a new laptop today. It was on sale.*

One should not, however, draw the conclusion that all direct objects bring their referents into focus, as direct objects, unlike subjects, are not brought into focus simply by virtue of their syntactic position. Moreover, the cognitive status of referents mentioned as direct objects may vary cross-linguistically based on parameters such as word order, as well as within the same language based on features such as the number of arguments or the order of information components. In Kumyk, objects most often occur in pre-verbal position, which is known as a favored position for information focus in Turkish and other SOV languages (İşsever 2003). Nevertheless, in order to eliminate subjectivity, I code all referents introduced as direct objects as at most activated (in the corpus study and the questionnaire), but discuss the possibility of some of these referents being in focus in my analysis in this chapter. Referents of other non-subject arguments, for example dative objects or nominals which occur in prepositional phrases, are assumed to be activated but not in focus even though this could possibly vary depending on other factors, particularly

pragmatic factors, the number of arguments represented in the sentence, or word order variations which increase the prominence of the dative object (Gundel et al.1993).[75]

As stated in the introduction to this chapter, the questionnaire also tests whether or not any of the forms said to signal activation are in fact signaling that the referent is at most activated. Such a restriction would violate the Givenness Hierarchy prediction that forms that signal a particular status can always be used with referents of a higher (entailing) status, but it is important to test, in particular due to the fact that few (4 out of 33) of the referents for *sho* are coded as higher than activated. In order to form a conclusion about possible restrictions, I use questionnaire data to test whether or not *sho* can be used with a referent that is definitely in focus according to the coding protocol. Determining that a referent is definitely in focus is much more straightforward than determining that a referent is activated but not in focus, as any referents which meet the in focus criteria of the coding protocol are considered to be in focus. In the questionnaire I primarily use an entity introduced as the subject of a previous sentence or the event introduced by the previous sentence as an in focus referent for testing. Any instances where the use of *sho* or *o* is judged acceptable with these types of referents in sentence continuations or fill-in-the-blank portions are assumed to be evidence that there is no

---

[75]Gundel et al. 1993 cite the following example where *it* is used in reference to an entity mentioned as an NP within a prepositional phrase that is analyzed as in focus:
"However, the government of Barbados is looking for a project manager for a large wind energy project. I'm going to see the man in charge of it next week." (280, example 11a).

constraint that prohibits using these forms with entities in focus—in other words that this is an implicature and not a necessary inference.

Finally, in the questionnaire I use some different kinds of data to provide evidence of ranking effects and scalar implicatures. Sentence continuations suggest that certain forms prefer a particular type of referent, while fill-in-the-blank exercises with multiple pronouns provide indirect evidence of ranking effects. Section 4 of the questionnaire, moreover, is designed to provide direct evidence that the choice of pronominal form influences the choice of referent in an ambiguous context. This section tests sets of sentences in which the pronouns *o* and *sho* alternate and can potentially refer to multiple possible referents of different statuses. Respondents are asked to identify the referent of the pronoun, and the results are analyzed in order to determine the strength of the correlation between the choice of pronoun and the referent type.

5.2. Questionnaire data on null arguments

The primary goal of the questionnaire in relation to null arguments is to confirm the hypothesis that this form signals the status 'in focus' and, thus, is unacceptable with referents that are at most activated. Also, while the results of the corpus study indicate that there are no restrictions on the use of nulls in terms of animacy features or subjecthood (of previous mention), they do indicate a possible restriction on reference to entities not introduced by nominals, as only one out of 229 tokens in the corpus refers to a clausally-introduced entity. The questionnaire is designed to determine if this is an absolute restriction or simply a consequence of the fact that clausally-introduced entities

are typically not sufficiently salient to be brought into focus or simply that events and states are not common referent types in this text corpus.

5.2.1. Comments on potential category restrictions

The majority of referents of null arguments in the questionnaire are mentioned as previous subjects and, thus, are accepted without question to be in focus. The questionnaire data also demonstrates that null arguments can refer to events introduced by the previous clause (also in focus)—a category of referent not sufficiently represented in the corpus study. The use of nulls to refer to events demonstrates the acceptability of use both with entities not introduced by nominals and with inanimate entities (For reference to nominally-introduced inanimate entities, see examples (101) and (102) in section 5.2.2.). The acceptability of null arguments in reference to events and entities introduced as subjects is demonstrated in three types of exercises in the questionnaire: sentence continuations, fill-in-the-blank, and grammaticality judgments. [76] In sentence continuations, nulls are used only with referents mentioned as the subject of the previous sentence, which is a clear case of a referent in focus. An example of this type is shown in (98). In the fill-in-the-blank section of the questionnaire, null arguments are used with both types of in focus referents (e.g. previous subjects and events denoted by the preceding sentence) as in (99). The use of null arguments in reference to an event

---

[76] There is no evidence for the following types of in focus referents of null arguments in the questionnaire: referents mentioned earlier in the same sentence, referents mentioned in the syntactic focus position of the preceding clause, higher level topics which are part of the interpretation of the preceding clause and referents mentioned in the two immediately preceding clauses (without being mentioned as the subject of the previous clause).

denoted by the previous sentence is also confirmed by grammaticality judgments, as in (100), which is judged acceptable by three out of four respondents. (The judgments of each of the four respondents completing this section are listed in parentheses after the free translation.)

(98) *Ibrag'im restorandan shorpa aldy. (Ø)*
Abraham restaurant.ABL soup take.3S.PST
'Abraham ordered soup in the restaurant.'

Ø *Bitgende, ofitsiantny chak"yrdy.*
finish.PST.PRT.LOC waiter.ACC invite.3PST
'When he finished, he called the waiter.'

(99) *Bolat atyna minip barag"anda ŭyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Ø *bolma iaramas!*
happen.INF allow.NEG
'Bolat fell off while riding his horse. That's not possible! (Lit. 'For that to happen is not allowed.'
(2 respondents use a null argument)

(100) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

Ø *Song muallim studentni k"olun aldy.*[77]
after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave a book to the student. After (that) the teacher shook the student's hand.'
(1,1,3,1)

[77] I propose that there is a null in this sentence because of the parallel with the sentences with overt pronouns that start with "*Ondan song*…". It is also possible that there no null argument in this case.

5.2.2. Use of nulls with referents coded as activated

The questionnaire provides evidence related to the use of null arguments with two categories of referents that are at least activated but not necessarily in focus: entities mentioned as direct objects or those mentioned as dative objects of the previous sentence. Null references to entities introduced by non-subjects do not occur in sentence continuations and are infrequent in the fill-in-the-blank portion of the questionnaire, with a total of five references to an entity introduced by a direct object and one reference to a an entity introduced by a dative object, all but one of which are by the same respondent (orange). The grammaticality judgment portion of the questionnaire indicates that the use of nulls with entities introduced by direct objects is acceptable, but that the use of nulls with entities introduced by dative objects is unacceptable. In this section I discuss the evidence for null references to entities introduced by direct objects and those introduced by dative objects separately.

Examples (101) and (102) illustrate two test sentences involving reference to entities mentioned as direct objects in the fill-in-the-blank section. (Three of the five null references to direct objects come from these pairs.)

(101) *Ibrag'im* *restorandan* *shorpa* *aldy.*
Abraham restaurant.ABL soup take.3S.PST

O ***Ø*** *dag"y* *da* *alarmy* *edi* *eken?*
3.2 again also take.3FUT.INTER AUX.3PST COP
'Abraham got soup at the restaurant. Would he get it again?'
(1 respondent uses a null argument)

(102) A. *Ibrag'im* *restorandan* *shorpa* *aldy.*
Abraham restaurant.ABL soup take.3S.PST

B. ***Ø*** *tatyvlumu edi?*
sweet.INTER be.3PST
'Abraham got soup at the restaurant. Was it good?'
(2 respondents use a null argument)

In example (101), one out of five respondents uses a null argument to fill-in-the-blank assumed to refer to the soup, while others use the overt pronoun *o* or *sho*. In response to this example, one might claim that the use of the null argument in reference to the soup is acceptable due to the parallelism in argument structure, which contributes to the salience of the referent (cf. Chapter 4, section 4.1.3), as the two sentences both mention Abraham in subject position and the soup in object position. In the response shown in (102), however, two out of five respondents fill in the blank with a null argument (while the others use *o* or *sho*), yet there is no parallelism here, as the soup is mentioned as the object of one sentence and as the subject of a stative verb in the next sentence.

A third sentence pair in which the orange respondent uses a null to fill in a blank representing the direct object of the previous sentence is shown in (103). This example is presented here because the first sentence in the pair is parallel to the sentence used in the grammaticality judgment portion.

(103) *Muallim studentge kitap berdi. O* ***Ø*** *chalt okhuma*
teacher student.DAT book give.3PST 3.2 quickly read.INF

taryg"yn aĭtdy.
necessary.3POSS say.3PST
'The teacher gave the student a book. (S)he said ((s)he) needed to read (it) quickly.'
(1 respondent uses a null argument)

If one assumes that null references to entities introduced as direct objects are acceptable, the negative grammaticality judgment represented in sentence (104) provides an unexpected contrast to the acceptability of the null argument used in (103) if one only considers the argument position of previous mention.

(104) *Muallim studentge kitap berdi.* ****Ø*** *Tangcholpan edi.*
teacher student.DAT book give.3PST Tangcholpan be.3PST
'The teacher gave a book to the student. (It) was Tangcholpan.'
(3, 4)

Although the soup in (101) and (102) and the book in (103) and (104) are mentioned in the same syntactic position in the previous sentence, the use of the null argument in (104) is judged unacceptable by all respondents (note that this example only occurs in two questionnaires). However, I propose that this difference is not related to the cognitive status of the referent or the argument position of the previous mention but to the difference in the phonological focus structure between the identificational sentence in (104), where an overt argument is required, and the interrogative and declarative sentences in (101) - (103), where the phonological focus structure does not require an overt pronoun.[78] The grammaticality judgment of unacceptability in (105), which is structured similarly to (104), provides supporting data for this claim.

[78]The greatest similarity is between sentences (104) and (102), which differ structurally only in the addition of the interrogative suffix *–mu*, on the predicate adjective *tatyvlu* 'delicious'. A second difference between the two sentences is that (102) is structured as an interchange between two speakers A and B, while (104) is structured as a continuing utterance by the same speaker; however, this does not seem to be the relevant factor in relation to the acceptability of the null argument.

(105) *Bolat atyna minip barag"anda ĭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

***Ø** *Tiunegiun boldu.*
yesterday happen.3pst
'Bolat fell off while riding his horse. (It) was yesterday.'
Ø (3,3)
[*sho* (2,1,2,1); *bu* (2,3,2,3); *o* (2,4 (one respondent does not use *o* for non-humans)]

Although the null argument in (105) refers to an event, which can be assumed to be in focus, and such references are shown to be acceptable in example (99), both respondents judge this example to be unacceptable. I propose that the unacceptability of nulls in both (104) and (105) is related to the fact that the verb *bolmak* 'to be' or 'to happen' is a copular verb that requires two elements that can bear phonological focus (stress) , while (102) is acceptable because the interrogative morpheme serves as a "host" for phonological focus. Taken together with the corpus results, the evidence in (101) - (105) suggests that null reference to a previous direct object can be acceptable and not dependent on parallelism, as long as the focus structure of the sentence in which the null occurs does not require an overt form. This data supports the hypothesis that mention in direct object position brings a referent into focus, but further testing of null reference to direct objects is needed.

In contrast to the probable acceptability of some null arguments referring to previous direct objects, the questionnaire results indicate that null reference to the dative object of a preceding sentence is unacceptable. First of all, in the section on grammaticality judgments, in two cases where the dative object is assumed to be

activated, (106) and (107), the use of the null argument is judged unacceptable by all participants.

(106) **Muallim studentge kitap berdi.* ***Ø*** *savbol dedi.*
teacher student.DAT book give.3PST thank.you say.3PST
'The teacher gave a book to the student. (S)he said thank you.'
(4,4,4,3)

(107) **Muallim studentge kitap berdi. Song muallim*
teacher student.DAT book give.3PST then teacher

Ø *ekzameni bulan k"utlady.*
exam.3POSS with congratulate.3PST
'The teacher gave a book to the student. Then the teacher congratulated (him/her) on the exam.'
(3,3,3,4)

In spite of what seems like a clear case of judgments of unacceptability for null references to dative objects, there is one fill-in-the-blank example in the questionnaire where the orange respondent uses a null for a previous dative object. In (108) a null argument is used to fill in a blank associated with the dative object of the previous sentence, *Patimatg"a* 'to Patimat'.

(108) *Muallim Patimatg"a kitap berdi.* *Sho* *birinchi savg'at bolg"an song,*
teacher Patimat.DAT book give.3PST 3.3 first gift be.PST.PRT since

Ø*(Patimat)* *onu(book)* *anasyna* *gërsetme ciuedi.*
3.2.ACC mother.3POSS.DAT show.INF want.3PST
'The teacher gave Patimat a book. Since it was (her) first gift, she wanted to show it to her mother.'

My analysis of this case is that the occurrence of *savg'at* 'gift' in the first clause of the second sentence includes an implicit mention of Patimat, the recipient, thus the referent

would satisfy a criterion for in focus status at the time of the occurrence of the null argument, as the referent has been mentioned once in each of the previous two clauses. Assuming this analysis of (108), the data consistently supports the unacceptability of null references to entities introduced as dative objects. One area for further study would be to test the acceptability of null reference to entities previously mentioned as dative objects in cases where there is parallelism in argument structure, as the corpus data indicates that parallelism may play a role in bringing an entity into focus in cases where syntactic position alone is insufficient.

In addition to looking at reference to nominally-introduced activated entities, one sentence in the grammaticality judgment portion of the questionnaire, shown in (109), is structured to test null reference to a clausally-introduced entity considered to be at most activated.

(109) *Bolat atyna minip barag"anda ǐyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

Ø *baryn da tamasha etdi.*
everyone surprise do.3PST
'PRO surprised everyone.'
Ø (2,3,2,3)
[*o* (1,3,1,1); *bu* (2,4,2,2); *sho* (1,3,2,3); *shu* (1,3,2,3)]

While I intended for this example to create a context in which the null argument would be understood to refer to the fact that Bolat fell off his horse, responses from a similarly structured example in another section of the questionnaire, shown in (110), indicate that the referent in (109) is more likely to be understood as an entity introduced as a nominal

argument of the previous sentence—that is, Bolat. In (110), all respondents identify the referent of the form they use to fill in the blank as an entity introduced by a nominal in the previous sentence (that is, the book or the teacher), rather than a fact. By analogy, I conclude that the acceptability judgments for (109) cannot be interpreted as evidence for the acceptability of the use of particular forms in reference to a fact.

(110) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

___ *baryn da tamasha etdi.*
everyone surprise do.3PST
'The teacher gave a book to the student. ___ surprised everyone.'
o'ziu (teacher); *ol* (teacher); *o* (teacher); *bu* (book)

Based on the combined results described above, one possible conclusion is that, since nulls can refer to non-subject arguments (including both direct and indirect objects) which are coded as at most activated based on the protocol alone, nulls are not restricted to referents in focus—a conclusion that would disconfirm the hypothesis based on the results of the corpus study. On the other hand, the delineation between the acceptability of the use of nulls with direct objects versus the unacceptability of use with indirect objects by all but the orange respondent provides evidence that mention in certain less prominent argument positions may not necessarily be sufficient to bring a referent into focus. This data suggests that there is a distinction in the salience of direct objects mentioned in pre-verbal position and indirect objects (either in general or at least those not mentioned in pre-verbal position); therefore I assume that mention as a direct object in pre-verbal position brings an entity into focus in most cases (especially, but not only,

in cases where there is parallelism in the argument structure), while mention in other non-subject argument positions does not. Within this interpretation of the data, the results confirm the hypothesis from the corpus study that null arguments require a referent that is in focus. Furthermore, the data suggests that there is no restriction on the use of null arguments in referring to clausally-introduced entities that are in focus—that is, events—based on the hypothesis that some negative judgments of references to events are related to focus structure rather than to the cognitive status of the referent.

5.3. Reflexives

As most of the corpus data on the reflexive *o'ziu* involves reference to entities mentioned previously within the same sentence, the questionnaire is used to provide data on reflexives which refer to entities in the previous sentence and possible constraints on these uses. The intent is to verify that reflexives can refer at least to the subject of the previous sentence and to judge the acceptability of their use with referents of other arguments of the previous sentence which may be at most activated—that is, direct objects or dative objects. The data on reflexives comes from fill-in-the-blank exercises, and grammaticality judgments in the questionnaire. No other category restrictions are tested in the questionnaire, as the corpus study already indicates the acceptability of the use of *o'ziu* with inanimate entities (although I assume that the reflexive cannot be used with clausally-introduced entities).

In the fill-in-the-blank section, *o'ziu* is used in nine different sentence pairs (by one respondent each time, except for example (111), where two respondents use this form). For seven of the nine sentences, *o'ziu* is used in reference to a previous subject, as

in (111). One use of *o'ziu* is in reference to a previous direct object, as shown in (112); however, follow-up testing of this example shows that this use is at best marginal. In the other exceptional case illustrated in (113) *o'ziu* is used to refer to a previous dative object, but further testing of this example shows it to be unacceptable, as well.

(111) *Ibrag'im restorandan shorpa aldy.*
Abraham restaurant.ABL soup take.3S.PST

Adatly g'alda o pilavnu ***o'ziune/o'ziu*** *aldy.*
customary case.LOC 3.2 pilaf.ACC self.3S.DAT/self.3S take.3S.PST
'Abraham got soup from the restaurant. Usually he got pilaf for himself./Usually he(himself) got that pilaf.'
(2 respondents)

(112) ?*Ibrag'im restorandan shorpa aldy.* ***O'ziu*** *tatyvlumu edi?*
Abraham restaurant.ABL soup take.3S.PST self.3S sweet.INTER be.3PST
'Abraham got soup at the restaurant. Was it good?'
(1 respondent)

(113) ?*Muallim studentge kitap berdi.* *Ol* ***o'ziune*** *chalt*
teacher student.DAT book give.3PST 3.2 self. 3S.DAT quickly

okhuma taryg"yn aĭtdy.
read.INF necessary.3POSS say.3PST
'The teacher gave the student a book. (S)he told (him/her) (s)he needed to read it quickly.'
(1 respondent)

In the grammaticality judgments *o'ziu* is judged to be unacceptable when its referent is mentioned as the dative object of the previous sentence, assumed to be at most activated. (Note that these sentence pairs are identical to those in (106) and (107), except for the alternation between the null argument and the reflexive.) Neither of these sentences involves parallel argument structure.

(114) **Muallim studentge kitap berdi.* ***O'ziu*** *savbol dedi.*
teacher student.DAT book give.3PST self.3S thank.you say.3PST
'The teacher gave a book to the student. (S)he said thank you.'
(3,4,3,4)

(115) **Muallim studentge kitap berdi. Song muallim* ***o'ziunu***
teacher student.DAT book give.3PST later teacher self.3S.ACC

ekzameni bulan k"utlady.
exam with congratulate.3PST
'The teacher gave a book to the student. Then the teacher congratulated (him/her) on his/her exam.'
(4,4)

In summary, the evidence suggests that reference to previous non-subjects is unacceptable; however, further testing of the use of reflexives in reference to animate entities mentioned as direct objects is needed.[79]

5.4. Questionnaire data on bu

The primary issue the questionnaire data addresses in relation to *bu* is whether or not its use requires a referent that is in focus as concluded on the basis of the corpus data. Following the analysis of nulls and reflexives, particular attention is given to whether or not *bu* is acceptable when used with entities introduced by arguments of the previous sentence which are neither subjects nor direct objects. While the questionnaire is

[79] While separate work on long-distance reflexivization in Kumyk (Humnick 2007) indicates that it is unlikely that a reflexive can be coreferent with a non-subject argument of a previous sentence, further work is required to test the acceptability of reference to a non-subject argument that is clearly in focus because it has been mentioned in two or more sentences (criterion 5 of the coding protocol). If such uses of the reflexive are acceptable, it would clearly indicate that the restriction is related only to cognitive status and not to syntactic position. Otherwise, I assume that the apparent cognitive status restriction is a secondary effect of a general restriction on non-subject antecedents of the previous sentence.

primarily structured to provide evidence of cognitive status constraints, there is also data on the question of whether or not *bu* is acceptable when used in reference to events or other types of clausally-introduced entities. This question is intended to follow up the corpus analysis, where *bu* is documented as being used only with clausally-introduced entities with a degree of nominalization. The corpus analysis already provides strong evidence that there are no restrictions on *bu* related to grammatical role or animacy features; therefore, data related to these features is not specifically discussed in this section, though the acceptable use of *bu* with non-subjects and inanimate entities is evident from the discussion of its use with activated entities.

5.4.1. Use of *bu* with referents coded as activated

Sentence continuations and fill-in-the-blank exercises provide the most natural evidence of possible acceptable uses of the pronoun *bu* with at most activated entities. There are two sentence continuations for which the respondent is asked to write a sentence using *bu*. The first of these is listed as example (116). While three out of five respondents use *bu* for the referent of *Ibrahim*, the subject of the previous sentence and, hence, a referent which is in focus, one response, illustrated in (a), uses this form in reference to the soup, an entity which is mentioned as the direct object of the previous sentence and is coded as activated, while another response (b) uses *bu* in reference to the restaurant, which is mentioned as a locative argument and not likely to be in focus. For the sentence continuation shown in (117), three respondents use *bu* to refer to the student, an entity mentioned as the dative object of the previous sentence, and one of these cases is illustrated in (a). One respondent uses *bu* to refer to the book, mentioned as the direct

object of the previous sentence, as illustrated in (b). Though the latter case is coded as activated, it could easily represent a referent in focus, as described previously in section 5.1.

(116) *Ibrag'im restoranda shorpa ashady. (bu)*
Abraham restaurant.LOC soup eat.3PST
'Abraham ate soup in the restaurant.'

a. ***Bu*** *tuzlu edi.*
3.1 salty be.3PST
'It was salty.'

b. ***Mundag"y*** *ashlar aĭrycha tatyvlu edi.*[80]
3.1.LOC.MOD food.PL especially delicious be.3PST
'The food here was especially delicious.'

(117) *Muallim studentge kitap berdi. (bu)*
teacher student.DAT book give.3
'The teacher gave the student a book.'

a. ***Bu*** *shossag'at okhuma bashlady.*
3.1 immediately read.INF begin.3PST
'(S)he started to read immediately.'

b. ***Bunu*** *betleri këp edi.*[81]
3.1.GEN page.PL.3POSS many be.3PST
'It had a lot of pages.'

The fact that *bu* can be used in reference to the direct object of the previous sentence in free sentence continuations is particularly strong evidence that this type of use is acceptable. As discussed in relation to nulls and reflexives, however, it is likely that

[80] This is the respondent for whom the data on grammaticality judgments had to be disregarded.
[81] *Bunu* is understood as an alternation of *munu*, the genitive form.

referents introduced as direct objects should at least sometimes be analyzed as in focus, in spite of the fact that they are only coded as activated (cf. 5.2.2 and 5.3). The use of *bu* with a locative argument and a dative object in sentence continuations is significant, as there is less likelihood that mention in these argument positions can bring an entity into focus.

In the fill-in-the-blank section of the questionnaire, eleven out of twenty-one tokens of *bu* occur with referents originally coded as activated.[82] Six of these refer to an entity mentioned as a direct object in the previous sentence and five to an entity mentioned as a dative object (either of a matrix clause or of an embedded clause). A comparison of these uses with responses from the grammaticality judgment portion confirms the acceptability of *bu* with both of these two categories.

The sentences, shown in (118) - (121), all of which are continuations of a similarly-structured sentence (either fill-in-the-blank, grammaticality judgment, or both), indicate the acceptability of *bu* in reference to an entity mentioned as a direct object—in each case, the book.

(118) *Muallim Patimatg"a kitap berdi.*
teacher Patimat.DAT book give.3PST
'The teacher gave Patimat a book.'

Bu *birinchi savg"at bolg"an song,* *ol* *shonu* *anasyna*
3.1 first gift be.3PST.PRT since 3.2 3.3.ACC mother.3POSS.DAT

[82] It is interesting that 15 out of 21 uses of *bu* in the fill-in-the-blank section are by the same (blue) respondent. Though not all of the activated cases are from this respondent, a majority of them are (others are from black respondent); however, grammaticality judgments for similar sentence pairs using *bu* do not suggest that acceptable judgments are limited to one or two respondents.

gërsetme siuedi.
show.INF want.3PST
'Since it was her first gift, she wanted to show it to her mother.'
(2 respondents)

(119) *Muallim studentge kitap berdi. Student* ***munu*** *chalt okhudu.*
teacher student.DAT book give.3PST student 3.1.ACC quickly read.3PST
'The teacher gave the student a book. The student read it quickly.'
(1 respondent)

(120) *Muallim studentge kitap berdi. Student,* ***munu*** *alyp,*
teacher student.DAT book give.3PST student 3.1.ACC take.GER

u'ĭge getdi.
home.DAT go.3PST
'The teacher gave the student a book, The student, taking it, went home.'
(2,2)

(121) *Muallim studentge kitap berdi.* ***Bu*** *"Tangcholpan" edi.*
teacher student.DAT book give.3PST 3.1 Tangcholpan be.3PST
'The teacher gave a book to the student. It was Tangcholpan.'
(2 respondents)

Grammaticality judgments of the same sentence: (3,1,2,3)
[*Ø* (3,4); *o* (2,4); *shu* (2, 3;3,4); *sho* (1,1,2,1)]

In the fill-in-the-blank response, respondents use *bu* to refer to the book two times in (118) and once in (119). In a similarly structured sentence pair in the grammaticality judgment portion, shown in (120), both respondents judge *bu* to be acceptable. (This pair was given only to 2 respondents.) In example (121) in the grammaticality judgment portion, however, the response to this sentence pair was inconsistent, with two respondents (blue and red) judging the sentence to be unacceptable and two, acceptable. It is interesting that one of respondent who judges the sentence to be unacceptable (blue)

is the same as one of those who used *bu* for the same pair in the fill-in-the-blank section. The other (red) judges both *o* and *bu* to be "3", which seems to indicate that red prefers *sho* in reference to an inanimate entity (see data on *sho*) rather than any type of restriction based on cognitive status. (Note that the same sentence pair (cf. example (104)) is judged to be unacceptable (by 2 out of 2 respondents) when a null argument is used in the same position as *bu*, but I propose that this restriction is related to particular sentence types and phonological focus.)

Sentences (122) - (129) illustrate the use of *bu* with referents mentioned as the dative object of the previous sentence. In (122) and in the first uses of *bu* in (123) and (124) this form is used to refer to the student/Patimat by one respondent for each pair in the fill-in-the-blank portion of the questionnaire.

(122) *Muallim studentge kitap berdi. Shunu **bu** chalt*
teacher student.DAT book give.3PST 3.4.ACC 3.1 quickly

okhuma taryg"yn aĭtdy.
read.INF necessary.3POSS say.3PST
'The teacher gave the student a book. ((S)he) told (him/her) (s)he (*bu*) needed to read it (*shunu*) quickly.'
(1 respondent)

(123) *Muallim Patimatg"a kitap berdi. **Bu** shonu këp ushatdy .*
teacher student.DAT book give.3PST 3.1 3.4.ACC much be.pleased.3PST

Song, bu u'ĭge getdi.
later 3.1 home.DAT go.3PST
'The teacher gave Patimat a book. She was very pleased with it. Then she went home.'
(1 respondent)

(124) *Muallim Patimatg"a*$_j$ *kitap berdi.*
teacher Patimat.DAT book give.3PST

<u>O</u> birinchi savg"at bolg"an song, ***<u>bu</u>***$_j$ *<u>munu</u> anasyna*
3.2 first gift be.3PST.PRT since 3.1 3.1.ACC mother.3POSS.DAT

gërsetme siuedi.
show.INF want.3PST
'The teacher gave Patimat a book. Since it was her first gift, she wanted to show it to her mother.'
(1 respondent)

In similar examples in the grammaticality judgment portion the use of *bu* in reference to the student/Patimat is judged to be acceptable by all respondents for (125) and all but one respondent for (126), while the use of null arguments or reflexives in each case is judged unacceptable by all respondents. This provides strong evidence that the use of *bu* is less restricted than that of nulls and reflexives.

(125) *Muallim studentge kitap berdi.* ***Bu*** *savbol dedi.*
teacher student.DAT book give.3PST thank.you say.3PST
'The teacher gave a book to the student. (S)he said thank you.'
(2,2,2,2)
[*shu* (4,4,2,4); *o'ziu* (3,4,3,4); *Ø* (4,4,4,3); *o* (2, 2); *sho* (3,4,2,4)
cf. example (106)]

(126) *Muallim studentge kitap berdi. Song muallim*
teacher student.DAT book give.3PST then teacher

munu *ekzameni bulan k"utlady.*
3.1.ACC exam.3POSS with congratulate.3PST
'The teacher gave a book to the student. Then the teacher congratulated him/her on the exam.'
(1,2,2,3)
[*o'z* (4,4); *Ø* (3,3,3,4); *o* (2,1); *sho* (3,4,3,4); *shu* (3,4;3,4); cf. example (107)]

An interesting aspect to the data set represented in (122) - (126) is that the dative object is human, while the accusative object is inanimate. Later in this chapter I show that animacy ranking can also play a role in the choice of pronoun; therefore in these cases, the acceptability of the use of *bu* with the dative object may also be explained by the fact that the human referent outranks the inanimate referent in terms of salience.

Similar results occur for a second set of sentences in which a pronoun refers to a non-human entity mentioned as the dative object of the subordinate clause in the preceding sentence, which must be considered at most activated according to the coding criteria. In (127) and (128) *bu* or *munu* (the accusative of *bu)* are used to fill in a blank in reference to the horse by one respondent for each example. Note that the acceptability of *bu* in (127) does not depend on parallel argument structure, as in (122) - (126).

(127) *Bolat atyna minip barag"anda ĭyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Sho saialy, o ***munu*** *satma khyĭal etdi.*
3.3 because 3.2 3.1.ACC sell.INF decision make.3PST
'Bolat fell off while riding his horse. Because of that, he decided to sell it.'

(128) *Bolat atyna minip barag"anda ĭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

G'ali ***bug"ar*** *dag"y minip barmazhak"man dep aĭtdy.*
now 3.1.DAT again ride.GER go.NEG.FUT.1S COMP say.3PST
'Bolat fell off while riding his horse. Now (he) said that (he) won't ride it anymore.'

In the grammaticality judgment portion, three out of four respondents judge (129) to be acceptable, confirming the results for (127) and (128). The one respondent (red) who

judges this example to be unacceptable does so for all overt pronouns tested except *sho*; thus I conclude that the issue of acceptability in this one specific case is related to the respondent's preference for *sho* with non-human referents rather than cognitive status restrictions on the other forms.[83]

(129) *Bolat atyna minip barag"anda ǐyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Bolat Ø(Bolat) dag"y ***bug"ar*** *minmezhegin aǐtdy.*
Bolat more 3.1.DAT ride.3FUT.3POSS say.3PST
'Bolat fell off while riding his horse. Bolat said he won't ride it anymore.'
(2,2,2,3)
[*o* (1,2,2,4); *sho* (1,2,1,2)]

Since the acceptable uses of *bu* to refer to the horse do not involve reference to the highest-ranking participant in terms of animacy (Bolat, the human, is ranked higher), the data in this section shows that the acceptability of reference to dative objects is not conditional on the added salience resulting from reference to a human.

To summarize this section, the testing of *bu* with entities which are not coded as in focus by the criteria of the coding protocol shows that this pronoun is acceptable with entities introduced in the previous sentence as direct objects. In addition to this, a number of examples in the questionnaire indicate acceptability with entities introduced by other arguments of the previous sentence, including dative objects and locative arguments of both matrix and embedded clauses. One result of the grammaticality

[83] This respondent is not entirely consistent and uses *bu* for non-human/inanimate referents in some cases yet also indicates a constraint against using *o* for non-humans.

judgments is the finding that *bu* is judged to be acceptable in cases where neither null arguments nor reflexives are judged to be acceptable—more specifically, use with previous arguments other than subjects or direct objects—showing that nulls and reflexives are more restricted than *bu*. My conclusion is that the use of *bu* is not constrained to referents in focus, in spite of the fact that it is rarely used with referents not in focus in the corpus; and it is thus analyzed as signaling activation.

5.4.2. Use of *bu* with clausal entities

Due to the absence in the corpus study of uses of *bu* with entities denoted by a preceding sentence, the questionnaire tests the use of *bu* to refer to events denoted by the preceding sentence to determine whether this is indeed a constraint. The data related to this issue, however, proves to be inconclusive. For the fill-in-the-blank sentence pairs in (130), responses (a) and (b) (by two different respondents) indicate that *bu* is acceptably used with a clausally-introduced entity (event) in each case.

(130) *Bolat atyna minip barag"anda ǐyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

a. ***Bu*** *bolma iaramas!*
3.1 happen.INF allow.NEG.GER
'That's not possible!' (Lit. 'For that to happen is not allowed.) (blue)

b. ***Bu*** *nechik bolma bola?*
3.1 how happen.INF happen.PR
'How could it be/happen?' (red)

On the other hand, the use of *bu* to refer to an event in example (131) from the grammaticality judgment section is judged unacceptable by all but one respondent, while

sho is judged acceptable by all respondents and *o* and null are judged acceptable by all but one respondent. Likewise, in example (132), two respondents judged the use of *bu* in reference to the event introduced by the previous sentence to be unacceptable.[84]

(131) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

?***Mundan*** *song muallim studentni k"olun aldy.*
3.1.ABL after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave a book to the student. After that the teacher shook the student's hand.'
(2,4;3,4)
[*sho* (1,1; 2,1); *shu* (3,3;2,2); *Ø* (1,1;3,1); *o* (1,1, 2,4 (this respondent does not use *o* for non-human))]

(132) *Bolat atyna minip barag"anda ĭyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

?***Bu*** *tiunegjun boldu.*
3.1 yesterday happen.3PST
'Bolat fell off while riding his horse. (It) happened yesterday.'

(2,3,2,3)

[*sho* (2,1,2,1); *Ø* (3,3); *o* (2,4 (this respondent does not use *o* for non-human))]

Note that (131) and (132) are the only type of examples in the grammaticality judgment section in which *bu* was viewed as less than acceptable by the non-red respondents. Red, however, is somewhat restrictive in the use of *bu* and *o* with non-human referents, and

[84] There is some question as to whether the referent of *bu* in example (131) actually refers to an event or, more specifically, to the time of the event. If the referent is the time, it may be considered to be introduced indirectly in association with the event, which would affect its cognitive status and, thus, the choice of referring form.

this may also be the factor affecting black's response in (132). (Other respondents, such as blue, use *bu* with non-human referents – cf. (127).)

In the analysis of these facts, one must also consider that there is a significant difference between the sentence pairs shown in (130) and those in (131) and (132): the former pairs represent an exchange between two speakers in which the second speaker gives an emphatic response, while the latter pairs represent continuations by one speaker, with no emphatic feature. The mixed reaction to the second set might indicate that the use of *bu* in a context without the emphatic feature is marginal based on pragmatic features other than cognitive status. Another consideration is that the sentences in (131) and (132) may also be non-referential uses of *bu* rather than actual references to events, which can be common in conventionalized exclamations.

One limitation of the questionnaire is that it does not test the acceptability of *bu* when used with a clausally-introduced entity from the previous sentence that is *not* brought into focus—that is, a fact, proposition, situation or speech act. As described in relation to (109) and (110) in section 5.2.2, this is due to an unexpected ambiguity in the pairs targeted for testing this type of referent. Further testing of this type of referent would provide further information on both the limits of acceptability of the use of *bu* with clausally-introduced entities and confirmation that *bu* is acceptable in use with all types of entities which are considered to be at most activated. If one can demonstrate that *bu* can be used acceptably in reference to a speech act or a clausally-introduced entity that is indirectly introduced by the previous clause (such as a situation or a proposition), then it clearly cannot be restricted to referents in focus.

5.5. Questionnaire data on o

Since the corpus analysis shows significant numbers of tokens of *o* with both at most activated and in focus referents, the questionnaire does not directly test restrictions related to either category of cognitive status; however, the data about *o* from the questionnaire does provide some interesting information and also serves as a control for some of the other tests for constraints. The fact that *o* is used in reference to both in focus entities and at most activated entities in the questionnaire corroborates findings based on the corpus analysis that this form only signals activation and confirms that this form does not conventionally signal 'not in focus' (which would have been inconsistent with the GH model), while still allowing for the possibility that it implicates 'not in focus'. A possibly significant observation from the questionnaire is that one respondent appears to have a constraint against using *o* for non-human/inanimate entities that is not exhibited by the other respondents.

In (133) and (134), the two test sentences which require continuations that use *o*, the pronoun is used to refer to a human referent which is in focus and is mentioned as the subject of the previous sentence (either Bolat (133) or Abraham (134)) in all except two cases. In the latter two cases, (134), *o* is used to refer to the soup, the direct object of the previous sentence.

(133) *Bolat atyna minip barag"anda ŭyg'yldy. (o)*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'
(In all continuations of this sentence, *o* refers to Bolat.)

(134) *Ibrag'im restorandan shorpa aldy.*
Abraham restaurant.ABL soup take.3S.PST
'Abraham ordered soup in the restaurant.'
(In three sentence continuations *o* refers to Abraham; in two, *o* refers to the soup.)

In the fill-in-the-blank section of the questionnaire, most respondents use *o* to refer to both human and non-human/inanimate entities. The red respondent, however, only uses *o* for human entities. Similarly, the red respondent judges *o* to be unacceptable when referring to a non human entity, as in (135) - (137), while the remaining respondents judge these uses to be acceptable.

(135) *Bolat atyna minip barag"anda ĭyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Bolat Ø dag"y ***og"ar*** *minmezhegin aĭtdy.*
Bolat more 3.1.DAT ride.3FUT.3POSS say.3PST
'Bolat fell off while riding his horse. Bolat said he won't ride it anymore.'
o (1,2,2,4)

(136) *Muallim studentge kitap berdi.* ***O*** *"Tangcholpan" edi.*
teacher student.DAT book give.3PST Tangcholpan be.3PST
'The teacher gave a book to the student. It was Tangcholpan.'
o (2,4)

(137) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

Ondan *song muallim studentni k"olun aldy.*
3.2.ABL after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave a book to the student. After that the teacher shook the student's hand.'
o (1,1,2,4)

The red respondent's responses are consistent across all sections of the questionnaire with a constraint against using *o* with non-human referents. In section 4, in particular for two responses for which *o* could only refer to a non-human entity, this respondent writes "Ø" when asked to identify the referent. Evidence from the other respondents, however, does not support such a constraint.

Except for the respondent who appears to have a constraint against using *o* with non-human referents, all except two sentence pairs containing *o* are judged to be acceptable. In one exceptional case shown in (138), which only appears in two questionnaires, one respondent judges all overt pronouns to be acceptable, while the other judges all overt pronouns to be unacceptable; thus, one may conclude that some aspect of the sentence pair other than the pronoun choice is unacceptable. A similar conclusion may be drawn about (139), where one out of four respondents judge *o* and all other forms to be unacceptable, while the remaining respondents judge *o* to be acceptable.

(138) *Murat o'ziuniu mashini bulan tiukenge barma tok"tashdy.*
Murat self.3S.GEN car.3POSS with store.DAT go.INF decide.3PST

Song o ***onu*** *dëgerchegi boshalg"anny bildi.*
later 3.2 3.2.ACC tire.3POSS empty.PST.PRT.ACC know.3PST
'Murat decided to go to the store with his(own) car. Then he(o) found out that his/its tire was flat.'
o (2,3)
[*sho* (2,3); *bu* (1,3)]

(139) *Bolat atyna minip barag"anda ŭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

O baryn da tamasha etdi.
3.2 everyone surprise do.3PST
'It/he surprised everyone.'
o (1,3,1,1)
Ø (2,3,2,3); *bu* (2,4,2,2); *sho* (1,3,2,3); *shu* (1,3,2,3)

5.6. Questionnaire data on sho

Since the number of corpus referents for *sho* that are coded as in focus is small, one of the goals of the questionnaire is to test whether or not the use of *sho* is restricted to referents that are not in focus. The questionnaire provides further data that *sho* occurs most frequently with at most activated entities, but it also specifically tests the acceptability of *sho* with two types of in focus referents which do not occur with *sho* in the corpus study: events and subjects of the previous sentence. In the analysis of the questionnaire data, I also consider whether or not two more specific constraints might apply to *sho*: a constraint against individual human referents or a constraint against reference to previous subjects.

The questionnaire data supports the generalization that *sho* prefers (or most commonly refers to) entities that are not in focus—more specifically, non-human entities that are not the subject of the previous sentence. Sentence continuations provide the most natural data on the types of entities which can be referred to with *sho*. The three sentences used to elicit sentence continuations with *sho* are given in (140) - (142), with four of the response sentences illustrated in (142) (a) – (d). [85]

---

[85] One response in this section was not considered because it used *sho* as a determiner rather than a pronoun.

(140) *Muallim studentge kitap berdi. (sho)*
teacher student.DAT book give.3PST
'The teacher gave the student a book.'
(4 responses use *sho* to refer to the book; 1 response uses *sho* to refer to the giving of the book)

(141) *Ibrag'im restoranda shorpa ashady. (sho)*
Abraham restaurant.LOC soup eat.3PST
'Abraham ate soup in the restaurant.'
(3 responses use *sho* to refer to the restaurant; 2 use *sho* to refer to the soup)

(142) *Bolat atyna minip barag"anda ĭyg"yldy. (sho)*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

(a) ***Shonu*** *aiag"y avurtdu.*
3.3.GEN leg.3POSS be.injured.3PST
'His leg was injured.'

(b) *Bolat* ***shog"ar*** *dag'y minmezhek.*
Bolat 3.3.DAT more ride.3FUT
'Bolat will not ride him anymore.'

(c) *Bolat* ***sholaĭ*** *ĭyg"ylag"any birinchi gezik.*
Bolat 3.3.MOD fall.PST.PRT.3POSS first time
'It was the first time he had fallen like that.'

(d) *Ol* ***shog"ar*** *k"amuchu iaman urg"an edi.*
3.2 3.3.DAT whip bad hit.PST.PRT AUX.3PST
'He whipped him hard.'

In (140) and (141), the continuations use *sho* to refer to entities that are inanimate and not mentioned as the subject of the previous sentence: the book (4 times) or the giving of the book (once) in (140) and the restaurant (3 times) or the soup (twice) in (141). In each of these cases, there is no coding criterion by which the referent of *sho* is automatically coded as in focus, though it is possible, as explained above, that at least the entities mentioned as direct objects may be in focus. The majority of responses to sentence

(142), show a similar preference for non-human entities that are not coded as in focus, as shown in (b) – (d), where the referents are the horse (2 responses) or the act of riding a horse (1 response).[86] The continuation in (a), however, uses *sho* to refer to Bolat, a human entity that is mentioned as the subject of the previous sentence and thus meets a sufficient criterion of a referent in focus. The responses from this section alone indicate that *sho* refers to a non-human entity that is not in focus much more frequently than to a human entity that is in focus, but that this is not an absolute restriction.

Additional evidence for the absence of a constraint against entities in focus is the use of *sho* with events introduced by a previous sentence. In the questionnaire data, the acceptability of *sho* in reference to events is illustrated both by fill-in-the-blank cases (where the blank is associated with an event referent) and grammaticality judgments of sentences in which *sho* refers to an event mentioned by the previous sentence. In fill-in-the-blank responses, like those illustrated in (143), *sho* is not only acceptable, but is used more often than any other pronominal form in reference to the event denoted by the previous sentence.[87] In grammaticality judgments, every use of *sho* in reference to an event is rated as acceptable by all participants, as illustrated in (144) and (145).

---

[86] This activity is coded as 'activated' rather than 'in focus' due to the fact that this activity is introduced, not by the matrix verb phrase, *ŭyg"yldy* 'fell' but by the subordinate clause verb phrase, *atyna minip barag"anda* 'when riding (his) horse'.

[87] Though the referent of *sho* in (143) and (144) is interpreted here as an event which is introduced directly by the previous clause and, thus, in focus, in later stages of the research a question has arisen as to whether these are references strictly to the time of the event and that the time is not necessarily introduced directly by the previous clause. However, even without these two examples, there is still the evidence from (145), which is a clear reference to an event and illustrates the acceptability of *sho* with this type of referent.

(143) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

shondan *song muallim studentni k"olun aldy.*
PRO.ABL after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave the student a book. After that, the teacher shook the student's hand.'
shondan (4 times); *ondan* (1 time)

(144) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

Shondan *song muallim studentni k"olun aldy.*
3.3.ABL after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave a book to the student. After __ the teacher shook the student's hand.' [corresponds to the fill-in-the-blank example in (143).]
sho is good (1,1,2,1)

(145) *Bolat atyna minip barag"anda ŭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Sho *tiunegiun boldu.*
3.3 yesterday hapen.3PST
'Bolat fell off while riding his horse. It happened yesterday.'
sho (2,1,2,1)

Though the use of *sho* in reference to events is enough to confirm that *sho* can be used with entities in focus, another possibility which should be explored is whether or not the use of *sho* for some speakers is more specifically limited to entities which are not introduced as subjects in the previous sentence. As mentioned in the beginning of this section, one response in the sentence continuation portion of the questionnaire already suggests the lack of a subject constraint (green respondent). In the analysis of questionnaire data from other sections, the absence of *sho* in reference to previous subjects in the fill-in-the-blank section and the results of example (146) in the

grammaticality judgment portion suggest that use with previous subjects is not acceptable for all speakers.

(146) *Mariam anasyna alma berdi.*
Mariam mother.3POSS.DAT apple give.3PST
'Mariam gave an apple to (her) mother.'

O ***shog"ar*** *raziligin bildirdi.*
3.2 3.3.DAT pleasure.3POSS.ACC know.CAUS.3PST
'She(o) expressed her pleasure to/at her.'
(2, 3)

In this example, where *sho* most likely refers to Mariam, one respondent judges the acceptability as "2", while the other gives a "3".[88] (Note that the respondent who judges this example as "2" is not the same respondent who gives a sentence continuation with *sho* referring to a previous subject.) Further analysis, however suggests that these judgments are difficult to distinguish in this data set from an apparent constraint for some, but not all, speakers against using *sho* with human referents.[89] (Note that blue is one of two respondents who doesn't use *sho* with humans.)

Section 4 of the questionnaire provides less direct evidence that *sho* is acceptable, at least for some speakers, when used in reference to previous subjects. In a number of examples in this section in which the pronoun *sho* is used and multiple interpretations are

[88] This was an alternative example that did not appear on green's and red's questionnaires, and the orange respondent's answers on the grammaticality judgments were disqualified. There is possibly some unexpected abiguity which would allow for the referent to be understood as the apple or the giving of the apple.
[89] Unfortunately the questionnaire does not provide any direct data on the acceptability of *sho* in reference to a non-human subject of the previous clause. Section 4 has indirect data that reference to previous non-human subjects is acceptable.

possible, respondents select a referent which is mentioned as the subject of the previous sentence. Example (147) illustrates one of these cases where two respondents identify the referent of *sho* as the teacher.

(147) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST
'The teacher gave the student a book.'

Sho *baryn da tamasha etdi.*
3.2 everyone surprise do.3PST
'It/(s)he surprised everyone.'
(2 respondents identify the teacher as the referent of *sho*; 2 respondents identify the book or the giving of the book as the referent of *sho*)

For this type of data, however, identifying the referent of a pronoun is not necessarily the equivalent to a judgment of acceptability or an indication that the pronoun would be used in the same way in natural speech.

Unexpectedly, the questionnaire data provides evidence that some, but not all, speakers use *sho* only with non-human referents. Two different respondents (red and blue), use *sho* only for non-humans, regardless of cognitive status or the grammatical role of the previous mention. These respondents do not use *sho* to refer to humans in any continuations or fill-in-the-blank sentences and judge all examples which use *sho* in reference to a human individual as unacceptable.[90] This is illustrated for both respondents by (148) and (149), where the referent is activated, and for one respondent in

[90] Since the corpus data includes tokens of *sho* that refer to human groups but not to human individuals and the grammaticality judments do not test the acceptability of *sho* with human groups, there is enough evidence to suggest a constraint against human individuals, but not enough to suggest a constraint against groups of humans.

(146), where the referent is in focus. The unacceptability of *sho* with the activated, non-subject referent suggests that only the human parameter affects the acceptability of *sho* for these two respondents.[91]

(148) *Muallim studentge kitap berdi. Song muallim* ***shonu***
teacher student.DAT book give.3PST then teacher 3.3.ACC

ekzameni bulan k"utlady.
exam.3POSS with congratulate.3PST
'The teacher gave a book to the student. Then the teacher congratulated him/her on (his/her) exam.'
sho (3,4,3,4)
[*o* and *bu* are judged to be acceptable: cf. (126) *bu* (1,2,2,3); *o* (2,1)]

On the other hand, as shown in example (149) from the grammaticality judgment portion of the questionnaire, one respondent (green) judges the use of *sho* with a human individual to be acceptable (in this case, the referent is mentioned as the dative object of the previous sentence). The same respondent also uses *sho* in reference to a human in the continuation test shown previously in (142) (a).

(149) Muallim studentge kitap berdi.
teacher student.DAT book give.3PST
'The teacher gave the student a book.'

[91] It is possible that judgments of unacceptability are related to the ranking of participants. In other words, if there is a human referent in the previous sentence, and *sho* specializes in indicating the lower-ranked of two or more entities, as subsequent sections of this study claim, then the judgment of unacceptability could relate to the fact that *sho* should not be used with a higher-ranking entities (in terms of cognitive status and/or animacy) rather than a general constraint against use with humans.

Sho *savbol* *dedi.*
3.3 thank.you say.3PST
(2,3,4,4)

The data from the grammaticality judgment in (149) could be considered somewhat questionable, as the other three respondents judge this use of *sho* here to be unacceptable. In addition to this case, however, there are two different answers from two other respondents that indicate that using *sho* to refer to individual human referents is possible. In the fill-in-the blank response in (150), where the respondents are also asked to identify the referent of the pronoun, one respondent (orange) uses *sho* in reference to Patimat (object of the previous sentence).

(150) Muallim Patimatg"a kitap berdi.
teacher Patamat.DAT book give

Sho*(Patimat)* *Ø(kitap)* *këp* *ushatdy.* *Song,* *ol* *u'ĭge* *getdi.*
3.3 much please.3PST later 3.2 home.DAT go.3PST
'The teacher gave Patimat a book. She was very pleased (with it). Later she went home.'

In addition to the evidence from (150), in (146) another respondent (black) judges *sho* to be acceptable in reference to Mariam. Note that black is inconsistent and judges *sho* as unacceptable in (149) and (148), where the referent is human and activated, suggesting no consistent parameter.[92] Based on these three examples from three different respondents, there is no evidence of a constraint against human referents among all

[92] If the black respondent understands *sho* to refer to a referent other than Mariam in (146), the responses would be more consistent..

speakers. I assume that this is a preference which is not fully grammaticized as a constraint or perhaps varies by age or region. The unacceptability of certain examples where *sho* is used for a human may also depend solely on the number and types of activated entities and ranking tendencies discussed in the next section (5.8).

To summarize this section, the questionnaire data indicate that, while there is a strong tendency against the use of *sho* for referents that are in focus, especially when these are human, this cannot be considered an actual constraint, but it is consistent with the presence of a strong implicature that the referent is not in focus.

5.7. Questionnaire data on shu

The questionnaire data shown in sentences (151) and (152) provides evidence that the pronoun *shu* is acceptable for at least some speakers. In the sentence continuation portion, all respondents use *shu* in a continuation of the sentence illustrated in (151) to refer to the book (one representative response is provided).

(151) *Muallim studentge kitap berdi. (shu)*
teacher student.DAT book give.3PST

Student ***shundan*** *tapmuruvlar etezhek.*
student 3.4.ABL assignment make.3FUT
'The student will do an assignment from it.'

In the fill-in-the-blank section *shu* is used two times by the same (blue) respondent. One of these cases is illustrated in (152).

(152) *Bolat atyna minip barag"anda ŭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST
'Bolat fell off while riding his horse.'

Shu *nechik bolma bola?*
3.4 how happen.INF happen.PR
'How could it be/happen?' (blue)

In the grammaticality judgment portion, however, the use of *shu* receives a significant number of unacceptable judgments. One respondent (black) judges all examples with *shu* to be "3" or "4"—that is 'possible, but people don't usually say it that way', or 'no Kumyk person would say it that way'. The use of *shu* in reference to human entities is judged unacceptable in all but one case, which is shown in the green response to example (153).[93] Note that, in this example, the judgments for *shu* are very similar to those of *sho*.

(153) *Muallim studentge kitap berdi.* ***Shu*** *savbol dedi.*
teacher student.DAT book give.3PST 3.4 thank.you say.3PST
'The teacher gave a book to the student. (S)he said thank you.'
shu (4,4,2,4)
[*o'ziu* (3,4,3,4); *bu* (2,2,2,2); *Ø* (4,4,4,3); *o* (2,2); *sho* (3,4,2,4)]

The use of *shu* in reference to non-human entities is somewhat more acceptable, though the majority of responses (8 out of 14) still judge these references to be unacceptable. Two examples of judgments of *shu* in reference to an inanimate entity are illustrated in (154) and (155), where the grammaticality judgments are 2-3-3-4 and 3-3-2-2, respectively.

[93] Though responses from the orange respondent are not included in the discussion of acceptability judgments in this section because only one sentence was judged unacceptable, it is worthy of note that the one unacceptable sentence was this case of the use of *shu* in reference to the student in (153).

(154) *Muallim studentge kitap berdi.* ***Shu*** *"Tangcholpan" edi.*
teacher student.DAT book give.3PST 3.4 Tangcholpan be.3PST
'The teacher gave a book to the student. It was Tangcholpan.'
shu (2,3,3,4)
[*bu* (3,1,2,3); *Ø* (3,4); *o* (2,4); *sho* (1,1,2,1)]

(155) *Muallim studentge kitap berdi.*
teacher student.DAT book give.3PST

Shundan *song muallim studentni k"olun aldy.*
3.4.ABL after teacher student.GEN hand.3POSS.ACC take.3PST
'The teacher gave a book to the student. After __ the teacher shook the student's hand.'
shu (3,3,2,2)
[*sho* (1,1,2,1); Ø (1,1,3,1); *o* (1,1,2,4); *bu* (2,4,3,4)]

In comparing these two examples, it is hard to determine any criteria for degree of acceptability based on the context, as the blue respondent judges *shu* to be acceptable in (154) but not in (155), while the red and green respondents judge *shu* to be acceptable in (155) but not in (154). (The black respondent judges *shu* to be unacceptable in all cases.) It is also interesting that *sho* is consistently acceptable in all uses with inanimate entities, while *shu* is not. Given the variability of results, there is not enough evidence to predict when the use of *shu* is acceptable; nevertheless, there is enough evidence to suggest that there is no restriction to entities in focus, and, thus, one can claim that *shu*, to the extent that it is used at all, signals that its referents are activated.

5.8. Evidence for implicatures

This section discusses evidence from the questionnaire that the use of forms hypothesized to signal 'activated' on the basis of the corpus study, that is, *o*, *sho*, and *shu*, sometimes give rise to the implicature 'not in focus'. In addition to this, based on the

questionnaire results described in section 5.4, the revised prediction for the pronoun *bu* is that it signals 'activated', rather than 'in focus'; therefore this section includes a discussion of any evidence that *bu* also implicates 'not in focus'. Since implicatures typically arise only when the information signaled by a stronger form is relevant, they generally arise with pronominal forms when they function to differentiate between two or more possible referents of the same minimal status (e.g. both at least activated) by implicating that one of them is not in focus. This phenomenon is particularly noticeable in utterances in which two or more entities are referred to with pronouns. For example, in a context where a null argument and *sho* are used to refer to different entities, one of which is in focus and one of which is (at most) activated, the null argument necessarily refers to the referent in focus (since nulls require a referent in focus), while *sho* is used to refer to the referent that is at most activated, even though, in principle, it could refer to either one. Even in a context in which the speaker refers to only one of multiple possible activated entities, the use of any of the forms which only signal activation, that is, *bu*, *o*, *sho*, or *shu*, can implicate 'not in focus'.[94]

Evidence both from previous studies (Mulkern 2003) and from the current analysis shows that the implicature 'not in focus' is sometimes extended to differentiate between two referents of the same maximal status (e.g. both in focus or both at most activated) by implicating that one is less salient (that is, less in focus or less activated) than the other. In a context with multiple pronominal references, if this ranking effect is

[94] Note, however, that implicatures do not necessarily arise in contexts which require an overt form for an independent reason, such as phonological focus or other focus contexts (cf. example (104)).

extended in order to differentiate between two referents in focus one would expect, for example, that a null argument (or any form signaling in focus) could be used with the referent of greater salience and *sho* (or any form signaling activated) could be used with the referent of lesser salience. In the case of reference to two entities that are at most activated, *sho* or any other overt pronoun could be used to refer to the less salient one, but a pronoun signaling in focus cannot be used with an entity that is at most activated; thus, another overt pronoun would be used. If all overt pronouns can implicate 'not in focus' (or less salient), then there is no logical basis for differentiating the relative salience of two entities that are both at most activated. I propose that, for this reason, one of the four overt pronouns in Kumyk (*sho*) has become specialized in communicating the implicature 'not in focus' and, by extension, designating the less salient of two entities. Both the corpus study and questionnaire data support the idea that the use of *sho* has a strong correlation with entities that are of lower relative cognitive status or less prominent. In particular, section 4 of the questionnaire demonstrates that *sho* is much more likely than *o* to implicate 'not in focus'.[95]

Finally, the reanalysis of *bu* raises the question of the degree to which this form is used to implicate 'not in focus'. Following the argumentation above about the relationship between implicatures and the relative salience of referents, this analysis looks at the use of *bu* compared to other pronominal forms in contexts with multiple

---

[95] The goal of this section of the questionnaire was primarily to explore the difference in distribution between *o* and *sho*, two forms hypothesized to signal 'activated'. The questionnaire did not test *sho* in relation to *bu* in the same way, as *bu* was originally hypothesized to signal 'in focus' based on the corpus results. *Shu* was not tested in this way due to the overall rarity of the use of this form.

possible referents. The results show that, while the use of *sho* often implicates the less salient entity, *bu* appears to have a specialized function of indicating the most salient of two or more entities with the same minimal cognitive status. This would account for the fact that *bu* almost exclusively refers to entities in focus in the corpus, even though, as the questionnaire data indicates, it is not absolutely restricted to entities in focus, and for the fact that *bu* is rarely associated with the implicature, 'not in focus'.

5.8.1. Overt pronouns, topic-shift, and implicatures

In analyzing the use of *bu*, *o*, *sho* and *shu* to create scalar implicatures, it is important to explore facts about the distribution and interpretation of these forms within simple and complex clause structures. While these facts are generally considered to be syntactic, they may, in fact, be derived from the implicatures created by the use of overt forms versus null arguments.

In Kumyk, non-reflexive overt pronouns are always interpreted as not coreferent with a subject within the same clause, as shown in the second sentence of (156).[96]

(156) *Bolat$_i$ o'ziuniu$_i$ uchun kitap aldy.*
Bolat self.3POSS.GEN for book buy.3S.PST
'Bolat bought a book for himself.'

[96] *O* is used as a representative of an overt pronoun, as it is syntactically interchangeable with *bu*, *sho*, or *shu*. Note that a non-reflexive pronoun can be used in reference to a nominal within the same noun phrase—such as *Katja va onu k"ardashlar* 'Katy and her friends'.

Leĭla$_j$ ***og"ar***$_{i/*j}$ *da* *kitap* *aldy.*
Leyla 3.2.DAT EMPH book buy.3.PST
'Leyla$_i$ bought a book for him /*herself, too.'

The second sentence of (156), shows that the pronoun *og"ar* (dative form of *o*), cannot be interpreted as coreferent with *Leĭla*, the subject of the clause in which it occurs, but can be interpreted as coreferent with Bolat, the entity mentioned as the subject of the previous sentence. In the same type of context, a reflexive can be coreferential with the subject, as shown in the first sentence of (156), where *o'ziuniu* is coreferent with the subject, *Bolat*.

Another restriction on coreferentiality is that an overt non-reflexive subject pronoun in one clause is always interpreted as not coreferent with the subject of a conjoining clause of the same sentence, as illustrated in (157).

(157) *[**Naima***$_i$ *getgende,]* ***o***$_{*i/j}$ *ĭylady.*
Naima go.PST.PRT.LOC 3.2 cry.3.PST
'When Naima left, he/she(not Naima) cried.'

In (157), the overt pronoun *o* must be interpreted as not coreferent with *Naima*, the subject of the preceding non-finite clause. In the same context, either a null argument or a reflexive could be used to express a coreferent subject, as illustrated in (158). [97]

[97] Interestingly, the use of null arguments, on the other hand, is difficult to describe in terms of coreferentiality restrictions. As shown in (i), a null argument in one clause of a complex clause construction does not necessarily have to be interpreted as coreferent with the subject of the other clause. Whether the null occurs in the matrix clause, as in (a), or in the preceding non-finite clause, as in (b), it can be interpreted as referring either to Bolat or to Naima.

(158) *[**Naima**$_i$ getgende,]* *__Ø__ $_i$ /**o'ziu**$_i$* *ĭylady.*
Naima go.PST.PRT.LOC 3S cry.3.PST
'When Naima left, she(Naima) cried.'

The restrictions on coreferentiality in complex clause constructions do not, however, apply to cases where a non-subject of one clause is coreferent with a subject of an adjoining clause, as illustrated in (159), where the pronoun *onu* can be interpreted as referring to Naima or someone other than Bolat, but not to Bolat.

(159) *[**Naima**$_i$ getgende,]* *Bolat$_j$* ***onu*** $_{i/*j}$ *arysyndan* *ĭylady.*
Naima go.PST.PRT.LOC Bolat 3S after cry.3.PST
'When Naima left, Bolat cried after her.'

As these examples show, within the unit of the clause/sentence, *o* is understood as not coreferent with the nearest subject—whether within the same clause, as in (156), or within an adjoining clause as in (157).[98] Since the mention of a referent in subject position is associated with greater relative salience, the use of the overt pronoun can be said to implicate 'not the most salient referent'. In (159), however, the pronoun *onu* does

---

(i) *Bolat$_i$ Naimany$_j$ këp siuedi.*
Bolat Naima.ACC much love.3PST

a. [***Naima**$_i$* *getgende,*] *__Ø__ $_{i/j}$* *ĭylady.*
Naima PST.PRT.LOC cry.3.PST
"When Naima left, (Naima)/(Bolat) cried."

b. [*__Ø__ $_{i/j}$* *getgende,*] *Naima* *ĭylady.*
PST.PRT.LOC Naima cry.3.PST
"When Naima left, (Naima)/(Bolat) cried."

[98] Note that a non-reflexive pronoun can be used in reference to a nominal within the same noun phrase—such as *Linda and her friends*.

not implicate 'not in focus', since it can refer to Naima, a referent which is at least in focus.

While overt pronouns have certain coreferentiality restrictions within sentences, their use is not constrained in the same way in intersentential contexts. For example, in (160) (repeated from Chapter 1), the use of an overt pronoun in the second sentence can refer either to the referent of the dative object *erine* (Salimat's husband) or to the referent of the subject of the previous sentence (Salimat).[99]

(160) *Salimat*$_i$ *erine*$_j$ *ash ete. Song,* ***o***$_{i/j}$ *savut-saba zhuĭe.*
Salimat husband.3POSS.DAT food make.PR later 3.2S dishes wash.PR
'Salimat fixes her husband a meal. Later s(he) washes the dishes.

Based on examples like (160), the pronoun *o* is not automatically interpreted as not coreferent with the subject of the previous sentence. Corpus data, likewise, provide evidence of a number of overt subject pronouns that have referents that are mentioned as subjects of the previous sentence.[100] (While 'not the subject' is by definition a subset of 'not in focus', the Givenness Hierarchy category 'not in focus' is broader and is not solely defined on the basis of syntactic criteria.) Such data is, however, consistent with the idea that overt pronouns sometimes implicate that their referents are not in at the center of attention (in focus) and, hence, not the previous topic.

---

[99] It is also possible here that the pronoun refers to an entity not mentioned in the previous sentence.
[100] It should be noted that overt pronouns are used much more often than nulls in a non-subject position in reference to an entity which is mentioned as the subject of the previous sentence, as in (156), where *og"ar* refers to Bolat.

Within the GH model, the apparent association in a number of languages, including Kumyk, between overt pronouns and topic-shift is one of the natural consequences of the implicature 'not in focus' which can be attributed to any form which signals 'activated' as part of its conventional meaning. In a study of pronominal subjects in Turkish, Enç (1986) also suggests that the signaling of topic shift is not the most basic function of overt or 'semantically redundant' pronouns, but that the indication of contrast is the primary function of overt pronouns and topic shift is a sub category of this function (204-206). [101] Enç explores the idea that the effect of the use of an overt pronoun when a null argument would be more efficient ("parsimonious") conveys "extra pragmatic information" which can be derived based on Grice's Maxim of Quantity, but rejects it, presumably because she does not have an overall framework within which it can be situated. Viewed within the GH model, the idea that overt forms versus nulls can be associated with the scalar implicature 'not in focus' based on the Maxim of Quantity follows as a consequence of the unidirectional entailment of cognitive statuses within the theoretical model. In cases where the implicature is blocked, the use of a lower form on the hierarchy (in this case, an overt form) may still be said to convey added contextual effects (see also Mulkern 2003) such as contrast or unexpectedness.

[101] Enç calls thess pronouns semantically redundant since their referent is recoverable from person and number marking on the verb.

5.8.2. Implicatures, frequency, and specialization

While the implicature 'not in focus' may be attributed to any form signaling activated, my analysis of the distribution of these forms suggests that this implicature is more often associated with one particular pronoun – that is the form *sho* – than any of the other forms signaling 'activated'. While *sho* is rarely used for in focus referents, *o* and *bu* are often used in such contexts, including contexts where there are competing referents.[102] The pronoun *o* is frequently used in the corpus study and in various parts of the questionnaire (including sentence continuations) in reference to an in focus entity including those introduced by the subject of a previous sentence (see section 5.5), as well as other referents that are clearly in focus. The form *sho*, on the other hand, refers only once to the subject of a previous sentence in the corpus and rarely refers to a referent categorized as in focus based on any other criteria. The questionnaire data likewise suggests that the use of this form with referents mentioned previously as subjects is limited or even unacceptable with some speakers (cf. section 5.6). The fact that *sho* consistently prefers entities that are mentioned as non-subject arguments of the previous sentence and are not in focus is strong evidence that this form frequently gives rise to the implicature 'not in focus', even suggesting that the association between *sho* and this implicature has become conventionalized for some speakers.

---

[102] The assumption based on the interaction of the GH model with pragmatic principles is that, scalar implicatures are most likely in cases with competing referents; hence it is unusual for a form which signals 'activated' to not create the implicature 'not in focus' in a context with multiple referents with the same minimal status.

I use Section 4 of the questionnaire to verify respondents' intuitions about what the use of *sho* (in contrast to *o*) implicates about the referent in contexts with multiple possible referents of different statuses. In four pairs of sentences constructed to allow for reference ambiguity, if a form creates an implicature 'not in focus', then respondents should consistently pick out the entity which is not in focus as the referent of that form. For each sentence pair there are two versions which differ only in whether *o* or *sho* is used in the second sentence, and respondents are asked to identify the referent of the pronoun. (Although these are shown here in summary form, in the questionnaire the versions using *o* versus *sho* were interspersed in random order rather than shown side by side.) In examples (161) - (164) each pair includes at least one possible referent coded as in focus and at least one that is at most activated. Some of these examples also include referents which have the same cognitive status but differ in animacy features. In example (165), the two possible referents have the same cognitive status, but one is introduced by a clause while the other is introduced by a nominal.

Different examples differ in the degree to which there is a correlation between the use of *sho* and the identification of the referent as one of lower cognitive status. In (161), for example, the pronoun in the second sentence can refer either to Bolat, mentioned as the subject of the previous sentence and in focus, or to the horse, mentioned as the dative object of a subordinate clause of the previous sentence and considered to be at most activated.

(161) *Bolat atyna minip barag"anda ĭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

Og"ar/Shog"ar *bolg"an* *zat* *boldumu?*
PRO.DAT happen.3PST.PRT thing happen.3PST
'Bolat fell off while riding his horse. Did anything happen to PRO?'

When *o* is used, all five respondents identify the referent as Bolat, but when the pronoun *sho* is used, three respondents identify the referent as Bolat, and two, as the horse. The fact that three respondents identify the referent of *sho* as a referent in focus show that this form is not restricted to referents that are not in focus—in other words, that 'not in focus' is an implicature, not part of the conventional meaning of the form. The fact that the implicature does not arise in these cases may be due to a strong natural tendency for the reader to expect the second sentence to be about Bolat, the subject of the previous sentence and the most likely object of concern in most people's reading of the sentence. In light of the tendency for topic continuity, the two responses which identify the referent of *sho* as the horse are significant, showing the strong association between this form and the implicature 'not in focus' (even though they do not represent the majority response to this test sentence).

The responses to example (162) show only a slightly stronger correlation between the use of *sho* and the entity with lower cognitive status (activated).

(162) *Muallim studentge* *kitap* *berdi.*
teacher student.DAT book give.3PST
'The teacher gave the student a book.'

O/Sho *baryn da* *tamasha etdi.*
3.2/3.3 everyone surprise do.3PST
'It/(s)he surprised everyone.' (Alternately, '(S)he was surprised by everything.')

In (162) four respondents identify the referent of *o* as the teacher, while one respondent (red) identifies the referent as the student (see alternate free translation). Only two respondents identify the referent of *sho* as the teacher, with the remaining three identifying the referent as the book (twice) or the giving of the book.

In (163), where the pronoun can refer either to the previous subject, *Ansar* (in focus), or the situation involving the giving the apple (activated), the use of *o* versus *sho* correlates strongly with the cognitive status of the referent.

(163) *Ansar anasyna alma berdi.* ***O/Sho*** *rag"mulu edi.*
Ansar mother.3POSS.DAT apple give.3PST PRO s weet be.3PST
'Ansar gave his mother an apple. PRO was sweet.'

All five respondents identify Ansar, an in focus referent, as the referent of *o*, and four out of five identify the referent of *sho* as the activated entity—the giving the apple (the fifth respondent identifies the referent of *sho* as Ansar).

Example (164), which is based on an excerpt from the text corpus, is a bit more complicated in the sense that the context contains four possible entities for reference at the time of utterance of the final clause: 1) the woman (in focus), 2) the storekeeper (in focus), 3) the accordion (in focus), and 4) the money (activated).

(164) *Song k"atyngishi tiukenge gire. Tiukenchi bulan sĕĭleĭ;*
later woman store.DAT come.in.PR storekeeper with speak.PR

og"ar arg"anyn sata. Ø(accordian) Satyp,
3.2.DAT accordian.3POSS.ACC sell.PR sell.GER

ak"chany da alyp, ***og"ar/shog"ar*** *k"aramady.*
money.ACC also take.GER PRO look.at.3PST
'Later the woman goes into the store. (She) talks with the storekeeper; (she) sells him the accordion. Having sold it, taking the money, she did not look at it(organ or money)/him.'

In testing the pronoun alternation in the final sentence, I assume that the referent will be identified as one of three previously mentioned entities: the storekeeper, the accordion, or the money.[103] When the pronoun *o* is used in the final sentence, all four respondents identify the referent as the storekeeper, the only human referent that is in focus. When the pronoun *sho* is used, the responses are divided between the storekeeper (once), the money (once) and the accordion (twice).[104] While there is only one respondent who identifies the referent of *sho* as an activated entity, it is interesting that, of the two possible referents in focus, two respondents select an inanimate entity rather than the human entity, suggesting that the use of *sho* communicates a lower-ranking entity in terms of animacy or cognitive status for at least three of the four respondents.

The questionnaire includes one additional sentence pair in which the referent of the form *o* or *sho* could be ambiguous but in which the context is biased towards two possible referents of identical cognitive status (both in focus). In (165) one possible referent of *olany/sholany* (plural accusative forms) is the food, which is introduced as a nominal (in subject position) in the previous sentence and the other is a situation introduced by the previous clause – the ruining of the food.

---

[103] Note: In one interpretation, which is not considered in the results, the respondent assumes that the storekeeper was selling the accordian and the storekeeper is also the subject of the last clause.
[104] The one respondent who selected the money put the accordian as a second choice in parentheses.

(165) *U'ĭune girip k"arasa: onda këp ullu k"almag"al bar.*
home.DAT go.in.GER look.PR.3COND 3.2.LOC much big mess exist.PR

G'ar tiurliu ashlar tashlang"an, buzulg"an.
each various food.PL be.thrown.down.3PST be.ruined.3PST

Olany/sholany *o'ziuniu anasy etgen.*
PRO self.GEN mother.3POSS do.3PST
'Coming into the home if (he/she) looks, there is a big mess there. Various foods are thrown down, ruined. His/her own other did/made them/these things.'

Four respondents identify the referents of both *o* and *sho* as the food, while one respondent (red) identifies the referent of both forms as the ruining of the food. These responses indicate no difference in reference identification due to the alternation of *o* and *sho*, even though the two possible referents are mentioned among the responses. These responses suggest that the alternation in form is not associated with implicature because the possible referents are not ranked in relation to one another.

Though the results are somewhat variable, there is enough data to confirm that *sho* often implicates a referent with lower cognitive status in relation to another possible referent. At the same time, the data described in this section indicates that *o* is not consistently used to implicate that a referent is not in focus. In fact, given a choice between a referent in focus and a referent that is at most activated, *o* is more likely to refer to the referent in focus.

Though the pronoun *bu* is not tested in the same way as *o* and *sho* (primarily due to my hypothesis based on the corpus study that *bu* signals 'in focus') there is interesting data in the questionnaire related to sentences with multiple pronouns which shows that *bu*, like *o*, is more likely to refer to an entity in focus, given a focus-activated referent

pair. Data from section 3 of the questionnaire shows that, in sentences in which more than one overt pronoun is used, *bu* consistently refers to an entity that is higher ranking than the entity referred to by *o* or *sho*, suggesting that *bu* has a particular association with referents of higher relative prominence. Out of 10 sentences with multiple blanks, there are 8 sentence pairs which contain at least one referent in focus and one that is coded as at most activated, one sentence pair where both referents are in focus, and one sentence pair where both are at most activated.[105] The responses to these 10 test sentences contain 11 instances of the use of *bu*, and, for each instance, I look at each possible pairing of the referent of *bu* with another referent. In other words, in a sentence with three blanks in which *bu* is used once, the referent of *bu* can be paired with two different referents. There are 16 such pairings of *bu*, and the data on these pairings is summarized in Table 7, where the first column indicates the example number in the questionnaire, the second column gives the pair of forms that one or more respondents used to fill-in-the-blanks, the third and fourth indicate the cognitive status and animacy features of the referent of the form, and the last column summarizes the type of ranking based on cognitive status or, where cognitive status is equal, based on animacy features.

[105] As previously described, all entities mentioned as subjects in the control sentence are coded as in focus and non-subjects mentioned in the control sentence are coded as at most activated.

Example # in questionnaire	Pronouns	Cognitive Status	Animacy	Ranking
Different Maximum Cognitive Status				
30	*bu*	FOC	Human	*bu* > *o*
	onu	ACT	Human	FOC > ACT
32	*bu*	FOC	Human	*bu* > *o*
	o	ACT	Human	FOC > ACT
34 b (blue)	*bu*$_2$	FOC	inanimate	*bu* > *o*
	o	ACT	inanimate	FOC > ACT
34 b	*bu*	ACT	Inanimate	**o* > *bu*
(2 - red + black)	*sho*	FOC	Inanimate	FOC > ACT
37	*bu*	FOC	Human	*bu* > *sho*
	sho	ACT	Inanimate	FOC > ACT (also hum > inanim)
39 a	*bu*	FOC	Human	*bu* > *o*
	o	ACT	animate (horse)	FOC > ACT (also hum > anim)
38 a	*bu*	ACT	animate (horse)	**o* > *bu*
	o	FOC	Human	FOC > ACT (also hum > anim)
Same Maximum Cognitive Status				
31	*bu*	ACT	Human	*bu* > *shu*
	shu	ACT	Inanimate	hum > inanim
33	*bu*	ACT	Human	*bu* > *sho*
	sho	ACT	Inanimate	hum > inanim
34 a (blue)	*bu*$_1$	ACT	Human	*bu* > *o*
	o	ACT	Inanimate	hum > inanim
34 a	*bu*	ACT	Inanimate	**o* > *bu*
(2 - red + black)	*o*	ACT	Human	hum > inanim
38 b	*bu*	ACT	animate (horse)	*bu* > *sho*
	sho	ACT	inanimate (fact)	anim > inanim
39 b	*bu*	FOC	Human	*bu* > *sho*
	sho	FOC	Inanimate (event)[106]	hum > inanim
41	*bu*	FOC	Human	*bu* > *sho*
	sho	FOC	inanimate	hum > inanim

Table 7: *Bu and ranked entities*

[106] This entity could alternately be interpreted as a situation, which would change the pair to the category 'different maximum cognitive status' but would support the hypothesized ranking.

In 8 of the form pairs from the data, the referent of *bu* has the same maximum cognitive status as the other referent but different animacy features, while in 8 of these pairs the referent of *bu* has a different maximum cognitive status. Of the latter cases *bu* refers to the entity in focus 5 times, while in 3 cases (two of which are different responses to the same example) *bu* refers to the lower-ranking or activated entity. In one of the pairs (example 38) where *bu* refers to an entity that is lower ranking in cognitive status, the entity is higher ranking in terms of animacy.

In the 8 pairs where referents have the same cognitive status but can be ranked in terms of animacy features (human > animate > inanimate), *bu* refers to the higher-ranking entity in 6 cases (5 of which are human vs. animate/inanimate and one of which is animate vs. inanimate). The two cases where this ranking does not hold are the same responses (example 34, red and black) which resulted in two of the cases which do not follow cognitive status ranking, suggesting that there is possibly some other factor associated with the use of *bu* in this example. To summarize this analysis, the majority of uses of *bu* in sentences with multiple pronominal forms correlate with a higher-ranking referent in terms of cognitive status where the maximum cognitive status is different, or in terms of a hierarchy of animacy where maximum cognitive status is equal.

The proposal that the use of *bu* prompts the hearer to select the higher-ranked among multiple possible referents is somewhat outside the structure of the Givenness Hierarchy model. The GH model provides for the indication of the higher-ranking referent only via the use of forms restricted to a higher status. In other words, if *bu* is restricted to use with forms in focus, then it makes sense that it is used with the higher-

ranked of two forms, one of which is in focus and one of which is activated. On the other hand, if, as the questionnaire data indicates, *bu* is not restricted to referents in focus, then the preference of *bu* for higher-ranking entities can only be described in terms of the implicature created by the use of a different form (such as the strong correlation of *sho* with the implicature 'not in focus'). Since most of the uses of *bu* are paired with *o* rather than *sho* in my data, I would need to claim that the use of *o* in a sentence where *bu* is also used implicates that the referent is not in focus or is the lower-ranked of two entities of the same status, while the use of *o* in conjunction with *sho* does not implicate the same thing (since previous analysis indicates *sho* is the form most likely to be associated with the implicature 'not in focus'). An alternate and more reasonable possibility is that *bu* has a specialized function as an indicator of the higher-ranking of two entities which are at most activated (which is not contradictory to the GH model) and that this function is sometimes generalized to include indication of actual cognitive status ranking (or perhaps in cases where nulls or reflexives cannot be used for some reason).

To summarize this section, I claim that the overt pronouns *bu*, *o*, *sho*, and *shu* can all, to some extent, implicate 'not in focus' at the intersentential level. This implicature, however, is most often associated with the use of *sho*. The form *bu* appears to have a specialized function of indicating the higher-ranked entity among entities with the same minimal status, which leaves *o* as the neutral or unmarked form.

5.8.3. Ranking data

As a conclusion to this chapter, and, in particular, as a follow-up to discussion of scalar implicatures, this section provides a more complete discussion of the ranking

effects created by the use of different forms of referring expressions. In natural Kumyk discourse, it is rare for two identical pronoun forms to be used to refer to two different entities in the same clause or sentence.[107] While, in many cases, a pronoun is used for one entity and a full noun phrase for the other, there are also a number of instances where two different pronoun forms are used for two different entities, and it is often the case that the forms chosen provide some clue which allows the speaker to disambiguate among potential referents of the pronoun. Assuming that nulls and reflexives signal that a referent is in focus, that *bu*, *o*, *sho*, and *shu* signal that the referent is at least activated and to varying degrees implicate 'not in focus', and that *sho* specializes in indicating the lower-ranked of a set of entities while *bu* specializes in indicating the higher-ranked of a set of entities, I hypothesize that speakers use a combination of the juxtaposition of forms associated with different statuses, implicatures, and specializations in order to create ranking effects. More specifically, if one referent is in focus and the other at most activated or if two referents are in focus, a ranking effect is created by using nulls or reflexives (based on cognitive status restrictions) to refer to the highest-ranked—that is the referent with the higher cognitive status or the more prominent of two referents in focus, while *o*, *sho*, and *bu* are used for the lower-ranked referent (in terms of cognitive status) or the less prominent of the entities in focus. The ranking effect could also be

---

[107] In English, it is more common to use two different pronouns in the same sentence, perhaps due to the fact that different pronominal forms allow the hearer to disambiguate referents based on gender or animacy features, as in the following sentence:

> *Paul told Barbara to bring her cat inside because he wanted her to feed it.*

created by using *sho* to refer to the less prominent entity in juxtaposition with another form that signals 'activated'—that is *o*, *shu*, or *bu*—because of the specialized function of *sho* in implicating 'not in focus'. Assuming that *bu* also specializes in indicating the more prominent of two possible referents, then the juxtaposition of *bu* with another form that signals 'activated' would also create a ranking effect. If two referents are at most activated, the ranking effect can only be created by juxtaposing one of the two specialized forms, *sho* or *bu,* with another form that signals 'activated', since the in focus forms cannot be used in reference to entities whose maximum status is activated. Figure 6 summarizes the pairs which can create ranking effects based on these principles. Note that pairs which use both a form restricted to in focus referents and *sho*, which specializes for lower-ranked referents, are somewhat redundant. Also, the pairs which consist of a form that signals in focus and *bu* are somewhat anomalous, and, thus, not expected or attested in the data, since one would be signaling the more restricted cognitive status and the other is associated with the higher-ranking entity—in other words, both forms would be pointing to the same entity.

Ranked Pair	Basis for Ranking
Ø - bu	Cognitive status restriction + specialization for higher-ranking referent
o'ziu – bu	
Ø - o	Cognitive status restriction
o'ziu - o	
Ø - sho	Cognitive status restriction + specialization of *sho* for lowest-ranking referent
o'ziu - sho	
o - sho	specialization of *sho* for lowest-ranking referent
bu – sho	
bu – o	specialization of *bu* for higher-ranking referent

Figure 6: *Ranked pairs*

To test these ranking claims, I analyze ten fill-in-the-blank exercises in Section 3 of the questionnaire with more than one pronoun (blank)[109]. (A portion of this analysis is discussed previously in the discussion of *bu* in section 5.8.2.) The questionnaire results support the idea that Kumyk speakers infrequently use two identical pronouns to refer to different entities in the same sentence. For the six cases with two blanks, out of the 24 responses examined, only two use identical pronouns for different referents. Likewise, out of the 16 responses for the four cases with three blanks, only three use identical pronouns for different referents.[110] In all ten responses there are 37 cases where 2

[109] In this portion of the questionnaire I consider it crucial that the respondents identify the referent of the pronoun; therefore I eliminate one set of data where the respondent does not do this.
[110] There are 2 additional cases in which the same form is used for two different mentions of the *same* referent.

different forms are used for a pair of referents with different maximum cognitive statuses (including the 2 cases where the same referent is mentioned twice and has a different cognitive status at the time of the second mention) and 23 cases where 2 different forms are used with referent pairs with the same maximum cognitive status (all of which have differences in animacy features). Of the 37 pairs of forms with referents ranked for cognitive status, 26 (70.3 %) correlate with the possible rankings listed in Figure 6, with the most common form pair being *o – sho* (13 instances, or 35.1 %). Of the 23 pairs of referents with the same maximum status, 19 (82.6 %) correlate with the following animacy ranking: human>animate>inanimate. In the latter case, the *o – sho* pair is also the most frequent (11 instances, or 47.8 %). Interestingly, there are 5 unexpected uses of null arguments with entities assumed to be at most activated (3 from pairs where one referent is in focus and one at most activated and 2 from pairs where both referents are activated).

The results from section three of the questionnaire provide evidence for ranking effects derived from the signaling of cognitive status, implicatures and specializations (as described above). The analysis of pairs of referents with the same maximum cognitive status indicates that animacy features also play a role in ranking and that the rankings which derive from form-status correlations and their related implicatures can be extended to other types of categories which involve some type of relative prominence – in this case animacy categories. It is also possible that, rather than being an extension of cognitive status rankings, rankings that correlate with animacy features actually indicate that

animacy contributes to salience and, therefore, cognitive status (i.e. the likelihood that something will be in focus).

To demonstrate these generalizations, I discuss the following types of examples: those in which two referents have different maximum statuses but have identical animacy features, and those in which pairs of referents have different maximum cognitive statuses and different animacy features. Finally, I discuss pairs of referents that have the same minimum and maximum cognitive status, but have different animacy features.

Example (166) illustrates the use of four different form pairs in reference to two entities that are both human but have different cognitive statuses.

(166) *Muallim studentge kitap berdi. Song* ___ ___
teacher student.DAT book give.3PST then

ekzameni bulan k"utlady.
exam.3POSS with congratulate.3PST
'The teacher gave a book to the student. Then (s)he congratulated him/her on the exam.' [numbered as #30 in questionnaire]

Responses	
teacher (FOC)	Student (ACT)
bu	*onu*
o'ziu	*onu*
Ø	*onu*
ol	*onu*

All four respondents select the teacher as the referent associated with the first blank and the student as the referent associated with the second blank. Based on the coding of these referents according to the Givenness Hierarchy criteria, the teacher is in focus and the student is likely to be at most activated. In two out of four responses, respondents use a

pronoun from the focus set (one each of null, reflexive) to refer to the teacher and the pronoun *onu* (accusative case of *o*) to refer to the student. One respondent uses *bu* juxtaposed with *o*, which is also one of the predicted rankings based on the proposed specialization of the activated form *bu*. The fourth respondent uses *o/onu* for both references. The responses in this example support the proposed ranking effects.

Example (167) also illustrates two mentions (associated with the first and third blanks) of referents with different maximum cognitive statuses but identical animacy features. This example is noteworthy in the sense that it illustrates a ranking effect between the first and second pronominal mention of the same referent.

(167) Muallim Patimatg"a kitap berdi. __ __ këp ushatdy .
teacher Patimat.DAT book give.3PST very pleased

Song, __ *u'ïge getdi.*
later home.DAT go.3PST
'The teacher gave Patimat a book. She was very pleased with it. Then she went home.' [numbered example #33 in questionnaire]

Responses		
Patimat$_1$ (ACT)	(book)	Patimat$_2$ (FOC)
bu	(*shonu*)	*bu*
sho	(*Ø*)	*ol*
o	(*shonu*)	*Ø*
ol	(*shonu*)	*ol*

The responses to this example are as follows: 2 respondents use the same form, and 2 use ranked forms. The fact that ranked pairs of forms are used in reference to the same entity is strong evidence that the rankings derive from cognitive status differences, since the cognitive status of the entity would be higher at the time of the second mention. It is also

interesting (and somewhat unexpected) that, among all the responses where two references to humans are juxtaposed, *sho* is used only once in reference to a human entity. In most cases involving two human referents with different cognitive statuses, *o* is used for the lower-ranking entity.

Example (168) also contains 2 mentions of the same referent—in this case the book associated with the first and third blanks. As in the previous example, the book is considered to be at most activated at the time of mention associated with the first blank and in focus at the time of mention associated with the second blank.

(168) *Muallim Patimatg"a kitap berdi.*
teacher student.DAT book give.3PST

___ *birinchi savg"at bolg"an song,* ___ ___ *anasyna*
first gift be.3PST.PRT since mother.3POSS.DAT

gërsetme siuedi.
show.INF want.3PST
'The teacher gave Patimat a book. Since it was her first prize, she wanted to show it to her mother.' [corresponds to #34 in questionnaire]

Responses		
book$_1$ (ACT)	(Patimat) (ACT)	book$_2$ (FOC)
o	*(bu)*	*munu*
sho	*(Ø)*	*onu*
bu	*(o)*	*shonu*
bu	*(ol)*	*shonu*

In the pairs of forms associated with these mentions there is some evidence for ranking and some counter-evidence. The responses of one respondent support the expected ranking pairs *sho* (referent at most activated) – *o* (referent in focus), and the response of

another suggests *o* (referent at most activated) – *bu* (referent in focus) ranking; however, two respondents use *bu* for the first pronominal mention of the book and *sho* for the second. If one accepts the hypothesis that humanness contributes to prominence, then the unexpected use of *sho* in the second mention of the book could be explained by the fact that it contrasts with a pronominal reference to a human in the same clause.[111]

There are three cases in the questionnaire in which the cognitive status ranking coincides with an animacy ranking (human > non-human). Each of these cases is based on an example with the same first sentence, shown in (169) (a), so the referent ranking is identical: Bolat, the human referent, is in focus; the horse, a non-human animate entity is activated; and the fact that Bolat fell from the horse is also activated. Whether the comparison is between Bolat and the horse or Bolat and the fact that he fell, in each case Bolat is a referent that is ranked higher in terms of both cognitive status and animacy. Out of twelve possible pairings of Bolat and the horse (4 of which are displayed in (169) (b), eight use ranked form pairings which correspond to the cognitive status/animacy ranking, three use the same form (*o*) for both referents, and only one uses a form pairing which contradicts the cognitive status/animacy ranking (*o* – *bu*). When Bolat is paired with the fact that Bolat fell, four out of four responses have ranked pairs of forms which correspond to the cognitive status and animacy rankings.

(169) a. *Bolat* *atyna* *minip* *barag"anda* *ĭyg'yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

[111] Alternately, the use of *bu* in the first clause of the second sentence could possibly be attributed to other pragmatic factors, such as the focus structure of the equative clause.

b. ___ *saialy,* ___ ___ *satma khyial etdi.*
because sell.INF decision make.3PST
'Bolat fell off while riding his horse. Because of that, he decided to sell it.'
[corresponds to #38 in questionnaire]

Responses		
fact – falling (ACT)	Bolat (FOC)	horse (ACT)
sho	*o*	*munu*
sho	*Ø*	*onu*
sho	*o*	*onu*
sho	*ol*	*shonu*

Example (170) involves two non-human referents which are ranked in terms of cognitive status and also differ in terms of animacy: one is an event (inanimate), which is in focus, and the other is a horse (animate), which is activated.[112]

(170) *Bolat atyna minip barag"anda ĭyg"yldy.*
Bolat horse.3POSS.DAT ride.GER go.PR.PRT.LOC fall.3PST

___ *song* ___ *dag"y* ___ *minip barmag"an.*
after again ride.GER go.NEG.3PST
'Bolat fell off while riding his horse. After that he wouldn't ride it anymore.'
[corresponds to #39 in questionnaire]

[112]The respondents' identification of the referent in this example allows for some ambiguity between an event (in focus) and a situation, which would be considered at most activated. I have treated this as a reference to an event, but if the reference is, in fact, a situation that is introduced indirectly by the mention of an event, then this example would not be a counter-example to my ranking hypotheses, as this would be a ranking of two entities with the same maximum status and, thus, animacy ranking would be expected to play the primary role here.

responses	
event (FOC)	horse (ACT)
sho	*o*
Ø	*sho*
sho	*o*
sho	*sho*

If the only feature considered is cognitive status, two of these responses would be a counter-example to the ranking prediction, as *sho* is used in reference to an event (in focus), and a higher-ranking form is used in reference to the horse (at most activated). These responses, however, could be explained in terms of ranking if one proposes that the animacy ranking takes precedence over the cognitive status ranking, in which case the animate horse would be ranked higher than the inanimate event.

In the example illustrated previously as (168) the second and third blanks (associated with Patimat and the book, respectively) represent a pairing of mentions of two referents that are ranked one way in terms of cognitive status and ranked the opposite way in terms of animacy. In this case, the book, an inanimate entity in focus, is paired with a human entity assumed to be at most activated—that is, Patimat. While one response uses the same form (*bu*) for both entities, the remaining three form pairs support the animacy ranking rather than the cognitive status ranking, suggesting that animacy rankings can be more important than cognitive status rankings in some cases.

The correlation between ranked pairs of forms and animacy rankings is also supported by the analysis of pronominal pairs representing entities of the same maximum status (both in focus or both at most activated). In most of the examples in this category a human referent contrasts with an inanimate referent introduced by a nominal, as

illustrated in (171). (See also the book in contrast with Patimat in examples (167) and (168), as well as the *shu – bu* contrast in (122).)

(171) *Muallim Patimatg"a kitap berdi. U'ĭge gelgen song* ___
teacher student.DAT book give.3PST home.DAT come.PST.PRT after

g'ak"ynda anasyna aĭtdy.[113]
about mother.3POSS.DAT say.3PST
'The teacher gave Patimat a book. After (s)he came home, ((s)he) told his/her mother about PRO.'

Responses	
Patimat (ACT)	book (ACT)
ol	*shonu*
o	*shonu*
ol	*shonu*

In one case a human referent contrasts with an inanimate referent introduced directly by a clause (the event of Bolat falling off his horse. Example (169) illustrates a case where an animate entity (horse) contrasts with an inanimate entity introduced indirectly by a clause (the fact that Bolat fell off the horse).

As previously stated, of the 23 responses that use ranked pairs of forms, all but 4 support the ranking of entities according to animacy. Of the 4 exceptions, 2 involve a null argument in reference to the entity ranked lower in terms of animacy (cf. (167) and (170)) and 2 cases involve the use of *bu* in reference to the book in (168), (discussed previously). One interesting observation is that, of the 19 pairs which correlate with animacy ranking, 17 use *sho* for the lower-ranking entity in contrast with another form.

[113] The blue respondent fills in the second blank with the form *munu* and identifies the teacher as its referent.

Even though *sho* is not restricted to use with inanimate entities for all speakers, this fact could explain why, for certain speakers, the contrast between *sho* and *o* might become conventionalized as a non-human – human contrast (cf. section 5.6).

In summary, the data discussed in this section support the use of ranked form pairs based on form-status correlations, implicatures, and the specialization of *bu* and *sho*. The rankings of the entities to which these forms refer are demonstrated to correlate, not only with cognitive status differences, but also with differences in animacy features. As a matter of fact, the correlation between form pairs and animacy rankings for entities with the same maximum cognitive status is more consistent than the correlation between form pairs and cognitive status rankings, though the fact that the data set for the former category is more limited than the latter category also plays a role in the statistical outcome. One cannot, however, draw the conclusion that animacy rankings are the only or primary factor involved, as there are clear examples where ranked form pairings correlate with referents which share the same animacy features but have different cognitive statuses. Furthermore, animacy rankings can be interpreted as affecting the inherent prominence of an entity, while cognitive status can potentially be interpreted as a result of both inherent prominence and contextual prominence. The role of animacy features in relation to cognitive status is an important area for future study in Kumyk and in other languages where animacy distinctions are not grammaticized as different pronoun sets.

An interesting observation in this portion of the questionnaire analysis is the frequency with which *o* - *sho* pairs occur in the data: 24 (or 40.0 %) out of the total of 60

ranked form pairs. With the addition of null – *sho* and *bu* – *sho* pairs, there are 33 (or 55 % of the total) ranked form pairings involving the use of the form *sho*. In all cases except one where *sho* is used with referents of different cognitive status, it refers to the referent of lower status. In the one exceptional case, animacy features seem to play a stronger factor, as *sho* represents an event, while *bu* and *o* represent animate entities (human, horse, see (170)). In situations with referents of the same status, whether both are in focus or both are at most activated, in all but one case *sho* represents the entity which is lower on the animacy hierarchy (human, animate non-human, inanimate). This data strongly supports the hypothesis that *sho* is used to implicate the lower of two ranked entities.

While the questionnaire provides evidence for these implicatures primarily in other contexts with multiple pronominal references, I assume that these implicatures are not dependent on whether or not more than one pronoun is used. The "specializations or strong implicature associated with *sho*, even in cases where there are not necessarily two pronouns used, is confirmed by evidence from sentence continuations. Kaiser and Trueswell 2004 suggest that continuation tests provide evidence that the choice of proform (from among the set of proforms possible in the language) allows the hearer to identify the referent of that form from among a set of multiple possible referents. In other words, when a respondent is given a context sentence and asked to write a continuation using a particular proform, the choice of form is assumed to influence the topic of the continuing text. This type of data is particularly interesting in Kumyk, where there is no morphological distinction between animate and inanimate entities. The

detailed results of the continuation tests in Kumyk, which are discussed in section 5.6 (cf. examples (140) - (142), show that continuations with the pronoun *sho* overwhelmingly select the referent of the pronoun as an inanimate referent that is not likely to be in focus (more specifically, not the previous subject).[114]

5.9. Conclusions

The evidence from the questionnaire confirms the hypothesis based on the corpus study that at least two pronominal forms in Kumyk are restricted to referents in focus: reflexives and nulls. As for the form *bu*, while the corpus study suggests this form is restricted to referents in focus, the questionnaire data disconfirm this hypothesis. Nevertheless, the data indicate that *bu* is more restricted in use than the other demonstrative forms and exhibits a preference for reference to the more prominent in a set of entities.

The questionnaire data also clearly demonstrate the varying degrees to which the overt demonstrative pronouns give rise to the implicature 'not in focus'. The analysis shows that *sho* has a strong association with this implicature in comparison to *bu* and *o*. Moreover, in spite of the strong association with the implicature 'not in focus', evidence from the questionnaire demonstrates that there are no absolute constraints on its use with higher status forms, supporting the predictions of the Givenness Hierarchy model.

---

[114] Unexpectedly, in the sentence continuation portion of the questionnaire, the results for *bu* are somewhat variable or show a weaker correlation with the most prominent referent if one judges mainly by the syntactic prominence of previous mention.

Like the corpus study, the questionnaire data contain very few mentions of the pronoun *shu*. Speaker judgments indicate that its use is somewhat unacceptable (for some, but not all, speakers). In the few cases available for analysis, the use of this form is characterized in a very similar way to that of *sho*. The evidence both from frequency and from speaker judgments suggests that, as in some other Turkic languages, the distinction between *sho* and *shu* may soon be lost.

One of the most important contributions of the questionnaire is that it provides evidence of the psychological reality of scalar implicatures created by the differences in cognitive status restrictions, as well as the specialization of *bu* for higher-ranking entities (which cannot necessarily be derived from cognitive status differences based on the questionnaire evidence). The discussion both of the use of ranked forms and the frequency with which sentence continuations for pronouns signaling 'activated' pick a lower-ranking referent provides evidence of ranking effects in referent disambiguation.

While previous research within the GH model focuses on the role of previous mention and features related to the speech context in relation to the cognitive status of an entity, data from the questionnaire suggest that an inherent feature of an entity—that is, animacy characteristics—may also play a role in cognitive status (or at least relative prominence) and choice of form. Unlike English, where the gender and animacy categories each have their own pairs of pronouns representing 'activated' and 'in focus', in Kumyk, where animacy is not grammatically marked in pronoun form, one set of pronouns must be used to represent both cognitive status rankings and potential rankings derived from differences in animacy features. In my analysis of ranked pairs, the

correlation with animacy rankings is even stronger than the correlation with cognitive status, which suggests that animacy features may play a role in the cognitive status of an entity that is not reflected in the coding criteria of the GH model.

The relationship between animacy features and the use of different categories of referring forms is addressed to some extent by Dahl and Fraurud 1996 and Fraurud 1996. While both studies address the idea that human referents are more likely to be referred to with pronouns than non-human referents, Dahl and Fraurud relate this propensity to the fact that humans are more likely to be topics than non-humans, as well as the fact that narrative point of view represents a human perspective (1996:59-62). Fraurud 1996 takes the idea a step further by giving some statistical evidence for the claim that, given the same degree of topicality or cognitive status, humans are more likely than non-humans to be referred to with pronouns. Specifically, Fraurud's data suggests that humans are more likely than non-humans to be referred to with pronouns when a referent is mentioned in the text earlier than the same or preceding sentence (1996:66-67).[115] A subsequent study by Prat-Sala and Branigan (2000) provides further evidence of the independent influence of so-called *inherent salience*, which is a result of animacy or other ontological features, and *derived salience* or contextual givenness—thought the latter is considered to be the stronger of the two (169, 177). Dahl 2008 justifies the relative cognitive influence of different ontological categories in terms of a developmental process which begins when a

[115] This could be related to the degree of topicality in larger units of text, as with the criteron from the coding for 'in focus' status, which refers to the role of an entity as a 'higher-level topic' that can be part of the interpretation of a sentence even when not overtly mentioned (Gundel 2004).

member of a species first becomes aware that there are other beings like itself in the world and eventually develops into a theory of mind. In Dahl's words, "the self is the model for other animate individuals, which are in their turn models for inanimate objects when understood as individual 'things'" (149).[116]

My analysis, likewise, assumes that both ontological features and contextually-derived features contribute to the cognitive status of an entity, which I support with a comparison of reference to entities with the same contextual salience (based primarily on the syntactic status of previous mention) but different animacy features. Moreover, as with Prat-Sala and Branigan's study, the evidence primarily shows that, in cases where ranking in terms of animacy features parallels ranking in terms of contextual salience, the correlations with ranked pairs of forms are more consistent than in cases in which animacy rankings are in the reverse order from rankings associated with the syntactic prominence of previous mention. However, further study in this area is needed—in particular experimental data which has been specifically structured to establish the independent affects of ontological features versus the form of previous mention in the context.

One of the limitations of the analysis of the questionnaire data is that analyzing constraints on the use of forms based primarily on the category of the argument position of previous mention is not very satisfactory and yields inconclusive results in testing the

---

[116] Note that Fraurud addresses the human-non-human distinction in terms of an ontological difference between individuated entities and "instantiations of types" (1996:72). In such case, one would expect the form *a woman* to represent an entity that is an instantiation, while *War and Peace* represents an individuated entity.

hypothesis that certain forms are constrained to referents in focus. As previously discussed, the criteria of the coding protocol are considered to be minimum criteria; thus, for example, even if a referent is mentioned in a non-subject position in the previous sentence, it is not easy to determine that a referent is at most activated. Since I cannot determine definitely that a referent mentioned in a particular non-subject argument position in the preceding sentence is at most activated, I look at where natural distinctions between the use of forms fall. For example, the fact that nulls are acceptable with referents mentioned previously as subjects and direct objects while not acceptable with referents mentioned previously as dative objects leads me to operate under the assumption that dative objects are at most activated. In light of this limitation of the methodology, I believe that, while I am able to reach sound conclusions about the relative degree of restrictiveness, it is possible that the line which separates the forms that signal 'in focus' and those that signal 'activated' is not entirely accurate. Such a statement is not intended to challenge the overall validity of the Givenness Hierarchy model, but merely to indicate that there are aspects of the coding process which need further consideration and which have already undergone some revision since the 2004 version used in my research.

Chapter 6: Implications of the Analysis

An integrated analysis of the results from the corpus study and questionnaire demonstrates both the restrictions and the preferences of each pronominal form in Kumyk. Section 6.0 summarizes the results in terms of the form-status correlations of the GH model and the predictions about scalar implicatures which follow from the unidirectional entailment of the hierarchy. In section 6.1, I discuss the adequacy of the theoretical model and methodology in demonstrating the psychological reality of form-status correlations and the ranking effects derived from them. Section 6.2 summarizes the results in terms of a generalized hierarchy of degree of restrictiveness, which takes into consideration multiple factors demonstrated to relate to the cognitive status of referents—both those which derive from contextual use and those inherent to the entity. In section 6.3 I discuss the ranking effects associated with contrasts in pronominal forms, and section 6.4 relates how the ranking effects associated with particular forms contribute to imposed salience. Finally, section 6.5 addresses the significance of the research to a number of issues of wider interest, as well as suggesting several areas of possible future research.

6.0. Form-status correlations and predictions of the GH model

The combined analysis of the corpus study and questionnaire data supports the following form-status correlations as part of the lexical meaning of pronominals: null and *o'ziu* signal that the referent is in focus, while the overt demonstrative pronouns *bu*, *o*,

sho, and *shu* signal that the referent is activated. Figure 7 provides a summary of proposed form-status correlations.

FOC	**ACT**
Ø, o'ziu,	***bu, o, sho, shu***

Figure 7: *Final Form-Status Correlations*

The most difficult form in the analysis is *bu*, which appears from the corpus data to be restricted to referents in focus and often indicates the higher of two or more ranked entities, but, in the questionnaire data, does not exhibit the same level of restriction as nulls and reflexives. On the other hand, *bu* patterns like nulls in being restrictive to a certain degree in the area of reference to clausally-introduced entities, unlike *o*, *sho*, and *shu*, which appear to be unrestricted in this area. The restriction in relation to clausally-introduced entities, however, could be partially explained by the fact that *bu* is usually associated with high relative prominence and, given the same cognitive status, entities introduced by clauses are typically of lower relative prominence than entities introduced by nominals. In the end, the proposal that *bu* signals the status 'activated' is more consistent with the data, although its association with entities of relatively greater prominence cannot be derived directly from this categorization.

In addition to the form-status correlations, there is evidence of further specialization in the distribution of demonstratives. Though in principle, any

demonstrative form can give rise to the implicature 'not in focus', the pronoun *sho* has a strong association with this implicature. Based on both the analysis of referents of *sho* in the corpus and the referent preferences of this form demonstrated in the questionnaire, I claim that this form specializes in indicating the lower of two or more ranked entities. Such a specialization is easily derived from the scalar implicatures inherent in the unidirectional entailment of the GH model. Finally, even though the majority of referents of *sho* are at most activated, the testing of this form demonstrates that, as predicted by the GH model, it can also be used with entities that are in focus.

In contrast to *sho*, the form *o* is often used without giving rise to the implicature 'not in focus' and is the least specialized of the demonstratives. This pronoun appears to be somewhat neutral and, in fact, occurs three times as frequently as *sho* or *bu*.

The pronoun *bu*, though not restricted to entities in focus, most often refers to the more prominent of two or more entities that are at least activated and can be considered to specialize in indicating higher relative ranking. This specialization cannot, however, be derived directly from the structure of the GH model. I assume that, given multiple possible pronouns that signal the status 'activated', it is a natural linguistic process for forms to become specialized for particular functions.[118] In particular, given the fact that pronouns that are restricted to referents in focus cannot be used to indicate the more

---

[118] Theoretically, my claim is similar to that of Gundel, Hedberg, and Zacharski 1993, who claim that the English pronoun *this* specializes in indicating speaker activation, not just activation in general.

prominent of two entities that are at most activated, it seems natural that one of the pronouns that signal activation would become specialized in signaling higher relative prominence. Also, in Kumyk, given the fact that pronoun referents cannot be disambiguated on the basis of animacy categories, it seems even more likely that there would be a more detailed way to disambiguate referents on the basis of relative cognitive status rankings—in addition to the categorical distinctions between the set of forms which signal 'in focus' and the set that signals 'activated'.

6.1. Adequacy of the GH model in this analysis

One of the prominent features of the GH model is that it addresses differences in the use of referring forms on the basis of cognitive status, while still allowing for other possible distinctions to be encoded by differences in form—for example, gender, animacy, or entity class/noun class. For Kumyk, where pronominals have no categorical differentiation based on gender or animacy, the evidence from this study suggests that choice of pronominal is based almost entirely on relative cognitive prominence; thus the GH model provides an excellent means of accounting for distributional differences in pronominal form.

In Chapter 2, I argue that the GH model is preferable to other models that explain the distribution of forms solely on the basis of syntactic criteria or linear distance because it also takes into consideration other factors that can contribute to the cognitive status of an entity, such as whether or not it is a higher level topic, whether it is directly or indirectly introduced, as well as non-linguistic factors such as gesture and eye gaze. Evidence from the corpus study suggests a few additional contextual criteria that

contribute to cognitive status, such as the repetition of parallel argument structures. One of the limitations of the GH model is that it does not specifically address the role of features that are inherent to an entity type, in particular, whether or not animacy characteristics contribute to the cognitive status of an entity.

One of the methodological challenges of analyzing the distribution of pronouns within the GH model is the difficulty in determining both the minimum and maximum status of the referent a form on the basis of the current coding criteria alone. More specifically, 99.8 % of pronoun uses occur with entities that are at least activated, with the only possible distinction being between entities that are at most activated and those that are in focus. Methodologically, it is very difficult to determine in a consistent way what the maximum status of a referent is, as the criteria for coding a referent as 'in focus' are intended to be sufficient, not necessary. The differences between the coding criteria for the statuses 'activated' and 'in focus' suggest that referents may possibly be at most activated if they are mentioned in the pre-previous sentence, but not mentioned in the previous sentence.[119] However, in my corpus, only 3 out of 410 coded pronoun tokens

---

[119] Based on the coding criteria alone (2004 version), referents that are possibly at most activated (referents satisfy a condition for the status 'activated' but do not meet the criteria for 'in focus') include the following:

> 1. Mentioned in the preceding sentence , but not mentioned overtly previously in the same sentence, not a higher-level topic that is part of the interpretation of a previous clause in the same sentence, not mentioned as a subject in any clause of the preceding sentence or as a higher-level topic that is part of the interpretation of the previous matrix clause--in short, a non-subject argument of the previous sentence that is not a higher level topic.
>
> 2. Mentioned in the pre-preceding sentence, but not mentioned overtly in the previous sentence or previously in the same sentence and not a higher-level topic that is part of the interpretation of the previous clause or sentence.

were in reference to entities not mentioned in the previous sentence and not coded as 'in focus' by some other criteria (such as higher level topic or my addition of script repetition). Since almost all referents of pronouns are mentioned in the previous sentence, the remaining possible distinction between those that are at most activated and those that are in focus lies in differences of syntactic position or the difference between an event, which is directly introduced by a clause, and facts, situations, propositions, and speech acts, which are most often introduced indirectly by a clause. For this reason, much of my questionnaire focuses on the different levels of acceptability of each form with referents mentioned in various argument positions. The Kumyk data suggest that a point of difference in acceptability for certain forms versus others (for example, nulls versus overt pronouns) occurs with entities mentioned in argument positions other than that of subject or direct object. While I make this distinction on the basis of empirical data alone, it also correlates with claims about the syntactic prominence of particular arguments in SOV languages.[120] While it is challenging to determine the line between forms requiring a referent in focus and those which only require an activated referent, the fact that acceptability judgments indicate different degrees of restrictiveness at all gives evidence that there are (at least) two different cognitive statuses on the hierarchy that may be encoded by pronouns..

---

[120] Note, however, that I cannot necessarily generalize this claim to other languages, particularly those with other word order categories.

Another challenge in my analysis is that, in principle, one should also be able to distinguish between forms that require entities in focus and those that require activation on the assumption that forms which require referents in focus can refer to a much narrower range of entities introduced by a clause than those that only require activation. More specifically, a form which requires a referent in focus can only refer to an event or state introduced by a sentence, while forms which require only activation can also refer to entities introduced indirectly, such as facts, propositions, or speech acts. The difficulty in my analysis is that both nulls and the overt pronoun *bu*, are very restricted in use with all clausally-introduced entities. In the case of *bu*, if this form is not restricted to referents in focus, one would have expected that it could refer to entities introduced indirectly by a clause; however, the fact that it is generally restricted to entities introduced by a nominal or by a clause with some degree of nominalization means that it is not possible to test this differentiation.

One of the contributions of this analysis is further exploration of a number of criteria in determining the cognitive status of an entity. More specifically, this study discusses several potential categories of syntactic prominence in addition to subjecthood that bring a referent into focus: objects in SOV languages, topicalized elements, and the possessor in a possessive clause. My analysis also raises the question of whether human animacy features and parallelism in argument structure can be incorporated into the sufficient criteria for 'in focus' status. In terms of the technical application of the coding protocol, this study addresses some of the issues that arise in the coding of plural entities and determining the role of eye gaze in bringing an entity into focus.

Finally, another contribution of my study is that it addresses the methodological aspect of what constitutes evidence of scalar implicatures. The use of sentence continuations, exploring the correlation between reference identification and the use of alternating forms, and the exploration of ranked pairs all provide evidence that the implicature, 'not in focus' plays a role in referent identification.

6.2. Degrees of restrictiveness

The integrated analysis of the corpus and questionnaire data provides evidence for gradient degrees of restrictiveness that both support distinct cognitive status categories 'in focus' and 'activated' and provide evidence of finer levels of difference in the use of forms. The idea of gradient distinctions is consistent with the evidence that the choice of pronominal often indicates relative cognitive prominence, which I call "ranking effects", even for referents of the same cognitive status sub-category—that is, those that share the same minimum and maximum status.[121]

In this section I summarize the distribution of pronouns in relation to multiple features demonstrated to be associated with the difference between an activated versus an in focus referent (either from the coding protocol of the GH model or from my own research): argument position and linear distance of previous mention, direct mention (introduced by a nominal or event introduced by a clause) versus indirect mention (non-

[121] Note that while *sho* could, in a general sense, be considered to be more restrictive due to its preference for referents that are not in focus or less prominent in terms of relative salience or animacy features, I am defining restrictiveness in the sense of the GH model—that is the degree to which a form is restricted in representing entities of the highest level of prominence or higher relative prominence.

event entity type introduced by a clause), and inherent features such as the distinction between human, animate, and inanimate entities. Note that, while animacy features are not specifically discussed in the coding protocol, my analysis suggests that animacy features play a role at least in relation to the relative prominence of entities, and perhaps even their cognitive status. Based on these features, the forms can be ranked as follows according to descending degree of restrictiveness: *o'ziu*, *null*, *bu*, *o*, and *sho*/*shu*.[122] Figure 8 provides a visual summary of the rankings and the most notable distinctions of restrictiveness. Figures 9 and 10 illustrate different degrees of restrictiveness related specifically to nominally-introduced entities and clausally-introduced entities, respectively.

The form *o'ziu* is the most restricted of the six pronominal forms, and all of its referents in the data are clearly in focus. While, unlike reflexive anaphors in some languages, reference is not restricted to subjects of the local clause, across sentence boundaries *o'ziu* is restricted to reference to entities mentioned as subjects of the previous sentence. Though a very high percentage of uses of this form in the corpus data are in reference to human entities, there is no evidence that *o'ziu* is restricted to use with animate or human entities. The use of this form, however, is restricted to entities introduced by nominals. In brief, reflexives are not only restricted to reference to entities in focus, but also have additional restrictions related to the argument position of previous mention and mention by a nominal.

---

[122] Due to the limited evidence related to *shu* and the fact that it mostly parallels the information about *sho*, I will not discuss this form separately here.

Null pronouns are also restricted to use with referents in focus. More specifically, the majority of nulls refer to entities previously mentioned as subjects (either in a preceding clause of the same sentence or the previous sentence). Unlike the reflexive, however, nulls can be used with referents mentioned as non-subject arguments in the previous sentence, particularly when the reference occurs in the same argument position in both sentences (parallelism). On the other hand, there is evidence that nulls are restricted to reference to entities mentioned either as subjects (assumed to be in focus) or as direct objects (also likely to be in focus), as most null references to entities mentioned as indirect objects (which are much less likely to be in focus) are judged unacceptable. Though there is only one example in the corpus of null reference to an entity in focus not introduced by a nominal—in this case, an event—the analysis of the questionnaire data provides evidence that some null reference to events are acceptable, while others may be judged unacceptable for independent reasons. Namely, a number of examples in the data suggest that null pronouns are not acceptable in certain contexts where phonological stress is required. Note that such a distributional difference between stressed and unstressed pronouns would be expected in any language.

Bu exhibits no clear restrictions in the area of the argument position of previous mention; however, there appears to be a degree of restriction in reference to clausally-introduced entities. In the case of the one corpus token for which *bu* refers to an activity, the clause which introduces the event is nominalized. In the questionnaire, the acceptability of the use of *bu* with clausally-introduced entities appears to be limited to certain types of emphatic expressions which may be conventionalized exclamations. In

terms of relative prominence, *bu* most often indicates the more prominent of two or more entities.

The form *o* shows no outstanding preference for previous subjects or entities in focus, even though more of its referents are in focus than activated. There are no documented restrictions on use with entities not introduced by nominals, and *o* can be used with either entity in a ranked set of entities referred to with pronominals. Finally, *o* is less likely than *sho* to give rise to the implicature 'not in focus', a factor which contributes to the impression that this form is neutral.

The form *sho* (and possibly *shu*, as well) demonstrates a strong preference for entities that are at most activated and those not mentioned as previous subjects. *Sho* is the form most often used with entities introduced by clauses, and shows no restrictions in that area. This pronoun occurs more frequently with non-human and animate entities than human entities, and the human entities to which it refers in the corpus are primarily plural entities. In terms of relative prominence, *sho* usually indicates a lower-ranking entity in a set. Overall, this form is the least likely to refer to an entity that has any degree of prominence related to inherent features such as animacy or contextual features such as the argument position of previous mention.

Finally, there is an interesting correlation between the description of gradient restrictiveness derived from both questionnaire and corpus data and the percentage of referents of a given form that are in focus in the corpus study. Based on gradient restrictiveness, the ranking of pronominals is as follows: *o'ziu* > *Ø* > *bu* > *o* > *sho*. As shown in Figure 11, the most restrictive form, *o'ziu* has 100 % of its referents in focus,

while the form *sho* has only 18.7% of its referents in focus. Except for nulls and *bu*, which have virtually the same percentage of referents in focus (though technically *bu* has more by 0.4%), the descending percentage of referents in focus represents the descending degree of restrictiveness described in this section.

o'ziu
Requires a referent in focus
Intersentential reference restricted to previous subjects
Not used with clausally introduced referents

Ø
Requires a referent in focus
Majority of referents mentioned as previous subjects
Use with in focus entities not introduced by nominals most likely acceptable
Reference to entities mentioned as direct objects, but not dative objects okay
Some entities not automatically coded as in focus okay if parallelism involved
Not acceptable in phonologically-marked contexts

bu
Majority of referents previous subjects
Reference to entities mentioned as direct objects and dative objects okay, regardless of parallelism
Limited use with focus entities not introduced by nominals
Usually indicates higher-ranked in a set of entities referred to with pronominals

o
No outstanding preference for previous subjects
Majority use with focus entities
No restriction with entities not introduced by nominals
Can be used with either entity in a ranked set of entities referred to with pronominals

sho (shu)
Strong preference for entities not mentioned as previous subjects
Strong preference for entities at most activated
No restriction on entities not introduced by nominals
Usually indicates lowest entity in a ranked set of entities referred to with pronominals
More frequently occurs with non-human/inanimate entities

Figure 8: *Degree of Restrictiveness*

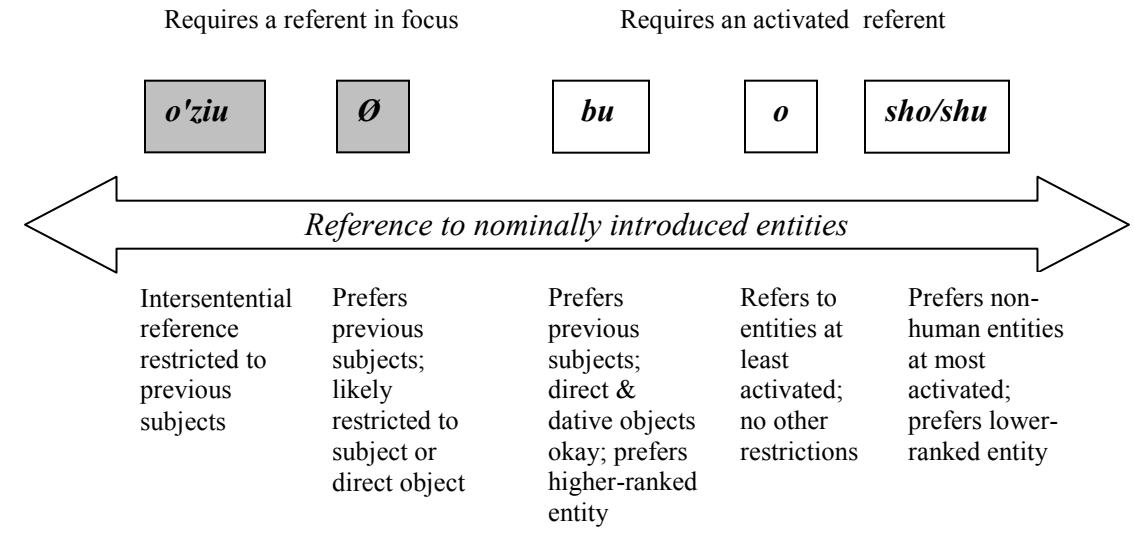


Figure 9: *Reference to nominally introduced entities*

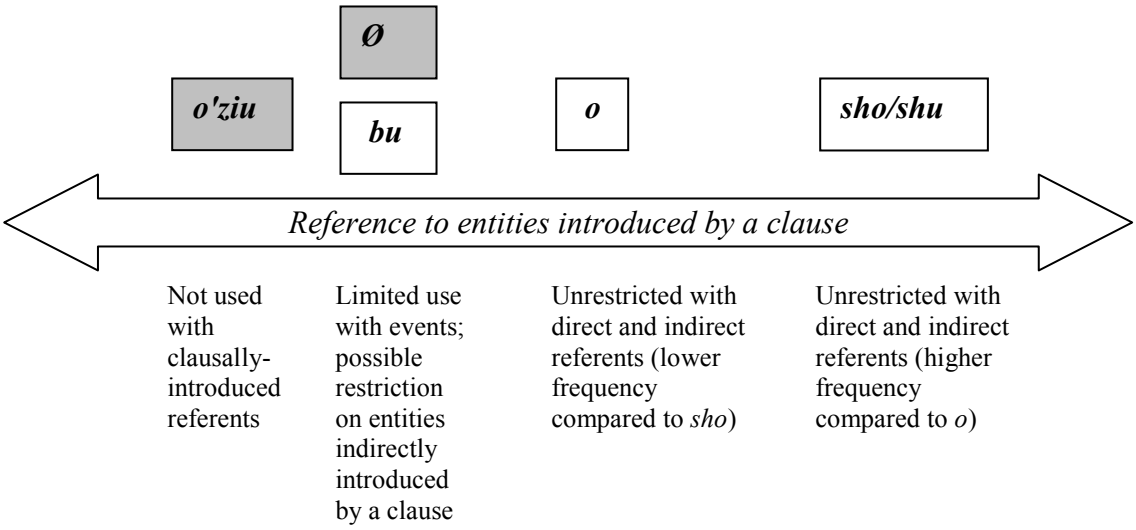


Figure 10: *Reference to entities introduced by a clause*

Figure 11: *Percent of corpus referents in focus* (cf. Table 6)

*The placement of *shu* is not really significant, since it represents the coding of only one form.

6.3. Cognitive status and ranking effects

An important question in this study is whether or not there is a distinction between cognitive status and relative salience (ranking). While some theories, such as Centering Theory, focus only on relative salience based on factors such as argument position or animacy, the GH model focuses on how these same factors all contribute to allow a referent to attain a particular cognitive status. The distinct categories of cognitive status provide an explanation for why certain forms are unacceptable with referents of a certain maximum status, yet still allow for different degrees of relative salience for referents of the same cognitive status (Mulkern 2003:23). Accordingly, it is plausible that, even though the use of forms signaling different levels of cognitive status can be used to create a ranking effect, there are still other ways to indicate finer degrees of relative salience. Another possible conclusion is that the categories of 'in focus' and 'activated' are prototype categories with fuzzy boundaries (rather than discrete categories) mapped onto gradient distinctions of cognitive prominence.

In Kumyk, I propose that the distinction between the (inherent) cognitive statuses 'in focus' and 'activated' is primarily encoded by the use of a null versus an overt

pronoun, while the different demonstratives primarily encode finer degrees of relative salience. I would further suggest that in most languages, the contrast between a set of forms with lower phonological prominence and one with greater phonological prominence (e.g. nulls versus overt forms or unstressed pronouns versus stressed pronouns) is used to signal different levels of inherent salience. This fits with the suggestion of Givón (1983) and others that unstressed pronouns and nulls (ellipsis—or generally, phonologically reduced form) are associated with the highest level of natural cognitive prominence or "givenness", as well as a high degree of "expectedness" or predictability.

If I claim the focus-activated distinction in Kumyk is encoded primarily by the null-overt contrast, what can I then claim about the reflexive, since the data analysis also indicates that this form can only be used with referents in focus? In all languages, even languages such as Kumyk, which has a relatively unrestricted use of nulls (purely in terms of syntax), there are still reasons a null form cannot be used to refer to a highly prominent entity. In Kumyk, for example, a null argument in a relative clause is always understood as coreferent with the head of the relative clause, and any other argument of the relative clause must be expressed as an overt form, even highly salient entities. In addition to this, null arguments are not acceptable in contexts which require phonological focus. In Kumyk these contexts include equative clauses, expressions of contrastive focus, and other emphatically-marked expressions (with *de*, for example). In some of these contexts, the reflexive seems to act as a specialized overt form which signals the addressee to block the implicature 'not in focus' and search for the referent among the

entities at the center of attention. I claim, moreover, that the contrast between *o'ziu* and nulls (unlike the four different forms that signal the status 'activated') is not used to indicate the relative salience of two entities in focus. This is a logical consequence of the fact that, since there is no status higher than 'in focus', there is no way an implicature of lower ranking can be derived, and, as expected, there is no evidence for such a functional contrast in the data. (Note that, though the contrast between null and reflexive is not used for relative salience, the contrast between *o'ziu* and any demonstrative can be used to indicate relative salience.)

Based on my analysis, I claim that the primary function of the four demonstratives in Kumyk is to encode the relative salience of entities that have the same minimum cognitive status—that is, entities that are at least activated, but not necessarily at most activated. As previously stated, the use of any form from this category potentially gives rise to the implicature 'not in focus' based on the unidirectional entailment of the GH model, yet the form *sho* seems to have a specialized association with this implicature. The form *bu*, on the other hand seems to have a specialized function of indicating the more salient of one or more entities with the same minimum cognitive status. Though this association does not directly derive from the structure of the GH model, it is not surprising that there is a need for some means of indicating the more salient of two entities that are both at most activated, a function which, by definition, cannot be fulfilled by a form which requires an entity in focus—that is, a null or reflexive.

The psychological reality of both the ranking effect resulting from focus-activated distinctions and the finer distinctions of the demonstrative contrast are adequately

supported by my data analysis. I predict that pairs of ranked forms can be used to indicate the salience of one entity in relation to another, and an analysis of sentences with multiple pronouns shows a correlation with the salience ranking (of entities) which is derived from either cognitive status differences or animacy differences. Sentence continuations comparing the four demonstrative forms provide further evidence that contrasts in form (ranked pairs) guide anaphora resolution. When asked to read a sentence and write a second sentence using a particular demonstrative form, respondents tend to use the pronoun *bu* to refer to the entity assumed to be the highest-ranking entity and consistently used *sho* to refer to a lower-ranking entity. One point of interest is that the consistent preference for *sho* referring to lower-ranked entities seems much more consistent than correlations between *bu* and higher-ranking entities (in various types of questionnaire data). The most likely explanation for this phenomenon is that *bu* is a form which can be used to impose salience on an entity (Mulkern 2003).

A broader question related to the claim that the demonstratives encode relative salience is the relationship between relative salience and the categorization of demonstrative use primarily in terms of distance features. As I explain in Chapter 2, following the GH model, I assume that the cognitive status of an entity is a more basic feature than spatial distance, though the two are perhaps related. While a number of recent studies characterize the use of demonstratives in terms of attention states (Enfield 2003, inter alia), it is interesting to look at possible correlations between traditional descriptions of the Kumyk demonstratives and the specializations I claim in this study. The correlation between the form *bu*, which is described as a proximal demonstrative,

and its specialization in indicating the higher-ranking of one or more entities is rather iconic. On the other hand, the form *sho*, which specializes in indicating the lower-ranking entity in a set, is described as a medial demonstrative, which is less iconic, as one might have expected the distal demonstrative here. It is interesting, however, that in Kumyk, as in a number of languages, the distal demonstrative *o* seems to function as the default/neutral personal pronoun, whereas *bu* and *sho* are much less frequent and more specialized.

6.4. Inherent versus imposed salience

Following Mulkern 2003 and Levinsohn 2008 this study assumes a distinction between the natural cognitive prominence of an entity, which is related to inherent features of entity type and contextual features (linear distance of previous mention, argument position, eye gaze, introduction by a nominal or clause, etc.) and salience which can be attributed to an entity via the linguistic form. Mulkern (2003) describes salience attributed via a linguistic form as *imposed salience* that reflects "how the speaker intends the hearer to rank [or rerank] discourse entities relative to one another" (25). Levinsohn, on the other hand, describes this type of salience as thematic prominence, or the attention drawn to an already established topic, with the assumption that "normally" attention is drawn to new information (2008:61).

What happens if a speaker wants to impose salience on an entity by the choice of linguistic form—here, choice of referring expression? Imposed salience is impossible with nulls and generally not likely to work with phonologically reduced forms; hence it is associated with stressed forms. Functionally, imposed salience is associated with both

reranking, such as topic shift or other types of unexpectedness or "heightened emotion" (Mulkern 2003:25, Levinsohn, 2008:62)).

Which types of forms can be used to impose salience? In Kumyk, nulls and reflexives both signal 'in focus'. Nulls, however, by definition, cannot impose salience in any form. Reflexives are restricted to entities which are already highly salient (most likely the highest-ranking entity) and the use of these forms to impose salience is not usually associated with reranking or imposed ranking, but with an added contextual effect. A Relevance-theoretic interpretation of this phenomenon might be that the use of an overt form with an entity that is already in focus/highly salient is unexpected and harder to process than a null argument, but has a greater potential for contextual effect (Sperber and Wilson 1986/95). This is based on the idea that, given the basic contrast between nulls and overt forms (primarily non-reflexive forms), the use of null arguments is restricted to the set of possible referents in focus, which makes referent identification easier than for most overt forms, which signal only activation and, thus can be associated with a wider set of possible referents. In other words, the use of a reflexive communicates something to the effect of, "even though I'm using an overt form (instead of a null), which usually means a less salient referent, search for a highly salient referent." Specific contextual effects often mentioned in the literature on reflexives include logophoricity, indication of point of view shift (e.g. shift from narrator as the deictic center to the protagonist as deictic center), and contrastive focus.

While Mulkern suggests that a speaker can impose salience on an entity by "using an expression that signals a higher cognitive status than a referent actually has for the

addressee at a particular point in the discourse" (2003:27), I suggest that the speaker can also impose salience on an entity by using a form that signals 'activated'—more specifically, the specialized form *bu*. The pronoun *bu* is the form naturally associated with imposed salience, since it specializes in indicating the more prominent entity within a set. The pronoun *o*, on the other hand, is both more frequent and more neutral (which follows ideas of typological markedness), while *sho* can be said to impose lower relative salience.

Evidence that *bu* is associated with imposed salience includes the association between *bu* and reference to a highly salient entity that involves some type of unexpectedness, particularly the cases in which *bu* is associated with topic shift and elaboration in the corpus study (c.f. section 4.3.5). Topic shift involves unexpectedness because subject continuity is expected; thus, as Centering Theory predicts, topic shift requires more processing effort (Walker, Joshi and Prince 1998). Elaboration is an unexpected adding of information about a topic rather than continuing the chronological narrative, which likely also requires more processing effort.[123] Particularly in cases of topic shift, the use of *bu*, given its specialization in signaling the highest-ranking referent (which, in most cases would be the previous topic), would ease the processing effort by signaling the hearer to search for the most salient referent of the previous utterance. Note

[123] Mulkern 2003 also predicts that a lower form (e.g. activated form for in focus referent) can be used for elaboration,which is, to some extent exemplified in Kumyk by the use of demonstratives like *o* in inter-sentential elaboration. I propose that the form *bu* is more specialized for the function of elaboration within the same noun phrase, while *o* and other demonstratives are associated with intersentential elaboration.

that, especially in the case of topic shift, the use of a reflexive would not create the same effect because it is likely to be assumed to be coreferent with the subject of the current sentence rather than with the subject of the previous sentence.

In other types of cases, the use of *bu*, especially in contrast to another form, can say, "pay attention, this entity is extra salient for a reason primarily associated with speaker emotion." In one text, the use of *bu* is associated with the introduction of the VIP participant (also the 1st person narrator) into the narrative. Cataphoric reference is another case where the speaker subjectively imposes salience on a piece of new information. The fact that *bu* is the only demonstrative associated with cataphoric reference is further evidence of its function in imposing salience. The contextual effect of imposed salience that is associated with speaker emotion is to communicate something about the speaker's point of view, thus forms such as *bu* can easily be perceived as point of view markers.

6.5. Questions of wider interest and areas for further study

The data presented in this study offer insight into a number of questions of wider interest to linguists and raise a number of questions that are relevant both to theories of reference and to other areas of linguistics.

The description of Kumyk pronominals in this study provides basic information about the syntactic distribution of overt pronouns, nulls and reflexives. Of particular interest in this description is the distributional evidence in the corpus analysis and speaker judgments that supports the categorization of the reflexive as a pronominal in the binding theory sense. Though a number of languages have long-distance uses of

reflexive anaphors, few languages have reflexive forms for which the distribution is unrestricted enough for those forms to be classified as pronominals (cf. Cole et al. 2001, Kornfilt 2001), thus the Kumyk data should be of interest to this body of research.

The categorization of null reference in this study also complements various descriptions of null anaphora in Turkic languages (Erguvanli-Taylan 1986, Turan 1995, inter alia). The corpus analysis of null arguments and the resulting hypothesis that these forms require a referent in focus supports the idea of pragmatic or discourse restrictions on null anaphora and should be of particular interest to those working within the syntax-pragmatics interface. Moreover, the evidence for the role of parallelism and degree of expectedness in zero anaphora in Kumyk relates to theories of ellipsis, not only nominal ellipsis, but also to VP ellipsis, where parallelism is cited as a factor that contributes to the felicity of this phenomenon.

An important question of interest to a number of sub-disciplines of linguistics—particularly typology—is the degree to which animacy features affect the inherent salience of an entity. Croft (2003) claims that animacy features are a parameter in a number of typological patterns. For example, differences in animacy features (plus definiteness features) can correlate with overt versus implicit object/patient case marking as well as different grammatical treatments of noun incorporation and indexation (167-168). Other examples include the fact that nominals with human referents can have a grammaticized form of number marking not found among those with inanimate referents (129). More specific to theories of reference, Fraurud 1996 and Prat–Sala and Branigan 2000 provide evidence that inherent salience based on animacy features influences

referential choice. The key to these claims is developing a method that allows the investigator to differentiate between the role of salience derived from contextual features and salience derived from ontological features.

While animacy features likely contribute to cognitive status cross-linguistically, the degree and manner in which animacy features affect pronominal choice obviously differs in a language in which animacy categories are grammatically marked (e.g. English) versus languages like Kumyk in which animacy categories are not grammatically marked. Since, in the latter case, one can compare the correlation between pronominal choice and cognitive status to the correlation between pronominal choice and animacy features, further studies of languages of this type would serve to elucidate the role of animacy in relation to cognitive status.

A related question is whether or not events, states, activities, facts, propositions, speech acts are naturally more or less salient than other types of inanimate entities. If inherent features of an entity such as animacy contribute to cognitive status, it is not unreasonable to suppose that other types of inherent features could contribute to salience, as well. While the GH model makes a distinction between entities that are directly versus indirectly introduced by clauses, the model does not make a distinction in salience between the types of entities directly introduced by clauses (namely, events and states) and entities directly introduced by nominals. Nevertheless, certain aspects of this study, such as the restrictiveness in forms such as *bu* in reference to clausally-introduced entities, suggest that there may be a distinction, raising questions about how entity class affects inherent salience.

Finally, one of the more interesting typological questions this study indirectly addresses is the relationship between cognitive status distinctions and traditional categorizations of demonstratives in terms of distance features. This study does not take the position that cognitive status distinctions are derived from what some would perceive as more basic distinctions in terms of distance features. In fact, the results of the study show that it is the so-called medial demonstrative *sho* in Kumyk, not the distal demonstrative *o* that is most closely associated with the implicature 'not in focus' and the indication of lower relative salience. Some recent studies have suggested classifications of demonstratives based on joint attention (Enfield 2003, inter alia). For example, Özyürek and Kita 2001 (as cited in Enfield 2003) classify the Turkish demonstrative ş*u*, previously described as the medial demonstrative, as a form which indicates a referent that the addressee is not attending to. This classification of Turkish ş*u* would fit better with the current study's evidence about the Kumyk "medial" demonstrative than an explanation that proposes that distance features are directly mapped onto the discourse use of demonstratives Taking into consideration the fact that different languages have different numbers of demonstratives with variations related to the linguistic encoding of animacy, number, gender, or other features, further cross-linguistic comparison of discourse features of demonstratives and the cognitive status of their referents should provide interesting data on what is basic to these forms.

References

Abdulaeva, A. Z. 1973. Slozhnopodchinennoe predlozhenie v sovremennom Kumykskom iazyke. Dissertation in fulfillment of the degree "Kandidat Filologicheskie Nauk". Baku.

Abdurakmanova, Pasikhat Dzhalilovna. 2005. Osnovnye voprosy istoricheskoĭ morfologii Kumykskogo iazyka: Morfemika, imennye I glagol'nye kategorii. Disseration in fulfillment of the degree of "Doktora Filologichesie Nauki." Makhachkala.

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24:65-87.

Ariel, Mira. 1990. *Accessing noun-phrase antecedents.* London: Routledge.

Asher, Nicholas. 1993. *Reference to abstract objects in discourse.* Dordrecht: Kluwer Academic Publishers.

Baker, C.L. 1995. Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English. *Language* 71:63-101.

Bosch, Peter, Graham Katz and Carla Umbach. 2007. The Non-Subject Bias of German Demonstrative Pronouns. In *Anaphors in text: Cognitive, formal and applied approaches to anaphoric reference*, ed. Monika Schwarz-Friesel, Manfred Consten, and Mareile Knees, 145–164. Amsterdam: John Benjamins.

Branco, Antonio, Tony McEnery and Ruslan Mitkov, eds. 2005. *Anaphora processing: Linguistic, cognitive and computational modelling*. Amsterdam: John Benjamins.

Brown-Schmidt, Sarah, Donna K. Byron, and Michael K. Tanenhaus. 2005. Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language* 53(2):292-313.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and topic*, ed. C. Li, 25-55. New York: Academic Press.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11:1-46.

--1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass: The MIT Press.

--1986. *Barriers*. Cambridge, Mass: The MIT Press.

--1995. *The minimalist program*. Cambridge, Mass: The MIT Press.

Clinton, Paul. 1997. The Grammar of the Kumyk language. Unpublished manuscript.

Cole, Peter, Gabriella Hermon, and C.T. James Huang. 2001. *Syntax and semantics volume 33: Long-distance reflexives*. San Diego: Academic Press.

Cornish, Francis. 1999. *Anaphora, discourse, and understanding: Evidence from English and French*. Oxford: Clarendon Press.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Dahl, Östen. 2008. Animacy and egophoricity: Grammar,ontology and phylogeny. *Lingua* 118:141–150.

Dahl, Östen and Kari Fraurud. 1996. Animacy in grammar and discourse. In *Reference and Referent Accessibility*, ed. T. Fretheim and J.K. Gundel, 47-64. Amsterdam: John Benjamins.

Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization: Typological studies in language 42*. Amsterdam: John Benjamins.

Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17: 463-489.

Dik , S.C. 1997. *The theory of functional grammar, I: The structure of the clause*. Berlin: Mouton de Gruyter.

Dixon, R.M.W. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language* 27(1):61-112.

Djanmavov. Iu. D. 1967. *Deeprichastiia v Kumykskom literaturnom iazyke*. Moscow.

Dmitriev, Nikolaĭ Konstantinovich. 1940. *Grammatika kumyksogo jazyka*. Moscow: Izdatel'stvo Akademii nauk SSSR.

Enç, Murvet. 1986. Topic switching and pronominal subjects in Turkish. In *Studies in Turkish linguistics* (*Typological studies in language, volume 8*), ed. Dan Isaac Slobin and Karl Zimmer, 195–208. Amsterdam: John Benjamins.

Erguvanlı-Taylan, E. 1986. Pronominal versus zero representation of anaphora in Turkish. In *Studies in Turkish Linguistics*, ed. D.I. Slobin and K. Zimmer, 209-231. Amsterdam: John Benjamins.

Erkü, Feride and Jeanette K. Gundel 1987. The pragmatics of indirect anaphors. In *The Pragmatic Perspective*, ed. Jef Verschueren and Marcella Bertuccelli Papi, 533-545. Amsterdam: John Benjamins.

Enfield, Nick. 2003. Demonstratives in Space and Interaction. *Language* 79(1):82-117.

Fretheim, Thorstein and Jeanette Gundel, eds. 1996. *Reference and referent accessibility*. Amsterdam: John Benjamins.

Gadzhiakhmedov, N. E. 1987. *Grammaticheskie kategorii glagola v Kumykskom iazyke*. Makhachkala.

Gadzhiakhmedov, N. E. 1998. Slovoizmenitel'nye kategorii imeni i glagola v kumykskom iazyke. "Avtoreferat" of the Dissertation in fulfillment of the degree of "Doktora Filologicheskie Nauki." Makhachkala.

Givón, T., ed. 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Gordon, Raymond G., Jr. (ed.), 2005. *Ethnologue: Languages of the world, fifteenth edition*. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/. (Accessed 11-10-06.)

Grice, H.P. 1975. Logic and Conversation. In *Syntax and Semantics 3: Speech Acts,* ed. P. Cole and J.L. Morgan, 41-58. New York: Academic Press.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. *Proceedings for the 21st Annual Meeting of the Association for Computational Linguistics, Cambridge, Mass*: 44-50.

Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistic*s 21:203-225.

Grüning, André, and Andrej Kibrik. 2005 Modelling referential choice in discourse: A cognitive calculative approach and a neural network approach. In *Anaphora processing: Linguistic, cognitive and computational modelling*, ed. A. Branco, T. McEnery, and R. Mitkov, 163-198. Amsterdam: John Benjamins.

Gundel, Jeanette, Mamadou Bassene, Bryan Gordon, Linda Humnick, and Amel Khalfaoui. Reference and cognitive status: Results from corpus studies of Eegimaa, Kumyk, Ojibwe, and Tunisian Arabic. Paper presented at the 10th International Pragmatics Conference, Göteborg, Sweden, July 8-13, 2007.

Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274-307.

Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 2005. Pronouns without NP antecedents: How do we know when a pronoun is referential? In *Anaphora processing: Linguistic, cognitive and computational modelling*, ed. A. Branco, T. McEnery, and R. Mitkov, 351-364. Amsterdam: John Benjamins.

Gundel, Jeanette K. and Anne Mulkern. 1998. Quantity implicatures in reference understanding. *Pragmatics and Cognition* 6:21-45.

Gundel, J.K., K. Borthen, and T. Fretheim. 1999. The role of context in pronominal reference to higher order entities in English and Norwegian. In *Modeling and Using Context. Proceedings from the Second International Interdisciplinary Conference, CONTEXT '99* (*Lecture Notes in Artificial Intelligence* 1688), ed. P. Bouquet et al., 475-478 Berlin: Springer-Verlag.

Gundel, Jeanette K., Michael Hegarty, and Kaja Borthen. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information* 12: 281-299.

Gundel, Jeanette. 2003. Information structure and referential givenness/newness: How much belongs in the grammar? *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University, July 18-20, 2003* (invited talk), ed. Müller, Stefan, 122-142. CSLI Publications. http://csli-publications.stanford.edu/. (Accessed 10-26-2006.)

Gundel, Jeanette.[124] 2004. Coding protocol for statuses on the Givenness Hierarchy. Ms. University of Minnesota.

[124]This is a revision of the 1993 version of the coding protocol to which Nancy Hedberg and Ron Zacharski contributed. Ann Mulkern, Tonya Custis, Bonnie Swierzbin, also contributed to the 2004 revision.

Gundel, Jeanette.[125] 2006. Coding protocol for statuses on the Givenness Hierarchy. Ms. University of Minnesota.

Gundel, Jeanette K., Dimitrios Ntelitheos, and Melinda Kowalsky. 2007. Children's use of referring expressions: Some implications for theory of mind. In *ZAS Papers in Linguistics, Nr. 48. Special Issue on Inter-sentential Pronominal Reference in Child and Adult Language*, ed. D. Bittner and N. Gagarina, 1-22. Berlin.

Halliday, M.A.K., and R. Hasan. 1976. *Cohesion in English*. London: Longman.

Hankamer, Jorge and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7:(3):391-426.

Haspelmath, Martin. 1995. The Converb as a cross-linguistically valid category. In *Converbs in cross-linguistic perspective*, ed. M. Haspelmath and and E. König, 1-56. Berlin: Mouton de Gruyter.

Hedberg, Nancy, Jeanette K. Gundel, and Ron Zacharski. 2007. Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles. *Proceedings of DAARC 2007* (the Sixth Discourse Anaphora and Anaphora Resolution Colloquium, Lagos, Portugal. March 29-30, 2007):31-36.

Hegarty, Michael. 2003. Semantic types of abstract entities. *Lingua* 113:891-927.

Hegarty, Michael, Jeanette K. Gundel, and Kaja Borthen. 2001. Information structure and the accessibility of clausally introduced referents. *Theoretical Linguistics* 27:163-186.

Huang, Yan. 2000. *Anaphora: A cross-linguistic approach*. Oxford: Oxford University Press.

Humnick, Linda. 2002. Some categories of participant reference in Kumyk Narrative. In *Comments on discourse structures in ten Turkic languages*, ed. John M. Clifton and Deborah A. Clifton, 105-133. Dallas: SIL International.

[125] Additional contributors to the 2006 version of the coding protocol include Amel Khalfoui, Linda Humnick, Bryan Gordon, Mamadou Bassene, and Shana Watters.

Humnick, Linda. 2005. Pronominal references and the encoding of cognitive status in Kumyk. In *Chicago Linguistic Society 41: The Main Session,* ed. Rodney Edwards, Patrick Midtlyng, Colin Sprague, and Kjersti Stensrud, 149-163. Chicago: The Chicago Linguistic Society.

Humnick, Linda. 2008. Null arguments in Kumyk adverbial clauses. In *Subordination and coordination in North Asian languages: Current issues in linguistic theory, Vol. 300*, ed. Edward J. Vajda, 47-62. Amsterdam: John Benjamins.

Humnick, Linda. 2007. Pronominal reflexives in Kumyk. Paper presented at the Conference on the Languages of the Caucasus, December 7-9, 2007, Leipzig, Germany.

İşsever, Selçuk. 2003. Information structure in Turkish: The word order – prosody interface. *Lingua* 113:1025–1053.

Kadyradzhiev, K. S. 1999. Istoricheskaia morfologiia imeni v kumykskom iazyke. "Avtoreferat" of the Dissertation in fulfillment of the degree of "Doktora Filologicheskie Nauki." Makhachkala.

Kaiser, Elsi and John Trueswell. 2004. The referential properties of Dutch pronouns and demonstratives: Is salience enough? In *Proceedings of the Conference "sub8 – Sinn und Dedeutung"* (Arbeitspapier Nr. 177, FB Sprachwissenschaft), ed. Cécile Meier, Matthias Weisgerber, 137-149. Universität Konstanz, Germany.

Kaiser, Elsi. 2005. Different forms have different referential properties: Implications for the notion of 'salience'. In *Anaphora Processing: linguistic, cognitive and computational modelling*, ed. A. Branco, T. McEnery, and R. Mitkov, 261-282. Philadelphia/Amsterdam: John Benjamins.

Khangishiev, Zh. M. 1985. Prichastie b Kumykskom Iazyke. Makhachkala.

Khangishiev, Zhangishi M.. 1995. *K"umyk" til: morphologiia*. Makhachkala.

Kibrik, Andrej. 1999. Reference and working memory: Cognitive inferences from discourse observations. In *Discourse studies in cognitive linguistics: Selected papers from the Fifth International Cognitive Linguistics Conference, Amsterdam, July 1997*, ed. Karen Van Hoek, Andrej A. Kibrik and Leo Noordman, 29-52. Amsterdam: John Benjamins.

Kornfilt, Jaklin. 2001. Local and long-distance reflexives in Turkish. In *Syntax and semantics volume 33: Long-distance reflexives*, ed. P. Cole, G. Hermon, and C.T.J. Huang, 197-225. San Diego: Academic Press.

Kuno, Susumo. 1987. *Functitonal syntax: anaphora, discourse and empathy*. Chicago: University of Chicago Press.

Küntay, A.C. and A. Özyürek. 2002. Joint attention and the development of the use of demonstratives in Turkish. *Boston University Conference on Language Development* 26:336-347. Somerville: Cascadilla Press.

Lambrecht, Knud. 1994. *Information structures and sentence form*. Cambridge: Cambridge University Press.

Levinsohn, Stephen. 2008. Self-instruction materials on narrative discourse analysis. Online URL at https://mail.jaars.org/~bt/narr.zip. (Produced for the Summer Institute of Linguistics at the University of North Dakota, copyright 2008 by SIL International.)

Levinson, Stephen C. 2004. Deixis. In *Handbook of Pragmatics*, ed. G. Ward and L. Horn, 97-121. Oxford: Blackwell.

Lewis, G.L. 1967. *Turkish grammar*. Oxford: Oxford University Press.

Lyons, John. 1977. *Semantics: Volume 2*. Cambridge: Cambridge University Press.

Mulkern, Ann. 2003. Cognitive status, discourse salience, and information structure: Evidence from Irish and Oromo. Ph.D. Dissertation, University of Minnesota.

Muratchaeva, Farida. 2001. Deĭktichieskie sredstva v sintaksicheskoĭ organizatsii predlozheniia kumykskogo iazyk. Paper presented at LENCA1: International symposium on deictic systems and quantification in languages spoken in Europe and North and Central Asia, , May 21-25, 2001, Udmurt State University, Izhevsk, Udmurt Republic. Abstracts ed. Valery Kel'makov, Pirkko Suihkonen, Pjotr Voroncov, Bernard Comrie, and Maria Beznosova. http://www.ling.helsinki.fi/uhlcs/symposium-izhevsk/abstraktit-rtf.pdf . (Accessed 10-7-2006.)

Ol'mesov, N. Kh. 1994. *Morfonologiia kumykskogo iazyka.* Makhachkala: Dagestanskii Gosuldarstvenoy Universitet.

Özyürek, Asli and Sotaro Kita. 2001. Interacting with demonstratives: Encoding of joint attention as a semantic contrast in Turkish and Japanese demonstrative systems. Ms. Istabul and Nijmegen: Koç University and Max Planck Institute for Psycholinguistics. (Cited by Nick Enfield, 2003 in "Demonstratives in Space and Interaction" *Language* 79(1): 82-117.)

Pica, Pierre. 1987. On the nature of the reflexivization cycle. *Proceedings of NELS* 17(2):483-499.

Poesio, Massimo and Nissim, Malvina. 2001. Salience and possessive NPs: The effects of animacy and pronominalization. *Proceedings of AMLAP* (Poster Session), Saarbruecken, September.

Prat-Sala, M. and H. P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42:168-182.

Prince, E. F. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. P. Cole, 223-255. New York: Academic Press.

Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Silverstein, M. 1976. Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, ed. Dixon, R.M.W., 112–171. Australian Institute of Aboriginal Studies, Canberra.

Sperber, Dan and DeirdreWilson. 1986/1995. *Relevance : Communication and cognition*. London: Blackwell.

Sperber, Dan and DeirdreWilson. 2004. Relevance theory. In *Handbook of Pragmatics*, ed. G. Ward and L. Horn, 607-632. Oxford: Blackwell.

Sultanmuradov, A. M. 1997. Sposoby vyrazheniia aspektualnykh znachenii v Kumykskom iazyke. Dissertation in fulfillment of the degree "Kandidat Filologicheskie Nauk".

Turan, U. D. 1995, Null versus overt subjects in Turkish discourse: A centering analysis. Ph.D. dissertation, University of Pennsylvania.

Turan, Ümit. 1998. Ranking forward-looking centers in Turkish: Universal and language-specific properties. In *Centering theory in discourse*, ed. M. Walker, A. Joshi, and E. Prince, 139-160. Oxford: Clarendon Press.

Vendler, Z., 1967. *Linguistics in philosophy*. Cornell University Press, Ithaca.

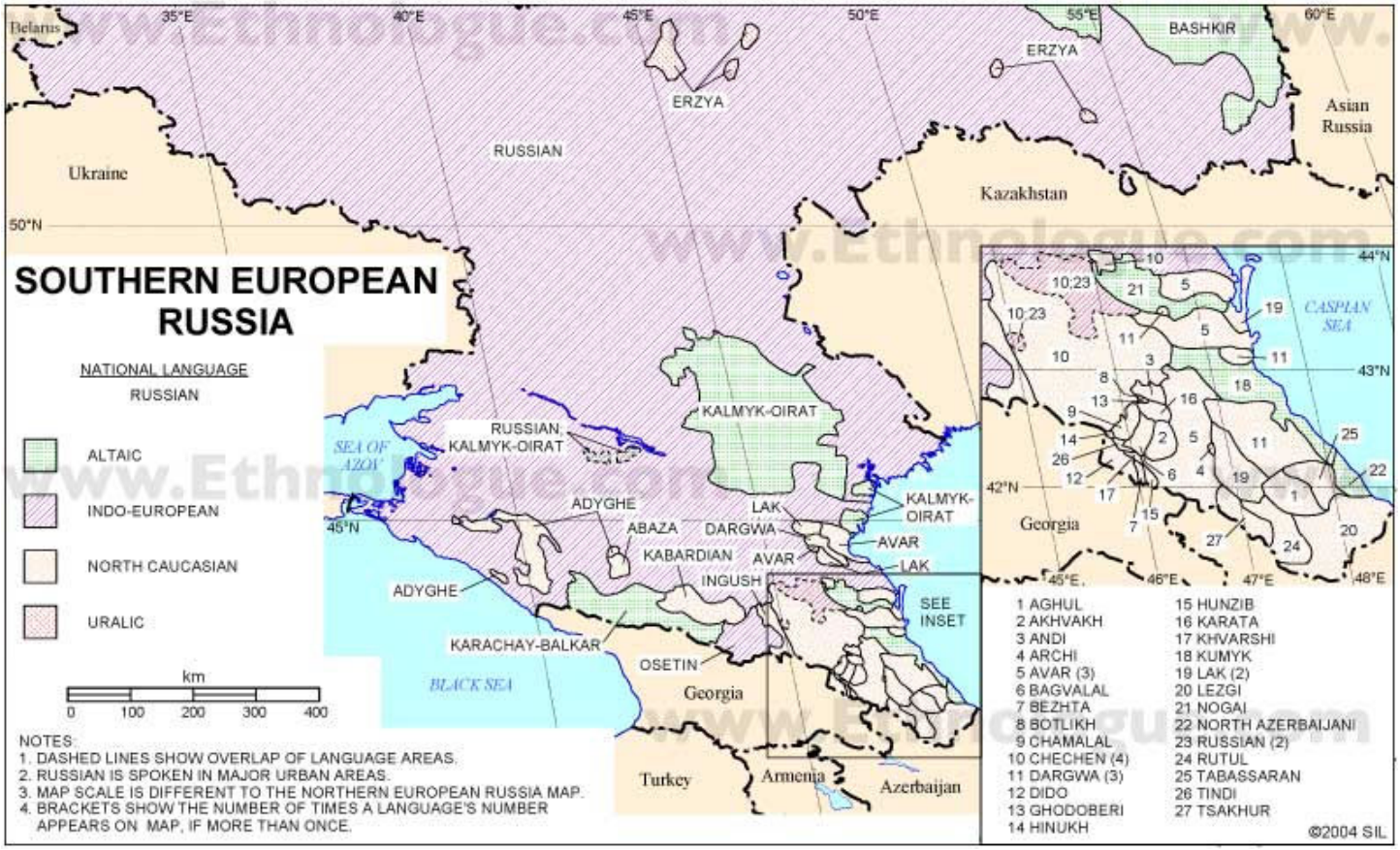
Walker, Marilyn, Aravind Joshi, and Ellen Prince. 1998. Centering in naturally occurring discourse: An overview. In *Centering theory in discourse*, ed. M. Walker, A. Joshi, and E. Prince, 1-28. Oxford: Clarendon Press.

Watters, Shana and Jeanette Gundel. 2007. An empirical investigation of the relation between coreference and quotations: Can a pronoun that is located in quotations find its referent?. In *Anaphora, Analysis, Algorithms and Applications*, *6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos (Algarve), Portugal, March 29-30, 2007* (*Lecture Notes in Artificial Intelligence* 4410) ed. Antonior Branco, 94-106.

Zribi-Hertz, A. 1995. Emphatic or reflexive? On the endophoric character of French lui-même and similar complex pronouns. *Journal of Linguistics* 31:333-374.

Appendix 1: Geographical Distribution of the Languages of Dagestan

Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/. Accessed 11-10-06. Reprinted with permission from SIL, International.



Appendix 2: Text corpus

Text	**Title/ID**	**Mode/ Genre**	**Source**	**# pages** (glossed)
1	*I Am Guilty, Marian,* Introduction, Excerpt from Chapter 1 (ID Marian)	Written Narrative	Abukov, Kamal. 1973. Men gunag'lyman, Mar'ian. In *Siuĭse de siuĭmese de*. Makhachkala: Dagestanskoe knizhnoe izdatel'stvo.	5
2	The importance of MT Literature (ID Mother Tongue)	Written Expository	Elicited	8
3	Hibat (Slander) and Ekiiuzliuliuk (Hypocrisy) (ID Character)	Written Hortatory	Akaev, Abusupiĭan. Reprinted in 1993. *Paĭkhammarny ël bulan*. Makhachkala: Dag'ystan kitap basmakhanasy.	7
4	Khalk"ybyzny aburu artsyn, o'r bolsyn. *(*ID Editorial)	Written Hortatory	G'amzatov, Ag'mat. Khalk"ybyzny aburu artsyn, o'r bolsyn. *Yoldash.* April 20, 2001:8.	9
5	Aihalay's Slippers (ID Aihalay)	Written Narrative	Aĭhalaĭny Machiĭleri. *Tangcholpan* [Kumyk language journal]	9
6	Birth (ID Birth)	Oral Expository	Elicited	6
7	Obituary (ID Obit)	Written Narrative/ expository	Hasayev, Amet. *Yoldash*. September 7, 2004:8. [newspaper]	7
8	The Prophet's Daughters (ID Prophet)	Written Narrative	Pajxammarny k"yzlary (1993. *Tangcholpan* 4:18-19. No author listed.) [Kumyk language journal]	12
9	Oral narratives of video [ID Oral]	Oral Narrative	Elicited	9
10	Agriculture text (ID Agr)	Written Expository	D. Zhavathanov. Ihshlamegen - Tishnemes. *Yoldash*. Aug. 31, 2004:3. [newspaper]	15
11	Oral narrative of wordless book (ID Oral)	Oral Narrative	Elicited	8
12	Oral narrative of publication event (ID NT)	Oral Narrative (historical)	Elicited	4

Appendix 3: Sample Text

Text Title: **The Prophet's Daughters**

Text Code: Prophet
Language: Kumyk
Primary Coder: Linda Humnick
Date: 25 September 2006
Coding Guidelines: 7/04

This text, entitled "Pajxammarny k"yzlary" was originally printed in *Tangcholpan*, a Kumyk literary publication and is reproduced here by permission from the publisher. *(Tangcholpan* 1993, volume 4:18-19, no author listed)

While this text primarily uses the transliteration guidelines suggested by ALA-LC 1997, the following variations apply: "j" is used intead of "ĭ" to represent the Cyrillic symbol "й", "ja" is used instead of "ia" to represent the symbol, "я" and "jo" is used instead of "ë" to represent "ë".

The following abbreviations are used in glossing: ABL=Ablative, ACC=Accusative, COND=Conditional, CON=Conjunction, COP=Copula, DAT=Dative, DET=Determiner, EMPH=Emphatic, FUT=Future, GEN=Genitive, GER=Gerund/Converb, INF=Infinitive/Verbal Noun, LOC=Locative, MOD=Modifier, NOM=Nominalizer, PASS=Passive, PL=Plural, POSS=Possessive, PR=Present, PRT=Participal, PST=Past, RFLX=Reflexive (verbal), S=Singular, SPEC=Speculative. 3.1, 3.2, 3.3 and 3.4 indicate the four distinct third person pronouns, *bu*, *o*, *sho*, and *shu*, respectively.

Each form coded in this text is enclosed in square brackets with a corresponding superscript number, e.g. [*bir k"yzy*][x]. The coding information appears in the corresponding footnote, where the first column represents the referring form, the second column identifies the referent, and the third column provides the coding information. The following abbreviations are used in coding: FOC=in focus, ACT=activated, FAM=familiar, UNQ=uniquely identifiable, REF=referential, TYP=type identifiable, and UNDET=coding undetermined. Numbers following the status indicate the first criterion of the coding protocol which applies (sometimes more than one criterion is relevant).

#Prophet.0001

Бир пайхаммарны бир къызы, бир эшеги, бир мишиги, бир де ити болгъан.

[[Bir pajkhammarny][1] *bir k"yzy]*[2] *[bir esheki]*[3]
one prophet.GEN one girl.3.POSS one donkey.3.POSS

[bir mishiki][4] *[bir de iti]* [5] *bolg"an.*
one cat.3.POSS one EMPH dog.3.POSS be.PST

'A prophet had a daughter, a donkey, a cat and a dog.'

#Prophet.0002

Къызны эрге берме ярайгъан вакътиси гелип, бугъар гелечи геле.

[[[K"yzny][6] *[erge]*[7] *berme jarajg"an*
girl.ACC husband.DAT give.INF possible.PST.PRT

vak"tisi][8] *gelip [bug"ar]*[9]
time.3.POSS] come.GER 3.1.DAT

[gelechi][10] *gele.*
one.who.requests.engagement come.PR

'When it comes time for the daughter to be given in marriage,
a person requesting engagement comes to her.'

#Prophet.0003

Къызны атасы да рази болуп, къызын бережек бола.

[[K"yzny][11] *atasy]*[12] *da razi*
[girl.GEN] father.3.POSS] and agreed

---

[1] *Bir pajkhammarny;*	prophet;	REF1
[2] *Bir pajkhammarny bir k"yzy*	daughter;	UNQ1
[3] *bir esheki*	donkey;	UNQ1
[4] *bir mishiki*	cat;	UNQ1
[5] *bir de iti*	dog;	UNQ1
[6] *K"yzny*	daughter;	ACT1
[7] *erge*	husband;	TYP1
[8] *K"yzny erge berme jarajg"an vak"tisi;*	time;	UNQ1
[9] *bug"ar*	daughter;	FOC1
[10] *gelechi*	suitor1 (individual);	REF1
[11] *K"yzny*;	daughter;	FOC5
[12] *K"yzny atasy;*	prophet;	ACT1

bolup, ***[Ø]***[13] *[k"yzyn]*[14] *berezhek* *bola.*
Be.GER % girl.3.POSS give.FUT be.PR

'The girl's father, agreeing, will give the girl.'

#Prophet.0004

Экинчи гюн башгъа ерден гелечи геле,огъар да бережек бола.

[ekinchi giun][15] *[bashg"a erden]*[16] *[gelechi]*[17]
second day other place.ABL one.who.requests.engagement

gele,
Come.PR

[Ø][18] ***[****og"ar]*[19] *da* ***[Ø]***[20] *berezhek* *bola.*
% 3.2S.DAT EMPH % give.FUT be.PR

'The second day a person requesting engagement comes from another place; [he] will give [her] to him, also.'

#Prophet.0005

Уьчюнчю гюн бирдагъысы геле, огъар да сёз бере, дёртюнчю гюн бирдагъысы геле,огъар да къызын бережек болуп йибере.

[U'chiunchiu giun][21] ***[Ø]***[22] *[birdag"ysy]*[23] *gele,*
Third day % one.more.3S.POSS come.PR

---

[13] *Ø;* prophet FOC1/2
[14] *k"yzyn;* girl; FOC2/1
[15] *ekinchi gun;* second day UNQ1
[16] *bashg"a jerden;* another place TYP1
[17] *gelechi;* suitor2 (individual) REF1
[18] *Ø;* prophet; FOC4
[19] *og"ar* suitor2 FOC1
[20] *Ø;* daughter; FOC4/5
[21] *U'chiunchiu giun;* third day; UNQ1
[22] *Ø;* suitor (type) FOC5
[23] *birdag"ysy;* suitor3 (individual) REF1

[Ø][24] ***[****og"ar]*[25] *da* *[sjoz]*[26] *bere,*
% 3.2S.DAT EMPH word give.PR

[djortiunchiu giun][27] ***[Ø]***[28] *[birdag"ysy]*[29] *gele,*
fourth day % one.more.3S.POSS come.PR

[Ø][30] ***[****og"ar****]***[31] *da* *[****Ø****]*[32] *k"yzyn]*[33] *berezhek bolup* *jibere.*
% 3.2S.DAT EMPH % girl.POSS give.FUT become.GER send.PR

'The third day one more [person requesting engagement] comes; [he] promises him, also; the fourth day one more [person requesting engagement] comes, [he] sends the girl to be given to him also.'

#Prophet.0006

Гече ятгъанда,оьзю этген терс ишге оьзю гьёкюнюп, тобагъа тюше туруп бирдагъы намаз къыла.

[Geche][34] ***[Ø]***[35] *jatg"anda* *[[o'ziu]*[36] *etgen*
night % lie.down.PST.PRT.LOC self.3.POSS do.PST.PRT

ters ishge][37] *[o'ziu]*[38] *g'jokiuniup*
wrong work.DAT self.3.POSS regret.GER

[tobag"a][39] ***[Ø]***[40] *tiushe* *turup*
repentance.DAT % descend.PR stay.GER

---

[24] *Ø*; prophet; FOC4
[25] *og"ar;* suitor3; FOC1
[26] *sjoz;* word; TYP1
[27] *djortiunchiu giun;* fourth day; UNQ1
[28] *Ø*; suitor (type) FOC5
[29] *birdag"ysy;* suitor4 (individual) REF1
[30] *Ø*; prophet; FOC4
[31] *og"ar;* suitor4; FOC1
[32] *Ø*; prophet; FOC2
[33] *k"yzyn;* girl; FOC4
[34] *Geche ;* night; UNQ1
[35] *Ø*; prophet; FOC1
[36] *o'ziu;* prophet; FOC1
[37] *ters ishge;* bad thing FAM
[38] *o'ziu;* prophet; FOC1
[39] *tobag"a;* repentance; TYP1
[40] *Ø*; prophet; FOC1

[Ø][41] *[birdag"y namaz]*[42] *k"yla.*
% one.more prayer do.PR

'[That] night when [he] lay down, he regrets the wrong thing that he had done, [he] falls into repentance, [and] prays one more prayer.'

#Prophet.0007

Бу вакътиде Жабраил малайик гелип,
-Аллагьутаъала сагъа этген ишинге гьёкюнме,берген сёзюнгде табуларсан деп айтды,
- дей.

[Bu vak"tide][43] *[Zhabrail malajik]*[44] *gelip,*
3.1.DET time.LOC Gabriel angel come.GER

\- ***[****Allag'uta"ala]*[45] *[sag"a]*[46] *[[****Ø]***[47] *etgen ishinge]*[48]
\- Most.High.God 2S.DAT % do.PST.PRT work.2S.DAT

[Ø][49] *g'jokiunme,* ***[[Ø]***[50] *bergen sjoziungde]*[51]
% regret.NEG.PR % give PST.PRT word.2S.POSS.LOC

[Ø][52] *tabularsan]*[53] *dep ajtdy,*
% keep.FUT.2S] COMP say.3S.PST

\- ***[Ø]***[54] *dej.*
\- % say.PR

'At this time the angel Gabriel comes and (he) says,"The Most High God said not to regret the thing you have done, that you will keep your word."'

---

[41] *Ø*;	prophet;	FOC1
[42] *birdag"y namaz;*	prayer;	TYP1
[43] *Bu vak"tide*;	time;	ACT3
[44] *Zhabrail malajik;*	angel Gabriel;	FAM2
[45] *Allag'uta"ala;*	God;	NA Direct speech
[46] *sag"a;*	prophet;	NA Direct speech
[47] *Ø*;	prophet;	NA Direct speech
[48] *etgen ishinge;*		NA Direct speech
[49] *Ø*;	prophet;	NA Direct speech
[50] *Ø*;	prophet;	NA Direct speech
[51] *Bergen sjozjungde;*		NA Direct speech
[52] *Ø*;	prophet;	NA Direct speech
[53] *etgen ishinge g'jokiunme, bergen sjoziungde tabularsan;*		NA Direct speech
[54] *Ø*;	Angel Gabriel;	FOC1

#Prophet.0008

Пайхаммар да намазын къылып, юрегин де парахат этип ята.

[Pajkhammar][55] *da* *[namazyn]*[56] *k"ylyp*
prophet EMPH prayer.3.POSS do.GER

[Ø][57] ***[****iurekin]*[58] *de* *[parakhat]*[59] *etip* ***[Ø]***[60] *jata.*
% heart.3.POSS EMPH peace do.GER % lie.down.PR

'The prophet finishes the prayer, makes his heart peaceful, and goes to bed.'

#Prophet.0009

Эртен тургъанда тилегенлер гелип къызны элтелер.

[Erten][61] ***[Ø]***[62] *turg"anda* *[tilegenler]*[63] *gelip*
morning % get.up.PST.PRT.LOC request.PST.PRT.PL come.GER

[Ø][64] *[k"yzny]*[65] *elteler.*
% girl.ACC] lead.PR.PL

'In the morning, when [he] gets up, the requesters come and take away the girl/daughter.'

#Prophet.0010

Къыз гетип, гечесинде пайхаммарны уьюнде оьзюню къызындан да исбайы къыз бола.

[K"yz][66] *getip* *[gechesinde]*[67]
girl leave.GER night.3S.POSS.LOC

---

[55] *Pajkhammar;* prophet FOC4
[56] *namazyn;* prayer ACT1
[57] *Ø*; prophet; FOC1
[58] *yurekin*; heart (prophet's) UNQ1
[59] *parakhat* peace; TYP1
[60] *Ø*; prophet; FOC1
[61] *Erten*; morning; TYP1;
[62] *Ø*; prophet; FOC1
[63] *tilegenler;* suitor1+friends/relatives; FAM1
[64] *Ø*; suitor1+friends/relatives; FOC1
[65] *k"yzny*; girl; FAM1
[66] *K"yz;* girl; ACT1
[67] *gechesinde;* evening/night UNQ1

[[pajkhammarny][68] *u'iunde]*[69] *[[o'ziuniu]*[70]
prophet.GEN at.3POSS.home self.3POSS.ACC.3.POSS

k"yzyndan][71] *da [isbajy k"yz]*[72] *bola.*
girl.3.POSS.ABL and slender girl be.PR

'After the girl/daughter leaves, that evening, in the prohet's home, besides his own daughter there appears a slender girl.'

#Prophet.0011

О - уьйдеги мишиги болгъан.

[O][73] - *[**[Ø]***[74] *[u'jdegi]*[75] *mishiki]*[76] *bolg"an.*
3.2S - % home.LOC.MOD cat.3.POSS be.PST

'She had been [his] house cat.'

#Prophet.0012

Эртен гелип гелешегенлер ону да элте.

[Erten][77] ***[Ø]***[78] *gelip [geleshegenler]*[79]
morning % come.GER ask.for.in.marriage.PR.PRT.PL]

[onu][80] *da elte.*
3.2S.ACC EMPH lead.PR

'Coming in the morning, the ones asking [for her] in marriage lead her [away], too.'

---

[68] *pajkhammarny;* prophet; ACT1
[69] *pajkhammarny u'iunde*; prophet's home; UNQ1
[70] *o'ziuniu;* prophet; FOC2
[71] *o'zuniu k"yzyndan;* girl; FOC1 / 2
[72] *isbajy k"yz;* slender girl (formerly cat); REF1 (possibly FAM)
[73] *O*; girl (cat); FOC1
[74] *Ø*; prophet; FOC5
[75] *u'jdegi* house TYP1;
[76] *u'jdegi mishiki*; cat; FAM1 (possibly FOC)
[77] *Erten*; morning; TYP1
[78] *Ø*; suitor2(+friends/relatives); UNDET
[79] *geleshegenler;* suitor2(+friends/relatives); FOC1
[80] *onu;* cat-daughter; FOC1

#Prophet.0013

Уьчюнчю гече ити къыз бола, эртенинде гелип о къызны да алып гете.

[U'chiunchiu geche][81] *[**[Ø]**[82] iti]*[83] *[k"yz]*[84] *bola,*
third night % dog.3.POSS girl become.PR

[erteninde][85] *[**Ø]**[86] gelip [o k"yzny]*[87] *da*
morning.3.POSS.LOC % come.GER 3.2.DET girl.ACC EMPH

*[**Ø]**[88] alyp gete.*
% take.GER leave.PR

'The third night [his] dog becomes a girl, in the morning [they] come and take that girl, too.'

#Prophet.0014

Арадан бир-эки гюн гетип бир гече эшеги къыз бола.

Aradan [bir - eki giun][89] *getip [bir geche]*[90]
Interval.ABL one - two day pass.GER one night

*[[**Ø]**[91] [esheki]]*[92] *[k"yz]*[93] *bola.*
% donkey.3.POSS girl become.PR

'After a couple of days pass, one night [his] donkey becomes a girl.'

#Prophet.0015

Эртенинде ону да элтелер.

---

[81] *U'chiunchiu geche;* third night; UNQ1
[82] *Ø;* prophet; ACT1
[83] *iti;* dog; FAM1
[84] *k"yz;* dog-girl TYP1
[85] *erteninde* ; morning; UNQ1
[86] *Ø;* suitor3(+friends/relatives); ACT1
[87] *o k"yzny ;* girl (previously dog) FOC1
[88] *Ø;* suitor3; FOC1
[89] *bir - eki gjun* 1-2 days; TYP1
[90] *bir geche;* one night; TYP1
ACT null will go here (ref=prophet)
[91] *Ø;* prophet; ACT1
[92] *esheki;* donkey; FAM1
[93] *k"yz;* girl; TYP1

[Erteninde][94] *[onu]* [95] *da* ***[Ø]***[96] *elteler.*
Morning.3.POSS.LOC 3.2S.ACC EMPH % lead.PR.PL

'In the morning, [they]take her, too.'

#Prophet.0016

Арадан 4 - 5 ай гетип пайхаммар къызларын, гиевлерин гёрме гете.

Aradan *[4-5 ai]*[97] *getip* *[pajkhammar]*[98]
interval.ABL 4-5 month pass.GER prophet

*[[**Ø**]*[99] *[k"yzlaryn]]*[100] *[[**Ø**]*[101] *[gijevlerin]*[102]
% girl.PL.3.POSS.ACC % son.in.law.PL.3.POSS.ACC

gjerme *gete.*
see.INF1 go.PR

'A period of 4 - 5 months having passed, the prophet goes to see his daughters and sons-in-law.'

#Prophet.0017

Башлап эшегине бара.

Bashlap *[[**Ø**]*[103] *[eshekine]]*[104] ***[Ø]***[105] *bara.*
Begin.GER % donkey.3S.DAT % go.PR

'(He) begins by going to the donkey.'

---

[94] *Erteninde;* morning; UNQ1
[95] *onu* donkey; FOC1
[96] *Ø*; suitor4(+friends/relatives) ACT1
[97] *4-5 ai* ; 4-5 months; TYP1
[98] *pajkhammar;* prophet; FAM1
[99] *Ø*; prophet; FOC2
[100] *k"yzlaryn;* daughters; FAM1
[101] *Ø*; prophet; FOC2
[102] *gijevlerin*; sons-in-law; FAM (coreferent w/suitors)
[103] *Ø*; prophet; FOC1
[104] *eshekine;* daughter/donkey; ACT1
[105] *Ø*; prophet; FOC1

#Prophet.0018

Ахшам ашап-ичип, ондан-мундан лакъыр башлап пайхаммар гиевюне яшав гьалдан сорай.

[Akhsham][106] ***[Ø]***[107] *ashap - ichip ondan - mundan*
evening % eat.GER – drink.GER from.there - from.here

[lak"yr][108] ***[Ø]***[109] *bashlap [pajkhammar]*[110]*[[**Ø**]*[111] *[gijeviune]]*[112]
conversation % begin.GER prophet % son.in.law.3.POSS.DAT

[yashav g'aldan][113] *soraj.*
life situation.ABL ask.PR

'In the evening, having eaten and taken drink, beginning the conversation about this and that, the prophet asks his son-in-law about [his] life situation.'

#Prophet.0019

Гиевю: -- Барыны да эби табулар, тек бир зат яман - бир гьайванны уьйретеген заман болду.

*[[**Ø**]*[114] *[Gijeviu]]*[115] -- *"[Baryny_da]*[116] *ebi tabular*
% son.in.law.3.POSS -- everything okay keeps.FUT2

tek [bir zat][117] *jaman - [[bir g'ajvanni]*[118] ***[Ø]***[119] *u'jretegen*
but one thing bad - one animal.GEN % teach.PR.PRT

---

[106] *Axsham;* evening; TYP1
[107] *Ø;* prophet; FOC1
[108] *lak"yr;* conversation; TYP1
[109] *Ø;* prophet; FOC2
[110] *pajkhammar;* prophet; FOC1
[111] *Ø;* prophet; FOC2
[112] *gijeviune;* son-in-law; FOC4
[113] *yashav g'aldan;* life situation; UNQ1
[114] *Ø;* prophet; FOC1
[115] *Gijeviu;* son-in-law; ACT1
[116] *Baryny_da;* NA Direct speech
[117] *bir zat;* NA Direct speech
[118] *bir g'ajvanni;* NA Direct speech
[119] *Ø;* NA Direct speech

zaman][120] *boldu.”*
time be.3S.PST

'Son-in-law: “Everything’s okay, but one thing is bad – enough time has passed to teach an animal.”'

#Prophet.0020

Бугъар айтгъан да бир, айтмай къойгъан да; къуллукъдан-ишден бек парахат, - деген.

“[Bug"ar][121] *ajtg"an da bir, ajtmay k"ojg"an*
3.1.DAT say.PST.PRT EMPH one say.NEG.PR leave.PST

da; [k"ulluk"dan][122] *- [ishden]*[123] *bek parakhat” - degen.*
EMPH service.ABL - work.ABL very peacefully - say.PST

'”Whether you say something to her or not, it’s the same; very peaceful in attitude towards work.”'

#Prophet.0021

Эшек ишни бажармай экен деп пайхаммарны эсине геле.

[[Eshek][124] *[ishni]*[125] *bazharmaj eken]*[126] *dep [[pajkhammarny]*[127]
donkey work.ACC know.how.NEG.PR COP COMP prophet.GEN

esine][128] *gele.*
memory.3.POSS.DAT come.PR

'”A donkey will not know how to work,” the prophet thought.'

#Prophet.0022

Гече буларда къонакъ да болуп, эртен итине бара.

---

[120] *bir g'ajvanni u'jretegen zaman;*; NA Direct speech
[121] *Bug"ar;* NA Direct speech
[122] *k"ulluk"dan;* NA Direct speech
[123] *ishden;* NA Direct speech
[124] *Eshek;* donkey (type); TYP1;
[125] *ishni;* work; UNQ1 (the work of a donkey)
[126] *Eshek ishni bazharmay eken;* thought of the prophet; ACT3
[127] *pajkhammarny;* prophet; ACT/FOC (addressee in dialogue)
[128] *pajkhammarny esine;* prophet’s mind; UNQ1

Geche ***[Ø]***[129] *[bularda]*[130] *[k"onak"]*[131] *da* *bolup* *erten*
night % 3.1.PL.LOC guest thus be.GER morning

[[Ø][132] *[itine]]*[133] ***[Ø]***[134] *bara.*
% dog.3.POSS.DAT % go.PR

'At night he is a guest at their place; in the morning he goes to the dog.'

#Prophet.0023

Шунда да шолай геч болгъунча яшав гьалдан сорашып, гиевю де бугъар кант эте: "Бир зат айтма амал ёкъ, ит йимик гьап-гьазир, сен бирни айтсанг, сагъа онну айтажакъ."

[Shunda][135] *da* *[sholaj]*[136] *gech* *bolg"uncha*
3.4.LOC EMPH 3.3.ADV late become.before.GER

*[[yashav]****[137]*** *g'aldan]*[138] ***[Ø]***[139]***[Ø]****[140]* *sorashyp*
life situation.ABL % % greet.GER

*[****[Ø]***[141] *[gijeviu]]*[142] *de* *[bug"ar]*[143] *[kant]*[144] *ete:*
% son.in.law.3.POSS EMPH 3.1.DAT complaint do.PR

"[Bir zat][145] *ajtma* *amal* *jok"* *[it]*[146] *yimik*
one thing say.INF1 situation exist.NEG dog like

---

[129] *Ø*; prophet; FOC1 (experiencer of psych. verb)
[130] *bularda;* son.in.law + daughter; FOC4
[131] *k"onak";* guest; TYP1
[132] *Ø*; prophet; FOC2
[133] *itine;* dog-girl; FAM1
[134] *Ø*; prophet; FOC2
[135] *Shunda;* dog's home; ACT1
[136] *sholaj;* NA: intensifier
[137] *yashav;* life; TYP1;
[138] *yashav g'aldan;* life situation; UNQ1
[139] *Ø;* prophet; FOC1
[140] *Ø;* son-in-law FOC2/4 (via *shunda*=dog's home)
[141] *Ø;* prophet; FOC2
[142] *gijeviu;* son-in-law; FOC2
[143] *bug"ar;* prophet; FOC1
[144] *kant;* complaint; TYP1
[145] *Bir zat;* NA: DIALOGUE
[146] *it;* NA: DIALOGUE

g'ap-g'azir *[sen]*[147] *[birni]*[148] *ajtsang* *[sag"a]*[149] *[onnu]*[150]
ready 2S one.ACC] say.2S.COND 2S.DAT ten.ACC

[Ø][151] *ajtazhak."*
% say.FUT1

'At that place, thus (as before), before it got late, (he) asks about the situation of life, and his son-in-law answers him: "It's not possible to say a single word. If you say one word, she's]ready like a dog and will say ten words."'

#Prophet.0024

Авзун жыйса багьанасы ёкъ эди, - дей.

[Avuzun][152] *djyjsa* *bag'anasi*
[mouth.3.POSS] hold.3.COND **reason**.3.POSS

jok" *edi* - ***[Ø]***[153] *dej.*
exist.NEG 3.PAST.AUX - % say.PR

'"If (she) had held her tongue there would have been no reason," (he) says.'

#Prophet.0025

Ит гьапламаймы да деп атасыны эсине геле.

[It][154] *g'aplamajmy* *da* *dep*
dog bark-NEG-INTER EMPH COMP

*[[**[Ø]***[155] *[atasyny]]*[156] *esine]*[157] *gele.*
% father.3.POSS.GEN memory.3.POSS.DAT come.PR

'It came to her father's mind, "don't dogs bark?"' (form of indirect speech)

---

[147] *sen;* NA: DIALOGUE
[148] *birni;* NA: DIALOGUE
[149] *sag"a;* NA: DIALOGUE
[150] *onnu;* NA: DIALOGUE
[151] *Ø;* daughter (previously dog) NA: DIALOGUE
[152] *Avuzun;* NA: DIALOGUE
[153] *Ø;* son-in-law; FOC1
[154] *It;* dog (generic); TYP1
[155] *Ø;* son-in-law; FOC1
[156] *atasyny;* prophet; FOC4
[157] *atasyny esine;* father's mind; FOC (coreferential with prophet)

#Prophet.0026

Буларда гече де къалмай, мишигине бара.

[Bularda][158] *[geche]*[159] *de* ***[Ø]***[160] *k"almaj* *[****[Ø]***[161] *[mishikine]]*[162]
3.1.PL.LOC night EMPH % stay.NEG.PR % cat.3.POSS.DAT

[Ø][163] *bara.*
% go.PR

'Not staying with them for the night, (he) went to the cat.'

#Prophet.0027

Хош-беш этип сорашып ашгъа олтуралар.

Hosh-besh ***[Ø]***[164] *etip* *sorashyp* *[ashg"a]*[165] ***[Ø]***[166] *olturalar.*
Greeting % do.GER greet.GER food.DAT % sit.3PL.PR

'Greeting each other, (they) sit down to eat.'

#Prophet.0028

Ашап битгенде, бу яшны къатыны барып караватда ятып юхлай.

[Ø][167] *Ashap* *bitgende* *[[bu* *yashni]*[168]
% eat.GER finish.PST.PRT.LOC 3.1 child.GEN]

katiny][169] *baryp* *[karavatda]*[170] ***[Ø]***[171] *jatyp* *jukhlaj.*
wife.3.POSS go.GER bed.LOC % lie.down.GER sleep.PR

'When [they] finished eating, this child's wife goes and lies down to sleep in the bed.'

---

[158] *Bularda;* dog-daughter+son-in-law; FOC5
[159] *geche;* night UNQ
[160] *Ø;* prophet; FOC1 (experiencer of psych. verb)
[161] *Ø;* prophet; FOC2
[162] *Mishikine;* cat; FAM1
[163] *Ø;* prophet; FOC1/2
[164] *Ø;* prophet+cat +son-in-law ACT
[165] *ashg"a;* food TYP1
[166] *Ø;* prophet+cat+son-in-law; FOC1/2;
[167] *Ø;* prophet+cat+son-in-law; FOC1;
[168] *bu yashni;* son-in-law (w/cat); ACT (mentioned in plural)
[169] *bu yashni katiny;* wife/cat; ACT (mentioned in plural)
[170] *karavatda;* bed; TYP1
[171] *Ø;* cat/daughter; FOC1

#Prophet.0029

Пайхаммар гиевюне сорай: - Яшав гьал нечикдир? - деп.

[Pajkhammar][172] *[**[Ø]**[173] **[**gijeviune]]*[174] *soraj: - [yashav g'al]*[175]
prophet % son.in.law.3.POSS.DAT ask.PR - [life state

nechikdir? - dep.
how.COP - say.GER

'The prophet asks his son-in-law, "How's the state of life?"'

#Prophet.0030

Гиевю айта: - Муна сен гёреген кюйде, иш-къуллукъ булан бир аварасы да ёкъ.

[Gijeviu][176] *ajta: - Muna [[sen]*
son.in.law.3.POSS say.PR - behold 2S

*gjorgen kuyde**]*[177] *[ish - k"ulluk"]*[178]
see.PST.PRT way.LOC work – service

*bulan **[Ø]***[179] *[bir avarasy]*[180] *da jok".*
with % one concern.3.POSS EMPH exist.NEG

'The son-in-law says, "As you see, she has no concern for the work."'

#Prophet.0031

Ашап битсе, тепсини де жыймай, барып ятып къала.

[Ø][181] *Ashap bitse [tepsini]*[182] *de*
% eat.GER finish.3.COND tray.ACC EMPH

---

[172] *Pajkhammar;* prophet; ACT1
[173] *Ø;* prophet; FOC2
[174] *Gijeviune;* son-in-law; FOC5
[175] *yashav g'al;* life condition UNQ (your life condition)
[176] *Gijeviu;* son-in-law; FOC5;
[177] *sen gjorgen kuyde;* NA Direct speech
[178] *ish - k"ulluk";* NA Direct speech
[179] *Ø;* NA Direct speech
[180] *Bir avarasy;* NA Direct speech
[181] *Ø;* NA Direct speech
[182] *tepsini;* NA Direct speech

[Ø][183] *zhyjmaj* *baryp* *jatyp* *k"ala.*
% clean.up.NEG.PR go.GER lie.down.GER stay-PR

'"If [she] finishes eating, [she] does not clean up the dishes; [she] goes and lies down."'

#Prophet.0032

Юху буса дагъы зат тарыкъ тюгюл, - дей.

[Iukhu][184] *busa* *[dag"y* *zat]*[185] *taryk"* *tugul* -
sleep however more thing necessary not -

[Ø][186] *dej.*
% say.PR

'"Other than sleep, [she] does not need anything," [he] says.'

#Prophet.0033

Гече буларда да къалып, пайхаммар оьзюню къызына бара.

Geche ***[Ø]***[187] *[bularda]*[188] *da* *k"alyp* *[pajkhammar]*[189]
night % 3.1.PL.LOC EMPH stay.GER prophet

[[o'ziuniu][190] *k"yzyna]*[191] *bara.*
self.3POSS.GEN girl.3.POSS.DAT go.PR

'[After] staying with them at night, the prophet goes to his own daughter.'

#Prophet.0034

Атасын ариден гёргенде къызы да, гиевю де алдына барып хош-беш этип, атасын гётерип уьйге гийирелер.

---

[183] *Ø;* wife; NA Direct speech
[184] *Iukhu;* NA Direct speech
[185] *dag"y zat;* NA Direct speech
[186] *Ø;* son-in-law; FOC1
[187] *Ø;* prophet; UNDET (backwards anaphora)
[188] *bularda;* son-in-law FOC1 (plural is idiomatic)
[189] *pajkhammar;* prophet; FOC1
[190] *o'ziuniu;* prophet; FOC2;
[191] *o'ziuniu k"yzyna;* girl (own daughter); FAM1;

*[**[Ø]**[192]* *Atasyn]*[193] *ariden* ***[Ø]***[194] *gjorgende*
father.3.POSS.ACC further.ABL % see.PST.PRT.LOC

*[**[Ø]**[195]* *k"yzy**]***[196] *da* *[**[Ø]**[197]* ***[****gijeviu]]*[198] *de*
% girl.3.POSS and % son.in.law.3.POSS and

[Ø][199] *aldyna* *baryp* ***[Ø]***[200] *[khosh-besh]*[201] *etip* ***[Ø]***[202]
% before.DAT go.GER % greeting do.GER %

[[Ø][203] ***a****tasyn]*[204] *gjoterip* ***[Ø]***[205] *[u'jge**]***[206] *gijireler.*
% father.3.POSS.ACC bring.up.GER % home.DAT bring.in.3PL.PR

'Seeing their father from far away, his daughter and son-in-law go before him and greet him, bringing their father up and bringing (him) into their home.'

#Prophet.0035

Аш гьазир болгунча юрек авуртмас деп гиев, къатынгъа къычыра: - Ариги уьйден бир харбуз алып гел чи деп.

[Ash][207] *g'azir* *bolguncha* *[iurek]*[208] *avurtmas*
food ready be.before.GER heart cause.to.be.heavy.NEG

dep *[gijev**]***[209] *[k"atyng"a**]***[210] *k"ychyra:*
COMP son.in.law wife.DAT cry.out.PR

---

[192] *Ø;* daughter/husband; UNDET (backwards anaphora)
[193] *Atasyn;* prophet; FOC1
[194] *Ø;* daughter/husband; FOC2 (via POSS suffix)
[195] *Ø;* prophet; FOC2
[196] *k"yzy;* girl; FOC2
[197] *Ø;* prophet; FOC2
[198] *gijeviu;* son-in-law; FOC2
[199] *Ø;* prophet; FOC2
[200] *Ø;* husband/wife; FOC1/2
[201] *khosh-besh;* greeting; TYP
[202] *Ø;* husband/wife; FOC1/2
[203] *Ø;* husband/wife; FOC1/2
[204] *atasyn;* prophet; FOC5
[205] *Ø;* husband/wife FOC1/2
[206] *u'jge;* home UNQ
[207] *Ash;* food; TYP
[208] *iurek;* heart TYP
[209] *gijev;* son-in-law ACT (part of plural subject)
[210] *k"atyng"a;* wife; ACT (part of plural subject)

- [Arigi u'jden][211] *[bir kharbuz]*[212] ***[Ø]***[213] *alyp*
- further.MOD home.ABL one watermelon % take.GER

gel chi dep.
come EMPH say.GER

'So that the heart won't be heavy while waiting for the food to be ready, the son-in-law shouts to his wife, saying, "Bring a watermelon from the back room."'

#Prophet.0036

Ариги уьйде аш этеген къатыны бир харбуз алып геле.

[[Arigi u'jde][214] *[ash]*[215] *etegen* *[**Ø**]*[216] *k"atyny]]*[217]
Further.MOD home.LOC food do.PR.PRT % wife.3.POSS]

[bir kharbuz][218] *alyp gele.*
one watermelon take.GER come.PR

'The wife, who was making food in the back room, brings one watermelon.'

#Prophet.0037

Эри харбузгъа къарап: - Бу чу бечиген, башгъасын гелтир - дей.

*[**Ø**]*[219] *Eri]*[220] *[kharbuzg"a]*[221] *k"arap:* *- "[Bu]*[222]
% husband.3.POSS watermelon.DAT look.GER - 3.1

---

[211] *Arigi u'jden;* NA Direct speech
[212] *bir kharbuz;* NA Direct speech
[213] *Ø;* wife NA Direct speech
[214] *Arigi u'jde;* back room; FAM1
[215] *ash*; food; ACT1;
[216] *Ø;* husband; FOC1
[217] *Arigi u'jde ash etegen k"atyny;* wife; ACT1
[218] *bir kharbuz*; watermelon; ACT1 (via direct speech)
[219] *Ø;* wife; FOC5
[220] *Eri;* husband/son-in-law; FOC1;
[221] *kharbuzg"a;* watermelon1; FOC5
[222] *Bu;* watermelon1; NA Direct speech

chu *bechigen* *[bashg"asyn]*[223] ***[Ø]***[224] *geltir," - **[Ø]***[225] *dej.*
EMPH wither.PST.PRT other.3.POSS.ACC % bring - % say.PR

'The husband, looking at the watermelon, says, "This one is withered, bring another one."'

#Prophet.0038

Катыны о харбузну элтип, башгъасын алып геле.

*[**[Ø]***[226] *K"atyny]*[227] *[o* *kharbuznu]*[228] *eltip*
% wife.3.POSS 3.2.DET watermelon.ACC bring.GER

*[**[Ø]***[229] *bashg"asyn]*[230] ***[Ø]***[231] *alyp* *gele.*
% other.3.POSS.ACC % take.GER come.PR

'The wife takes away that watermelon and brings another one.'

#Prophet.0039

Эри даъы да харбузгъа къарап: - Не къалын къабукълусун гелтирдинг, адамшавлусу ёкъму, - дей.

*[**[Ø]***[232] *Eri]*[233] *[dag"y* *da* *kharbuzg"a]*[234] *k"arap:*
% husband.3.POSS another yet watermelon.DAT look.GER

- "Ne *[k"alyn* *[**[Ø]***[235] *k"abuk"lusun]]*[236] ***[Ø]***[237]
- what thick % rind.3.POSS.ACC %

---

[223] *bashg"asyn;* another one(watermelon2);NA Direct speech
[224] *Ø;* wife; NA Direct Speech
[225] *Ø;* husband; FOC1/2
[226] *Ø;* husband; FOC1
[227] *K"atyny;* girl/daughter/wife; FOC5
[228] *o kharbuznu;* watermelon1; FOC5 (direct speech + narrative)
[229] *Ø;* watermelon (type) FOC4
[230] *bashg"asyn;* another one(watermelon2);ACT1 (direct speech)
[231] *Ø;* wife; FOC1/2
[232] *Ø;* wife; FOC1
[233] *Eri;* husband; FOC5
[234] *dag"y da kharbuzg"a;* watermelon2; FOC5
[235] *Ø;* NA Direct speech
[236] *k"alyn k"abuk"lusun;* NA Direct speech
[237] *Ø;* NA Direct speech

*geltirding, [**[Ø]**[238] adamshavlusu]*[239] *jok"mu, - **[Ø]***[240] *dej.*
bring.2S.PST % decent.3.POSS exist.NEG.INTER - % say.PR

'The husband, again looking at the watermelon, says, "What a thick-skinned one you brought, don't you have a decent one?"'

#Prophet.0040

Къатын даъы да о харбузну элтип, бир башгъасын гелтире.

[K"atyn][241] *dag"y da [o kharbuzn**u]***[242]
wife again and 3.2.DET watermelon.ACC

*eltip [bir [**[Ø]**[243] bashg"asyn]]*[244] ***[Ø]***[245] *geltire.*
bring.GER one % other.3.POSS.ACC % bring.PR

'The wife again takes that watermelon away and brings another one.'

#Prophet.0041

Эри харбузну алып:

*[**[Ø]**[246] Eri]*[247] *[kharbuzn**u]***[248] *alyp:*
% Husband.3.POSS watermelon.ACC take.GER

'The husband, taking the watermelon...'

#Prophet.0042

- Муна шулай бизин яшавубуз да, арадагъы татывлукъда.

- *Muna [shulaj]* [249] *[bizin yashavubuz]*[250] *da*
behold 3.4.ADV 1PL.GEN life.1PL.POSS EMPH

---

[238] *Ø;* NA Direct speech
[239] *adamshavlusu;* watermelon3; NA Direct speech
[240] *Ø;* husband; FOC1/2
[241] *K"atyn;* wife; FOC4 (via direct speech)
[242] *o kharbuznu;* watermelon2; FOC5
[243] *Ø;* watermelon (type) FOC4
[244] *bir bashg"asyn;* watermelon3; ACT3 (via direct speech)
[245] *Ø;* wife; FOC2
[246] *Ø;* wife; FOC1
[247] *Eri;* husband; FOC5
[248] *kharbuznu;* watermelon3; FOC5
[249] *shulaj;* NA direct speech
[250] *bizin yashavubuz;* NA direct speech

[aradag"y tatyvluk”da][251]*.*
interval.even.so agreement.LOC

'"Behold, our life, our mutual agreement (relationship), is like this.”'

#Prophet.0043

Бизин уьйде биргине бир харбузубуз бар эди.

[Bizin u'jde][252] *[birgine bir kharbuzubuz]*[253] *bar edi.*
1PL.GEN home.LOC one.MOD.what one watermelon.1PL.POSS exist 3.PAST.AUX

'"In our house there was only one watermelon.”'

#Prophet.0044

Мен ......нече керен къайтардым, о яман бириси бечиген деп, бу да ариги уьйге гетип, къайтып шо харбузну алып геле эди.

[Men][254] *[neche keren]*[255] ***[Ø]***[256] *k"ajtardym* *[o]*[257]
1S how.many instance % cause.to.return.1S.PST 3.2S

jaman ***[Ø]***[258] *[birisi]]*[259] *bechigen dep* *[bu]*[260] *da*
bad % one.of wither.PST.PRT say.GER 3.1 EMPH

[arigi u'jge][261] *getip kajtyp*
further.MOD home.DAT leave.GER return.GER

---

[251] *aradag"y tatyvluk”da;* NA direct speech
[252] *Bizin u'jde* NA direct speech
[253] *birgine bir kharbuzubuz;* NA direct speech
[254] *Men;* NA direct speech
[255] *neche keren;* NA direct speech
[256] *Ø;* wife; NA direct speech
[257] *o;* watermelon; NA direct speech
[258] *Ø;* watermelon (type) FOC4
[259] *birisi;* one (watermelon) NA direct speech
[260] *bu;* wife; NA direct speech
[261] *arigi u'jge;* NA direct speech

[sho kharbuznu][262] ***[Ø]***[263] *alyp* *gele* *edi.*
3.3 watermelon.ACC % take.GER come.PR 3.PAST.AUX

'"However many times I would have sent (her), saying that one is bad, one is withered, she would have gone to the back room and brought this watermelon."'

#Prophet.0045

Дагъы харбуз ёкь деген затны айтмады.

[[[Dag"y kharbuz][264] *jok"]*[265] *degen* *zatni]*[266]
more watermelon exist.NEG say.PST.PRT thing.ACC

[Ø][267] *ajtmady.*
% say.NEG.3.PST

'"(She) did not say that there are no more watermelons."'

#Prophet.0046

Муна шулай татывлу турабыз, - деген.

Muna *[shulaj]*[268] *tatyvlu* ***[Ø]***[269] *turabyz*
behold 3.4.ADV delicious % stay.PR.1PL

\- ***[Ø]***[270] *degen.*
\- % say.PST

'"Look how we stay in such a sweet state," (he) said.'

#Prophet.0047

Гече буларда къонакъ да болуп, пайхаммар эртен уьюне къайта.

---

[262] *sho kharbuznu;* NA direct speech
[263] *Ø;* wife; NA direct speech
[264] *Dag"y kharbuz;* NA direct speech
[265] *Dag"y kharbuz jok";* NA direct speech
[266] *Dag"y kharbuz jok" degen zatni;* NA direct speech
[267] *Ø;* wife; NA direct speech
[268] *shulaj;* NA direct speech
[269] *Ø;* wife + husband; NA; direct speech
[270] *Ø;* husband; FOC1 (clause b/f quoted speech)

Geche ***[Ø]***[271] *[bularda]*[272] *[k"onak"]*[273] *da* *bolup*
night % 3.1.PL.LOC guest EMPH be.GER

[pajkhammar][274] *erten* *[****[Ø]***[275] *[u'iune]]*[276] *kajta.*
prophet morning % home.DAT return.PR

'The prophet stays at their place for the night, and, in the morning, returns to his home.'

---

[271] *Ø;* prophet; UNDET (backwards anaphora)
[272] *Bularda;* husband (+wife); FOC1
[273] *k"onak";* guest; TYP1;
[274] *pajkhammar;* prophet; FOC1/2
[275] *Ø;* prophet; FOC2
[276] *u'iune;* prophet's home; FAM1

Appendix 4: Questionnaire

The questionnaire was administered in written form to 5 respondents.[1] The directions for each section are written in Russian.[2] All other material (including sample sentences) is in Kumyk. The responses are coded according to colors, and the order presented here is as follows: blue, orange, black, red, green.

I. Для каждого предложения А, напишите предложение Б, который является продолжением. В предложении Б, используйте слово, которое отмечено в круглых скобках. Форма (падеж) этого слово может измениться, в зависимости от предложения. Если возможно, избегите предложений, в которых это слово сопровождается существительным, например, 'Студент <u>огъар</u> савбол деди.', а ни 'Студент <u>о муаллимге</u> савбол деди.'

For each sentence A, write a sentence for B which would be a continuation of A. In the sentence in B, use the word in parentheses. The form (case) of this word can change according to the sentence context. If possible, avoid sentences in which this word is followed by a noun, for example 'The student thanked <u>him</u>.' And not 'The student thanked <u>that</u> teacher. (Though in English these sentences are translated with two different forms, *him* and *that,* in Kumyk, the form *о* is used in both sentences, since this form can be either a pronoun or a determiner.)

Например:
For example

А. Муаллим студентге китап берди.
'The teacher gave the student a book.'
Б. (о)

Response: О савбол деди.
'(S)he said thank you.'

Или…(Or…)

---

[1] This research is carried out under NSF grant #0519890 to Jeanette Gundel and received IRB exemption under the same grant.
[2] The choice of Russian for the directions is related solely to the process of my developing the questionnaire with limited access to Kumyk speakers.

А. Мариям анасына алма берди.
‘Mariam gave her mother an apple.’

Б. (о)

Response: Анасы **огъар** разилигин билдирди.
‘Her mother expressed her appreciation to her.’

The respondents give complete sentences. Here, I list the referent of the target pronoun for each respondent.

1. А. Ибрагьим ресторанда шорпа ашады.
‘Abraham ate soup in the restaurant.’

   Б. (бу)

   Responses: Ibragim; restaurant; Ibragim; Ibragim; soup

2. А. Болат атына минип барагъанда йыгъылды.
‘Bolat fell off while riding his horse.’

   Б. (о)
   Responses: Bolat; Bolat; Bolat; Bolat; Bolat

3. А. Муаллим студентге китап берди.
‘The teacher gave the student a book.’

   Б. (шо)
   Responses: book; book; book; book; book or giving of book

4. А. Ибрагьим ресторанда шорпа ашады.
‘Abraham ate soup in the restaurant.’

   Б. (о)
   Responses: Ibragim; Ibragim; Ibragim; Ibragim; soup

5. А. Муаллим студентге китап берди.
‘The teacher gave the student a book.’

   Б. (шу)
   Responses: student; book; book; book; book

6. А. Ибрагьим ресторанда шорпа ашады.
   ‘Abraham ate soup in the restaurant.’

   Б. (шо)
   Responses: restaurant; soup; restaurant; restaurant; soup

7. А. Болат атына минип барагъанда йыгъылды.
   ‘Bolat fell while riding (his) horse.’

   Б. (шо)
   Responses: horse; horse; DEM; thus; (ambiguous between Bolat or the horse)

8. А. Муаллим студентге китап берди.
   ‘The teacher gave the student a book.’

   Б. (бу)

   Responses: student; incorrect response; student (presentational); student; book

В (9) и (10), думайте, как вы закончили бы предложение с неявным подлежащим вместо местоимения. Мы используем символ Ø, чтобы представить неявный подлежащий или объект.

In (9) and (10) think how you would finish the sentence with an implicit form instead of a pronoun. We use the symbol “Ø” to stand for an implicit subject or object.

Например:

А. Муаллим студентге китап берди.
‘The teacher gave the student a book.’

Б. Ø…..
Response: Сонг, гетди.
Later ((s)he) left.

9. А. Болат атына минип барагъанда йыгъылды.
   ‘Bolat fell off while riding his horse.’

   Б. Ø…..
   Responses: Bolat; incorrect form; Bolat; Bolat; Bolat

10. А. Ибрагьим ресторанда шорпа ашады.
‘Abraham ate soup in the restaurant.’

Б. Ø…..
Responses: Ibragim; Ibragim; Ibragim; Ibragim; Ibragim

II: Прочитайте каждую пару предложений. Заполнитесь в бланке во предложении Б с одном местоимением в соответствующем падеже. Если есть несколько возможных вариантов, выбирают одного из них. Если не является соответствующее использовать местоимение, заполняться в типе слова или слов, которые соответствовали бы контексту.

Read each pair of sentences. Fill in the blank in the second sentence with a pronoun in the appropriate case. If there are several possibilities, choose one of them. If it is not appropriate to use a pronoun, fill in the blank with the type of word or words which would be appropriate for the context.

Местоимении: бу, о, шо, шу, оьзю

11. Например: А. Муалим студентге китап берди.
Б. Студент …огъар.. савбол деди.

For example: ‘The teacher gave the student a book.
The student told <u>him/her</u> thank you. ‘

12. А. Муалим студентге китап берди.
Б. …….. „Тангчолпан” эди.

‘The teacher gave the student a book .
____ was Tangcholpan.’

Responses: bu; sho; bu; sho; ol

13. А. Муалим студентге китап берди.
Б. Сонг муаллим …..... экзамени булан къутлады.

‘The teacher gave the student a book.
Later the teacher congratulated ___ on the exam.’

Responses: onu; onu; onu; onu; onu

14. А. Муалим студентге китап берди.
    Б. …….дан сонг муаллим студентни къолун алды.

    ‘The teacher gave the student a book.
    After ___ the teacher shook the student’s hand.’

    Responses: shondan; shondan; shondan; shondan; ondan

15. А. Муалим студентге китап берди.
    Б. Студент …. ... чалт охуду.

    ‘The teacher gave the student a book.
    The student read ___ quickly.’

    Responses: munu; shonu; shonu; shonu; onu

16. А. Болат атына минип барагъанда йыгъылды .
    Б. Хали …….. барамагъанын деп айтды.

    ‘Bolat fell while riding his horse.
    Now (he) said that (he) will not ride ___.’ (Now (he) said that __ will not ride (it).)

    Responses: G’ali og”ar dag’y minip barma’ak”man dep ajtdy; o’ziu;
    G’ali bug”ar; dag’y minip barma’ak”man dep ajtdy; shog”ar dag’y minip barma’ak”man dep ajtdy; no answer[3]

17. А. Болат атына минип барагъанда йыгъылды.
    Б. ……. къолу сынды.

    ‘Bolat fell while riding his horse.
    ___ broke (his) arm.’

    Responses: munu; onu; onu; onu; onu

18. А. Бир къыш гюнде Болат рестаурандан уьйге исси шорпа алма сюеди.
    Б. Тек уьйге гелген сонг ….. субукъ эди.

[3] Some unnaturalness in the original construction of the test sentence was corrected in 3 versions.

‘One winter day Bolat wanted to bring hot soup home from a restaurant.
But after coming home ___ was cold.’

Responses: o; sho; sho; ol; o

19. А. Мурат оьзюню машини булан тюкенге барма токъташды.
    Б. Сонг о ….. дёгерчеги бошалгъанны билди.

‘Murat decided to go to the store in his own car.
Later he found out that ___’s tire was flat.’

Responses: shonu; song **onu** djegerchegi…; o’ziuniu; shonu; song **onu** djegerchegi…

20. А. Мариям ону анасына алма берди.
    Б. О ….. бек тизив татыву бар деди.

‘Mariam gave an apple to her mother.
She said that ___ was very delicious.’ (one response: (She) said that __ was very delicious.)

Responses: shonu; shonu; shonu; shonu; **Onu** bek tiziv tatyvu bar dedi.

В следующих парах предложения, предложение А представляет утверждение одним человеком, и предложение Б представляет вопрос или восклицание от различного человека. Заполнитесь в бланке во предложении Б с местоимением, если приспособлено. Если кажется лучше иметь неявную форму (Ø), отметьте это.

In the following pairs of sentences, the sentence in (A) represents an utterance by one speaker, and the sentence in B represents a question or exclaimation/statement by another speaker. Fill in the blank in sentence (B) with a pronoun, if appropriate. If it seems better to use a null form (Ø), note that.

Местоимении: Бу, О, Шо, Шу, Оьзю, Ø

21. А. Болат атына минип барагъанда йыгъылды.
    Б. …….. гёрдюнгмю?

‘Bolat fell while riding his horse.
Did you see ___?’

Responses: shonu; Ø; shonu; shonu; onu

22. А. Болат атына минип барагъанда йыгъылды.
Б. …….. нечик болду?

‘Bolat fell while riding his horse.
How did ___ happen?’

Responses: o; sho; sho; sho; o

23. А. Болат атына минип барагъанда йыгъылды.
Б. Мен…….. инанмайман.

‘Bolat fell while riding his horse.
I don’t believe ___.’

Responses: shog”ar; shog”ar; shog”ar; shog”ar; shog”ar

24. А. Болат атына минип барагъанда йыгъылды.
Б. …….. болма ярамас!

‘Bolat fell while riding his horse.
___ is not possible!’

Responses: bu; Ø; Ø; sho; olaĭ

25. А. Болат атына минип барагъанда йыгъылды.
Б. …….. нечик болма бола?

‘Bolat fell while riding his horse.
How did ___ happen?’

Responses: shu; olaĭ; sho; bu; o

26. А. Ибрагьим ресторандан шорпа алды.
Б. …… татывлуму эди?

‘Abraham bought soup from the restaurant.
Was ___ good?’

Responses: o’ziu; Ø; Ø; sho; o

27. А. Ибрагьим ресторандан шорпа алды.
Б. Адатлы гьалда о пилавну ….. ала.

‘Abraham bought soup from the restaurant.
He usually buys soup ___.’

Responses: o’ziune; Ø; shondan; shondan; o’ziu

28. А. Ибрагьим ресторандан шорпа алды.
Б. О ….. дагъы да алармы эди экен?

‘Abraham bought soup from the restaurant.
Will he buy ___ again?’

Responses: onu; Ø; shonu; shondan; onu

III. Следующие предложения могут иметь одно или многократные местоимения. Заполнитесь в каждом бланке местоимением (в соответствующем падеже) или пустой формой (Ø) и укажите, к кому или что обращается каждая форма.

III. The following sentences may have one or more pronouns. Fill in each blank with a pronoun in the appropriate case or a null form (Ø) and indicate to whom or what the form refers.

(Below the presentation of the responses, I include the coding of referents I use in my analysis. The responses to one questionnaire are not included in this summary because the repsondent failed to identify the referent.)

Местоимении: Бу, О, Шо, Шу, Оьзю, Ø

29. А. Муаллим студентге китап берди.
Б. ……. …….. кёп ушатды.
1 2

1 = …
2 = …

‘The teacher gave the student a book.
___ was pleased with ___.’
1 2

Responses:

1	2
O (student)	shonu (book)
O (student)	Ø (book)
O (teacher)	shonu (book)
Ol (student)	shonu (book)

(teacher=FOC; student=ACT; book=ACT)

30. А. Муаллим студентге китап берди.
Б. Сонг …….. …….. экзамени булан къутлады.
1 2

1 = …
2 = …

‘The teacher gave the student a book.
Later ___ congratulated ____ with (his/her) exam.’
1 2

Responses:

1	2
Bu (teacher)	onu (student)
O’ziu (teacher)	onu (student)
Ø (teacher)	onu (student)
Ol (teacher)	onu (student)

(teacher=FOC; student=ACT)

31. А. Муаллим студентге китап берди.
Б. …….. …….. чалт охума тарыгъын айтды.
1 2

1 = …
2 = …

‘The teacher gave the student a book.
___ said to read ___ quickly.’
1 2

Responses:

1	2
shunu (book)	bu (student)
o (teacher)	Ø (book)
O (teacher)	shonu (book)
ol (teacher)	o’ziune (student)

(teacher=FOC; student=ACT; book=ACT)

32. А. Муаллим Патиматгъа китап берди.
Б. ……. уьйге гелген сонг ……. гьакъында анасына айтды.
1 2

1 = …
2 = …

‘The teacher gave Patimat a book.
After ___ came home, (she) told (her) mother about ___.’
1 2

Responses:

1	2
O (Patimat)	munu (teacher)
Ol (Patimat)	shonu (book)
O (Patimat)	shonu (book)
Ol (Patimat)	shonu (book)

(teacher=FOC; Patimat=ACT; book=ACT)

33. А. Муаллим Патиматгъа китап берди.
Б. ……. …….. кёп ушатды. Сонг, ……. уьйге гетди.
1 2 3

1 = …
2 = …
3 = …

‘The teacher gave Patimat a book.
___ was pleased with ___. Later, ___ went home.’
1 2 3

Responses:

1	2	3
Bu (Patimat)	shonu (book)	bu (Patimat)
Sho (Patimat)	Ø (book)	ol (Patimat)
O (Patimat)	shonu (book)	Ø (Patimat)
Ol (Patimat)	shonu (book)	ol (Patimat)

(Patimat$_1$=1ACT; book=ACT; Patimat$_3$=FOC)

34. А. Муаллим Патиматгъа китап берди.
Б. ……. биринчи савгьат болгъан сонг, …..… …….. анасына гёрсетме сюеди.
1 2 3

1 = …
2 = …
3 = …

‘The teacher gave Patimat a book.
Since ___ was (her) first prize, ___ wanted to show ___ to her mother.
1 2 3

Responses:

1	2	3
O (book)	bu (Patimat)	munu (book)
Sho (book)	Ø (Patimat)	onu (book)
Bu (book)	o (Patimat)	shonu (book)
Bu (book)	ol (Patimat)	shonu (book)

(book$_1$=ACT; Patimat=ACT; book$_3$=FOC)

35. А. Муалим студентге китап берди.
Б. …….. барын да тамаша этди.
1
1 = …

'The teacher gave the student a book.
___ surprised everyone.'
1

Responses:

1
O'ziu (teacher)
Ol (teacher)
O (teacher)
Bu (book)

(teacher=FOC; book= ACT)

36. А. Болат атына минип барагъанда йыгъылды.
Б. …….. болгъан зат болдуму?
1
1 = …

'Bolat fell off while riding his horse.
Did anything happen to ___?'
1

Responses:

1
og"ar (Bolat)
og"ar (Bolat)
og"ar (Bolat)
o'ziune (Bolat)

(Bolat=FOC)

37. А. Болат атына минип барагъанда йыгъылды.
Б. Бир жума гетип, ……. ……. сатды.
1 2

1 = …
2 = …

'Bolat fell off while riding his horse.
After a week ___ sold ___.'
1 2

Responses:

1	2
bu (Bolat)	shonu (horse)
Ø (Bolat)	shonu (horse)
o (Bolat)	onu (horse)
ol (Bolat)	shonu (horse)

(Bolat=FOC; horse=ACT)

38. А. Болат атына минип барагъанда йыгъылды.
Б. …….. саялы, …….. …….. сатма хыял етди.
1 2 3

1 = …
2 = …
3 = …

'Bolat fell off while riding his horse.
Because of ___, ___ sold ___.'
1 2 3

Responses:

1	2	3
Sho (falling)	o (Bolat)	munu (horse)
Sho (falling)	Ø (Bolat)	onu (horse)
Sho (falling)	o (Bolat)	onu (horse)
Sho (falling)	ol (Bolat)	shonu (horse)

(falling (fact)=ACT; Bolat=FOC; horse=ACT)

39. А. Болат атына минип барагъанда йыгъылды.
Б. …….. сонг ……. дагъы …….. минип бармагъан.
1 2 3

1 = …
2 = …
3 = …

‘Bolat fell off while riding his horse.
After ___ ___ did not ride ___ anymore.’
1 2 3

Responses:

1	2	3
Shondan (falling)	bu (Bolat)	og”ar (horse)
Ø	o (Bolat)	shog”ar (horse)
Shondan (falling)	o (Bolat)	og”ar (horse)
Shondan (falling)	ol (Bolat)	shog”ar (horse)

(falling (event)=FOC; Bolat=FOC; horse=ACT)

40. А. Болат атына минип барагъанда йыгъылды.
Б. ……. къолун сындырды.
1

1 = …

‘Bolat fell off while riding his horse.
___ broke his arm/hand.’
1

Responses:

1
O’ziu (Bolat)
Ol (Bolat)
O (Bolat)
Ol (Bolat)

(Bolat=FOC)

41. А. Агьмат гъар гюн сют тюкенге бара. Бугюн де сют алмакъ учун барды.
Б. Уьйге гелген сонг, къатыны …….. …… неге алгъанын сорады. Уьйде эки шиша барын айтды. 1 2

1 = …
2 = …

‘Ahmat goes to the milk store every day. (He) went today to buy milk, as well. When (he) came home, (his) wife asked ___ why he bought ___. She said that there were two bottles at home.’ 1 2

Responses:

1	2
bug”ar (Amat)	shonu (milk)
Ø (Amat)	shonu (milk)
og”ar (Amat)	shonu (milk)
og”ar (Amat)	shonu (milk)

(Amat=FOC; milk=FOC)

IV. Следующие предложения имеют одно или многократные местоимения. Укажите, к кому или что обращается подчёркнутая форма.

IV. The following sentences have one or more pronouns. Indicate to whom or what the underlined form refers.

42. Сонг къатынгиши тюкенге гире. Тюкенчи булан сёйлей; огъар аргъанын сата. Сатып, акъчаны да алып, <u>огъар</u> къарамады.
1

‘Later the woman goes into the store. (She) talks with the storekeeper; (she) sells him the accordian. Having sold it, taking the money, she did not look at it/him.’

1 = …
Responses: store-person, store-person, store-person, store-person, woman

43. А. Болат атына минип барагъанда йыгъылды.
Б. <u>Шогъар</u> болгъан зат болдуму?
1

‘Bolat fell off while riding his horse. Did anything happen to him/it?’

1 = ……..
Responses: horse, Bolat, Bolat, horse, Bolat

44. Уьюне гирип къараса: онда кёп уллу къалмагъал бар. Гьар тюрлю ашлар ташлангъан,бузулгъан. <u>Оланы</u> оьзюню анасы этген.
1

‘Coming into the home if (he/she) looks, there is a big mess there. Various foods are thrown down, ruined. His/her own other did/made them/these things.’

1 = …….
Responses: food, food, food, the ruining of the food – the mess, food

45. А. Болат атына минип барагъанда йыгъылды.
Б. Огъар болгъан зат болдуму?
1

‘Bolat fell off while riding his horse. Did anything happen to him/it?’

1 = ……..
Responses: Bolat, Bolat, Bolat, Bolat, Bolat

46. Муалим студентге китап берди. О барын да тамаша этди.
1

‘The teacher gave the student a book. (S)he/that/it surprised everyone.’’

1 = ……..
Responses: teacher, teacher, teacher, student, teacher

47. Ансар анасына алма берди. Шо рагьмулу эди.
1

‘Ansar gave his mother an apple. That/he was sweet.’

1 = ……..
Responses: giving his mother the apple, *ish* (‘work’)/giving the apple, giving the apple, giving the apple, Ansar

48. Сонг къатынгиши тюкенге гире. Тюкенчи булан сёйлей; огъар аргъанын сата. Сатып, акъчаны да алып, шогъар къарамады.
1

‘Later the woman goes into the store. (She) talks with the storekeeper; (she) sells him the accordian. Having sold it, taking the money, she did not look at it/him.’

1 = …
Responses: organ/accordian, store-person, money; organ, store-person

49. Уьюне гирип къараса: онда кёп уллу къалмагъал бар. Гьар тюрлю ашлар ташлангъан,бузулгъан. Шоланы оьзюню анасы этген.
1

‘Coming into the home if (he/she) looks, there is a big mess there. Various foods are thrown down, ruined. His/her own other did/made them/these things.’

1 = …….

Responses: food, food, food, the ruining of the food – the mess, food

50. Муаллим студентге китап берди. Шо барын да тамаша этди.
1

'The teacher gave the student a book. (S)he/that/it surprised everyone.'

1 = ……..
Responses: book, teacher, giving the book, book, teacher

51. Ансар анасына алма берди. О рагьмулу эди.
1

'Ansar gave his mother an apple. That/he was sweet.'

1 = ……..
Responses: Ansar, Ansar, Ansar, Ansar, Ø (not sure what this means), Ansar

V. Оцените следующие предложения для того, как хорошо они звучат. Используйте следующий масштаб:

1 великолепно звучит
2 хорошо звука
3 возможный, но обычно не говорил бы так
4 никаких Кумыков не сказали бы так

Rate the following sentences according to how well they sound. Use the following scale:

1 Sounds great
2 Sounds good
3 Possible, but people don't usually say it that way
4 No Kumyk person would say it that way

(There are a number of sets of sentence pairs which are identical except for the form of pronoun used in the second sentence. The list of sentences in Kumyk are shown here for reference, but the summary of responses is given with the English translations which follow.)

Version A:

1. А. Муаллим студентге китап берди.
   Б. Бу „Тангчолпан" эди. 1 2 3 4

2. А. Муаллим студентге китап берди.
   Б. Шу савбол деди. 1 2 3 4

3. А. Болат атына минип барагъанда йыгъылды.
   Б. Ону къолу сынды. 1 2 3 4

4. А. Муаллим студентге китап берди.
   Б. Шондан сонг муаллим студентни къолун алды. 1 2 3 4

5. А. Болат атына минип барагъанда йыгъылды.
   Б. Барын да тамаша этди. 1 2 3 4

6. А. Муаллим студентге китап берди.
   Б. Сонг муаллим оьзюню экзамени булан къутлады. 1 2 3 4

7. А. Болат атына минип барагъанда йыгъылды.
   Б. Болат дагъы бугъар минмежегин айтды. 1 2 3 4

8. А. Муаллим студентге китап берди.
   Б. Оьзю савбол деди. 1 2 3 4

9. А. Болат атына минип барагъанда йыгъылды.
   Б. Бу дагъы огъар минмежегин айтды. 1 2 3 4

10. А. Муаллим студентге китап берди.
    Б. „Тангчолпан” эди. 1 2 3 4

11. А. Болат атына минип барагъанда йыгъылды.
    Б. Шо тюнегюн болду. 1 2 3 4

12. А. Муаллим студентге китап берди.
    Б. Сонг муаллим муну экзамени булан къутлады. 1 2 3 4

13. А. Болат атына минип барагъанда йыгъылды.
    Б. Къолу сынды. 1 2 3 4

14. А. Муаллим студентге китап берди.
    Б. Бу савбол деди. 1 2 3 4

15. А. Муаллим студентге китап берди.
    Б. Шундан сонг муаллим студентни къолун алды. 1 2 3 4

16. А. Болат атына минип барагъанда йыгъылды.
    Б. Бу барын да тамаша этди. 1 2 3 4

17. А. Муаллим студентге китап берди.
    Б. Сонг муаллим экзамени булан къутлады. 1 2 3 4

18. А. Муаллим студентге китап берди.
    Б. Савбол деди. 1 2 3 4

19. А. Болат атына минип барагъанда йыгъылды.
Б. Болат дагъы огъар минмежегин айтды. 1 2 3 4

20. А. Муаллим студентге китап берди.
Б. О „Тангчолпан” эди. 1 2 3 4

21. А. Муаллим студентге китап берди.
Б. Сонг муаллим ону экзамени булан къутлады. 1 2 3 4

22. А. Болат атына минип барагъанда йыгъылды.
Б. О барын да тамаша этди. 1 2 3 4

23. А. Муаллим студентге китап берди.
Б. Сонг муаллим студентни къолун алды. 1 2 3 4

24. А. Болат атына минип барагъанда йыгъылды.
Б. Бу тюнегюн болду. 1 2 3 4

25. А. Муаллим студентге китап берди.
Б. Шу „Тангчолпан” эди. 1 2 3 4

26. А. Болат атына минип барагъанда йыгъылды.
Б. Шо барын да тамаша этди. 1 2 3 4

27. А. Муллим студентге китап берди.
Б. Сонг муаллим шону экзамени булан къутлады. 1 2 3 4

28. А. Муаллим студентге китап берди.
Б. О савбол деди. 1 2 3 4

29. А. Муаллим студентге китап берди.
Б. Ондан сонг муаллим студентни къолун алды. 1 2 3 4

30. А. Болат атына минип барагъанда йыгъылды.
Б. Шу барын да тамаша этди. 1 2 3 4

31. А. Болат атына минип барагъанда йыгъылды.
Б. Тюнегюн болду. 1 2 3 4

32. А. Болат атына минип барагъанда йыгъылды.
Б. Оьзюню къолу сынды. 1 2 3 4

33. А. Муаллим студентге китап берди.
Б. Шо „Тангчолпан” эди. 1 2 3 4

34. А. Болат атына минип барагъанда йыгъылды.
    Б. О дагъы огъар минмежегин айтды. 1 2 3 4

35. А. Муаллим студентге китап берди.
    Б. Сонг муаллим шуну экзамени булан къутлады. 1 2 3 4

36. А. Муаллим студентге китап берди
    Б. Мундан сонг муаллим студентни къолун алды. 1 2 3 4

37. А. Болат атына минип барагъанда йыгъылды.
    Б. О дагъы шогъар минмежегин айтды. 1 2 3 4

38. А. Болат атына минип барагъанда йыгъылды.
    Б. О тюнегюн болду. 1 2 3 4

39. А. Муаллим студентге китап берди.
    Б. Шо савбол деди. 1 2 3 4

40. А. Болат атына минип барагъанда йыгъылды.
    Б. Болат дагъы шогъар минмежегин айтды. 1 2 3 4

41. А. Болат атына минип барагъанда йыгъылды.
    Б. О тюнегюн болду. 1 2 3 4

Sentences from Version B that do not appear in A:

42. А. Муаллим студентге китап берди.
    Б. Студент, ону алып, уьйге гетди. 1 2 3 4

43. А. Мурат оьзюню машини булан тюкенге барма токъташды.
    Б. Сонг о шону дёгерчеги бошалгъанны билди. 1 2 3 4

44. А. Мариям анасына алма берди.
    Б. О бугъар разилигин билдирди. 1 2 3 4

Combined results for Versions A and B

(B results don't show orange responses (--), which were eliminated because all but one response was judged acceptable.)

1. The teacher gave a book to the student.
__ was Tangcholpan. (cf. 1, 10, 20, 25, 33)

Responses:
bu (3, --, 1; 3, 2)
Ø (4, 3). [only version A]
o (4, 2). [only version A]
shu (2, --, 3; 4, 3)
sho (1, --, 1; 1, 2)

2. The teacher gave a book to the student.
__ said thank you. (cf. 2, 8, 14, 18, 28, 39)

Responses:
shu (4, --, 4; 4, 2)
o'z (3, --, 4; 4, 3)
bu (2, --, 2; 2, 2)
Ø (4, --, 4; 3, 4)
o (2, 2) [only version A]
sho (3, --, 4; 4, 2)

3. The teacher gave a book to the student.
__Then the teacher congratulated (him) on (his) exam. (cf. 6, 12, 17, 21, 27, 35)

Responses:
o'z (4,4)
bu (1, --, 2; 3, 2)
Ø (3, --, 3; 4, 3)
o (1, 2) [only version A]
sho (3, --, 4; 4, 3)
shu (3, --, 4; 4, 3)

4. The teacher gave a book to the student.
After __ the teacher shook the student's hand. (cf. 4, 15, 23, 29, 36)

Responses:
sho (1, --, 1; 1, 2)
shu (3, --, 3; 2, 2)
Ø (1, --, 1; 1, 3)
o (1, --, 1; 4, 2)
bu (2, --, 4; 4, 3)

5. Bolat fell off while riding his horse.
__ hand broke. (cf.. 3, 13, 32)

Responses:
onu (1, --, 1; 1, 2)
Ø (2, --, 3; 3, 2)
o'z (1, --, 3; 4, 2)

6. Bolat fell off while riding his horse.
Bolat said he(Ø) wouldn't ride __ anymore. (cf. 7, 19, 40)

Responses:
bu (2, --, 2; 3, 2)
o (1, --, 2; 4, 2)
sho (1, --, 2; 1, 2)

6. Bolat fell off while riding his horse.
He (bu) said he(Ø) wouldn't ride __ anymore. (cf. 9)

Responses:
o (1, --, 2; 3, 2)

6. Bolat fell off while riding his horse.
He (o) said he(Ø) wouldn't ride __ anymore. (cf. 34, 37)

Responses:
o (4, 2) [only version A]
sho (1, --, 1; 1, 2)

7. Bolat fell off while riding his horse.
__ was yesterday. (cf. 11, 24, 31, 38)

Responses:
sho (2, --, 1; 1, 2)
bu (2, --, 3; 3, 2)
Ø (3, 3) [only version A]
o (4, 2) [only version A]

8. 'Bolat fell off while riding his horse.
__ amazed everyone.' (cf. 5, 16, 22, 26, 30)

Responses:
Ø (2, --, 3; 2, 3)
bu (2, --, 2; 2, 4)
o (1, --, 1; 1, 3)
sho (2, --, 3; 1, 3)
shu (2, --, 3; 1, 3)

9. ‘The teacher gave the student a book
The student, taking __ went home.’ (cf. 42)

Responses:
o (1, --, 1) [only version B]
sho (1, --, 2) [only version B]
bu (2, --, 2) [only version B]
shu (2, --, 4) [only version B]

10. ‘Murat decided to go to the store with his(own) car.
Then he(o) found out that __’s tire was flat.’ (cf. 43)

Responses:
sho (2, --, 3) [only version B]
o (2, --, 3) [only version B]
bu (1, --, 3) [only version B]

11. Mariam gave an apple to (her) mother.
She made known her pleasure (at/to ___). (cf. 44)

Responses:
sho (3, --, 2) [only version B]
o (2, --, 2) [only version B]
bu (1,--, 2) [only version B]

Дата анкетного опроса:	Date of questionnaire
Возраст:	Age
Место рождения:	Place of birth
Места жительства:	Place of residence

Red – 28 yrs. born/resides in Buinask region
Blue 19 yrs. born/resides in Makhachkala
Orange 20 yrs. born/resides in Babayurt region
Black – 19 yrs. born/resides in Makhachkala
Green – 64 yrs. born/resides in Buinask region