The synchrony and diachrony of New Western Iranian nominal morphosyntax

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy
in the Graduate School of The Ohio State University

By
Shuan Osman Karim

Graduate Program in Linguistics
The Ohio State University
2022

Dissertation Committee:
Professor Brian D. Joseph, Advisor
Professor Andrea Sims
Professor Björn Köhnlein
Professor Ashwini Deo
Professor Robert Levine

© Shuan Osman Karim, 2022

Abstract

There is rich diversity in New Iranian nominal systems reflecting retentions from a common Old Iranian ancestor and many significant innovations. My primary aim is to engage in a discussion of the typological richness of inflection among these languages as comprehensively as possible. This work represents a combination of synchronic and diachronic linguistics, where a thorough theoretically anchored synchronic analysis feeds my otherwise diachronic study. I have divided the work into distinct sections that represent issues concerning the nominal morphology of the Iranian languages, focusing on those spoken in the Kurdish zone.[1] These sections are distinct, yet taken together, they demonstrate the breadth of issues concerning New Iranian nominal morphology.

I begin with a typological overview of nominal systems in New Iranian languages focusing on the interaction of case, number, gender, and attribution marking. At the intersection of these features exist several patterns that establish the issues addressed in subsequent sections. Perhaps the most well-studied phenomenon in Iranian nominal morphology is the ezafe (attribution marker). Here, I break from previous work on the ezafe phenomenon rooted in the minimalist tradition by describing the syntactic combinatorics of the various attribution/possession strategies in New Iranian languages in a categorial framework (HTLCG). This analysis unifies two facts of Iranian languages: (1) adjectives are both attributive and substantive, and (2) nouns are the marked entity in attributive constructions. I use the principles established in my discussion of the ezafe to unify the analysis of the Iranian noun-phrase syntax and nominal morphology based on foundational assumptions of

[1] I use the term "Kurdish zone" to refer to the parts of Iran, Iraq, Syria, and Turkey where the majority of the population consider themselves to be ethnic Kurds.

inferential-realizational morphology and Categorial Grammar. CGs assume that the lexicon stores syntactic and semantic functors. Inferential-realizational morphology assumes that morphology structures and organizes the lexicon.

Based on my analysis of the ezafe and the juxtaposition of nominal sub-systems within the New Iranian languages, I selected three substantial issues in the diachrony of these languages: (1) *why Kurmancî has no definite article* and (2) *the origin of the d-form ezafat in Zazaki*. These represent three open questions in Iranian and Kurdish-zone linguistics. It has long been assumed (at least since MacKenzie, 1960) that Kurmancî or Northern Kurdish and its neighbor Zazaki represent particularly conservative varieties preserving much of Old Iranian. However, I argue that much of their inflectional richness can be attributed to mutual (areal) innovation. Patterns in their inflectional systems and their syntactic combinatorics point to new solutions to well-established problems.

(1) *Why Kurmancî and Zazaki have no definite article* represents the first of the three historical vignettes. Here, I address the claim that the definite article *-eke* in Soranî (Central Kurdish) and Hewramî (Gorani) is an innovative feature in these two languages (MacKenzie, 1960). This assertion was based largely on the fact that Kurmancî and Zazaki, genetically related to Soranî and Hewramî, respectively, are missing the definite article despite their conservative appearance. I consider the syntactic combinatorics of the ezafe and its interaction with other inflectional features to make a case that both Kurmancî and Zazaki are highly innovative. Their innovations are coincidentally the source of their conservative appearance. Considering the nominal paradigms in Kurmancî and Zazaki (and in other ezafe-languages), I propose a univerbation between the original ezafe particles and the inherited definite article. This development is based on some circumstantial evidence: Kurmancî and Zazaki feature differential case marking; they feature group inflection, a phenomenon where only the last of a list of conjoined nouns are inflected; their combinatorics match a strategy of definite ezafe marking in other languages, and the ezafe has separate allomorphs for nondefinite nouns.

(2) *The origin of the d-form ezafat in Zazaki* represents a deep look at the *d*-form ezafat that appear in Southern Zazaki. These formatives occur when a noun is followed by a modifier (attributive adjective or nominal possessor), as do all ezafe allomorphs. However, they mark a noun that is indefinite, genitive, dative, allative, a prepositional complement, a denominal postpositional complement, a complement of a (non-denominal) postposition that is a peripheral argument (i.e. required by the verb), or topicalized.[2] I dispel some assumptions about these forms (e.g. that they are borrowings from Aramaic), showing that this is likely a language-internal development, and the seemingly strange distribution of the *d*-forms can be explained by a historical incompatibility between the proto-ezafe and indefinites and genitives. I expand on this idea by bringing in evidence from Northern and Central Zazaki, which did not develop these forms but maintained the original *d*-form pronouns cognate with the *d*-form ezafat.

This dissertation utilizes data from diverse languages traditionally classified as western Iranian. Among these languages, there appears to be a continuum of inflectional systems, including ezafe marking from the morphologically rich Kurmancî and Zazaki to the near inflection-free New Persian. The languages of the Caspian region and Baluchi have not developed ezafe marking in the same way. Instead, they feature explicit attributive marking on adjectives and genitive marking on substantives. This dissertation challenges the received wisdom regarding the origin of and relationships between these languages based on their nominal inflection systems alone. At the core of this challenge is the traditional conception of which languages are "archaic" and which are innovative. Once the equation of complexity and conservatism is dismissed, the solutions to long-standing Iranian historical linguistics issues can be resolved. I employ a mixture of synchronic study based on an intimate knowledge of the New Iranian languages and classical historical linguistic analysis to develop a deeper understanding of the prehistory and connections between these languages and their

[2]The peripheral arguments of the *d*-form ezafat were previously described as "optional" (following Paul, 1998). Topicalized functions were described as native-speaker errors (following Hadank, 1932).

nominal systems.

Acknowledgements

Thank you to my wife, Alicia, who has supported me since the day I told her I wanted to dedicate my life to the study of languages and the people who speak them.

Thank you to my son, Mithra, who was born the month after I started grad school. Mithra, we just celebrated your fifth birthday, and you have only known your baba from behind a desk. May we spend the following years of your life experiencing what the world has to offer.

Thank you to my adviser, Brian Joseph, who has been like a father to me over the past five years. I am truly blessed to have you in my corner as a mentor, adviser, and friend. I hope our relationship deepens as I become the scholar you have prepared me to be.

Thank you to Andrea Sims for helping me to stay grounded. I have been able to rely on you for an honest, direct, and constructive opinion since the day that I met you. Listening to your insights and working alongside you has made me a better linguist. I hope that our relationship continues in the years to come.

Thank you to all the members of my dissertation committee, Brian Joseph, Andrea Sims, Björn Köhnlein, Ashwini Deo, and Bob Levine. Without your guidance, I would not have succeeded. When I read my work and ponder new avenues of inquiry, I see your ideas in my text, and I hear your words in my thoughts. For all this and more, I am eternally grateful.

Vita

A.A.S in Japanese language and culture, Austin Community College, 2015.

B.A. in Linguistics with Special Honors, the University of Texas at Austin, 2015.

M.A. in Linguistics, the Ohio State University, 2021.

Benjamin A. Gilman International Scholarship, 2014.

Terry Foundation Transfer Scholarship, 2013-15.

Extended Dean's Distinguished University Fellowship, 2016-21.

Graduate Teaching Associate, The Ohio State University Department of Linguistics, 2019-20.

Publications

Shuan Osman Karim. Absolute Prepositions. To appear in *The Oxford Handbook of Kurdish Linguistics*, eds. Jaffer Shekholislami, Geoffrey Haig, Ergin Öpengin, and Haidar Khezri, forthcoming.

Shuan Osman Karim and Ali Salehi. An applicative analysis of Soranî "absolute prepositions." To appear in *Neglected syntactic functions and non-syntactic functions of Applicative Morphology*, eds. Sara Pacchiarotti and Fernando Zúñiga, forthcoming.

Shuan Osman Karim. The Ezafe and the Article. To Appear in *The Proceedings of the second North American Conference on Iranian Linguistics*, forthcoming.

Shuan Osman Karim. Akkadian [e]. To appear in *the Journal of the American Oriental Society*, 2021.

Micha Elsner, Andrea D. Sims, Alexander Erdmann, Antonio Hernandez, Evan Jaffe, Lifeng Jin, Martha Booker Johnson, **Shuan Osman Karim**, David L. King, Luana Lamberti Nunes, Byung-Doh Oh, Nathan Rasmussen, Cory Shain, Stephanie Antetomaso, Kendra V. Dickinson, Noah Diewald, Michelle McKenzie, and Symon Jory Stevens-Guille. Modeling morphological learning, typology, and change: What can the neural sequence-to-sequence framework contribute? *Journal of Language Modelling* 7.1, 126-170, 2019.

Fields of Study

Major Field: Linguistics

Table of Contents

Abstract ii

Acknowledgements vi

Vita vii

List of Figures xii

List of Tables xiii

1 Introduction 1
1.1 Notes on stylistic choices 2
1.2 Background on Iranian 5
1.3 The languages referenced in this work 9
1.4 The linguistic study of Central and Southwestern Iranian 10
1.4.1 Broad-scoping studies 13
1.4.2 The study of particular Northwestern Iranian languages 26
1.4.3 The study of particular Southwestern Iranian languages 49
1.5 Thoughts and Conclusions 53

2 New Iranian Nominal Morphology 65
2.1 Introduction 65
2.1.1 Background on the Iranian languages 66
2.2 Case systems 68
2.2.1 Cases systems in New Iranian 73
2.2.2 a Typology of New Iranian case 87
2.3 Definiteness marking strategies 91
2.3.1 The indefinite markers 92
2.3.2 The definite markers 95
2.4 Attribution marking strategies 103
2.4.1 The interaction of (in)definiteness and ezafe 104
2.4.2 Summary of forms 113
2.4.3 A typology of case, definiteness and attribution 115
2.5 Thoughts and conclusions 117
2.5.1 The definite ezafe 117
2.5.2 The indefinite ezafe 119
2.5.3 Resolution of the historical questions 120

3 An HTLCG account of the Ezafe phenomenon . . . 121
3.1 Introduction . . . 121
3.2 What is an ezafe . . . 124
3.3 Introduction to HTLCG . . . 130
3.4 A HTLCG approach to the ezafe . . . 134
3.4.1 The type adjective . . . 136
3.4.2 The ezafe and the Iranian adjective . . . 138
3.4.3 The Ezafe explored . . . 145
3.4.4 Reverse ezafe . . . 169
3.4.5 Notes on compounding strategies . . . 183
3.5 Typology of ezafat . . . 184
3.6 Diachronic patterns in the development of ezafat . . . 188
3.6.1 Directionality . . . 189
3.6.2 The development of nominalizers from the construct state . . . 192
3.6.3 Against Rießler's (2016) diachrony (DEM > NMLZ > ATTR) . . . 199
3.6.4 The role of syntactic functors in morpho-syntactic change . . . 201
3.7 Reflections . . . 204

4 Why Kurmancî has no Definite Article . . . 208
4.1 Introduction . . . 208
4.1.1 The development of the K-form in Kurmancî . . . 211
4.2 The loss of K-forms in Kurmancî . . . 212
4.2.1 The Proto-Iranian *-(V)ka- marker was inherited by the daughter languages . . . 213
4.2.2 The reflex of *-(V)ka-, used to express an evaluative meaning, gives way to deictic and recognitional functions and eventually to anaphoric function . . . 216
4.2.3 The Iranian languages rearrange themselves into multiple sprachbünde dividing Kurdish and Zaza-Gorani into a northern and a southern sprachbund. . . . 218
4.2.4 The northern sprachbund undergoes a subsequent change, where intervocalic *k is lenited and lost. . . . 222
4.2.5 Further reductions and syncretism result in case disappearing from most Kurdish varieties. It is paradoxically retained in the varieties in which the loss of *k obscured the definite article. . . . 223
4.3 Impact of this hypothesis/further research . . . 234

5 The Origin of the *d*-form Ezafe . . . 239
5.1 Introduction . . . 239
5.2 Background . . . 242
5.3 Paradigmatic Changes . . . 244
5.4 Displacement of the Proto-Ezafe . . . 248
5.4.1 The d-forms . . . 249
5.4.2 Displacement of the Proto-Ezafe by the Indefinite Article . . . 250

5.4.3 Displacement of the Proto-Ezafe by the Head Noun in Genitival Constructions . . . 253
5.4.4 Displacement of the Proto-Ezafe by a Postposition . . . 255
5.4.5 The Remaining Forms (Paul's Oblique IIa) . . . 256
5.4.6 An Alternate Theory . . . 261
5.5 Conclusion . . . 262

6 Big Picture: the place of this study in (Iranian) Linguistics . . . 265

A Language Family Tree: Central Iranian and Southwestern Iranian . . . 270

B List of Abbreviations . . . 275
References . . . 281

List of Figures

4.1 K-form definite articles, reduced forms, and reduced forms with IOM . . . 221

List of Tables

1.1 Articles on the ezafe 27

1.2 Bound Volumes on Central and Southwestern Iranian Linguistics 55

2.1 The loss of final codas between Old and Middle Iranian 70

2.2 Case syncretism due to accidental homophony with DAT-GEN and INST-ABL mergers 70

2.3 Parthian and Middle Persian declension 72

2.4 Suleymanî Soranî case(lessness) 74

2.5 Vafsi case (Stilo, 2008a, 703) 75

2.6 Stilo's (2008a) typology of Iranian case systems 76

2.7 Hewramî Luhon case 77

2.8 Southern Zazaki case 78

2.9 Standard Kurmancî case 82

2.10 Amadî Kurmancî (Behdînî) case 82

2.11 Muş Kurmancî case 83

2.12 Standard New Persian case 84

2.13 Colloquial New Persian case 85

2.14 Turkmen Balochi case 86

2.15 Koroshi Balochi case 87

2.16 Stilo's (2008a) typology of Iranian case systems (updated) 88

2.17 The relationship between the numeral "one" and the indefinite article . . . 92

2.18 Strategies for Marking Definiteness and Adjectival Attribution 114

2.19 Ezafe Syncretism 115

2.20 Soranî definiteness and attribution 115
2.21 Zazaki case, definiteness, and attribution 116

3.1 Baluchi adjectival inflection *mazan* 'big' 181
3.2 Typology of ezafat 185
3.3 Ezafe types by language 187

4.1 Grammatical convergence between Kurdish-zone languages (edited and expanded from Jügel, 2014, 136) 210
4.2 Kurdish morphological features (not shared by ZG) 219
4.3 Northern-belt features shared by Kurmancî 220
4.4 Case syncretism due to accidental homophony with DAT-GEN and INST-ABL mergers 224
4.5 Inherited case syncretism 225
4.6 *a*-stem and *aka*-stem forms 226
4.7 Hypothetical case system 226
4.8 The reorganized Kurmancî system 231

5.1 Blocking of the ezafe in Hewramî 246
5.2 Case and Number by Ezafe Type 247
5.3 The Indefinite Nominal Declension 251
5.4 The Demonstrative Circumposition in Kurdish Zone Languages 259
5.5 d-forms in Zaza-Gorani Languages 262

Chapter 1

Introduction

The current study consists of an overview of the nominal morphology of Western Iranian languages and a deep dive into several related topics, diachronic and synchronic. My ongoing research is geographically seated in the heart of the Kurdish zone, the area where the majority of the people living there consider themselves to be ethnic Kurds. However, in an effort to consider all relevant information, I cannot limit myself to just Kurdish; rather, I must consider data from related languages and develop a comprehensive account, especially on the diachronic side.

I begin with some background on the relevant languages and varieties, including a summary of linguistic research on some languages and essential topics within Western Iranian linguistics. One of these topics that has received a significant amount of attention is the ezafe, or modification marking on the head noun. Despite its long history of study and various theories regarding its nature, syntactic and morphological, I do not believe it has accurately been characterized theoretically. In chapter 3, I redefine the ezafe as a derivational morpheme that changes the syntactic category of a noun from a primitive to a functor looking to combine with a nominal. I do so in Hybrid Type-Logical Categorial grammar (HTLCG), a variant of Categorial Grammar (CG) developed by Bob Levine (see Kubota & Levine, 2020). In addition to my account of the ezafe, I believe that the CG approach may elucidate some aspects of the ezafe's diachronic development.

In chapter 2, I give an overview of the nominal morphology of Western Iranian languages. The core focus of this section is to present the relevant data from a variety of languages and to develop both a typological characterization of the various systems observed and a

diachronic account of how they arose. The main thread throughout the section is to show how case, number, gender, definiteness, and ezafe interact to create the diverse systems observed. Additionally, several issues arise due to how these features combine (or don't combine) that may shed light on their ultimate origins. This chapter is essentially an elaboration of Karim (2021c) and the intellectual successor of both Karim (2021c), an exploration of the interaction between definiteness and ezafe marking, and Stilo (2008a), an exploration of case in New Western Iranian.

A significant sub-section of this exploration is reserved for chapter 4, a deeper exploration of the interaction between definiteness and case with a focus on Kurmancî. In this chapter, I present evidence from across the Iranian world to build a case for the idea that Kurmancî has preserved the reflexes of an older inherited definite article in its case-system and ezafat (plural of ezafe). As the previous sections ask and answer several open questions about the ezafe, definiteness, and case marking in the Western Iranian languages, I complete this study with chapter 5, an exploration of the ezafe featuring the phoneme *d* in Zazaki. The ezafe with *-d-* perhaps the most significant counter-evidence to the unchallenged theory that the ezafe is the reflex of the Old Iranian relative pronoun *ya-. I assert that *ya- is the source of some but crucially not all ezafat.

Before I begin this exploration, it is essential to understand a few aspects of the development and state of Iranian studies and the relevant languages setting the stage for this discussion. I thus provide some background here on the history of Iranian studies and the languages and the people who speak them. I reserve all discussion of relevant grammatical points for chapter 2.

1.1 Notes on stylistic choices

There are several stylistic decisions I have made here. To facilitate understanding of the examples I provide throughout this dissertation, I provide a master list of all abbrevia-

tions and orthographic symbols I have employed. All glossing generally conforms to the Leipzig Glossing Rules. Deviations from Leipzig are only used for Iranian-specific terms. For instance, the term "ezafe" (EZ) is used to refer to head marking in a complex noun phrase. The ezafe is variably referred to as the construct state. Both terms are based on the Arabic grammatical tradition, where they are employed to describe a similar phenomenon. Likewise, the term "reverse ezafe" (REZ) is a term coined by Don Stilo to refer to a similar phenomenon in the languages of the Caspian region, which marks the modifier in a complex noun phrase. The terminology serves to remind the linguist that the reverse ezafe perhaps belongs to a constellation of related constructions. However, there is already existing terminology for both the attributive and possessive reverse ezafat (plural of ezafe), namely attributive marker (ATT) and genitive (GEN). For this reason, I use reverse ezafe in discussions of the reverse ezafe and genitive and attributive more generally. A complete list of glosses employed is included in appendix B.

I have chosen not to use abbreviations for languages in the text in most instances. The exception to this tendency is where data from multiple language varieties are referenced in succession. In these instances, I use the full name in the first instance, and just the initial corresponding to the varietal in further iterations, e.g. Rakhshani Balochi followed by R Balochi. To facilitate exposition in a few tables, I have fully abbreviated the names of some languages. A complete list of those abbreviations is included in appendix B.

Additionally, there are quite a few syntactic, semantic, and prosodic operators employed throughout this work and especially in section 3. In section 3, these operators and the corresponding orthographic conventions are explained in detail. A master list of these operators and variables is included in appendix B.

I employ data from various Iranian languages, each with its own history. These languages were studied at different times by different scholars. As a result, there are various scripts used for the native orthographies of these languages and a variety of systems used for their transliteration. Here, I use native orthography whenever said orthography is in a variant

of the Latin alphabet. When the native orthography is in a non-Latin alphabet, I defer to the transliteration system employed in the source material.

In some cases, my knowledge of the languages is good enough to convert the data into IPA in principle. However, I have chosen not to do so as my primary aims are morphological analysis and reconstruction, i.e. not phonological. There are several instances where I have chosen to transliterate certain examples: (1) If a language has two standard orthographies, one based on the Latin alphabet and one otherwise, I have transliterated these texts into the native orthography based on the Latin alphabet regardless of how the original data was transcribed/transliterated. For example, Soranî (Central Kurdish) is natively written in a modified Arabic script and occurs natively in the Hawar script based on the Latin alphabet. Despite the existence of Hawar, some scholars have chosen and even continue to choose to transliterate Soranî data in any number of orthographies. I use Hawar for all Kurdish languages and varieties except for Zazaki because Zazaki has its own version of the Latin alphabet. The Zazaki script is identical to Hawar except for two vowels: Hawar *î*, Zazaki *i*, IPA iː and Hawar *i*, Zazaki *ı*, IPA ɪ̆. (2) I have changed the transliteration in erroneous examples. Note that I do not correct any native speaker data, which has been a problem that has plagued Iranian studies from very early on. What I do correct are errors based on the mix of transliteration systems employed. As different authors use their own sometimes idiosyncratic transliteration system, multiple systems are often used in a single example, making the reading inconsistent or even impossible.[1] I have corrected the transliteration to match Hawar for Kurdish examples and the system of choice within the work cited for all others.

[1] Examples employing inconsistent transliteration systems are too numerous to cite and too infamous to warrant doing so. Additionally, I do not intend to "call out" any other scholar only to explain my orthographic choices.

1.2 Background on Iranian

Since Geiger (1901), the Iranian languages have been classified according to their geographic location, thus Northwestern, Southwestern, Northeastern, and Southeastern languages. This geographic designation was deemed problematic from very early on. However, "no other model has taken its place so far" (Korn, 2019a, 240). Despite Korn's (2019a) comment, it seems that her model of Central Iranian and peripheral sub-families is now likely to replace the traditional four-way geographic division of the Iranian languages. Korn's (2019a) view is the one that has now been adopted by Glottolog (Hammarström et al., 2020). The only resistance to her model comes from inertia and institutional memory.

In the Western-Iranian world, there is a long-held view that there is a binary genealogical division. This view is rooted in Tedesco's (1921) pioneering work on Iranian dialectology, which looked at the Middle Western Iranian text discovered in the Turfan expeditions. The Turfan expeditions were a series of four German archaeological explorations of Turfan, an oasis town in Chinese Turkistan (Xinjiang), between 1902 and 1914 (see Zhang & Rong, 1998, 24). These expeditions were organized by Albert Grïwedel and Albert von Le Coq. They brought back to Germany paintings, art, and more than 40,000 fragments of text. These fragments included texts in Old Turkish, Chinese, Sanskrit, Sogdian, Middle Persian, New Persian, Parthian, Tibetan, Mongolian, Prakrit, Tumšuq Saka, Tocharian A, Bactrian, Khotan Saka, Tocharian B, Hebrew, Syriac, Arabic, Tangut, Greek, Khitan, and other as-of-yet unidentified languages in a variety of scripts (der Wissenschaften, 2007). Among these were several previously unattested varieties. Based on a comparison between Middle Persian and the newly extant Parthian Tedesco (1921) developed his binary view of Iranian.

Tedesco's (1921) view was that there are (South-)Western and (North-)Eastern Iranian groups (Tedesco, 1921 apud Skjærvø, 2017a, 472). Another way to conceptualize this division is as Persian and non-Persian Iranian. Tedesco's (1921) classification was based on a series of phonological isoglosses not shared across the Western Middle Iranian languages

Middle Persian and Parthian, with Middle Persian as an exemplar of the Southwest and Parthian as an exemplar of the Northwest. Tedesco's (1921) original 30 isoglosses were mostly phonological[2] but also included lexical and morphological isoglosses. Paul (1998a) took twelve of Tedesco's (1921) original phonological isoglosses, the reflexes of Old Iranian *k̡, *ĝ, *k (/ _V$_{[+front]}$), *g (/ _V$_{[+front]}$), *k̡w, *tr, *dw, *rd, *rz, *sw, *tw, and *y (/ #_), as a basis for the comparison of New Iranian languages with the Middle Persian and Parthian exemplars. His ultimate conclusion was that the New Western Iranian languages do not form two coherent groups Northwestern and Southwestern Iranian but rather a continuum of northern-ness. This conceptualization marked an important point in Iranian dialectology. However, its methodology was soon questioned by Korn (2003). The issue with Paul's (1998a) (and ultimately Tedesco's (1921)) analysis is that many if not the majority of the phonological innovations considered are not significant. For instance, the shift from the palatal glide [j] to the affricate d͡ʒ is known to have occurred in languages across the world from different languages families at different times in history. At the core of historical comparative linguistics is the assumption that only significant shared innovations can reliably demonstrate subgroupings within an established genealogical relationship. These are features of languages that occur in cognate vocabulary, and are not likely to have occurred independently. Looking only at significant shared innovations, Paul's (1998a) continuum loses some of its neat geographical arrangement.

In the Eastern-Iranian world, there are additional problems; the languages traditionally classified as southeastern–Ormuri, Parachi, Pashto, Saka, Sanglechim, Ishkashimi, Shughni, Yazgulami, Wakhi, Yidgha, Munji, and their varieties–do not constitute a sin-

[2] Tedesco's (1921) original phonological isoglosses are (1) the reflex of PIE *ĝ (N(W): *z*, S(W): *d*), (2) the reflex of PIE *ĝ (N(W): *s*, S(W): *h* (< θ)), (3) the reflex of Old Iranian *j (N(W): *ž*, S(W): *z*), (4) the reflex of Old Iranian *č (N(W): *č/ž*, S(W): *č/z*), (5) the reflex of Old Iranian *y (N(W): *y*, S(W): *ǰ*), (6) the reflex of intervocalic *d (N(W): *d*, S(W): *i*), (8) the reflex of intervocalic *g (N(W): *g*, S(W): *y/'*), (10) the reflex of Old Iranian *θr (N(W): *hr*, S(W): *s*), (11) the reflex of Old Iranian *θu̯ (N(W): *f*, S(W): *h*), (12) the reflex of Old Iranian *du̯ (N(W): *b*, S(W): *d*), (13) the reflex of Old Iranian *θy (N(W): *h*, S(W): *š*), (15) the reflex of Old Iranian *rd (N(W): *rd/l(r)*, S(W): *l(r)*), (16) the reflex of Old Iranian *rz (N(W): *rz*, S(W): *l(r)*), (17) the reflex of Old Iranian *xu̯ (N(W): *wh*, S(W): *xw*), and (19) the reflex of Old Iranian *m (N(W): *w*, S(W): *m*).

gle group but rather fall into seven distinct groups (Hammarström et al., 2020). These groups do not share enough isoglosses to be linked as such. Likewise, Northeastern Iranian languages–Ossetian, Sogdian, Yaghnobi, and Bactrian–form at least two distinct groups. Korn (2016a) showed that Bactrian, in particular, possessed both prototypically-eastern and prototypically-western features. In other words, Bactrian is not obviously western or eastern. This led her to propose a Central Iranian with Bactrian at its core, perhaps the most significant recent development in Iranian dialectology. Although the possibility of a three-way branching of Iranian into Southwestern, Northeastern, and Central groups was proposed by Schmitt (1989, 27-28) based on phonology, Korn's (2016a) proposal culminating in her (2019a) "Isoglosses and subdivisions of Iranian" established the boundaries of the Central Iranian sub-family.

The tree of Central Iranian (PBS) contains three sub-branches that can be classified according to their known Middle Iranian members Parthian, Bactrian, and Sogdian. Note that Bactrian and Sogdian were formerly considered to be Northeastern Iranian languages along with Ossetian. In Central Iranian, Sogdian(/Yaghnobi) and Ossetian form a sub-branch; Bactrian is terminal on its own sub-branch, and Parthian exists on a sub-branch with Adharic, Balochic, Caspian, Central Iranian Kermanic, Komisenian, Laki, Kurdish, Semnani, and Biyabuneki. Note that this group corresponds to what was traditionally labeled Northwestern Iranian, and the moniker has been retained. In addition to the Central Iranian sub-branch of Iranian, there are nine more sub-branches: Ormuri-Parachi, Pashto, Saka, Sanglechi-Ishkashimi, Shughni-Yazgulami, Wakhi, and Yidgha-Munji, which were traditionally thought of as Southeastern, as well as Southwestern Iranian, corresponding to Persian and its close sisters, and Avestan which does not easily fit into any grouping. These nine groups are to be understood as independent branches without any known hierarchy among them. Of course, some similarities may have developed among them due to shared geographical space. Like Northwestern Iranian, the moniker Southwestern Iranian has survived the reconceptualization of the Iranian branch of the Indo-European language

family.

In summary, the four-way geographical designation into Northwest, Southwest, Northeast, and Southeast (Geiger, 1901) has proven to be problematic. This view has always had a few flaws: (1) the geographic terminology is not descriptive. Based on significant shared innovations alone, Balochi is a Northwestern Iranian language.[3] However, it is spoken in the far southeast of the Iranian world, spanning the borders of Iran, Afghanistan, and Pakistan. Likewise, Ossetic, a Northeastern Iranian language, is spoken in the far northeast of the Iranian world, on the border between Georgia and the Russian Caucasus. An additional flaw with this terminology is that (2) there may not be any significant groupings that correspond to the original geographic terminology. Although it is relatively straightforward that the Northwestern Iranian languages form a coherent group, and the same can be said for Southwestern Iranian, the same is not true of the sub-families in the eastern Iranian world. Korn's (2019a) model showing a large Central Iranian sub-branch and many other distantly related peripheral sub-branches is beginning to replace the older geographically based classification system. Perhaps the only resistance to this change is institutional memory.

Featured in this study are the languages of the Kurdish zone; these languages belong to the Northwestern Iranian and the Southwestern Iranian branches of Iranian. Northwestern itself is but one significant part of the Central Iranian branch. I reference data from the Sogdian and Bactrian branches of Central Iranian only when relevant to a particular point about Northwestern Iranian. However, their place in this study is tangential. Although Korn's (2019a) update to the classification of Iranian has placed Northwestern Iranian and Southwestern Iranian on separate branches, they have nonetheless shared many innovations that make their mutual study relevant. Shared features such as ezafe, modification-marking on the nominal head, and a definite suffix, etc., at the heart of this study, may be seen

[3]Balochi shares many features indicative of Northwestern Iranian languages including the past tense stem of the verb 'to come' Balochi: *āxt* ∼ Kurdish: hāt < OIr. ā=gm̥-ta- (Paul, 2003, 62), shared exclusively with Kurdish, and the verb 'to fall' Balochi: kapt ∼ Kurmancî *keft* ∼ Gorani: *kewt* (Paul, 2003, 64).

as independent innovations, that is to say, innovations shared by languages for some reason other than borrowing or inheritance. The reason for such a change might be areal, shared by neighboring languages with or without a genetic relationship due to proximity and multilingualism (i.e. a sprachbund). Additionally, the reason for such a change could be a confluence of tendencies of grammatical change and shared building blocks, so that they are *cooking with the same ingredients*. The remainder of this chapter consists of an overview of what information is available concerning the relevant languages. It should serve both as an introduction to my source material for this dissertation and an introduction to (Western) Iranian linguistics.

1.3 The languages referenced in this work

Here I consider the Western Iranian languages, which have converged on many points. I do not weigh in on the reason for this convergence. Although there is some evidence pointing to the separation of Central Iranian, including Northwestern Iranian from Southwestern Iranian languages, from the oldest attested documents, e.g. reflexes of PIE laryngeals in New Iranian languages but not Old Persian (Kümmel, 2014), these languages have developed similar nominal and verbal inflection systems. One explanation for this is parallel innovation, producing the same constructions from cognate constructions inherited from a common ancestor. Another explanation is pattern borrowing, producing the same constructions from similar material under the influence of another language with or without cognacy. One example of such a convergence is the recruitment of a stative participle form to express the perfect tenses and subsequently becoming a generalized (ergative) past tense (aorist drift). This change has been observed independently across the world's languages.[4] In addition to the possibility of parallel innovation, there are several cases where borrowing seems to be a foregone conclusion. For instance, there is a Persianate ezafe construction

[4]See Condoravdi & Deo (2014) for an explanation of how historical semantics contribute to make aorist drift a commonly observed change.

in Iranian Balochi varieties alongside the genitive and attributive marked constructions. Likewise, in Gilaki, there is a Persianate ezafe construction alongside the reverse ezafe construction.

The Southwestern Iranian languages are generally located in the southwest of Iran. However, their most prolific member, New Persian, has spread across the Iranian world. The members of the family are thought to be divided into six subgroups: (1) Bashkardi, (2) Fars Dialects, (3) Farsic-Caucasian Tat, (4) Kumzari, (5) Larestani, and (6) Luric-Dezfulic. Both New Persian and Şirvan Tat, which I dedicate significant attention to in chapters 2 and 3, belong to group (3) Farsic-Caucasian Tat. Of the Central Iranian languages, I focus on the Northwestern Iranian sub-branch. The Northwestern Iranian languages are generally spoken all across the Iranian world. The members of the family are thought to be divided into eight subgroups: (1) Adharic, (2) Balochic, (3) Caspian, (4) Central Iran Kermanic, (5) Komisenian, (6) Laki-Kurdish, (7) Semnani-Biyabuneki, and (8) Parthian. Of these sub-branches, Parthian was last attested in the Middle Iranian period. The New Iranian languages examined in this paper are limited to branches (1), (2), (3), and (6). I have reluctantly chosen to neglect some of these branches because of the scarcity of reliable descriptions of these languages and varieties. This scarcity becomes apparent in the following section 1.4. These languages are Adharic: Zazaki (Northern, Southern, and Central), Hewramî (Luhonî, Paweyane), Tati (Çali, Takestani), Balochic: Balochi (Koroshi, Turkmen, Rakhshani), Caspian: Gilaki, and Laki-Kurdish: Kurdish (Northern, Central, and Southern). For a full view of the relationships between these languages and the Central and Southwestern sub-branches of Iranian see appendix A.

1.4 The linguistic study of Central and Southwestern Iranian

Here, I present an overview of the linguistic study of Central and Southwestern Iranian. The nature of Iranian studies is such that most scholarly work has been published in bound vol-

umes and conference proceedings. In my coverage of this subject, I restrict the discussion to works that I have deemed primarily linguistic (i.e. not literature, or culture). Additionally, I restrict my coverage here to Central and Southwestern Iranian languages. As a result, there may be works that should be included here that I have missed. Nonetheless, I hope this literature review serves as a document that can be continually updated as a resource for all scholars who wish to work in Iranian linguistics.

The scholarly study of Iranian came into its own with the publication of the first comprehensive overview of the Iranian languages *Grundriss der iranischen Philologie* (Geiger, 1895b). This volume was originally published between 1895 and 1904, and it contains a variety of articles in German. These articles cover a wide range of disciplines, including Iranian languages, history, art, literature, and culture.[5] This work contains linguistic articles on New Persian (Horn, 1895), Balochi (Geiger, 1895a), Kurdish (Socin, 1895), Old Iranian (Bartholomae, 1895, Avestan and Old Persian), Middle Persian (Salemann, 1895), and on Iranian dialectology (Geiger, 1895c). There were several iterations and updates to *Grundriss der iranischen Philologie*. The intellectual heir to Geiger's (1895b) is undoubtedly the *Compendium Linguarum Iranicarum* (Schmitt, 1989). This volume is focused more narrowly on the Iranian languages and their dialectology. There are sections on Western Middle Iranian (Middle Persian and Parthian Sundermann, 1989c,b,a, in German), New Persian (Lazard, 1989, in French), Balochi (Elfenbein, 1989), the Caspian languages (Le Coq, 1989a, in French), Ossetic (Thordarson, 1989), and two sections on the Western Iranian languages and their dialectology more generally Windfuhr (1989a,b).

The last of these comprehensive overviews of the Iranian languages to appear in print is The Routledge guide to *The Iranian languages* (Windfuhr, 2009b). This volume can be seen as an update to what was presented more generally in *Compendium Linguarum Iranicarum* (Schmitt, 1989). Here, each section is a complete grammatical sketch with examples

[5]See Schmitt (2002) for a comprehensive review of *Grundriss der iranischen Philologie* including the history of its publication.

published in English. Furthermore, Windfuhr's (2009a) "introduction, Dialectology, and Topics" marks the most complete approach to the subject to date, although it is distilled from pieces of other articles (e.g. Windfuhr, 1989b, 1991, etc.). Additionally, Windfuhr (2009b) contains articles on Old Iranian (Skjærvø, 2009b), Middle West Iranian (Skjærvø, 2009a, Middle Persian and Parthian), Kurdish (McCarus, 2009), Balochi (Jahani & Korn, 2009), Zazaki (Paul, 2009), Sogdian (Yoshida, 2009), Persian, and Tajik Windfuhr & Perry (2009) along with languages from other sub-branches of Iranian. There have been other overview articles on several Iranian languages appearing in larger volumes with a broader scope, such as Woodard (2004) *The Cambridge Encyclopedia of the World's Ancient Languages*, which includes sections on Old Persian (Schmitt, 2004), Avestan (Hale, 2004a), and Pahlavi (Hale, 2004b).

Another volume that must be mentioned in this regard is Haig & Khan (2019) *The Languages and Linguistics of Western Asian: an Areal Perspective*. This volume is a collection of grammatical sketches and research overviews similar to Windfuhr (2009b). However, it differs from Windfuhr (2009b) in several regards. First, it differs in its scope; Haig & Khan (2019) focuses not on the Iranian language family as a whole but the languages of Western Asia more generally. This collection includes languages spoken in Anatolia (Turkey), Iran, Iraq, and the Caucasus, including Azerbaijan from Multiple sub-branches of Indo-European (Iranian, Armenian, and Greek), Semitic (Arabic and Neo-Aramaic), Caucasian (Laz), and Turkic. Second, this volume differs from Windfuhr (2009b) in the availability of data; significant improvements in our knowledge of the Iranian language have led the contributors to Haig & Khan (2019) to include previously unavailable data from languages and varieties. Haig (2019b) includes data from varieties of Northern Kurdish that have come to light since the establishment of the Manchester Corpus (Matras et al., 2016). Another aspect of this volume that represents a significant addition to the scholarly literature is the inclusion of the two Southwestern Iranian languages, Bakhtiari (Anonby & Taheri-Ardali, 2019) and Kumzari (Anonby, 2019), which have not previously received

this level of attention. The third difference is the prior state of the intellectual tradition. For instance, Zazaki is not featured in this volume due to the substantial part it plays in Windfuhr (2009b) (Paul, 1998b). Unlike the contribution on Kurdish (McCarus, 2009), there has not been any significant update in our knowledge of Zazaki. It has come to my attention that there are many Zazaki scholars in Germany and in Turkey that are adding to the field of Zazaki studies (p.c. with mahîr Dogan, Otto-Friedrich-University Bamberg). I anticipate a future update to Paul (1998b) especially as our knowledge of Northern and Central Zazaki is illuminated. In addition to the absence of Zazaki, Domari, the Indo-Aryan language of the Doms ("Gypsies") of the Middle East (Matras, 2012; Herin, 2012), is missing along with Southern Kurdish (Fattah, 2000; Belleli, 2016).

Another source that must be mentioned here is Yarshater (1982) the *Encyclopaedia Iranica*. This source is a work in progress that features articles on diverse topics, including art, literature, religion, folklore, history, archaeology, geography, ethnography, science and philosophy, and linguistics. These articles are too numerous for me to mention here. However, I have included many in the following sections, particularly when they are the only available source for a particular language.

1.4.1 Broad-scoping studies

A variety of studies cover topics in Iranian linguistics such as genealogical sub-grouping, pronouns, agreement, argument indexing, typology, contact, alignment, and historical development from a pan-Iranian perspective. The nature of these studies requires that they consider data from diverse languages across the Iranian world. Therefore, I discuss them here and not in the following sections about the individual languages and their scholarly study.

classification and genealogy

There have been several studies concerning the topics of classification of Iranian languages and historical linguistics. For the historical linguist, the contributions to Klein et al. (2017) provide a baseline for the study of Iranian; the articles on documentation (Skjærvø, 2017a), syntax Jügel et al. (2017), morphology (Skjærvø, 2017b), and phonology (Cantera, 2017) provide a nearly comprehensive view of Old Iranian. As the focus of the volume as a whole is comparative Indo-European linguistics, the termination of the overview at the Old Iranian period represented by Avestan and Old Persian makes sense. However, an additional three sections contain more data from or even a focus on Middle and New Iranian languages. Of these, the one that focuses most on New Iranian languages is Korn's (2017b) "the evolution of Iranian." This section focuses on phonological developments that culminated in the New Iranian languages as they are spoken today. Korn (2017b) considers phonological and morphological changes, and she relates these changes to the typology of Iranian languages, e.g. concerning fusionality, inflectional periphrasis, and with some consideration of the role of borrowing and multilingualism in these changes. Typological approaches abound in Iranian linguistics. However, it is worth mentioning Stilo's (2005) "Iranian as a Buffer Zone between Turkic and Semitic" in particular. This article examines the Iranian languages as a compromise between the canonically left-headed Semitic (V>(S)O, N>Adj, etc.) languages and the canonically right-headed Turkic languages ((S)O<V, Adj<N, etc.). Likewise, Turkic is largely agglutinative, and Semitic is largely fusional, though this is out of the scope of Stilo (2005). Korn (2017b) serves to be the most important article in Klein et al. (2017) for a scholar studying New Iranian languages, especially in terms of their historical development.

Of the sections of Klein et al. (2017) on Iranian, Sadovski's (2017) "the lexicon of Iranian" compares the shared vocabulary of Iranian languages in the Old, Middle, and New Iranian periods. To this end, Sadovski (2017) references numerous dictionaries, including etymological dictionaries. These are too numerous to cite here. However, it is worth mentioning Cheung's (2007) *Etymological dictionary of the Iranian verb*, which focuses on

the comparison of verb forms across the Iranian world and connecting them to other Indo-European languages as well. It is also worth mentioning the absence of Hasandust's (2011b) *Farhang-i tatbiqi-mawzu'i-i zubanha u guvishha-i irani-i naw*, a thematically arranged Iranian etymological dictionary. This work and Cheung's (2007) represent to most complete sources for comparative Iranian vocabulary to date. However, Hasandust (2011b) was published in Farsi by Nashar Asar, Tehran. It may be due to this fact that it has not received as wide a circulation as Cheung (2007) published by Brill.

Huyse's (2017) "The dialectology of Iranian" focuses on the known Old and Middle Iranian languages with the most attention paid to the Old Iranian period. For the Old Iranian languages Old Persian and Avestan, Huyse (2017) gives notes on attestation and their ethnolinguistic context. Only for Median, an Old Iranian language only known through borrowings in attested languages, does he give background on phonological features by which relationships may be determined. Additionally, he includes one sentence on Scythian to say that it is only known through onomastic evidence. Concerning languages in the Middle Iranian period, Huyse (2017) give a little background on the phonological and morphological reasons for splitting the languages into Eastern and Western groups. This section is followed by a background on the attestation and ethnolinguistic context of Parthian, Middle Persian, Bactrian, Sogdian, Choresmian, Alanic, Khotanese, and Tumshuqese. The section on New Iranian languages is merely a short list of some languages with large numbers of speakers and their presumed affiliations. However, in the final section, he points the reader to the bound volumes I discussed in the previous section 1.4 (Windfuhr, 2009b; Schmitt, 1989; Geiger, 1895b, e.g.). This work is the most cursory look at what these languages are, where and when they were spoken, and by whom. However, there are many more comprehensive explorations of Iranian dialectology that look at features of these languages and their genealogical affiliations. These works include Kleinere Dialekte und Dialektgruppen (Geiger, 1895c) from Geiger (1895b), the contributions to Schmitt (1989) on Western Iranian languages and their dialectology (Windfuhr, 1989a,b), Windfuhr's contribution to

Yarshater (1982) on "DIALECTOLOGY" (Windfuhr, 1995), and the pioneering work of Agnes Korn on Central Iranian "a partial tree of Central Iranian" (Korn, 2016a) and the expanded "Isoglosses and subdivisions of Iranian" (Korn, 2019a). Korn (2019a) presents the significant shared innovations of Iranian languages that she proposes to be meaningful isoglosses. These include traditional phonological innovations following Tedesco (1921) (e.g. the reflex of *tr/tl: θ*r* or *s(s)*) and morphological innovations third-person pronominal clitic from *-šai or *hai (see Korn, 2019a, 279-281 for a complete list of isoglosses.).

It has been a long-standing issue in Iranian languages that borrowing and migration have served to obscure genealogical relationships (particularly with Balochi Korn, 2019a, 2005). This problem is robust with Persian borrowing, affecting Iranian and Indo-Aryan alike (p.c. with Paul Heggarty, Max Planck). This issue has led some to propose a multidimensional approach to language classification combining both contact phenomena and traditional genealogical classification; see "A multidimensional approach to classification of Iran' s languages" (Anonby et al., 2020a), and the *Atlas of the Languages of Iran* (Anonby et al., 2020b).

Geographic areas

There have been several broad overviews of languages by region, including "The Iranian languages of Northern Iraq (Haig, 2019c)," "KERMANSHAH vii. LANGUAGES AND DIALECTS" (Borjian, 2017b), "Judeo-Iranian Languages" (Borjian, 2015), and many others that have a narrow enough focus to be included in the following section 1.4.2 on the scholarly study of particular language groups. Haig (2019c) covers the Northern Kurdish varieties known as Behdînî (Iraqi Kurmancî), Central Kurdish, Hewramî, Shabaki, and other Gorani varieties. Borjian (2017b) covers Southern Kurdish, several Gorani varieties, Persian varieties, Neo-Aramaic, and the Turkic variety spoken in the Sonqor valley. The majority of this article focuses on the Kurdish variety, which "forms the linguistic backdrop of the province."

Topics in historical linguistics

Several articles focus on contact phenomena and miscellaneous topics in historical linguistics. "Iranian Elements in Middle Aramaic: Some Particles and Verbs" (Shaked, 1993), "On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part I: Present-Based Paradigms" (Noorlander & Stilo, 2015), and "On the Convergence of Verbal Systems or Aramaic and its Neighbors. Part II: Past Paradigms Derived from Present Equivalents" (Stilo & Noorlander, 2015) address aspects of contact between Aramaic and Iranian. Shaked (1993) focuses on Iranian borrowings that were already attested in Middle Aramaic, while Noorlander & Stilo (2015) and Stilo & Noorlander (2015) focus on grammatical convergence that happened between the Middle period and New Iranian and Neo-Aramaic varieties and languages. The latter two articles prove to be important for Iranian studies even without the added interest in Aramaic. They convincingly show how the verbal systems of Neo-Aramaic, Armenian, Caucasian, and multiple branches have adopted similar strategies in the reorganization of their tense-aspect-mood systems. Perhaps the only flaw in their analysis comes from the lack of a clear etymon for many of the morphemes relevant to their typology. For instance, they classify the *æ(t)*-type imperfective adverbial markers as being from an adverbial origin from and adverbial origin following Windfuhr (2009a) (Noorlander & Stilo, 2015, 435). However, Windfuhr's (2009a) the assumption of an adverbial origin is based entirely on the fact that Persian has an imperfective marker of adverbial origin, with the supposition that many of the others must therefore have the same sort of source too. New Persian has the prefix *mī-*, which comes from the adverbial OIr. *hama-aiwa "same duration, time" (Windfuhr, 2009a, 26). Windfuhr (2009a) points to Stilo (2007) for a complete discussion of the origin of these affixes across Iranian. Stilo (2007), in a discussion of the "provensial dialects" of Isfahan, discusses these formatives in a pan-Iranian perspective, although he stops short of proposing the same etymon for the markers in Kurdish. However he does employ Windfuhr's (1991) original etymon (< OIran. *aiwa-da "at the same time, all the time") for the *æt*-type formatives in the central plateau dialects. Additionally,

many of the facts that lead Stilo (2007) to agree with this etymon have multiple interpretations. For instance, in some languages, these formatives occur variably as prepositions or postpositions in specified contexts. I believe that this more accurately reflects their original circumpositional nature as reflected synchronically by Southern Kurdish; see Karim (2021b) for a complete discussion. It is also worth mentioning Jeremiás' 1993 "On the Genesis of the Periphrastic Progressive in Iranian Languages," which despite its title looks quite narrowly at Southwestern Iranian, and Persian and Tajik in particular. Jeremiás (1993) does give fleeting mention to Mazandarani (Caspian) and Central Iranian Kermanic varieties. However, their mention is only regarding how they fit into the Persian-type system. The historical section includes only what is attested for Persian and not what is reconstructable for other Iranian languages.

Additionally, in the realm of Historical linguistics, there have been many etymological studies focusing on single words (e.g. Benveniste, 1964; Gershevitch, 1985e; Thiesen, 2005, etc.), single morphemes(e.g. Gershevitch, 1985f,i, etc.), or word classes (Gershevitch, 1985g, 1964; MacKenzie, 1999h). Further work on contact in the Old and Middle Iranian world comes from Gippert's (2007) "Albano-Iranica." This article focuses on Iranian elements in the so-called Caucasian Albanian texts. Scholars now believe that the Udi language, currently spoken in Azerbaijan, is the daughter language of Caucasian Albanian. This is notable because Udi is one of the languages considered to share features of their verbal system common between Kurdish, Armenian, and the Caspian languages.

alignment shift, ergativity, and DOM

Another subject that needs to be mentioned is alignment shift, ergativity, and DOM. Much attention has been paid to the differential marking of subjects and objects. A few works on the development and state of ergativity in Iranian languages are "Ergativity in Iranian" (Haig, 2015), "The emergence of ergativity in Iranian: reanalysis of extension?" (Haig, 2008b), *Alignment in Kurdish* (Haig, 2004), *Alignment change in Iranian languages : a*

construction grammar approach (Haig, 2008a), "Non-Canonical Subject Construction in Endangered Iranian Languages: Further Investigation into the Debates on the Genesis of Ergativity" (Dabir-Moghaddam, 2018), and many others. The articles by Haig tend to explore the subject of ergativity from a typologist's perspective. His definition of ergativity can be paraphrased narrowly as (S = O).[6] This viewpoint allows for the separation of the Central Kurdish systems, which only show ergative alignment in the way they morphologically index subject, agents, and objects, as post-ergative or remnant ergative (Jügel, 2009, the latter term following). Likewise, the double-oblique construction, where (A = O) is shared by all spoken Kurmancî varieties in certain ill-defined circumstances, is not ergativity. The notion that this is either a variation on ergativity or even not quite ergative is shared by many scholars (e.g. Dorleijn, 1996; Gündoğdu, 2017a, etc.).

An in-depth discussion of both differential subject and object marking can be found in Stilo's (2008a) "Case In Iranian: From Reduction and Loss to Innovation and Renewal." This article focuses on the loss of case in Iranian languages and the various strategies that Iranian languages have used to replenish their case systems. He discusses the typological angle, formatives, and historical developments. Much is packed into this discussion, and his discussion of differential object marking in Caspian languages serves to shed light on what may be represented by the facts presented by Haig for Kurmancî double oblique construction. This connection is an integral part of my discussion of definiteness in Kurmancî in chapter 4. Stilo (2008a) discusses many of the *rā*-marking as one of the renewal strategies used to rebuild the case systems of Iranian languages, which I discuss in chapter 2. *Rā*-marking is one of the most important forms of differential object marking in Iranian languages. This morpheme in Persian only occurs on definite direct objects. This is one subject that has received a disproportionate amount of attention, as discussed in the following section 1.4.2.

[6] S: subject of an intransitive verb; A: agent of a transitive verb; O: direct object of a transitive verb (sometime P: patient).

Pronouns

Deeply connected with differential case marking are the pronominal systems of Iranian languages. This subject is tied to the verbal morphological systems, which feature argument-indexing morphemes often from etymological pronouns (e.g. Kurdish (Central and Southern), Caspian (Takestani Tati), Gorani (Hewramî), etc.). Additionally, these formatives index oblique arguments and possessors in Applicative constructions such as Soranî "absolute prepositions" (see Karim, 2021a; Karim & Salehi, 2020, etc.) and the Şirvan Tat placeholder construction (see Suleymanov, 2020b, etc.). Some highlights of work on the pronominal systems of Iranian, including pronominal clitics, are as follows. Mohammadirad's (2020) dissertation *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*. This thesis is an in-depth exploration of the function and distribution of pronominal clitics in a variety of Iranian languages.[7] For many of the languages in this study, this is the most accessible source providing the most information available in any one place. For this reason, it is an essential work in the tool kit of anyone who wishes to study the Western Iranian languages. However, Mohammadirad's (2020) contribution is by no means limited to his exposition of data. He makes major contributions to the typological study of absolute prepositions, which were thought to be a quintessentially Kurdish phenomenon. Although they occurred in Middle Persian, MacKenzie (1999e) hypothesized that there might have been influence from (proto-)Kurdish speakers on late Middle Persian that was responsible. Perhaps the understanding of these constructions in Middle Persian and New Western Iranian languages that has recently developed thanks to Jügel (2016) and Mohammadirad (2020) will allow scholars to reassess the phenomenon as a parallel development, mutual inheritance, or borrowing. This topic is the subject of my

[7] Mohammadirad's (2020) study includes data from Badrudi, Bastaki, Central Kurdish (Baneh), Behbahani, Bandari, Southern Kurdish (Bijar), Chali, Central Plateau dialects, Central Taleshi, Dashti, Davani, Delijani Delvari, Gorani [Hewramî] (Takht), Gorani (Qal' eh), Jondani, Khansari, Koroshi, Laki (Harsini), Laki (Kakevandi), Lari, Meymei, Minabi, Nikabadi, Northern Kurdish, Nowdani, Semnani, Sivandi, Takestani, and Yazdi (Zoroastrian).

ongoing research. Another important article on pronominal clitics is Korn's (2016b) "Western Iranian Pronominal Clitics." This article is an etymological account of the various clitic pronominals in Western Iranian languages. One of the major conclusions of this article is that many of the pronominals have descended from the genitive pronouns, as has long been observed. However, some of the time, in some of the languages, the attested form is the reflex of the accusative.

An additional theme that commonly arises in the study of Western Iranian pronominal systems is the development of argument indexing from pronominal sources. The first of these studies that bears mentioning here is Korn's (2011) "Pronouns as Verbs, Verbs as Pronouns: Demonstratives and the Copula in Iranian." This article focuses on the development of the copula from pronominal sources, which is most notable as the Sogdian <(')xw> (Yaghnobi: *=x*) but has parallels in Ossetic (e.g. Iron *u* < OIr. *awa-) and Pashto (not Central of Southwestern Iranian) *dəy, da, di* (< OIr. *aita-). Korn (2011) does not mention the Zazaki forms *=o, =a, =ê* 'is [M], is [F], are [PL]', which have clear connections to the far demonstratives *o, a, ê* 'that [M], that [F], those [PL].' There likely is more to be said on this subject.

In addition to developing the copula from pronominal sources, verbal indexing of agents, objects, and subjects also have pronominal roots in some Iranian languages. Jügel & Samvelian (2020), "Topic agreement, experiencer constructions, and the weight of clitics," looks at Middle Iranian experiencer subject constructions, sometimes called non-canonical subjects, where the subject of a verb of feeling is marked in the oblique case. They show how the extraposition of the subject due to topicalization followed by a resumptive pronoun in the main clause is the likely source for the New Persian experiencer constructions. Likewise, topicalization, leading to the recruitment of a resumptive pronoun, is the likely source of mandatory past-tense agent indexing with formatives that reflect Middle/Old Iranian oblique pronominal clitics in some languages. In line with this exploration is Haig's (2020) "The pronoun-to-agreement cycle in Iranian: Subjects do, objects don't," explores the fact

that we don't see mandatory object indexing. This article follows the intellectual tradition of Siewierska (1999), who shows that, typologically speaking, languages do not show the mandatory indexing of objects the way that is sometimes true of subjects and agents. The subject/agent-object asymmetry described in these articles is what prompted Haspelmath (2013) to propose the dismissal of terminology like agreement in favor of terms like indexing, in conjunction with further specification as to the type of indexing. For instance, indexing could then be further sub-categorized as either a grammatical-index that must occur with a conominal, a cross-index that may occur with a conominal, or a pro-index that cannot occur with a conominal. In Haig's (2020) exploration of these phenomenon, he presents some interesting data. However, there is no attempt to explain these developments. Furthermore, since Siewierska (1999), it has been known that grammatical-indexing of objects does not occur; in other words, there is no language where verbs are systematically marked for object agreement and must occur with a co-nominal object.

The ezafe

Another topic that has received a disproportionate amount of attention is the ezafe phenomenon. This is a trend that I have contributed to in the past (e.g. Karim, 2021c, 2019, 2018), and I continue the trend in this study (chs. 2-5). As the core of the current study focuses on the nominal morphology of Western Iranian languages, and the ezafe, in particular, is featured throughout, I mention briefly here the themes that I do not cover in the following chapters. The first trend in the study of the ezafe is primarily (though not necessarily only) descriptive articles. One noteworthy article of this type is Haig's (2011) "Linker, relativizer, nominalizer, tense-particle On the Ezafe in West Iranian." This article covers the forms and functions of the ezafe in a range of Iranian languages. One criticism of this article is that it presents the so-called 'tense ezafe' as a progressive aspect marker following Blau's (1976) initial description "placé devant un verbe, renforce le sens de durée [placed in front of a verb, [it] reinforces the meaning of duration]" (Blau, 1976, 40). In a

p.c. with Agnes Korn, I learned that the current thinking following Jügel et al. (2017) is that these forms are markers of the prospective aspect. On the other hand, Zazaki does have a construction, where the ezafe, placed in front of a verb, marks the progressive aspect. One might assume that close contact between Zazaki and Kurmancî might have caused the adaptation of the Zazaki construction by Kurmancî. However, this is far from proven. As for the Zazaki construction, the first person singular formative $=a$ following Paul (1998b) ($=an$ following Todd (2002)) suggests that it is the copula and not the ezafe (expected $=\hat{e}$ masculine and $=a$ feminine) that is employed in this construction. It is unlikely that the Kurmancî construction works like the one in Zazaki or the way that Haig (2011) and Blau (1976) have presented it (e.g. as a progressive aspect marker).

Another article worth mentioning is Franco-Rita & Savoia's (2012) "Some call it article, some ezafe: "Linkers" as "agreement."" Like Haig (2011), this article considers data from many languages across the Iranian world. Additionally, Franco-Rita & Savoia (2012) tie the Albanian 'nyje' particle to the greater discussion of the ezafe phenomenon. One issue with the analysis here is that Franco-Rita & Savoia take certain liberties with their historical assumptions that are not grounded in the literature nor supported in their article. One example of this is the claim that the 3rd person singular pronominal clitic $=\hat{\imath}$ in Sorani ($<$ OIr. *hai/*(h)īm Korn, 2016b, 163) may be the root of the ezafe -$\hat{\imath}$ ($<$ OIr. *yat). Their work focuses on the Albanian and Kurmancî data to make a theoretical contribution. They claim that the ezafe is "an argument, satisfying a predicate (the modifier)." However, the greatest contribution of this paper is probably the juxtaposition of forms. My own forthcoming paper, "The ezafe and the Article" Karim (2021c) is a purely descriptive account of the forms of the ezafe in the context of definiteness, number, and gender from a synchronic and diachronic perspective. This paper forms the foundation of the more comprehensive account of Iranian nominal morphology in chapter 2. Werner (2018), "Forms and Meanings of the Ezafe in Zazaki" is yet another descriptive account. However, it is set apart from the other papers I have mentioned thus far in that it is a comprehensive look at the ezafe in Za-

zaki. Karim (2021c), Franco-Rita & Savoia (2012), and Haig (2011) are all works that look at the ezafe phenomenon broadly across many languages. By contrast, Werner (2018) looks narrowly at Zazaki but at the three dialect groups within Zazaki Northern, Southern, and Central. Each of these groups features significant differences within their ezafic systems. Although this study narrowly focuses on the ezafe, it represents the most substantial work featuring data from Central Zazaki. It is essential reading for Zazaki scholars and scholars studying the ezafe alike.

Another important thread in the scholarly study of the ezafe is synchronic syntax. One feature of this thread is that it is skewed toward the minimalist program. Many scholars who have devoted the most ink to the ezafe happen to be syntacticians following Chomsky's footsteps, whether in Iran (e.g. Yadgar Karimi) or the west (e.g. Vida Samiian, Rich Larson, etc.). The first of these of note is Samiian's (1994) pioneering work, "The Ezafe Construction: Some Implications for the Theory of X-bar Syntax." In this article, she puts forth the beginnings of the view of the ezafe as a case marker. The noun is put into the ezafe case by the following modifier. This account is different from Ghomeshi's (1997) account that claims the ezafe "serves to identify constituent-hood" of a noun, which does not project. "The NP node cannot dominate any phrasal material" (Ghomeshi, 1997, 786). This analysis builds on Karimi & Brame's (2012 originally 1986) assertion that the ezafe is triggered by 'nominal' elements. Following this same tradition are Karimi (2007), which focuses on Central Kurdish (Soranî), and Larson & Yamakido (2008), which focuses on typological parallels placing the ezafe in its place among attributive constructions among the world's languages. Likewise, Larson & Samian (2020) provide an overview of the ezafe phenomenon more generally, which was updated in Larson & Samiian (2020). Forthcoming work from Vida Samiian expands this exploration to include Middle Iranian (Middle Persian) data. Kahnemuyipour's (2014) puts forth the theory that "the ezafe is [to be] seen as a reflex of the roll-up movement, where the assumption is that "the base order of the noun phrase in Persian ... [is] head-final, with the surface order derived via phrasal movement in a roll-up

fashion" (Kahnemuyipour, 2014, 2). The ezafe is then a marker of syntactic movement. All of these theories are discussed in Haghkerdar's (2009) thesis "Western Iranian Ezafe: A Comparative Syntactic Analysis." This work is interesting as it attempts to unify the ezafe with other methods of attribution across Iranian, e.g. the reverse ezafe, attributive, and genitive constructions. The most significant contribution of this work is his comparison of the various syntactic accounts of the ezafe, their merits, and their shortcomings.

There are a series of articles by Polet Samvelian focusing on the ezafe phenomenon from a morphological perspective, e.g. "When morphology does better than Syntax: The Ezafe construction in Persian" (Samvelian, 2005), "A (phrasal) affix analysis of the Persian Ezafe," (Samvelian, 2007a), and "The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish" (Samvelian, 2008). Each of these articles builds on the previous, and the final extends her analysis from Persian to Kurdish. One highlight of this paper (Samvelian, 2005) is her assertion that the form of the ezafe that occurs with the definite suffix in colloquial Persian -$\emptyset$ or simple juxtaposition is merely the compound compounding strategy. In chapter 3 herein, I discuss this suggestion. Although I don't believe this is the correct conclusion, it does seem that there is a syntactic and semantic connection between these forms making Samvelian's (2005) observation particularly telling. I hope that the current study will become an important addition to the literature on the ezafe from a morphological perspective. My analysis presented in chapter 3 is essentially that the ezafe is a derivational affix that converts a noun into a functor that needs a modifier to be well-formed following a CG approach to syntax. In this sense, my approach is morphological, although chapter 3 focuses on the syntactic implications of that view.

The final sub-topic to mention here is the diachronic development of the ezafe. There seems to be only one dedicated study on this issue Haider & Zwanziger's (1984) "Relatively attributive: The 'ezäfe'-construction from Old Iranian to Modem Persian." However, there has been mention of the basic assumption, namely that the ezafe developed from the Old Iranian relative pronoun *yat* or perhaps Old Persian *(ha)ya*, in various articles that cover

the subject briefly amid a broader discussion (e.g. Moyne, 1971; Moyne & Carden, 1974; Karim, 2021c; Karimi & Brame, 2012; Samvelian, 2008; Jügel, 2017, etc.). There are two contributions to the diachronic study of the ezafe that I intend to make in the current study: (1) in chapter 4, I claim that the ezafe in Kurmancî is actually a reflex of the nominative form of the Old Iranian *(a)ka- stems. Furthermore, (2) in chapter 5, I claim that the *d*-form ezafe in Zazaki is the reflex of the Old Iranian demonstrative pronoun *aita-. The first of these assertions is the response to a question–why the Kurmancî ezafe inflects for case and definiteness–which has essentially gone unasked and unanswered. The latter solution seeks to unify the seemingly chaotic distribution of ezafat with the phoneme /d/.

Much attention has been paid to the ezafe in the literature, and even now, it is an essential part of the research programs of many Iranian linguists. However, I have only scratched the surface of the study of the ezafe here, and it is likely that as I write this, new articles are being published on the subject. See table 1.1 for a chronological list of scholarly studies on the ezafe. Note that studies on other topics include sections on the ezafe such as "Specific Features of Persian Syntax: The Ezâfe Construction, Differential Object Marking and Complex Predicates" (Samvelian, 2018) have been omitted from this table.

1.4.2 The study of particular Northwestern Iranian languages

There is a great deal of variation as to which of the Western Iranian languages have been documented or have been the object of scholarly inquiry. Persian has, of course, been the target of many studies as it is the national language of Iran and has a long history of attestation and usage across the Middle East and South Asia. In addition to Persian, many of the languages with significant refugee presence in western countries, such as Kurdish, have received more attention than other minority languages. This section is a summary of what is available for the languages relevant to this study.

The Northwestern Iranian Languages are subdivided into eight sub-branches, Parthian and the seven living sub-branches Adharic, Balochic, Caspian, Central Iran Kermanic,

Haider & Zwanziger (1984)	Relatively attributive: The 'ezäfe'-construction from Old Iranian to Modem Persian
Samiian (1994)	The Ezafe Construction: Some Implications for the Theory of X-bar Syntax
Ghomeshi (1997)	Non-Projecting Nouns and the Ezafe: Construction in Persian
Samvelian (2005)	When morphology does better than Syntax: The Ezafe construction in Persian
Samvelian (2007a)	A (phrasal) affix analysis of the Persian Ezafe
Karimi (2007)	Kurdish Ezafe construction: Implications for DP structure
Larson & Yamakido (2008)	Ezafe and the deep position of nominal modifiers
Samvelian (2008)	The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish
Haghkerdar (2009)	Western Iranian Ezafe: A Comparative Syntactic Analysis
Parsafar (2010)	Syntax, Morphology, and Semantics of Ezafe
Haig (2011)	Linker, relativizer, nominalizer, tense-particle On the Ezafe in West Iranian
Franco-Rita & Savoia (2012)	Some call it article, some ezafe: "Linkers" as "agreement"
Karimi & Brame (2012)	A Generalization Concerning the EZAFE Construction in Persian
Kahnemuyipour (2014)	Revisiting the Persian Ezafe construction: A roll-up movement analysis
Kahnemuyipour (2016)	The Ezafe Construction : Persian and Beyond
Larson & Samian (2020)	Ezafe, PP and the Nature of Nominalization
Werner (2018)	Forms and Meanings of the Ezafe in Zazaki
Larson & Samiian (2020)	The Ezafe Construction Revisited
Karim (2021c)	The Ezafe and the Article

Table 1.1: Articles on the ezafe

Komisenian, Laki-Kurdish, and Semnani-Biyabuneki (Hammarström et al., 2020). The Adharic family is divided into Zaza, Gorani, and Tati. However, Tati is often grouped in with the Caspian languages due to its geographic location despite genetic similarities with Zazaki and Gorani. Thus, one might call Zaza-Gorani Kurdish-zone Adharic and Tati Caspian Adharic. Because many comparative sources introduce Tati with the Caspian languages, I continue the trend here.

The New Southwestern Iranian languages are subdivided into six sub-branches Bashkardi, Fars Dialects, Farsic-Caucasian Tat, Kumzari, Larestani, and Luric-Dezfulic. However, the majority of research into Southwestern Iranian languages has focused on New Persian in the form of the National languages Farsi (Iran), Tajiki (Tajikistan), and Dari (Afghanistan). For this reason, I introduce research on Southwestern Iranian in two sections, one on Farsic-Caucasian Tat and one on other Southwestern Iranian languages.

Zaza-Gorani (Adharic)

The Zaza-Gorani languages (Iranian > Central Iranian > Northwestern Iranian > Adharic > Zaza/Gorani) are divided into two sub-branches, Zazaki and Gorani. Note that according to Hammarström et al. (2020), there is no such division. Instead, the Adharic branch of Northwestern Iranian has three sub-branches Zazaki, Gorani, and Tatic. It is my understanding that this grouping should reflect the fact that not enough is yet known about these languages to confirm a greater degree of closeness between Gorani and Zazaki than with Tatic. Therefore, I have grouped Zazaki and Gorani together here as the Adharic languages spoken in the Kurdish zone, although the precise sub-grouping has yet to be established without a doubt. The languages of this sub-branch have been known for their archaism since the pioneering work of Hadank (1930) and Hadank (1932). The richness of nominal inflection in Zazaki, in particular, led Paul (1998a) to remark that there is "a still unanswered question of West Iranian dialectology: why Parthian, spoken nearly 2000 years ago, should be in its noun morphology more modern than any of the closely related

Northwestern dialects spoken today" (Paul, 1998a, 172). This question is framed from a Parthian perspective as Parthian was the exemplar for Northwestern Iranian employed by Tedesco (1921). However, the question could just as easily be asked from a Zaza-Gorani perspective: why is the nominal morphology of the Zaza-Gorani languages more highly inflected than either of the attested Middle Western Iranian languages? This is a question to which I hint at an answer in chapter 4 and partially answer in chapter 5. Resources on the Zaza-Gorani languages are as follows.

Gorani The Gorani languages are first known from Christensen1921 *Les dialectes d'Awromān et du Pāwä.* In this work, Christensen & Benedictsen provides background on the language that I refer to in this study as Hewramî based on the label given to the languages by its speakers. It contains a collection of oral texts in Hewramî and a smaller number of texts in the closely related Paweyane. Paweyane is sometimes referred to by scholars as the Gorani dialect spoken in Pawe city, as Hewramî by people in the region, including native speakers with out-group individuals but only Paweyane with other native speakers. Since Christensen & Benedictsen's (1921) overview, the core Hewramî language that was his main focus received additional attention in MacKenzie's (1966) *The dialect of Awroman (Hawraman-i Luhon): Grammatical sketch, texts, and vocabulary.* This work provides an in-depth look into the phonology and morphology of Hewramî and includes numerous examples, full paradigms, and a significant literature section. It is essential reading for those that wish to study Gorani languages. It did, however, fall short in a few places. For instance, MacKenzie (1966) claimed that the typical remnant ergative construction, where the agent is in the direct case, and the verbal agreement marker (grammatical index) is etymologically an oblique clitic, could be supplanted by an ergative construction if the agent was impersonal. In other words, an impersonal agent "may be expressed by a noun in the oblique case" (MacKenzie, 1966, 51). Further work on this construction, e.g. Mahand & Naghshbandi's (2014) "The effect of discourse factors on case system in Hawrami," have shown

that these are actually topicalized constructions. The impersonal nature of the subjects that MacKenzie (1966) observed was merely epiphenomenal of his data. Additional work on Paweyane is sparse despite Pawe city being the capital of the Hewraman region. This work includes "The Noun Phrase in Hawrami*" (Holmberg & Odden, 2008), which exclusively looks at Paweyane, and a portion of Mahmoudveysi & Bailey (2019). Mahmoudveysi & Bailey's (2019) study on Hewramî spoken in Iran includes some data Paweyane.

Early work on other Gorani languages began with Hadank (1930) *Mundarten der Gûrân: besonders das Kändûläî, Auramânî und Badschälânî.* Although a good portion of this work in German focuses on Hewramî (here as Auramânî), it is also the first work on Kändûläî and Badschälânî. The former, Kändûläî, represents the majority of Hadank's (1930) study and has not since received further scholarly attention, although there is currently a project underway (p.c. with Zaniar Naghbendi, the University of Kordistan, Sanandaj). Badschälânî only makes up a few paragraphs of Hadank's (1930) 479 page document. However, this section has now been supplemented by MacKenzie (1999a) Bājālānī (originally published in BSOAS, 1956). Together with Hadank's (1930) few paragraphs, this amounts to 19 pages on Bajalani. However, the comparability of the data in MacKenzie (1999a) and Hadank (1930) is unclear. Hadank's (1930) came from a speaker of Bajalani in Iran, while MacKenzie's (1999a) data came from a single Bajalani tribesman in Nineveh (Iraq), who referred to his language as Shabaki. It is unclear if these are to be understood as different languages, although a hint to this question can be found in Leezenberg (1994). Leezenberg's (1994) *The Shabak and the Kakais: dynamics of ethnicity in Iraqi Kurdistan* is primarily an ethnographic study. However, there are two pages at the end that feature words and sentences in Maço, Shabaki, Bajalanî, and Hewramî for comparison (Leezenberg, 1994, 15-16). A cursory look at these forms suggests that they are, in fact, different. To my knowledge, no study of Maço has ever been conducted.

There are several studies of other Gorani varieties that need to be mentioned. Mahmoudveysi & Bailey's (2013) *The Gorani language of Zarda, a village of West Iran* covers

the language known by its speakers as Zerdeyane. This variety is firmly planted in the same Group with Hewramî and Paweyane. In a p.c. with Hişmet Şiyenî, a native speaker of Paweyane, I was told that "Zerdeyane is the same as Paweyane." However, when we began to read from Mahmoudveysi & Bailey's (2013) grammar, he admitted that it was, in fact, different though with significant overlap. Bailey's (2016) dissertation *A grammar of Gawraju Gurani* and *The Gorani Language of Gawraju, a village of West Iran: texts, grammar, and lexicon* (Mahmoudveysi et al., 2012) both represent an in-depth exploration of the Gorani language spoken in the village of Gewrecû. This variety differs from all the previously mentioned Gorani varieties. While it has preserved the inherited phonological features of Gorani and cognate vocabulary, it has adopted the morphology of neighboring Kurdish varieties. In this sense, its study is crucial to understand contact effects between Zaza-Gorani and Kurdish, which have been a subject of much interest since Mackenzie's (1961) Kurdish dialect studies and MacKenzie (1961). This question has also been approached from a Gorani perspective by Leezenberg (2015) (Gorani influence on central Kurdish: Substratum or prestige borrowing).

Zazaki Zazaki studies roughly began with Hadank's (1932) *Mundarten der Zâzâ: hauptsächlich aus Siwerek und Kor*, which provided a grammatical description of and oral texts from Southern Zazaki, sometimes referred to as Dımli (Dimlî, Dimilî) or zonê ma (our language) (Paul, 1998a, 163). Other work on Southern Zazaki includes Todd's (2002) *A Grammar of Dimili Also Known as Zaza*, a grammatical sketch originally published in 1985. This work is a fairly comprehensive sketch of the language. However, it is short on examples and provides no texts. Additionally, there are several mistakes; for instance, in the section on the *d*-form ezafat, Todd (2002) shows the masculine singular variant as *-dê*. Todd's (2002) account completely neglects the existence of the masculine singular inanimate/indefinite ad-adjectival variant *-do*. This is particularly strange as Todd (2002) give an example of just such a form *i'sot-ên-do tûn =o* 'is a hot pepper [pepperIND-EZ.IND./ATT

hot =COP.3SG.M],' which features the ezafe *-do* (Todd, 2002, 96). Additionally MacKenzie's (1995a) "Notes on Southern Zaza (Dimili)" provides only the briefest sketch of Southern Zazaki.

Paul's (1998b) *Zazaki: Grammatik und Versuch einer Dialektologie* represents a significant improvement upon both Hadank (1932) and Todd (2002). He provides an in-depth grammatical account of Southern Zazaki, texts, and shorter accounts of other Zazaki varieties. Its size and comprehensiveness make Paul (1998b) an essential source for Zazaki studies. However, there are a few drawbacks. Its overall structure, combined with many seemingly proprietary German abbreviations, makes the text difficult to interact with. This drawback is made up for somewhat by copious examples for which the reader may draw their own conclusions. In a few cases, Paul (1998b) may have over-complexified a particular point. For instance, he invents a new case Oblique IIa to handle examples where the motivation behind a particular oblique morpheme is unclear. As new data comes in, this type of solution will likely prove unwarranted (see ch. 5 for my motivations for this claim). Another criticism of Paul's (1998b) work is that the majority of the text is based on a single literary source, *Na xumxum a ...* (Berz & Malmîsanij, 1951), a collection of Southern Zazaki folk tales. There are a few issues with the use of a literary source as the basis for a grammar. One such issue is that discrepancies arise from the conflict between a literary register and a "real" spoken language. Another issue is that the folk tales gathered from a variety of sources that made it into Berz & Malmîsanij (1951) may reflect multiple dialects or even idiolects. The final criticism is that Paul continually cites examples from Berz & Malmîsanij (1951), but they are not always duplicated completely or in context. This is an issue because Berz & Malmîsanij (1951) has been out of print for some time, and it was likely never easily accessible outside of Sweden, where it was published. As a result, readers of Paul (1998b) may come to conclusions that are merely epiphenomenal based on Paul's (1998b) cited examples.

The dialectology section of Paul (1998b) includes a sketch of the grammar from the

variety spoken in Çermik-Siverek, Palu and Bingöl, Kulp, and the "Alevi-dialect" belonging to both Northern and Southern Zazaki. Additional work on Northern Zazaki includes Aygen's (2010) *Zazaki/Kirmanckî Kurdish*. This volume covers Northern Zazaki in a way comparable to Todd's (2002) study of Southern Zazaki. It is a short grammatical sketch with few examples and no texts. It is a good source for consulting paradigms and getting a brief overview of the language. However, it becomes clear after consulting Werner's (2018) exploration of the forms and meanings of the ezafe in Zazaki that Aygen (2010) does not go into sufficient detail about many of the most intriguing topics in Zazaki and Iranian linguistics. Likewise, *Rastnustena Zonê Ma* [the correct spelling of our language] (Jacobson, 1993), perhaps meant to be a style guide, provides a brief grammatical sketch of Northern Zazaki with some examples.

Additional linguistic work includes "Zur dialektalen Stellung des Zazaki" (Gippert, 2009) and "The position of Zazaki among West Iranian languages" (Paul, 1998a), as well as the chapter "Gliederung der Dialekte" (Paul, 1998b, 208-214), which approach Zazaki from a genealogical perspective. Paul (1998a) attempts to place Zazaki along a continuum of northernness based on Tedesco's (1921) original classification of Northwestern and Southwestern Iranian. Gippert's (2009) study has a similar time depth. However, its most meaningful contributions are in the form of minor etymological suggestions, such as the idea that the masculine singular oblique endings are likely the reflexes of nouns in *-(a)ka but not the feminine endings. This assertion is integral to my proposal in chapter 4 in this study. Another study that must be mentioned here is Werner's (2018) "Forms and Meanings of the Ezafe in Zazaki." This article is an in-depth study into the allomorphs of the ezafe and the factors that condition their use. It is crucial both for the study of the ezafe and for Zazaki dialectology as it is the only study of its kind to feature data from Northern, Central, and Southern Zazaki. Note that Hammarström et al. (2020) does not include Central Zazaki spoken around Bitlis and Van. This variety has significant influence from Northern Kurdish and is very much understudied and underdocumented. To my knowledge, there is no

grammar or sketch of this variety available.

The Caspian

The languages and varieties that border the Caspian sea scattered across Azerbaijan, Northern Iran, and Turkmenistan are sometimes referred to as "Caspian." Many of these languages share certain features such as reverse ezafat, or genitive and attributive marking on the modifier in modifier > Noun constructions, and a numeral "one" from Old Iranian *aiwa. However, according to Hammarström et al. (2020), there are, in fact, two separate sub-groups within the Caspian region. The Tatic group is a sub-branch of Adharic more closely related to Zazaki and Gorani, and the Caspian group proper refers only to Gilaki and Mazandarani. Unfortunately, each of these groups is woefully understudied despite a considerable amount of work by Ehsan Yarshater, Don Stilo, and Habib Borjian.

The study of the Caspian begins with Le Coq's (1989a) contribution to Geiger (1895b) "Les dialectes du caspians et les dialectes du nord-ouest de l'Iran." This article contains an overview of the five groups Mazandarani, Gilaki, Talyshi, Tatic, and Semnani. Mazandarani and Gilaki fit genealogically in the Caspian group; Talyshi is a subset of Tatic, and Semnani (along with Biyabunaki) forms a separate sub-branch of Northwestern Iranian. An important broad comparison of the languages of the Caspian region is Stilo's (2018b) contribution to Haig & Khan (2019) "The Caspian region and South Azerbaijan: Caspian and Tatic." This article is a deep dive into the features of Caspian languages. Although Stilo (2018b) is not consistent as to which languages he references from section to section, this article proves to be the only comparative work on the Caspian languages with this depth or breath. However, some texts provide a more comprehensive overview of just a single sub-group.

Tatic Within the Tatic group, the most comprehensive dialect study that exists is Yarshater's (1969) *Southern Tati Dialects*. This work contains a comparative grammar of the

Southern Tati varieties Chāli, Tākestāni, Eshtehārdi, Xiāraji, Ebrāhim-ābādi, Sagz-ābādi, Dānesfāni, Esfarvarini, and Xoznini. Some of these varieties have received individualized attention, e.g. Yarshater (1990) sketch on the village and language of Chal (Chali Tati). Additional Southern Tatic varieties not covered in Yarshater (1969) Alviri and Vidari (Yarshater, 1964), and Vafsi represent by two substantial texts including a grammatical sketch, examples, and texts (Stilo, 1971, 2004). The varieties spoken in Alvir and Vidar (Yarshater, 1969) are further south than many of the other Tatic languages spoken in the Caspian region. In contrast with Vafsi, one of the few languages to have retained a case distinction across gender and number categories with no syncretism in the paradigm, Alviri and Vidari have lost many distinctions. For instance, Aviri has retained gender only in the pronominal system, and Vidari has lost it completely. Similarly, Vidari has retained case-marking on definite direct objects, while Aviri has lost case marking nouns.

Work on Central Tatic varieties includes Ivanow's (1931) *The Dialect of Gozärkhon in Alamut*. This variety is not clearly placed within Northern, Central, or Southern Tatic. Within Central Tatic, there has been a short sketch of Kajali (Yarshater, 1960). Another important article to the study of Tatic is Stilo's (2018a) "Dikin Marāqei Tati of Alamut: an undocumented conservative Tati language." This article of 29 pages describes Dikin Marāqei, a Central Tatic language on its own branch. This language shows certain features like definite marking in the direct case only on feminine nouns. This pattern is the opposite of Zazaki, which shows definite distinctions only on masculine nouns in most varieties. Additionally, there has been work on Southern (Paul, 2011a; Nawata, 1982) and Northern-Central (Schulze, 2000; Miller, 1953; Stilo, 2008b) Talyshi as well as an overview of the state of research on Talyshi, Borjian's (2005) "Talish and the Talishis." Paul's (2011b) is a dialect study of Southern Talyshi including data from âstârâ, Haštpar, Rezvânšahr, Mâsâl, Fuman, and Šaft. The data is limited as this thesis is based on Pear stories narrated by native Talyshi speakers. The lack of diversity in the data is somewhat mitigated by the

rich semantic content captured by the Pear Story.[8] Another grammar to highlight here is Schulze's (2000) Northern Talysh, which is the most substantial grammar of Northern Talyshi published in English. It is brief and lacking texts as well as highly illustrative examples.

Northern Tatic is divided into two branches Karingani-Kalasuri-Khoynarudi and Harzani-Kilit. Kalasuri is known through Yarshater's (2005) "The Tati Dialect of Kalasur." This article included only 15 pages and only the briefest overview of forms in Kalāsuri. A similar study of Keringāni by Yarshater (2007) "The Dialect of Karingan" offers a 19-page overview. The Harzani-Kilit branch has no representation in English. However, there is some work in New Persian, e.g. *Tāti va Harzani, do lahja az zabān-i bāstān-e Āẕarbāyjān* (Kārang, 1953). The texts that exist for Tati, except for just a few (e.g. Stilo, 2004) are short summary articles. There are a few broad regional or dialectal studies that present more information (Stilo, 2018b; Yarshater, 1969). However, significant work needs to be done to ensure that the contribution of Tatic is felt in comparative Iranian linguistics.

Gilaki-Rudbari Within the Caspian languages proper, the Gilaki-Rudbari sub-branch has received mixed attention. The Rudbari sub-group has only had one brief overview article despite its being spoken in at least nine villages. Lazard's (1990) "Le Dialect Rudbar (Gilan)," totaling only 13 pages, provides less than three pages of grammatical description and nine pages of texts with their translations in French. Gilaki, on the other hand, has received more attention. Three major grammars exist for Gilaki *Gilyanskij Jazyk* (Rastorgueva, 1971) and *Severo-zapadnye iranskie Yazyki* (Rogova, 1999) in Russian and *The Gilaki language* (Rastorgueva et al., 2012) in English. In addition to these two more accessible grammars, there are several works in Persian, e.g. Jahāngiri's (2003) *Guyeš-e Gilaki-e*

[8]The Pear Story (Chafe, 1980) is a form of elicitation where participants watch a video of a man gathering Pears. Each participant then recites the story conveyed by the video they just watched. The resulting narrative captures certain nuances of the language's temporal and deictic system with some uniformity across data sets. In other words, the goal is to produce comparable data across languages and populations.

Lāhijān: vižegihā-ye āvāyi va sāxtvāžei [Gilaki dialect of Lahijan: a study on phonology, morphology, and lexicon].[9]

Mazandarani-Šahmirzadi The languages belonging to the Mazandarani and Şahmirzadi group have received very little attention in English. The Šahmirzadi language has been documented in a short sketch by Borjian (2019c) "The Caspian Language of Šahmirzād." Mazandarani proper is sub-divided into five groups: there is Central Caspian Mazandarani known through Borjian (2010) 'KALĀRESTĀQ ii. The Dialect." There are the varieties Gachsari and Velatru known through Borjian (2012e) "The Dialects of Velātru and Gachsar: The Upper Karaj Valley in the Caspian-Persian Transition Zone." This article, at 36 pages, is somewhat longer than Borjian's typical grammatical sketch. However, it contains data from both the varieties Gachsari and Velatru. Despite these languages having a geographic connection, their differences are significant enough to be classified, at least preliminarily, as two separate branches of Mazandarani. Borjian (2012e) cites Lambton (1938) as an important source for information on these varieties because the languages have become moribund since her collection efforts leading to publication in 1938. Lambton (1938) is quite confusingly entitled *Three Persian dialects*. It features some grammatical information, texts, and lists of words for Meyma'i and Joshaqāni (Kermanic > Kashanic > Soic), and Velāntruyi and Gachsari (Mazandarani); there is no data from any Persian (or even Southwestern Iranian) language, variety, or dialect.

Another branch of Mazandarani is Galeshi, known through Borjian & Borjian (2008) "The Last Galesh herdsman: Ethnolinguistic Materials from South Caspian Rainforests." This is an ethnographic account with texts and a glossary but does not include grammatical notes. The last branch of Mazandarani is known as Nuclear Mazandarani. The longest description of any Mazandarani variety is Yoshie's (1996) *Sāri dialect*. This is an

[9]It appears that there is some confusion with the nomenclature concerning Christensen's (1930) *Dialecte guilāki de Recht, dialectes de Fārizānd, de Yaran et de Natanz*. Here Christensen (1930) refers to the dialects Fārzāndi, Yarani, and Natanzi as Gilaki. However, their languages are properly classified as Central Iranian Kermanic (Kashanic > Natanzic).

account of the language of Sār, the capital of Mazandaran. Another article that should be mentioned in conjunction with Yoshie (1996) is Borjian's (2019e) "The Mazandarani Dialect of Kalijān Rostāq." Although this latter article is a short sketch, the language it describes is closely related to what is spoken in Sār. According to Borjian (2019e), the variety of Kalijān Rostāq, a village outside of Sār, differs from the language of Sār in that it has been less exposed to Persian encroachment. Additional works on Mazandarani include "The Caspian Dialect of Māhā" (Borjian, 2018a) and "The Caspian Dialect of Kujūr in Central Alborz" (Borjian, 2013c); the latter, which has significant influence from Central Caspian Mazandarani varieties. Additionally, Borjian (2019b) gives a comparative typology of Mazandarani and its varieties. For more information on Mazandarani see Borjian (2004) "Mazandaran: Language and People (The State of Research)."

Central Iranian Kermanic

Like the Caspian languages Gilaki and Mazandarani, there has been relatively little scholarly study on the Kermanic languages. Central Iranian Kermanic has had the benefit of several extensive dialectal overview articles, including "KERMAN xvi. LANGUAGES" (Borjian, 2017a), "Isfahan xxi. PROVINCIAL DIALECTS" (Stilo, 2007), and"CENTRAL DIALECTS" (Windfuhr, 1991). Each of these articles appeared in the Encyclopædia Iranica and covered the data with a slightly different focus. Borjian (2017a) and Stilo (2007) both have a geographic focus, although Stilo's (2007) focus on the so-called dialects allows him to be more directed in his comparative discussion of Kermanic. In this sense Stilo (2007) overlaps both with Borjian's (2017a) regional discussion and with Windfuhr's (1991) comparative look at the languages. Additionally, Stilo (2007) devotes some time to answering diachronic questions, including a discussion of imperfective marking strategies that elevate this article beyond being a simple juxtaposition of data.

In addition to the comparative contributions to Encyclopædia Iranica, there is a book length exploration of the Kermanic languages, Lecoq's (2002) *Recherches sur les dialectes*

kermaniens (Iran central): grammaire, textes, traductions et glossaires. This text contains data from Kermanic varieties, including the Kashanic varieties Qohrudi (Soic), Abuzeydābādi (Soic), Tāri (Soic), Abyānei (Natanzic), Bādrudi (Natanzic), Nāini (Nayini), Anāraki (Nayini), Varzenei (Nayini), and Ardestāni (Gazic). This collection of languages is relatively broad. The Kermanic languages are divided into three groups Kavir and Sivandi, which are not represented in this work, and the larger Nuclear Central Iran Kermanic group. This latter group is further divided into five sub-branches: Gazic, Judeo-Hamadani-Borujerdi, Kashanic, Khunsaric, and Yazdi-Kermani-Nayini. The Soic and Natanzic languages represent two sizable sub-branches of Kashanic, the most highly represented group in Lecoq (2002). The Yazdi-Kermani-Nayini sub-branch is further divided into Yazdi-Kermani and Nayini, the latter of which is represented by three varieties in Lecoq (2002). Finally, the Gazic sub-branch is represented by Ardestāni. Neither Judeo-Hamadani-Borujerdi nor Khunsaric is represented in this work. This work includes a grammatical description with many paradigms but few examples followed by texts and their translations in French.

Further work on Central Iranian Kermanic includes from the Kavir sub-branch the short overview articles "FARVI DIALECT" (Borjian, 2013a), "The Dialect of Khur" Borjian (2018b), and "Notes on the dialect of Khūr and Mihrijān" (Ivanow, 1929). For Nuclear Central Iran Kermanic, there have been two more significant (book-length) grammatical sketches *Meyma' i: A Central Iranian Plateau Dialect* (Borjian, 2012c) and *The Raji Dialect of Jowshaqan* (Borjian, 2013d). Both of these varieties belong to the Soic sub-branch of Kashanic, which was the best-represented group in Lecoq (2002). An additional overview of Kashanic is provided in "KASHAN ix. THE MEDIAN DIALECTS OF KASHAN" (Borjian, 2012b). From the Yazdi-Kermani group, there are several short sketches e.g. "Behdinān Dialect" (Windfuhr, 1989a) and *Sîändî, Yäzdî und Sôî* (Christensen, 1972), the latter which additionally contains data from Sîvandî, another higher-order branch of Kermanic equal to Kavir and Nuclear CIK. Additional work on the Kermanic languages include Bīdgoli (Yarshater, 1989), Zoroastrian Dari (Gholami, 2018b), Judeo-Kermanic (Lazard, 1981; Borjian,

2012a) , Keša'i (Borjian, 2017c), Nešalji (Borjian, 2016), Jarquya'i (Borjian, 2008a), and Kuhpâya'i (Borjian, 2011)

Komisenian

The Komisenian languages are only known through a few studies, sometimes as just a short section within a broader study. This sub-branch is further subdivided into three groups Lasgerdi, Sangisari, and Sorkhei-Aftari. The only grammatical sketch of the Lasgerdi language is known from Contributions à la dialectologie iranienne II: Dialects de la région de Semnan: sourkheī, lāsguerdī, sängesärī et chämerzadī (Christensen, 1935). This work contains a comparative overview of the Komesenian languages Lasgerdi, Sangesari, and Sorkhei, covering all three branches along with Šahmirzadi (Caspian). At 198 pages, it is not able to go in-depth into any one. A more substantial look at the Sangesari language in English is presented in *A dictionary of Sangesari with a grammatical outline* (Azami & Windfuhr, 1972). This work is primarily a dictionary that is supplemented with a grammatical sketch. Finally, there is a 47-page sketch of Aftari *The Komisenian Dialect of Aftar* (Borjian, 2008b). This covers the one Komisenian language not covered by Christensen (1935), although it is closely related to Sorkhei. This article is longer than Bojian's typical sketch and with copious examples; additionally, it serves as a major improvement from Windfuhr's (1982b) article AFTERĪ in Encyclopædia Iranica, which consists of just three paragraphs. The Komisenian languages are likely important for the study of Northwestern Iranian, but their lack of documentation makes their contribution speculative at best.

Balochic

The Balochic languages have received much more attention than many other Western Iranian groups. Due to a large amount of research into Balochi, it would be imprudent to attempt a comprehensive overview of the history of its scholarly study. Here I include some highlights. The Balochic sub-branch of Northwestern Iranian can be further subdivided into

Eastern Balochi, Koroshi, and Southern-Western Balochi (Hammarström et al., 2020). This division is in contrast with Jahani (2019), which also divides Balochi into three branches Eastern, Southern and Western, without commenting on the place of Koroshi or whether or not Southern and Western Balochi constitute a sub-group. Instead, Jahani (2019) gives the examples of Sarawan and Koroshi as Balochi varieties that don't easily fit into the three established categories. The status of Koroshi, as not-easily-categorized, is reflected in Hammarström et al. (2020). However, Sarawani is classified as Western Balochi in the Glottolog tree. Jahani's (2019) grammar, *A Grammar of Modern Standard Balochi*, is the result of a 2012 initiative to standardize Balochi orthography. This initiative, followed by an orthography conference in 2014 in Uppsala and a grammar conference in 2016, culminated in this grammar. As the project's goal was standardization, this grammar's value to comparative Iranian linguistics is likely marginal.

Of the Balochic languages, the Southern and Western sub-branch has received the majority of research attention, although that may be changing thanks to linguistic work by Maryam Nourzaei. Like many other languages, there was an early and substantial contribution to Balochi studies in Geiger (1895b) "Die Sprache der Balūtschen" (Geiger, 1895a). However, there has been a (not-insubstantial) grammatical tradition since Mockler's (1877) *A Grammar of the Baloochee Language as it is spoken in Makrān (ancient Gedrosia), in the Persi-Arabic and Roman characters.* The Makrani variety of Balochi belongs to the Southern (Southern-Western) Group. Other grammars of Makrani Balochi include *The grammar of Balochi language* (Nasīr Xan, 1984) published in Quetta, the capital of Pakistani Balochistan and the significantly shorter "A Description of the Mekranee-Beloochee Dialect" (Pierce, 1875). Perhaps the largest grammar of Balochi is Barker & Mengal's (2014) *A course in Baluchi* covering the Rakhshani variety of Balochi, which belongs to the Western sub-branch. This grammar is primarily a teaching text. However, its sheer size and coverage make it an important reference. In over 1200 pages, Barker & Mengal (2014) present lessons on grammar, orthography, music, art, and culture. Perhaps the only real

issue with their grammatical account is that the imperfective clitic *=a*, which is a preposed (it occurs before the verb) enclitic (it is prosodically attached to the preceding element), is systematically absent from the text. According to Barker & Mengal (2014), "a brief /ə/ vowel is heard between [the preceding noun] and the verb (Barker & Mengal, 2014, 149). This /ə/ is optional and has no discernible meaning. The role of this morpheme marking the imperfective aspect is now well known through dedicated study (e.g. Nourzaei & Jahani, 2013), and through mention in more recent articles and grammars (e.g. Axenov, 2006; Paul, 2003, etc.). Additionally, Axenov's (2006) *The Balochi Language of Turkmenistan: A corpus-based grammatical description* provides a description of the Balochi spoken in Turkmenistan. This language, like the Rakhshani Balochi described by Barker & Mengal (2014), belongs to the Western sub-branch. It is divergent enough in its nominal morphology to be a useful addition to work on Balochi. Additionally, Axenov (2006) provides a much more complete picture from a linguistics perspective with fully glossed examples and grammatical descriptions. Dames1881NBalochi

Koroshi Balochi has recently received some attention in the form of a brief grammatical sketch Borjian (2014) and a grammar *Koroshi A Corpus-based Grammatical Description* (Nourzaei et al., 2015). The Koroshi variety is divergent from other Balochi varieties in several respects, including the lack of a separate object case on nominals, a plural marker *obār*, and the use of the definite suffix *-ok*. The occurrence of a K-form definite suffix has parallels in Kurdish, Zaza-Gorani, and other Iranian varieties. These parallel drew the attention of Maryam Nourzaei and Geoffrey Haig leading to their joint project at the University of Bamberg and resulting in numerous articles and presentations on the subject (e.g. Haig, 2019a; Nourzaei, 2017, 2020, etc.). In addition, there are several broad overview articles, including Elfenbein (1989); Jahani & Korn (2009) and the phonological overview Elfenbein (1997). Perhaps the most important study for the comparison of Balochi varieties is Korn's (2003) *Toward a historical grammar of Balochi*. This work is primarily a phonological study first reconstructing a common Balochi from which all varieties branch out. Then, she traces

the developments in all the dialects based on the best data available to her. This work should be consulted as an overview of Balochi studies. Its thoroughness and specialization make it more than could be covered in this section.

There has been additional work on specific topics in Balochi linguistics that include ergativity (Korn, 2008b), diachrony (Shahbakhsh, 2004; Korn & Durkin-Meisterernst, 2009; Korn, 2008a; Bashir, 2008a; Barjasteh-Delforooz, 2003; Korn, 2003), contact (Elfenbein, 1982; Rzehak, 2009; Barjasteh-Delforooz, 2008; Farrell, 2003; Baranzehi, 2003), the pronominal system (Delforooz & Levinsohn, 2014; Dabir-Moghaddam, 2008), syntax (Jahani et al., 2010; Jahani, 2008b), semantics (Filippone, 1996), morphology (Korn, 2008c; Jahani, 2003), discourse analysis (Barjasteh-Delforooz, 2007), and many others. As is the general trend in Iranian linguistics, the majority of publication occurs in bound volumes. Generally these volumes result from conferences or workshops. There are two such volumes that need to be mentioned here Jahani et al. (2008) and Jahani & Korn (2003). These two volumes represent a large portion of modern linguistics study of Balochi language and culture. A full list of the contents of these volumes and other volumes, important to Iranian studies are give in table 1.2.

Laki-Kurdish

The Laki-Kurdish sub-branch of Northwestern Iranian has, like Balochi, been the beneficiary of much scholarly study. However, this attention has only been paid to the Kurdish side. The only dedicated (book-length) grammar of Laki was Belelli's (2016) doctoral thesis *A study on language and folklore in the city of Harsin (in Kermānshāh province of West Iran): Sketch grammar with texts and lexicon.* However, this thesis has not been officially published and remains inaccessible to scholars outside of her circle.[10] An additional small work on

[10]The inaccessibility of Belelli (2016) may be subject to change. It is rumored to appear in print from the University of Bamberg press in early 2021. I can only speculate that the COVID-19 pandemic has delayed its release.

Laki is "Le dialecte laki d'Aleshtar" (Lazard, 1992). This article contains a grammatical sketch with few examples and some texts with translations in French.

On the other hand, Kurdish has been studied more thoroughly, though not all varieties have received equal attention. There have been several overview type articles, "On the linguistic history of Kurdish" (Jügel, 2014) and "Introduction to special issue Kurdish: A critical research overview" (Haig & Öpengin, 2014) provide an important starting point for the comparative linguistic study of Kurdish. There are several grammatical overviews, e.g. Socin (1895) and McCarus (2009) as well as the phonological sketch McCarus (1997) and the morphological sketch McCarus (2007). Each of these introductions is flawed in their exposition of Kurdish as they combine information from several different "dialects." After separating from Laki, the Kurdish sub-branch of Northwestern Iranian is further subdivided into Northern, Central, and Southern Kurdish, and there is significant variation and subdivision within each of the three groups. This problem is abundantly clear in one of the earliest grammars *Grammar of the Kurmanji or Kurdish Language* (Soane, 1913), which confused forms in side-by-side exposition from languages across sub-groups. This, coupled with inconsistent and non-standard orthography, make this work opaque. The Northern group shows case, number, and gender marking on nouns, which is absent from the Southern group; The Northern group has a three-way stop contrast, while the Central and Southern groups only have a two-way contrast. These are just two examples of how the monumental divergence between Kurdish varieties in their phonology, morphology, and syntax (see chapter 4 for more on shared features of Kurdish varieties). Despite the diversity observed between Kurdish varieties, there is a strong ethnic bond between Kurds that makes reference to Kurdish varieties as separate languages anathema.[11]

The first look into the comparative dialectology of Kurdish comes from Mackenzie's (1961) *Kurdish Dialect Studies*. Mackenzie (1961) was a two-volume series consisting of the first book, which consisted of a comparative grammar, and the second book, which pro-

[11] For more on the Kurdish language ideologies see Sheyholislami (2018).

vided texts that Mackenzie (1961) had gathered from various consultants during his time in Iraqi Kurdistan. This text is still the best resource for many varieties of Kurdish from the Northern and Central groups. However, data from Southern Kurdish is notably lacking. Additionally, some of Mackenzie's (1961) assertions about the relationships between the groups would serve as fodder for future scholars as more data from each of these groups comes to light. Additionally, MacKenzie contributed several articles to the historical-comparative study of Kurdish including "The Origins of Kurdish" (MacKenzie, 1961), "Gender in Kurdish" (MacKenzie, 1999b, originally 1954), and "Pseudoprotokurtica" (MacKenzie, 1999d, originally 1963), as well as numerous philological works. Work on Kurdish and its many sub-varieties has recently experienced enhanced prospects due to the establishment of the Manchester corpus *the Dialects of Kurdish* (Matras et al., 2016). This corpus was produced by eliciting responses from a survey. The majority of the consultants were educated, young males. One criticism of this project is that the team often elicited dialect data in cosmopolitan cities, where consultants were not necessarily vetted for their ties to the region (p.c. with Jaffer Sheyholislami, Carlton University). Despite issues stemming from data elicited in one location but reflecting the speech patterns of another, this is the largest and broadest study of its kind. The more recent article "Revisiting Kurdish dialect geography: Findings from the Manchester Database" (Matras, 2019) is a status report. Additionally, Matras (2019) gives an account of the features that form the major divisions within Kurdish. This paper is essential reading for Kurdish dialectology.

Additional works that serve as essential resources for Kurdish studies include Alignment in Kurdish: a diachronic perspective (Haig, 2004) and *Alignment change in Iranian languages: a construction grammar approach* (Haig, 2008a), which are large-scale analyses of the development of ergativity in Iranian from the perspective of the extant sources. These works look at Old Iranian, as represented by Old Persian, Middle Iranian (Middle Persian), and Northern and Southern Kurdish. Two additional sources that are primarily ethnographic are *The Kurds: a concise handbook* (Izady, 1992) and *The Kurds: An Encyclopedia*

of Life, Culture, and Society (Maisel, 2018). These are both large volumes covering various topics, including art, culture, geography, and language. Izady (1992), in particular, is well-known for its rich maps illustrating general background on Kurdish varieties and their distribution. Here, I augment this discussion of research on Kurdish generally with some background on the sources available for the three branches of Kurdish.

Northern Kurdish The Northern Kurdish group can be further subdivided into Northern, Southeastern, Southern, and Western Kurmancî. Kurmancî is the name given to Northern Kurdish by its speakers, and it is pervasive except for the Northern-Kurdish-speaking region of Iraq, where the language is referred to as Behdînî or Badînî by some speakers as well as the non-Northern-Kurdish-speaking residents of the area. Two articles must be mentioned regarding the sub-varieties within Kurmancî: "Regional variation in Kurmanji: A preliminary classification of dialects" (Öpengin & Haig, 2014), which outlined several broad generalization and prospects for the use of the data and "Kurmanji Kurdish in Turkey: structure, varieties, and status" (Haig & Öpengin, 2018), is a much larger study outlining the sub-divisions within Kurmancî and the various features that define them. Added to these two articles is Haig's (2019b) "Northern Kurdish (Kurmanji)," a reasonably detailed overview of Kurmancî and its regional variation. This article is, at its core, an overview of Kurmancî and its varieties. As such, it by no means should be seen as a replacement for Haig & Öpengin (2018).

There have been many grammars of Northern Kurdish; some of them have been published relatively recently, e.g. Thackston (2006a) and Ekici (2007), which feature a grammatical sketch with copious examples and texts, as well as Bedirxan & Lescot (1986), which purports to represent a standard Kurmancî. Two grammars appeared in Russian featuring the Kurmancî variety spoken in Azerbaijan (Bakaev, 1965), which belongs to the Northern sub-group, and the variety spoken in Turkmenistan (Bakaev, 1962), which, paradoxically, belongs to the Southwest group. The Kurmancî of Turkmenistan is most closely related to

the language spoken in Aleppo (Syria) and Urfa (Turkey) despite its geographic location in Khorasan, at the eastern edge of the Iranian world. A grammar of the Southeastern varieties spoken in the city Amadî and around Mt. Sincar appeared in Blau (1976), which contains a grammatical sketch with examples and texts.

There has been some work on the Northern Kurmancî variety spoken in Muş, e.g. Gündoğdu (2015), Gündoğdu (2017a), Gündoğdu (2018), and Gündoğdu (2019). Although none of these studies is a grammar, the sheer size of the works and the diversity of the subject matter assure their place as a resource for the language as well as the specific linguistic topics they address. In addition to these grammars and articles, there has been work on specific topics in Northern Kurdish linguistics, e.g. ergativity and alignment (Toma, 2018; Dorleijn, 1996; Gündoğdu, 2017a; Mahalingappa, 2009), phonology (Barry, 2019), sociolinguistics (Haig & Mustafa, 2019; Herkenrath, 2019; Haig, 2007), and more.

Central Kurdish Due to the location of Central Kurdish mainly in Iraq, the varieties under this umbrella have received much attention. The majority of the data collection in Mackenzie (1961) took place in Iraq, and therefore, the majority of the varieties covered therein are Central Kurdish. Central Kurdish is sometimes referred to as Soranî, although this term initially referred to the variety spoken in the Soran emirate with its capital in Rewanduz (Iraq). According to Hammarström et al. (2020), Central Kurdish is divided into three sub-branches, Mukri(yanî), Sine'i, and Sorani. However, it is not clear to me that these divisions are actually coherent. For instance, Garmiyani shares the verbal inflection system of Sine'i, but it is classified under Sorani. Likewise, the nominal morphology of Rewandiz (Sorani) has more in common with Mukri than it does with Suleymanî Soranî, sometimes referred to as the standard language.[12] Additionally, Khanaquin is filed under Sorani. However, this variety falls under the Southern Kurdish group based on cognate vocabulary, alignment, and verbal morphology.

[12]See Sheyholislami (2018) for more information on the varieties of Kurdish and related terminology.

Several grammars of Central Kurdish, the most comprehensive of which is *The Mukri Variety of Central Kurdish* (Öpengin, 2016). This book focuses on Mukriyanî, spoken in the area around Mehabad in Iran. Mukriyanî is a significant variety in the way it diverges from other Central Kurdish varieties, particularly in its case system. It preserves oblique marking on nouns, the preposition *de* 'on,' and other features that place it on a continuum of transitional dialects on the way to Northern Kurdish. According to MacKenzie (1954), the zone around the Greater Zab river marks the true transition zone, where the language is neither Northern nor Central Kurdish. A source for the Central Kurdish spoken in Suleymanî city is *A Kurdish-English dictionary* (Wahby & Edmonds, 1966). Wahby & Edmonds (1966) provide a dictionary with texts in Central Kurdish. The variety spoken in Suleymanî also had the benefit of an early comprehensive grammar McCarus's (1956) *Descriptive analysis of the Kurdish of Sulaimaniya, Iraq.* Just like Kurmancî, Thackston (2006b) has produced a reference grammar with copious examples and texts in the literary register of several varieties.

There have been many studies of specific issues in Central Kurdish linguistics: pronouns/clitics/verbal inflection (Öpengin, 2019; Kareem, 2016; Öpengin, 2013; Haig, 2013), adpositions (Edmonds, 1955; Öpengin, 2013; Samvelian, 2007b), and definiteness (Haig, 2019a; Haig & Mohammadirad, 2019), and more. See "Subdialectal Differences in Sorani Kurdish" (Malmasi, 2016) as well as Mackenzie's (1961) *Dialect Studies* for more on Central Kurdish varieties and the specific features that distinguish them.

Southern Kurdish The Kurdish variety that has received the smallest amount of scholarly attention is Southern Kurdish. Southern Kurdish is spoken primarily in Iran in Kermanshah, Ilam, and Kordistan provinces. Hammarström et al. (2020) divides this group into three sub-branches Ilami, Lakic SK, and Peripheral Kermanshahic. However, there is considerable difficulty in drawing a genealogical tree for Southern Kurdish. These varieties have significant differences in syntactic alignment patterns, morphology, attribution

marking, and many other points. An additional problem is with the geographic distribution of Southern Kurdish. Many varieties correspond roughly to cities, towns, or villages, just like the rest of Kurdish. However, many of the distinct varieties are regional variants (e.g. Kolyai) spoken broadly within an area but not necessarily in the geographic centers. Likewise, some varieties are tribal-lects associated with a particular group and not an area (e.g. Malikshahi). The result is that often people in the same town speak different varieties.

One influential article analyzing the genealogy problem is Belleli's (2019) "Towards a dialectology of Southern Kurdish: Where to begin?" This article analyses and expands upon a preliminary classification of Southern Kurdish by Fattah (2000). Fattah's (2000) *Les dialectes kurdes méridionaux: étude linguistique et dialectologique* is not just the largest study of Southern Kurdish but the only one of its kind. Southern Kurdish is the most endangered sub-branch of the Kurdish group, and it may be the one that shows the most diversity.

1.4.3 The study of particular Southwestern Iranian languages

The southwestern Iranian languages are divided into six sub-branches: Bashkardi, Fars Dialects, Farsic-Caucasian Tat, Kumzari, Larestani, and Luric-Dezfulic. I do not go into great detail about the Standard Persian varieties Farsi (Iran), Dari (Afghanistan), and Tajiki (Tajikistan). There are certain grammatical features of the standard languages, like the lack of the definiteness marker *-(h)æ/e*, which suggest that these dialects are constructed and do not reflect the natural evolution of the languages. That being said, the majority of research into Iranian languages, in general, has been focused on standard New Persian. Such studies include a variety of subjects, grammatical sketches (Lazard, 1989; Paul, 2019; Windfuhr, 1979; Windfuhr & Perry, 2009), phonology (Windfuhr, 1997; Mazdeh, 2020), morphology (Perry, 2007; Bonami & Samvelian, 2015; Ciancaglini, 2011; Lazard, 2011; Jasbi, 2020b,a; Karimi & Smith, 2020; Lorenz, 1982, 1999; Key, 2008; Paul, 2008b; Taleghani, 2008; Paul,

2008a; Samvelian, 2005, 2007a, 2018; Parsafar, 2010), corpus linguistics (Seraji, 2015), syntax (Rasekhi, 2020; Ghomeshi, 2020; Gebhart, 2008; Ghomeshi, 2008; Karimi, 2008; Jahani, 2008a; Kahnemuyipour, 2014, 2016; Parsafar, 2010; Samiian, 1994; Ghomeshi, 1997; Moyne, 1971; Moyne & Carden, 1974; Samiian, 1983), diachrony (Orsatti, 2011; Hasandust, 2011a; Gershevitch, 1985d; Utas, 2013d; Haider & Zwanziger, 1984), semantics (Abdollahnejad & Storoshenko, 2020; Windfuhr, 1982a; Mahootian, 2008; Karimi-Doostan, 2008; Utas, 2013e; Parsafar, 2010), dialectology (Gershevitch, 1985c), lexicography (Rastegar, 1999), psycholinguistics (Jeremiás, 1999), sociolinguistics (Utas, 2013a), and many more. Thee remaining Southwestern Iranian languages are understudied. For many of them no information is available at all. For others, there is merely little that is known.

Bashkardi For Bashkardi, little is available. Skjærvø's (1989a) 7-page contribution to Schmitt (1989) entitled "Languages of Southeast Iran: Lārestānī, Kumzārī, Baškardī" contains a brief overview of Bashkardi. This work can be supplemented by Skjærvø's (1988) contribution on Bashkardi to Encyclopædia Iranica. This article gives significantly more information and is dedicated to Bashkardi and its varieties. However, it requires updating as the overall understanding of Bashkardi has changed since 1988. Other than this, there are only the contributions provided by a short section in Mohammadirad (2020), and Korn's (2017a) "Notes on the Nominal System of Bashkardi." The focus of Mohammadirad's (2020) dissertation was pronominal clitics in Iranian languages. As such, there is a significant contribution to the Bashkardi pronominal system and associated topics verbs, adpositions, etc. Korn (2017a) is a more comprehensive look at the nominal system of Bashkardi. Therefore, this article serves as an accessible and essential work for the comparative study of Iranian languages. Additionally, there are several works published in Farsi; from the University of Sistan and Balochistan in Zahedan, there is *Bar-resi-ye touṣifi-ye saxt-važe-ye guyeš-e bašakardi [A Descriptive Study of Bashakardi Dialect Morphology]* (Seddiqi Nezhad, 2010). In addition, there is *Barresi va tawsif-e zabānšenāxti-ye guyeš-e Minābi [A linguistic study*

and description of the Minabi dialect] (Mohebbi Bahmani, 2006) on the Minabi variety. Finally, there is *Rudān: Behešte Janub* (Mo3tamadi, 2001) on the Rudani variety. These three works are not easily accessed by western scholars. Despite the current lack of study, there may be hope for Bashkardi linguistics as scholars like Agnes Korn continue to work on related projects.

Fars "Dialects" Even though it is the Fars province that lends its name to Farsi, the name of the standard New Persian languages, the languages and varieties spoken in Fars are highly divergent from standard Persian. Like other (Southwestern) Iranian languages, these varieties have not been the subject of many studies. There are several high-level overviews, including "Les dialectes du sud-ouest de l'Iran" (Le Coq, 1989b) and the more recent "FĀRS viii. Dialects" (Windfuhr, 2012). The latter, a contribution to Encycopædia Iranica, is an essential source for the comparison of these varieties. However, it is not an in-depth study of any of them. Early work on these varieties was conducted by (Mann, 1909), *Die Tâjîk-Mundarten der Provinz Fârs*. This volume is a comparative grammar of the varieties spoken in Som ān, Māsaram, Pāpun, and Burenǰān. Additionally, there is a short grammatical sketch of "The Language of the Kharg Island," which does not categorize well into the already established groupings. As such, I include Borjian's (2019d) 23-page sketch here. Additionally, the Angali "dialect" is another language that is not easily categorized into the existing sub-branching. If this truly fits into the Fars group, then Angali's (2004) dissertation, *The Angali Dialect*, serves as the largest grammatical sketch of any one of the languages of Fars. This text includes a comprehensive grammar, examples, and texts.

Farsic-Caucasian Tat Farsic-Caucasian Tat is the sub-branch of Southwestern Iranian that contains New Persian. For this reason, I have already covered a portion of it. Some highlights that are worth mentioning here are as follows: Suleymanov's (2020a) *A Grammar of Şirvan Tat* represents the largest study in English of any Caucasian Tat variety. In contains a grammatical sketch with examples and texts. Additionally, the documentation

efforts of Suleymanov (2020a) have focused on several different varieties of Caucasian Tat falling into the two sub-branches Judeo-Tat and Muslim-Tat. Suleymanov (2020a) serves to be the most substantial result from this work. However, there have been a number of articles, e.g. Suleymanov (2020b), that present a picture of the typological variation among these varieties. This article is an in-depth examination of the functions of the case-marking clitic *=ra* in Tat varieties. This is particularly interesting because the cognate particle in New Persian has lost functionality, becoming a definite direct object marker. However, in Tat, it has expanded its functions perhaps due to contact with Azeri (Turkic).

The Farsic side of this sub-branch splits into and Eastern and Western group, with some Judeo-Persian varieties that are not easily categorized into either group. From western Farsic, there have been several short summary articles about the so-called Perso-Tabaric dialects (Borjian, 2013b,e,f, 2012d). These are languages spoken in Northern Iran in what Borjian (2013b) refers to as the transition zone between New Persian and the Caspian languages (e.g. Mazandarani). From Judeo-Persian, there is a broad over view Borjian (2015), which scopes over all Judeo-Iranian languages, regardless of their genealogical affiliation. Borjian, additionally, produced a short sketch of Judeo-Isfahani (Borjian, 2019a). However, at 41 pages, it is significantly more detailed than many of the short sketches he has produced. It contains a grammar with paradigms and texts with English translations. Paul's (2013) *A grammar of early Judaeo-Persian* is a somewhat longer sketch. In line with his grammar of Zazaki Paul (1998b), Paul (2013) features a dense grammatical description with many examples. However, there are no whole texts, and a lack of glossing diminishes some of its functionality.

Other SW Iranian It is the case that very little research has been conducted on the rmaining branches of Southwestern Iranian languages. The remaining branches include Kumzari, Larestani, and Luric-Dezfulic. Kumzari now has the substantial overview provided by Anonby (2019). This is a broad-scoping, look at grammar and ethnographic information.

Its level of comprehensiveness makes it essential reading. Likewise, there are acouple of volumes on Larestani Kamioka & Yamada (1979) and Kamioka et al. (1986). These are a pair containing basic vocabulary in the first volume and a glossary in the second. These are not grammatical sketches, but they are worth mentioning due to the fact that they constitute the most that is available in English on Larestani. However, there is at least one substantial grammar available in Farsi, e.g. Khonji (2000) *Dastur-e zabān-e lārestāni bar mabnā-ye guyeš-e Xonji*.

As for Luric-Delfuzic, there are two sizable sudies in English. The most recent is Anonby & Taheri-Ardali's (2019) overview of Bakhtiari (South Luri). Like other contributions to Haig & Khan (2019), it is the most comprehensive description of the language to date. In contrast, the best availible grammar of Northern Luri comes from *Tales from Luristan (Matalyâ Lurissu): tales, fables, and folk poetry from the Lur of Bâlâ-Garîva (Harvard Iranian Series 4)* (Amān Allāhī & Thackston, 1986). However, this was not at its core a linguistic study. Rather, it is a collection of poetry with translations in English. The grammatical supplement, written by Wheeler Thackston, is just that, a supplement. However, this completion should be viewed as a substantial primary source for Luri texts.

1.5 Thoughts and Conclusions

Iranian linguistics and the documentation of the Iranian languages have been held back by several trends. First, the field has been Perso-centric. The majority of linguistics research has been conducted in Persian and with a Persian lens. Many phenomena are shared between various branches of the Iranian languages, including ezafe marking and differential object marking (DOM). However, when these concepts are studied, they are most often focused narrowly on Persian. An additional problem is that the standard language and standard language ideology permeates these studies. It is the standard New Persian language based on classical New Persian that is often the object of study. It is often assumed

that the standard language is frozen in time, while colloquial varieties are transient "new" things. However, as I demonstrate in the following chapter (ch. 2, spoken varieties have preserved features like the definiteness marker *-(h)e* that likely have old roots despite their conspicuous absence in the Standard. For Iranian and Persian linguistics to progress, scholars need to pay more attention to natural living languages; study the languages, not the prescriptions.

In addition to the unequal distribution of research focusing on Persian, national languages, languages with a significant international presence or national movements, like Kurdish and Balochi, have received more study than the languages that are isolated or have few speakers. In other words, the languages most in need of scholarly (and other types of) attention are the least likely to receive it. Another problem in Iranian linguistics that I haven't discussed yet here is how Iranian linguistics articles are published. Although there are journals that focus on Iranian studies, such as *Iran and the Caucasus*, *Indo-Iranian Journal*, and *Iranian Studies*, most linguistics articles get published in bound volumes that are the result of a workshop or conference. Although the quality of these articles is often of high quality due to the level of peer review typical in our sub-field, these volumes tend to difficult to access. Recently, some of these volumes, e.g. Gündogdu et al. (2019), have gone open access, which may prove to be a welcome trend. Hopefully, efforts like this will serve to connect research in the west with our colleagues in Iran, Iraq, Turkey, and beyond. I have included a table with the significant bound volumes with substantial contributions to Iranian linguistics in table 1.2 below.

To conduct diachronic linguistic research, one needs to juxtapose data from every presumed sub-branch. The greater the diversity of the data, the greater the accuracy of the reconstructed result will be. It is unlikely that one could discover the full spread of any particular feature, looking only at the available source for the Western Iranian languages presented here. The rest of this study is dedicated to evaluating several open questions in Iranian linguistics. I answer these questions knowing full well that as more and better data

comes to light, my theories will be subject to reanalysis and reinterpretation. The goal is to build the best case for a particular conclusion. For some of these questions, like the morphosyntactic status of the ezafe (ch. 3), they are long-standing questions. For others, like why Kurmancî has no definite article (ch. 4), they are questions that no one has asked.

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics

Geiger (1895b)	*Grundriss der iranischen Philologie*
Horn (1895)	Neupersische Schriftsprache
Geiger (1895a)	Die Sprache der Balūtschen
Socin (1895)	Die Sprache der Kurden
Bartholomae (1895)	Awestasprache und Altpersisch
Salemann (1895)	Mittelpersisch
Geiger (1895c)	Kleinere Dialekte und Dialektgruppen
Morgenstierne (1964)	*Indo-Iranica; mélanges présentés à Georg Morgenstierne, à l'occasion de son soixante-dixième anniversaire*
Abaev (1964)	O Dialektax Osetinskogo Yazlyk
Emeneau (1964)	Linguistic Desiderata in Baluchistan
Boyce (1964)	The use of relative particles in Western Middle Iranian
Yarshater (1964)	The Dialects of Alvir and Vidar
Benveniste (1964)	La racine *yat-* en indo-iranien
Leumann (1964)	Altpersisch hagmatā
Morgenstierne (1982)	*Monumentum Georg Morgenstierne I and II*
Lazard (1981)	Le Dialecte des Juifs de Kerman
Lorenz (1982)	Die direkte Rede im Tāğischen
MacKenzie (1982)	Matalūna
Nawata (1982)	The Masal Dialekt of Talyshi
Rossi (1982)	Balōčī Miscellanea
Shaked (1982)	Pahlavi Notes
Thordarson (1982)	Preverbs in Ossetic
Windfuhr (1982a)	The verbal category of inference in Persian
Sims-Williams (1985)	*Philologia Iranica*
Gershevitch (1985e)	Genealogical Descent in Iranian (1973)
Gershevitch (1985g)	Iranian words containing -ăn- (1971)
Gershevitch (1985i)	The Crushing of the Third Singular Present (1970)
Gershevitch (1985c)	Dialect Variation in Early New Persian (1965)

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Gershevitch (1985d)	Etymological Notes on Persian mih, naxcīr, bēgāne, and bīmār (1964)
Gershevitch (1985a)	Ancient Survivals in Ossetic (1952)
Gershevitch (1985b)	Bactrian Inscriptions and Manuscripts (1967)
Gershevitch (1985f)	Iranian Notes (1949)
Gershevitch (1985h)	Sogdian Compounds (1949)

Schmitt (1989)	*Compendium Linguarum Iranicarum*
Windfuhr (1989a)	New West Iranian
Sundermann (1989c)	Wesrmitteliranische Sprachen
Thordarson (1989)	Ossetic
Sundermann (1989b)	parthisch
Sundermann (1989a)	mittelpersisch
Lazard (1989)	Le persan
Windfuhr (1989b)	Western Iranian Dialects
Le Coq (1989a)	Les dialectes du caspians et les dialectes du nord-ouest de l'Iran
Elfenbein (1989)	Balōčī

	Iranica Varia: Papers in Honor of Professor Ehsan Yarshater
Lazard (1990)	Le Dialect Rudbar (Gilan)
MacKenzie (1990)	Pahlavi compound abstracts
Thordarson (1990)	Old Ossetic Accentuation
Vahman & Asatrian (1990)	Gleanings from Zāzā vocabulary

Skalmowski & van Tongerloo (1993)	*Medioiranica: proceedings of the International Colloquium organized by the Katholieke Universiteit Leuven from the 21st to the 23rd of May 1990 (Orientalia Lovaniensia analecta 48)*
Bielmeier (1993)	Das Alanische bei Tzetzes
Degener (1993)	Zur Syntax des Khotanischen
MacKenzie (1993)	Clitics in Khwarezmian
Shaked (1993)	Iranian Elements in Middle Aramaic: Some Particles and Verbs
Jeremiás (1993)	On the Genisis of the Periphrastic Progressive in Iranian Languages

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Josephson (1993)	The Preverb be in some Late Book Pahlavi Texts
Kaye (1997)	*Phonologies of Asia and Africa (Vol. 2)*
Testen (1997a)	Old Persian and Avestan Phonology
Weber (1997)	Pahlavi Phonology
Windfuhr (1997)	Persian Phonology
McCarus (1997)	Kurdish Phonology
Elfenbein (1997)	Balochi Phonology
Testen (1997b)	Ossetic Phonology
Proceedings of the Third European Conference of Iranian Studies: held in Cambridge, 11th to 15th, 1995 (Beiträge zur Iranistik 17):	
Sims-Williams (1998)	Part 1 Old and Middle Iranian Studies
Melville (1999)	Part 2 Medieval and Modern Persian Studies
Paul (1998a)	The position of Zazaki among West Iranian languages
Emmerick (1998)	Khotanese ei
Rastegar (1999)	Abschlußbericht über das Neupersische Personennamenbuch (NpPNB) Zur Erstellung der Namenslemmata im NpPNB
Jeremiás (1999)	Grammar and linguistic consciousnss in Persian
Goodrick (1999)	The measurement of style in Persian texts: a question of validity
Lorenz (1999)	Partikeln in der modernen Tadshikischen Sprache
Cereti & Paul (1999)	*Iranica Diversa Vol. I and II*
MacKenzie (1999a)	Bājālānī
MacKenzie (1999e)	The 'Indirect affectee' in Pahlavi (1964)
MacKenzie (1999c)	Pahlavi compound abstracts
MacKenzie (1999h)	When is a postposition not a postposition (1990)
MacKenzie (1999g)	The vocabulary of the Lahore Tafsir (1972)
MacKenzie (1999b)	Gender in Kurdish (1954)
MacKenzie (1999f)	The language of the Medians (1959)
MacKenzie (1999d)	Pseudoprotokurtica (1963)
Jahani & Korn (2003)	*The Baloch and their neighbors: Ethnic and Linguistic Contact in Historical and Modern Times*
Farrell (2003)	Linguistic Influences on the Balochi Spoken in Karachi

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Paul (2003)	The Position of Balochi Among the Western Iranian Languages: The Verbal System
Baranzehi (2003)	The Sarawani Dialect of Balochi and Persian Influence on It
Rzehak (2003)	Some Thoughts and Material on Balochi in Afghanistan
Barjasteh-Delforooz (2003)	The Structure of the Present and Past Stems in Balochi Compared to Old, Middle and New Persian
Korn (2003)	Balochi and the Concept of North-Western Iranian
Jahani (2003)	The Case System in Iranian Balochi in a Contact Linguistic Perspective
Woodard (2004)	*The Cambridge Encyclopedia of the World's Ancient Languages*
Schmitt (2004)	Old Persian
Hale (2004a)	Avestan
Hale (2004b)	Pahlavi
Weber (2005)	*Languages of Iran: past and present: Iranian studies in memoriam David Neil MacKenzie (Iranica 8)*
Yarshater (2005)	The Tati Dialect of Kalasur
Thiesen (2005)	Eleven Etymologies
Skjærvø (2005)	Avestica III: Notes on the Avestan Locative Singular
Kaye (2007)	*Morphologies of Asia and Africa (Vol. 2)*
McCarus (2007)	Kurdish Morphology
Weber (2007)	Pahlavi Morphology
Perry (2007)	Persian Morphology
Skjærvø (2007)	Avestan and Old Persian Morphology
Macuch et al. (2007)	*Iranian languages and texts from Iran and Turan: Ronald E. Emmerick memorial volume (Iranica 13)*
Cantera (2007)	The Accusative of the i- and u- stems with Presuffixal Full or Large Grade in Avestan
Lazard (2007)	La versification en parthe et son heritage persan
Sims-Williams (2007)	The Sogdian potentialis
Yarshater (2007)	The Dialect of Karingan

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Gippert (2007)	Albano-Iranica
Paul (2007)	Zur Lage der Gorani-Dialekte im Iran und ihrer Erforschung
Jahani et al. (2008)	*The Baloch and others: linguistic, historical and socio-political perspectives on pluralism in Balochistan*
Dabir-Moghaddam (2008)	On Agent Clitics in Balochi in Comparison with Other Iranian Languages
Barjasteh-Delforooz (2008)	A Sociolinguistic Survey among the Jadgal in Iranian Balochistan
Korn (2008c)	The Nominal Systems of Balochi: How Many Grammars?
Jahani (2008b)	Restrictive Relative Clauses in Balochi and the Marking of the Antecedent -Linguistic Influence from Persian?
Bashir (2008a)	Some Transitional Features of Eastern Balochi: An Areal and Diachronic Perspective
Stilo et al. (2008)	*Aspects of Iranian linguistics*
Samvelian (2008)	The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish
Holmberg & Odden (2008)	The Noun Phrase in Hawrami*
Gebhart (2008)	Classifiers, plural and definiteness in Persian
Ghomeshi (2008)	Markedness and bare nouns in Persian
Karimi (2008)	Raising and Controll in Persian
Key (2008)	Differential object markingin a Medival Persian text
Mahootian (2008)	Inversion and topicalization in Farsi discourse: A comparitive study
Paul (2008b)	Some remarks on Persian -ra as a general and historical issue
Taleghani (2008)	Mood and modality in Persian
Korn (2008b)	Marking of Arguments in Balochi Ergative and Mixed Constructions
Haig (2008b)	The emergence of ergativity in Iranian: reanalysis of extension?

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Jahani (2008a)	Expressions of future in Classical and Modern New Persian
Karimi-Doostan (2008)	Event structure of verbl nouns and light verbs
Nilipour (2008)	Aspects of agrammatic langage in Persian
Paul (2008a)	The individuating function of the Persian 'indefinite suffix'
Stilo (2008b)	Two sets of mobile verbal person markers in the Northern Talyshi language
Windfuhr (2009b)	*The Iranian languages*
Windfuhr (2009a)	Dialectology and Topics
Skjærvø (2009b)	Old Iranian
Skjærvø (2009a)	Middle West Iranian
McCarus (2009)	Kurdish
Jahani & Korn (2009)	Balochi
Paul (2009)	Zazaki
Yoshida (2009)	Sogdian
Windfuhr & Perry (2009)	Persian and Tajik
Maggi & Orsatti (2011a)	*The Persian language in history (Beiträge zur iranistik 34)*
Ciancaglini (2011)	The formation of the periphrastic verbs in Persian and neighbouring languages
Josephson (2011)	Definiteness and deixis in Middle Persian
Orsatti (2011)	The deictic suffix yā-ye ešārat: a hypothesis on the origin of the relative -i in Persian
Lazard (2011)	Homonymie et polysemie: beève note à propos des deux enclitiques -i du Persan
Durkin-Meisterernst (2011)	The importance of the Middle Persian thexts from Turfan
Baghbidi (2011)	New light on the Middle Persian-Chinese bilingual inscription from Xi'an
Provasi (2011)	New Persian texts in Manichaean script from Turfan
Filippone (2011)	The language of the Qor'ān-e Qods and its Sistanic dialectal background
MacKenzie (2011)	An index to "An Early Jewish-Persian argument"

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Maggi & Orsatti (2011b)	Two Syro-Persian Hymns for Palm Sunday and Maundy Thursday
Hasandust (2011a)	Etymological notes on some Classical New Persian words
Korn et al. (2011)	*Topics in Iranian Linguistics (Beiträge zur iranistik 34)*
Korn (2011)	Pronouns as Verbs, Verbs as Pronouns: Demonstratives and the Copula in Iranian
Gholami (2011)	Definite Articles in Bactrian
Sims-Williams (2011)	Differential Object Marking in Bactrian
Vydrin (2011)	Counterfactual Mood in Iranian
Paul (2011b)	A Glance at the Deixis of Nominal Demonstratives in Iranian Taleshi
Belyaev & Vydrin (2011)	Participle-Converbs in Iron Ossetic: Syntactic and Semantic Properties
Erschler & Volk (2011)	On Negation, Negative Concord, and Negative Imperitives in Digor Ossetic
Ganjavi (2011)	On Direct Objects in Persian: The Case of the Non-râ-Marked DOs
Pirooz (2011)	Finite Control in Persian
Wendtland (2011)	The Emergence and Development of the Sogdian Perfect
Naderi & Van Oostendorp (2011)	Reducing the Number of Farsi Epentheric Consonants
Deravi & Dommergues (2011)	Bilingual Speech of Highly Proficient Persian-French Speakers
Utas (2013b)	*From Old to New Persian: collected essays (Beiträge zur Iranistik 38)*
Utas (2013d)	The Grammatical Transition from Middle to New Persian
Utas (2013c)	Old Persian Miscellanea (Orientalia Suecana 14-15 (1965-1966))
Utas (2013e)	Traces of Evidentiality in Early New Persian (2000)
Utas (2013a)	A Multiethnic Origin of New Persian

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Ghomeshi et al. (2016)	*Further topics in Iranian linguistics: proceedings of the 5th International Conference on Iranian Linguistics, held in Bamberg on 24-26 August 2013*
Anonby & Mohebbi Bahmani (2016)	Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in south-west Iran
Comrie (2016)	Ergativity in Iranian languages: a typological perspective
Jeremiás (2016)	The history of grammatical ideas in Persian: kitābatan-lafzan in Classical Persian sources
Josephson (2016)	The expression of modality in Late Middle Persian
Karvovskaya (2016)	Comparative constructions in Ishkashimi
Miller (2016)	Theoretical issues in counting Persian words
Rasekh-Mahand & Izadifar (2016)	Compensating ergative alignment loss in Tākestāni
Rasekhi (2016)	Missing objects in Persian
Shirtz (2016)	Indirect participants as core arguments in Middle Persian
Shokri (2016)	Volition and obligation in four Caspian linguistic varieties
Klein et al. (2017)	*Handbook of comparative and historical Indo-European linguistics (Band 1)*
Jügel et al. (2017)	The syntax of Iranian
Sadovski (2017)	The lexicon of Iranian
Skjærvø (2017a)	The documentation of Iranian
Skjærvø (2017b)	The morphology of Iranian
Korn (2017b)	The evolution of Iranian
Huyse (2017)	The dialectology of Iranian
Cantera (2017)	The phonology of Iranian
Gholami (2018a)	*Endangered Iranian Languages*
Werner (2018)	Forms and Meanings of the Ezafe in Zazaki
Stilo (2018a)	Dikin Marāqei Tati of Alamut: an undocumented conservative Tati lanuguage
Gholami (2018b)	Pronominal clitics in Zoroastrian Dari (Behdīnī) of Kerman

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Dabir-Moghaddam (2018)	Non-Canonical Subject Construction in Endangered Iranian Languages: Further Investigation into the Debates on the Genisis of Ergativity
Gershevitch (1964)	Iranian Chronological Adverbs
Haig & Khan (2019)	*The Languages and Linguistics of Western Asian: an Areal Perspective*
Stilo (2018b)	The Caspian region and South Azerbaijan: Caspian and Tatic
Anonby & Taheri-Ardali (2019)	Bakhtiari
Anonby (2019)	Kumzari
Haig (2019b)	Northern Kurdish (Kurmanji)
Haig (2019c)	The Iranian languages of Northern Iraq
Mahmoudveysi & Bailey (2019)	Hewramî
Paul (2019)	Persian
Gündogdu et al. (2019)	*Current Issues in Kurdish Linguistics*
Öpengin (2019)	Accounting for the combinations of clitic and person markers in Central Kurdish
Haig (2019a)	Debonding of inflectional morphology in Kurdish and beyond
Anonby et al. (2019)	Kordistan Province in the Atlas of the Languages of Iran: Research process, language distribution, and language classification
Barry (2019)	Pharyngeals in Kurmanji Kurdish: A reanalysis of their source and status
Gündoğdu (2019)	Asymmetries in Kurmanji morphosyntax
Haig & Mustafa (2019)	Language choice and patterns of usage among Kurdish speakers of Duhok: An empirical intergenerational study
Herkenrath (2019)	Temporal noun squishes in Kurmanji academic writing: From lexicality via NP-level junction to clausal subordination

Table 1.2: Bound Volumes on Central and Southwestern Iranian Linguistics (continued)

Doostan & Daneshpazhouh (2019)	Kurdish -râ as an Anti-Actor marker
Matras (2019)	Revisiting Kurdish dialect geography: Findings from the Manchester Database
Belleli (2019)	Towards a dialectology of Southern Kurdish: Where to begin?

Larson et al. (2020)	*Advances in Iranian Linguistics*
Larson & Samiian (2020)	The Ezafe Construction Revisited
Jügel & Samvelian (2020)	Topic agreement, experiencer constructions, and the weight of clitics
Anonby et al. (2020a)	A multi-dimensional approach to classification of Iran' s languages
Jasbi (2020b)	The suffix that makes Persian nouns unique
Jasbi (2020a)	The meaning of the Persian object marker rā: What it is not, and what it (probably) is
Karimi & Smith (2020)	Another look at Persian rā: A single formal analysis of a multi-functional morpheme
Suleymanov (2020b)	Oblique marking and adpositional constructions in Tat: A mosaic of dialectal convergence and divergence
Abdollahnejad & Storoshenko (2020)	Syntactic and semantic constraints on pronoun and anaphor resolution in Persian
Ghomeshi (2020)	The additive particle in Persian: A case of morphological homophony between syntax and pragmatics
Haig (2020)	The pronoun-to-agreement cycle in Iranian: Subjects do, objects don' t
Mazdeh (2020)	Quantitative meter in Persian folk songs and pop lyrics
Rasekhi (2020)	Stripping structures with negation in Persian

Chapter 2

New Iranian Nominal Morphology

2.1 Introduction

Within the New Western Iranian languages, nominal morphology is characterized by an interaction between some or all of the following properties: case, number, gender, animacy, definiteness, and attribution (i.e. ezafe). This chapter provides an overview of the nominal systems of New Western Iranian languages[1] and the current state of diachronic and synchronic research into them. The New Western Iranian languages tend to compartmentalize their nominal systems so that no category is maximally distinguished for all other features. For example, Zazaki masculine singular nouns are marked for case (direct/oblique), animacy (+/-), definiteness (definite/indefinite/absolute), and modifier type (possessor/attribute). In contrast, plural nouns do not distinguish animacy, definiteness, modifier type, or even gender. There are several nominal marking strategies employed across these languages. Perhaps the most well-studied phenomenon in Iranian morphology is the ezafe or ad-nominal attribution marker (For an in-depth discussion of the New Persian Ezafe and its syntax and functions see Kahnemuyipour, 2014; Samiian, 1994; Franco-Rita & Savoia, 2012; Haig, 2011; Larson & Samiian, 2020; Larson & Samian, 2020, etc.). Second to this is the New Persian particle *=rā*. This particle occurs in Şirvan Tat (Suleymanov, 2020b) albeit with a slightly

[1]Here I use the term New Western Iranian to refer to both the branch of Central Iranian known as Northwestern Iranian and the separate Southwestern Iranian branch (not Central Iranian). These two branches do not form a coherent subgroup of Iranian. However, They do share a number of features such as ezafe marking that warrant their inclusion in this discussion. These features are not present in the Old Iranian period, and ezafe and definite marking were likely in their infancy in the Middle Iranian period. I neither address the issues of whether the features at issue here are the result of parallel independant innovation or borrowing nor the validity of our current understanding of the genealogical relationships between these languages. See appendix A for complete representation of the these two branches of the Iranian languages.

different distribution than New Persian, and it has reflexes in Caspian languages such as Gilaki and Tati as well as in Balochi. In New Persian, *=rā* is responsible for differential object marking[2], or object marking only with definite direct objects. Standard New Persian does not have a more generalized definite article. This is, perhaps, unsurprising from a diachronic standpoint as the relative pronoun *yat* which was the predecessor or the ezafe–at least as Old Persian *haya*[3]–acted as an article in the Old Iranian period (Kent, 1944, 1950); *yat* can be observed through all attested stages of Persian eventually becoming the ezafe in New Persian. Absent a definite article, standard Persian innovated a new one recruiting material from other parts of the lexicon (i.e. *-rā* < *rādiy*). This chapter examines various strategies of definiteness, case, and attribution marking in the Western Iranian languages. It represents an expansion of Karim (2021c) (forthcoming) which focused primarily on the attributive function of the ezafe and ignored most issues of case. This chapter forms the basis for the following chapters, which explore the problems that become apparent from the juxtaposition of these data. In this chapter, I devote more space to issues that I do not address more completely in subsequent chapters.

2.1.1 Background on the Iranian languages

The New Iranian languages have traditionally been categorized geographically, Northwestern (e.g. Zazaki), Northeastern (e.g. Yaghnobi), Southwestern (e.g. Persian), and Southeastern (e.g. Pashto). The geographic classification holds for Western Middle Iranian languages (Middle Persian and Parthian) but less so for Eastern Middle Iranian because of languages like Bactrian, which seem to have Eastern and Western features. This categorization does not work for Old Iranian, Old Persian, and Avestan, which had not developed the

[2]Differential object marking (DOM) is a feature of many Iranian languages. A range of factors influences whether a direct object will be marked, including definiteness, specificity, and animacy. This led Bashir (2008b) to coin the term "identified object marking" (IOM).

[3]Kent (1950) argues for a restricted use of *haya* (< PIE. *sa(s) and *(H)yod (Jügel, 2017, 561);(Kent, 1950, §197)) as a definite article in Old Persian (Kent, 1950, §260.IV). This function was not highly grammatical and shown to only have a fraction of the semantic range of the definite article in English.

East/West isoglosses that characterize the New Iranian languages. The geographical terminology has come into question due to shared features of non-Persian Iranian languages that seem to transcend the geographical categories. The incoherence of the traditional classification has led Korn (2016a) to propose the classification Central Iranian; see Korn (2017b); Huyse (2017) for more on the classification of these languages. Central Iranian consists of three branches that historically were thought to have belonged to distinct groups: Parthian, Bactrian, and Sogdian (Sogdian-Ossetian). For a more up-to-date family tree incorporating Central Iranian, see Hammarström et al. (2020) the basis for the tree in appendix A.

The Old and Middle Iranian languages cannot be understood as the direct ancestors of any New Iranian languages. The Old Iranian languages Avestan and Old Persian were probably contemporaries of languages like Proto-Kurdish, which have disappeared without attestation in manuscripts or inscriptions. Therefore, Old Iranian languages are at best approximations of Proto-Kurdish or Proto-Zaza-Gorani. For instance, both Avestan and Old Persian show constructions that resemble an early form of the ezafe (Jügel, 2017). Among New Iranian languages, the ezafe phenomenon is restricted to Western Iranian. Likewise, New Iranian languages like Kurdish show reflexes of PIE laryngeal consonants, which are missing from Old Iranian texts (Kümmel, 2014).

Within the current conception of Central Iranian, the Bactrian branch contains no currently spoken languages. The Sogdian-Ossetian branch contains no languages which bear the features at the core of this study, namely ezafe and definiteness marking. They do, however, have rich nominal inflection patterns. In Ossetian in particular, the noun has developed an inflectional system reminiscent of the neighboring Caucasian languages. The third branch of Central Iranian according to Hammarström et al. (2020) is Northwestern Iranian. This category directly corresponds to its namesake in the old purely-geographic nomenclature. Due to this fact, there are no theoretical issues at stake using this terminology. Consequently, I focus here on the interactions between case systems, ezafe marking, and definiteness in Northwestern Iranian languages. Additionally, I bring in data from stan-

dard and colloquial New Persian as well as Şirvan Tat, which belong to the Southwestern Iranian language family. What is not immediately clear from the branch names alone is that Southwestern Iranian and Northwestern Iranian do not constitute a particularly close grouping. They do, however, represent two groupings that have developed similar strategies of definiteness and attribution marking. New Persian has been selected as it is the language for which a disproportionate amount of the research into the ezafe phenomenon has been conducted; Şirvan Tat has been selected because it is genetically close to Persian but geographically situated in the heart of the Northwestern-Iranian zone. Its nominal system reflects its socio-linguistic situation, including its proximity to Azeri (Turkic).

The migration of speaker communities is widespread in the region. It is possible to glean much about the history of these migrations from the mark they have had on the languages. Most notably, Balochi fits into the Northwestern group and is spoken in the far east of the Iranian zone (Korn, 2003). It has developed features reflecting contact with Indic, Dravidian, Caspian, Kurdish, and other language groups (Korn, 2019a). The New Iranian languages referenced in this chapter are Northwestern: Goranî[4] (Hewramî), Zazaki (Southern and Central), Caspian (Gilaki), Southern Tati (Chali and Takestani), North Kurdish (Kurmancî), Central Kurdish (Soranî), Southern Kurdish (Kermanşay), Southwestern: Persian (Standard and Colloquial), Northern Tat (Şirvan), Balochi (Turkmen and Rakhshani).

2.2 Case systems

If we for a moment exclude ezafe (attribution marking), the case marking systems of Northwestern Iranian languages take many forms. Most of these languages have a bicasual (two-

[4]The term Gorani is generally used by linguists to refer to a group of closely related varieties, including Hewramî, Paweyane, Zerdeyane, Gewrecûî, Şebekî, the (classical) Gorani Koine, etc. The data considered in this study come from MacKenzie (1966) and Holmberg & Odden (2008); the former gives data from Hewramî (Luhonî), and the latter from Paweyane (the Goranî language from Pawe city). Although these varieties differ on several counts, their use of the ezafe and definiteness marking are comparable.

case) system consisting of a direct and an oblique case. The oblique case is thought to be the reflex of the original genitive case ending, while the direct seems to be the reflex of all other cases. The reason for this consolidation of the original case system was that between Old and Middle Iranian, word-final codas were lost (e.g. Av. nominative *daēna* and accusative *daēnam* both become MP. *dēn*). Assuming a case system like Avestan (7 cases plus the vocative), which was not necessarily found in all the languages considered here already by the Old Iranian period, we can infer a reduction like the system in table 2.1. Here, the Avestan nouns *yasna* 'sacrifice,' *daēna* 'faith,' and *dātar*ə 'creator, giver' represent the *a*-stem masculine nouns, *ā*-stem feminine nouns, and agent/kinship nouns, respectively. For the *a*-stem masculine and *ā*-stem feminine nouns, the majority of inflected forms consist of a stem and a suffix (simple or complex) with a -V(C) phonological shape. In these forms, the suffixes were lost "accidentally," or by the regular (physiological) process of sound change alone. This is different from the merger of cases due to semantic convergence observed already by Old Persian.[5]

Table 2.2 shows the case system resulting from the loss of word final codas assuming the Old-Persian-style merger of the instrumental with the ablative and the dative with the genitive. Effectively, this merger is only relevant for the a stem nouns which would have a genitive-else opposition in the singular and dative-ablative syncretism in the plural with the instrumental plural grouping with the nominative. Regardless of how one groups the the semantic mergers (like Old Persian, like Sogdian, or with no merger), the result is a system unlike any observed in New Iranian languages. However, the specific formatives involved do have reflexes across the greater Iranian world: Pashto *-una* [M.PL.DIR] (< *-aŋhō *a*-stem [NOM.PL.M]); Kurmancî, Hewramî, Pashto, etc. *-i* [M.SG.OBL], and Ossetic *-i* [GEN] (< *-eh

[5]It is not clear based on the general understanding of the New Iranian languages exactly which forms merged in their Old Iranian ancestors. For instance, Sims-Williams (1982) shows that the Middle Iranian language Sogdian shows a merger of the instrumental, dative, and genitive with their roles subsumed by the form of the genitive. This is clear in Sogdian because it has preserved much of the Old Iranian case system due to the "Rhythmic Law," where PIE weak stems preserved their codas and, by extension, the case system in a subset of the grammar.

	INS	VOC	NOM	LOC	ABL	DAT	GEN	ACC
Avestan (a) [SG]	yas.na	yas.na	yas.nō	yas.ne	yas.nāt	yas.nāi	yas.ne.he	yas.nəm
Avestan (a) [PL]	yas.nāiš	yas.nå.ŋhō/ yas.nå	yas.nå.ŋhō/ yas.nå	yas.naē.šu	yas.naē.ibyō	yas.naē.ibyō	yas.nā.nąm	yas.ną
Avestan (ā) [SG]	daē.na.ya	daē.ne	daē.na	daē.na.ya	daē.na.yāt	daē.na.yāi	daē.na.yå	daē.nąm
Avestan (ā) [PL]	daē.nābiš	daē.nå	daē.nå	daē.nā.hu	daē.nā.byō	daē.nā.byō	daē.nā.nąm	daē.nå
Avestan (r) [SG]	dā.θra	dā.tarə	dā.ta	dā.ta.ri	dā.θrat	dā.θre	dā.θro	dā.tārəm
Avestan (r) [PL]	—	—	dā.tā.rō	—	dā.tərə.byō	dā.tərə.byō	dā.θrām	dā.tārō

Table 2.1: The loss of final codas between Old and Middle Iranian

<table>
<tr><th></th><th>LOC</th><th>ACC</th><th>VOC</th><th>NOM</th><th>INS</th><th>ABL</th><th>DAT</th><th>GEN</th></tr>
<tr><td>Avestan (a) [SG]</td><td colspan="5" rowspan="2">yasn</td><td colspan="2">yasneh</td></tr>
<tr><td>Avestan (a) [PL]</td><td>yasneš</td><td colspan="3">yasneb</td><td colspan="2">yasnān</td></tr>
<tr><td>Avestan (ā) [SG]</td><td>daēnay</td><td colspan="4" rowspan="2">daēn</td><td colspan="3">daēnay</td></tr>
<tr><td>Avestan (ā) [PL]</td><td>daēnāh</td><td colspan="2">daēnāb</td><td>daēnān</td></tr>
<tr><td>Avestan (r) [SG]</td><td colspan="2">dātăr</td><td colspan="2">dāt</td><td colspan="3">dāθr</td></tr>
<tr><td>Avestan (r) [PL]</td><td>dātərəš</td><td colspan="3">dātăr</td><td colspan="2">dātərəb</td><td>dāθr</td></tr>
</table>

Table 2.2: Case syncretism due to accidental homophony with DAT-GEN and INST-ABL mergers

< *-ahya *a*-stem [GEN.SG.M]); Wakhi -ev [OBL.PL] and Pashto -*o* [OBL.PL] (< *-eb < *-byah *a*-stem [DAT/ABL.PL]); Zazaki, Kurmancî, Pashto -*e*, Hewramî -*e* [F.SG.OBL], and Ossetic -*æi* [INST/ABL] (< *-ay- *ā*-stem [INST/ABL/DAT/GEN.SG]); Zazaki *ma* 'mother [DIR.SG],' *pi* 'father [DIR.SG],' etc. (< *-t < *-tā *r*-stem [NOM.SG]); and Zazaki *mar* 'mother [OBL.SG],' *per* 'father [OBL.SG],' etc. (< *-tar < *-tāram *r*-stem [ACC.SG]). Note also that there are preserved inflected forms that have decoupled their function from the original significance of case: e.g. Makrani[6] Baluchi *māt* 'mother,' *pit* 'father' (< *-t < *-tā *r*-stem [NOM.SG]) and Rakhshani Balochi *mās* 'mother,' *piss* 'father' (< *-θra/*-s < *-θra *r*-stem [GEN.SG]), which dialectally show the reflexes of the former case endings decoupled from case. Additionally, New Persian shows the reflex of the *r*-stem accusative as the base form within an otherwise caseless system *madær* 'mother,' *pidær* 'father' (< *-tar < *-tāram *r*-stem [ACC.SG]).

The phonological reductions that occurred on the way from old to Middle Iranian have, for the most part, obscured our knowledge of what semantic mergers between case functions occurred before that point. This raises the question of what (case-)convergences apparent in the New Iranian languages are due to accidental homonymy (due to sound change), semantic mergers in the Old Iranian period, or semantic mergers in their near history. There are only two views of the semantic mergers that happened preceding the Middle Iranian period: one is from the shift from the Old Iranian Old Persian, which already began to show case-syncretism, to Middle Persian; the second is from the Middle Iranian Sogdian, which partially preserved a case system featuring phonological as well as semantically conditioned syncretism. However, these attestations only obscure the view, as they conflict in terms of semantic mergers. The situation is further complicated because Sogdian and Old Persian may not be the direct predecessors of any modern spoken varieties despite their closeness to the New Iranian languages Yaghnobi and New Persian, respectively. It is not strictly speaking possible to say that the mergers in today's languages have their roots in an Old Iranian shift.

[6]My attribution of the forms *māt* and *pit* to the Makrani dialect is based on Elfenbein (1961).

The view of case from the perspective of Middle Iranian to the present is equally problematic. Only two attested Middle Iranian languages belong to the traditionally recognized category of Western Iranian, which are relevant to this inquiry. They are Middle Persian and Parthian. Middle Persian is the predecessor of New Persian. However, it may not be the direct predecessor of any of today's spoken varieties.[7] Parthian, on the other hand, is the best representative of the Northwestern Iranian group. Although we now know that these languages are more distantly related than previously recognized (i.e. before Korn's (2016a) *a partial tree of Central Iranian*), they have converged along many features. In terms of case, they both had settled on a bicasual system by the time of the earliest attestation. This is illustrated in table 2.3[8], where the section on early texts shows a merger between the masculine and feminine declensions. The common gender declension features a bicasual system and two numbers with marked oblique plural and all other case and number combinations unmarked. The kinship nouns are further differentiated with unique oblique plural and direct singular forms and syncretism between the oblique singular and direct plural.

	Early texts				Later texts			
	M/F		fam.		M/F		fam.	
	DIR	OBL	DIR	OBL	DIR	OBL	DIR	OBL
SG	-Ø	-Ø	-Ø	-ar	-Ø	-Ø	-ar	-ar
PL	-Ø	-ān	-ar	-arān	-ān	-ān	-arān	-arān

Table 2.3: Parthian and Middle Persian declension

The later stages of the language show convergence on a caseless system similar to what is observed in New Persian. The oblique plural suffix becomes generalized plural, a marked form, unlike the singular, which is unmarked. These patterns represent all that is known about related languages in the Middle Iranian period, which reflect the systems in some but

[7]The assertion that Middle Persian may not be the direct ancestor of any spoken variety is based on the fact that there are constructions in New Persian varieties that may be the reflexes of Old Persian constructions that are conspicuously absent from Middle Persian. See section 2.3.2 for a possible example, the New Persian definite article *-(h)e*.

[8]Table 2.3 is based on on Skjærvø (2009a), 205.

certainly not all of the New Iranian languages. Additionally, there is evidence that the case systems of New Iranian languages have undergone innovative expansion after converging on a bicasual or even a caseless system.

2.2.1 Cases systems in New Iranian

The New Iranian languages have developed various systems from the maximally expanded Ossetian nine-case system (nominative, genitive, dative, allative, ablative, inessive, adessive, equative, and comitative) to truly caseless Southern Kurdish. Here, I focus only on the Northwestern Iranian languages (i.e. Central Iranian without Sogdic-Ossetic or Bactrian) and the more distantly related Southwestern Iranian as represented by New Persian and Şirvan Tat. As I show in section 2.4, the nominal system in some of these languages varies greatly between simplex NPs and complex NPs consisting of a noun modified by attributive adjectives of nominal possessors through an ezafe construction. Here, I provide a sketch of the systems of simplex nouns in these languages.

I begin this discussion with the caseless systems; first, it is important to define what constitutes a caseless system. For instance, standard New Persian is often considered to be caseless. However, there is an innovative definite direct object marker *=rā* (< Old Persian *rādiy* 'for (the sake of)'). This particle is an essential part of New Persian's system of IOM (DOM). Stilo (2008a) includes *ra*-marking as one strategy employed within the Iranian languages for case renewal. Stilo's (2008a) thesis states that after various forces whittled away at the Old Iranian case system resulting in caseless or bicasual systems, new case markers are recruited from existing material. I show in the following sections how in most Iranian languages, there are systems built upon other systems or that contain subsets of the previous systems pointing to the development of these languages. This is of special importance to my ongoing research into the Zaza-Gorani languages and my analysis of IOM in Kurmancî and Zazaki in chapter 4. In the interest of parsimony, I classify systems by the

most highly differentiated interpretation of the simplex noun data. The focus is on simplex nouns, although the case systems can vary when attributed.

caseless systems

The language included in this study that is genuinely caseless is Suleymanî Soranî. In this variety, there is a single form of the noun undistinguished for case (or gender) for all functions regardless of number or definiteness. This is illustrated by the forms in table 2.4, where the word *hewrê* 'friend' is unmarked for gender (can refer to biological males or females) and can be marked with the definite suffix *-eke* without affecting case; this is the form of the noun used in all functions. Systems like the one observed in Suleymanî Soranî are referred to by Stilo (2008a) as Persian-type, although Persian only has a system like this in non-definite and non-specific contexts.

	M	F	PL
DIR OBL	*hewrê(-eke)*		*hewrê(-ek)-an*

Table 2.4: Suleymanî Soranî case(lessness)

bicasual systems

The bicasual languages included in this section are Hewramî, Zazaki, Takestani Tati, Vafsi, Tafreš (CPD), Kurmancî, and New Persian. Focusing on the formatives employed and not on the functions of the specific cases, several patterns emerge. As a basis for examining these systems, I reference Stilo (2008a), which presents a typology of bicasual systems in Iranian languages. Stilo's (2008a) study is essential to this one for two reasons: it focuses on bicasual systems as a (default) result of the reductions that occurred between Old and Middle Iranian and shows how these systems have been elaborated upon due to subsequent innovations. This latter innovation is responsible in part for the layered case systems observed in some of these languages. On the "reduction axis," Stilo (2008a) identifies eight types of languages

according to their syncretism along the axes of case, number, and gender. Among the bicasual Iranian languages, some have maximally two genders (masculine and feminine), although some divide the masculine further into animate and inanimate. All have a way to express two numbers (singular and plural), and there are two cases referred to as direct and oblique. Stilo (2008a) further breaks down the polyfunctionality of the two cases based on the thematic roles that license them: subject, agent, object, recipient, possessor, experiencer, goal, temporal, and location. This more nuanced description is addressed in the discussion of functions of the cases (§.2.2.2).

Of the systems described by Stilo (2008a), the most fully differentiated is what he refers to as the Vafsi type. This is shown in table 2.5, where there is a unique ending for each paradigm cell. Gender is collapsed in the plural as a rule, which is true of all the languages that Stilo (2008a) considers and all the languages featured in the current study. They are displayed in the table in such a way that makes it look as if plurality is a feature in competition with gender. However, no theoretical implications should be gleaned from this; this syncretism is merely a fact of these languages.

	M	F	PL
DIR	*æsb-Ø*	*kærg-æ*	textitæsb-e/kærg-e
OBL	*æsb-i*	*kærg-é*	*ésb-án/kærg-án*

Table 2.5: Vafsi case (Stilo, 2008a, 703)

Stilo's (2008a) typology classifies these languages based on which cells display syncretism within the group. Stilo's (2008a) charts have been reproduced here (table 2.6), except that I have added shading for the languages which show syncretism in the direct plural and oblique singular. This type of syncretism is a notable feature of many Iranian languages (Arkadiev, 2007, 694). Arkadiev (2007) describes syncretism between oblique singular and direct plural as a "more 'exotic' pattern of syncretism." The Vafsi-type system is the most fully distinguished as there is a unique ending in each paradigm cell. From there, each of the languages collapses at least one set of distinctions. The Kajali-type system collapses gender

in the context of the oblique case. The Pakistani-type system collapses case distinctions in the feminine singular and the plural. Kurmancî collapses gender and number in the direct case, and in some dialects features syncretism between masculine singular oblique and direct (for more on Kurmancî dialectal variation, see the Kurmancî section below). The Khoini-type loses gender distinctions but maintains case and number robustly. The Northern Talyshi system loses gender distinctions across the board, and it loses case in the context of plural. The Alviri-type system loses case across the board. The Persian-type loses case and gender across the board.

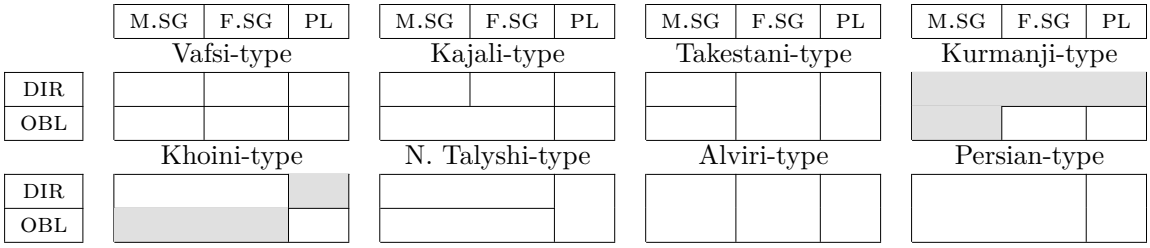


Table 2.6: Stilo's (2008a) typology of Iranian case systems

These systems present a fairly comprehensive view of the variety of bicasual systems in Iranian languages. However, several types are not included in this system. Furthermore, some systems occur as subsets of the systems described in Stilo (2008a) as part of what he calls the "innovation-axis." This is the phenomenon by which the languages referenced here have, after losing much of their case systems, sometimes including gender marking, innovated new case or case-like systems. In addition to repurposing morphology from more robustly inflected noun classes (e.g. kinship terms), these strategies including accusative markers, ezafat, agglutinative combination of existing morphemes, areal analogy (pattern borrowing), and adpositions (Stilo, 2008a, 706). Below I build upon Stilo's (2008a) typology adding to and updating table 2.6.

Hewramî Luhon In Hewramî Luhon, there are multiple declension classes. Nouns and adjectives can belong to one of three declension classes based on their termination;

consonant-final nouns and feminine nouns ending in *-i* belong to class I. Masculine nouns ending in *-æ* and feminine nouns ending in *-e* belong to class II,[9] and nouns and adjectives ending in a stressed *-a* belong two class III (MacKenzie, 1966, 14). As for how these systems pattern, classes I and III are alike, and class II features syncretism across a different axis. In table 2.7, I show the class I noun *pîr* 'old (person),' which can be masculine or feminine, and with the addition of the definite article *-ækæ* it is converted into a class II noun *pîrækæ* 'the old (person).' The Hewramî Luhon class I nouns have a nominal system maximally distinguished for case, number, and gender (Vafsi-type following Stilo, 2008a, 703). However, Hewramî Luhon class II nouns behave in a way that does not neatly fit into Stilo's (2008a) typology. Masculine class II nouns are maximally distinguished for case and number. On the other hand, feminine nouns collapse case in the singular and number in the direct. In other words, the is just one suffix that marks [DIR.SG.F], [OBL.SG.F], and [DIR.PL]. There is no equivalent in Stilo's (2008a) typology.

	M	F	PL	DEF.M	DEF.F	DEF.PL
DIR	*pîr-Ø*	*pîr-æ*	***pîr-ê***	*pîrækæ-Ø*	*pîræk-ê*	
OBL	*pîr-î*	***pîr-ê***	*pîr-a*	*pîrækæ-y*		*pîræk-a*

Table 2.7: Hewramî Luhon case

The system(s) observed in Hewramî build upon Stilo's (2008a) typology, adding a new type. Additionally, A pattern emerges from the Hewramî Luhon paradigms that appears all over the Iranian world and is not easily portrayed by the typological charts in Stilo (2008a). In particular, there is a syncretism across categories [OBL.SG(.F)] and [DIR.PL]. This pattern is observed in most languages that have preserved a case distinction in the plural, although this distinction is not always tied to the feminine gender (See Southern Zazaki below; cf. Arkadiev, 2007, 694, who states this pattern is "observed in many Indo-Iranian languages"). Another issue unrelated to the nominal formatives involved in Hewramî Luhon is that the oblique domain is defined differently depending on context. For instance, Hewramî Luhon

[9] According to Stilo (2008a), the Hewramî class II nouns may be the reflexes of Old Iranian nouns in *-i*.

nouns marked with the class II indefinite suffix *êw(æ)* do not take the oblique ending when acting as a direct object (MacKenzie, 1966, 16).

Southern Zazaki In Southern Zazaki, there are only really two declension classes, the general one and a special class that contains only kinship nouns (e.g. *ma* 'mother,' *pi* 'father,' *wa* 'sister,' etc.). However, in addition to these, inanimate or indefinite nouns collapse case, number, and gender (a unary system). This simplified system paradoxically creates a larger set of distinctions in ezafe-marked nouns (see §.2.4.1). In table 2.8, I show the animate noun *arwêş* 'rabbit,' which can be biologically (and grammatically) male or female in its unmarked (i.e. definite) and indefinite forms and the kinship nouns *bira* 'brother' and *wa* 'sister.' No kinship terms belong to this class that can be either male or female. Nonetheless, the forms are comparable, and as I show in the table (2.8), gender is not expressed by the declensional suffix in either the direct or the oblique. A case distinction is retained in both the singular and the plural, but no gender distinction is made. This is what Stilo (2008a) refers to as a Khoini-type system, although Khoini features the syncretism between [OBL.SG] and [DIR.PL], which is not featured in the Southern Zazaki kinship-noun declension.

	M	F	PL	INDF.M	INDF.F	INDF.PL
DIR	*arwêş-Ø*		***arwêş-i***	*arwêşê(n)-Ø*[10]		
OBL	***arwêş-i***	*arwêş(-er)-Ø*	*arwêş-an*			

	M	F	PL
DIR	*bira-/wa-Ø*		*bira-/wa-y*
OBL	*bira-/wa-r*		*bira-/wa-ran*

Table 2.8: Southern Zazaki case

[10] Todd (2002) shows the indefinite suffix *-ên* in use with a plural referent updating Hadank (1932) (e.g. (i))

(i) nê qeç-i qeç-ên-dê aqıl =î
these child-PL child-IND-EZ:/GEN clever =COP.PL
'These children a re some clever children.' (Todd, 2002, 42)

The basic declension in Southern Zazaki, like Hewramî class II nouns, does not fit well into Stilo's (2008a) typology. Looking at the masculine forms alone, there seems to be a system very much like the Khoini-type, which features the syncretism between [OBL.SG] and [DIR.PL], and unique forms for [DIR.SG] and [OBL.PL]. The feminine singular, however, complicates matters. Gender is collapsed in the direct singular, and case is collapsed in the feminine, and none of the feminine forms match what is observed in the plural. This is unlike any system observed in Stilo's (2008a) typology. Further complicating matters is the fact that there is an oblique marker that appears on feminine nouns *-er*. Its distribution is not fully understood (following Paul, 1998a). This issue is not unlike the classification of Hewramî indefinite nouns, which pattern like the class II nouns but do not include direct objects in the oblique domain. Is it not clear whether it is necessary to propose a tricasual system (direct, accusative, and oblique) for Hewramî with all nouns either being ACC = OBL or ACC = DIR. Similarly, with further research, the domain of the *er*-oblique in Zazaki may become clear. The precise number of paradigm cells should be determined by differentiation by function e.g. subject, agent, object, recipient, possessor, experiencer, goal, temporal, and location (following Stilo, 2008a).

Kurmancî (Northern Kurdish) As is true for all the languages in this study, Kurmancî is not monolithic. In this section, I include data from the nominal systems of three varieties of Kurmancî, Muş, Cizre-Botan (standard Kurmancî), and Amadî (Behdînî). Their inclusion is important because they show different patterns from each other. An additional issue in the Kurmancî nominal system is whether or not it has IOM (or DOM). Traditional grammars of the Cizre-Botan variety (e.g. Bedirxan & Lescot, 1986) do not include IOM as part of Kurmancî grammar. However, since the pioneering works of Celadet Bedirxan, standard Kurmancî grammar has been part descriptive and part prescriptive. According to Stilo (2008a), no variety of Kurmancî features DOM. However, an anonymous reviewer on Karim (2021c) provided me with examples (1) and (2). The first shows the expected

form based on standard descriptions of Kurmancî, which should, in principle, express both definite and generic readings. However, in the dialect referenced by the reviewer, it must be interpreted as definite. In contrast, the form in example (2) shows the direct object *pirtuk* in the direct case expressing the generic reading.

(1) Ez pirtuk-ê di-xwîn-im.
1SG.DIR book-F.SG.OBL
'I read/ am reading the book' (definite)

(2) Ez pirtuk di-xwîn-im.
1SG.DIR book-DIR
'I do book-reading/read books' (generic)

This juxtaposition reveals the generic form to be in violation of our expectations based on the standard dialect. It is paradoxically "marked" in the sense that it shows no marking when marking is expected. Kurmancî is known to be a canonically split-ergative language where intransitive subjects, present-tense agents, and past-tense objects take direct case marking (and control verbal agreement). In contrast, present-tense objects and past-tense agents take oblique marking. The question arises as to what form IOM would take in past tense constructions, where the unmarked direct form is expected. The result would be what is described in the Kurdish literature as a double-oblique construction (e.g., transitive alignment), where both the agent and object appear in the oblique. This pattern is widespread within Kurmancî, as observed in example (3). Here, the direct object *wan* 'them' appears for expected direct case *ew* 'they' without expected verbal agreement *kir-in* 'did [-3PL:O].'

(3) Gundi-yan wan bizor ji hev kir
villager-PL:OBL 3PL:OBL with.difficulty from each.other do:PST
'The villagers pulled them apart with difficulty' (Baksî, 1991, 31, apud Haig, 2004)

According to Dorleijn (1996), 58% of all past-tense transitive constructions occurred with oblique-marked direct objects, making it the majority pattern. This fact led her and sub-

sequent authors to conclude that the ergative pattern in Kurmancî is/was in decline in Kurmancî. However, these constructions are widespread and may even be reconstructible for a common Kurmancî (following Haig, 2008a). Furthermore, the percentages given by Dorleijn (1996) cannot demonstrate decline without evidence of a stage when the percentage of double-oblique constructions were fewer than they are today. To my knowledge, there has not been a study of this phenomenon as a type of IOM. It may even be likely that the peculiarity of double-oblique marking has distracted from Kurmancî's rich system of IOM (see ch. 4). In fact, Vafsi (Iranian: Caspian), which was the focus of much of Stilo's early work (e.g. Stilo, 1971) also shows double oblique marking. In his detailed study of the phenomenon, the agent's case is tense-sensitive direct in the present and oblique in the past; however, the case of the direct object is sensitive to IOM with specific and animate nouns occurring in the oblique case regardless of tense. Likewise, non-specific and inanimate nouns occur only in the direct case.

I am moving forward from the assumption that Kurmancî does have IOM, with the caveat that it is not well understood precisely how definiteness and specificity interact with animacy and the underlying alignment structure to produce the surface variants.[11] However, in the table of forms in standard Kurmancî and the Kurmancî of Amadî, I omit this aspect of the language as it is not a part of Bedirxan & Lescot (1986) or Blau's (1976) grammar.

The first aspect of Standard Kurmancî (table 2.9) nominal morphology that stands out is that number and gender are collapsed in the context of the direct case. This is true in the sense that they are zero-marked. However, the indefinite-plural marker *-(n)in* is different from the indefinite singular *-ek*. In the standard language, a three-way contrast is maintained between masculine-singular, feminine-singular, and plural in the context of the oblique case. However, the definite/generic form of the masculine singular oblique has

[11] I fundamentally reject the possibility that a particular construction can be "optional" in the sense that both occur without changing some nuance of meaning. Additionally, I reject the idea that native speakers are making mistakes that lead to the types complex distributions observed in many under-documented and endangered languages.

no ending and is therefore identical to the direct forms. It is generally assumed that the lack of an overt oblique marker is the result of a loss (e.g. Haig, 2008a, etc.). However, the masculine singular oblique marker *-î*, which always occurs in the variety spoken in Amadî (table 2.10), could in principle be reintroduced at any time from the form of the demonstrative circumposition; cf. *w-î heval=î* [that-M.SG.OBL friend=M.SG.OBL].[12]

	INDF.M	INDF.F	INDF.PL	(DEF.)M	(DEF.)F	(DEF.)PL
DIR	*hevalek-Ø*		*hevalin-Ø*	*heval-Ø*		
OBL	*hevalek-î*	*hevalek-ê*	*hevalin-an*		*hevalek-ê*	*heval-an*

Table 2.9: Standard Kurmancî case

This syncretism between the masculine singular oblique and masculine singular, feminine singular, and common plural direct is different from the system that Stilo (2008a) refers to as the Kurmancî-type system, which is reflected accurately in the definite/generic declension of the Kurmancî variety spoken in Amadî (Behdînî) (table 2.10). This system shows syncretism in the direct collapsing case and gender distinctions, robustly preserved in the oblique.

	INDF.M	INDF.F	INDF.PL	(DEF.)M	(DEF.)F	(DEF.)PL
DIR	*hevalek-Ø*		*hevalin-Ø*	*heval-Ø*		
OBL	*hevalek-î*	*hevalek-ê*	*hevalin-an*	*heval-î*	*hevalek-ê*	*heval-an*

Table 2.10: Amadî Kurmancî (Behdînî) case

In table 2.11, I show a hypothetical version of the Kurmancî variety spoken in Mu(s), where the system is maximally differentiated into generic, indefinite, and definite. This is by no means the result of a comprehensive study of the semantic and pragmatic conditions that license the oblique case. Instead, it is a hypothetical system that brings together the fact that there is some kind of DOM, albeit ill-understood. There are similar systems in closely related languages with a similar distribution. The first significant divergence between Muş and Standard Kurmancî is that it has lost most gender distinctions. There is a single zero-marked form for the generic reading regardless of case number or gender. The definite

[12]There are other varieties that use the strategy of ablaut to mark masculine singular oblique, which are not included here: cf. *heval* [friend.M.SG.DIR] ∼ *hevêl* [friend.M.SG.OBL].

forms collapse all distinctions in the direct case, and there is a distinction between singular and plural in the oblique. In the indefinite, the direct is zero-marked, although there is a different plural indefinite marker just as in standard Kurmancî. In the oblique, there can be a singular plural distinction in some cases; these presumably follow the rules of IOM with oblique marking only on specific indefinite direct objects; however, this is unconfirmed.

	M	F	PL	INDF.M	INDF.F	INDF.PL	DEF.M	DEF.F	DEF.PL
DIR	*heval-Ø*			*hevalek-Ø*		*hevalin-Ø*	*heval-Ø*		
OBL				*hevalek(î)*		*hevalin(a)*	*hevalek-ê*		*heval-a*

Table 2.11: Muş Kurmancî case

Muş Kurmancî provides evidence of yet another type that does not occur in Stilo's (2008a) typology, a collapse of all categories in the direct and a collapse of gender across the board. Some of the more speculative aspects of DOM in Kurmancî are addressed in ch. 4. What is important for this section is that there is some sort of DOM in Kurmancî. Whether or not it conforms to the type observed in other Iranian languages (i.e., IOM) remains to be seen. This phenomenon has been overlooked due to the resulting state, syncretism between direct-object marking and agent marking. This "optional" transitive alignment has proven to be more interesting to linguists than the semantic/pragmatic conditions that license it (e.g. Haig, 2004, 2008a; Dorleijn, 1996, etc.).

Standard New Persian Standard New Persian is a language that is known to have a caseless system. The bicasual system of early Middle Persian featuring morphological marking only on the oblique plural *-ān* gave way to a system where the old oblique plural was reanalyzed as a general plural. This system was retained in New Persian, where this form is preserved in highly animate nouns, and everything else was unmarked. An innovative collective suffix *-hā* was then recruited for plural. New Persian *-ha* was already available in Middle Persian as a plural suffix *-ihā* (Durkin-Meisterernst, 2014, 199). Sundermann proposed that this is a combination between the abstract suffix *-īh* and the plural suffix

ā(n) used as a collective marker (Sundermann, 1989c, 155). However, Sims-Williams (2004) has the Old Iranian abstract suffix *-iyaθwa, as the etmon. After case was lost, the enclitic *=rā* (< Old Persian *rādiy* 'for (the sake of)') was recruited to mark direct objects, first in addition to its use as an indirect object marker and later as a dedicated direct object marker. This development is part of what Stilo (2008a) refers to as the innovation axis, by which New Iranian languages have innovated case-marking systems anew. This axis is akin to what Friedman (1991) describes for Romani, where old designations are retained in currently spoken languages albeit through newly recruited material. In New Persian, the *=rā* marker is only used to mark definite direct objects.

	M	F	PL	DEF.M	DEF.F	DEF.PL
DIR	*ketāb-Ø*		*ketāb-hā*	*ketāb-Ø*		*ketāb-hā*
ACC				*ketāb-rā*		*ketāb-hā-rā*

Table 2.12: Standard New Persian case

Colloquial New Persian The same general facts can be stated about the development of case in colloquial Persian except that colloquial New Persian has retained[13] on the definite article. The result is that the standard New Persian forms in table 2.12 show syncretism between the definite and non-definite forms in the direct case. In contrast, colloquial New Persian (table 2.13) has a unique form for definite singular direct. In table 2.13, I show the more differentiated colloquial forms. The non-definite forms feature the system that Stilo (2008a) refers to as the Persian-type system. However, the system observed on definite nouns is like the Zazaki kinship terms collapsing gender and not case. However, the New Persian system has innovative accusative marking and cannot be equated with the oblique systems of other Iranian languages. The languages with oblique marking have polyfunctional oblique forms, while the colloquial New Persian forms are isofunctional.

[13]The claim that the definite article in colloquial New Persian is a retention was suggested in Karim (2021c) (forthcoming). It is supported by the fact that many closely related languages and dialects have such a marker. However, the article (suffix) is not observed in Standard New Persian or any extant Middle Persian texts. My thoughts on this are outlined in section 2.3.2.

	M	F	PL	DEF.M	DEF.F	DEF.PL
DIR	*ketāb-Ø*		*ketāb-ā*	*ketāb-e*		*ketāb-ā*
ACC				*ketāb-æ-ro*		*ketāb-ā-ro*

Table 2.13: Colloquial New Persian case

Turkmen Balochi

Syncretism across "collapsed" categories makes classification of the case systems presented here difficult. An additional layer of complexity is added when the inflected forms are examined from the perspective of polyfunctionality (see §.2.2.2 and Stilo, 2008a). The case system of Baluchi varies from subgroup to subgroup and between varieties within a subgroup. Here I show the system of Turkmen Baluchi, which belongs to the Western Branch, and features a five-case system (excluding the vocative[14]). Balochi's case system features a largely agglutinative paradigm. There is a core direct and oblique opposition seen in the other Western Iranian languages, and the rest of the case system is built upon that system. The oblique form becomes the host of the object marker *-rā* after vowels and *-ā* after consonants. Likewise, the genitive case is built from the oblique ending followed by the genitive suffix *-ī*. This suffix is thought to be descended from the original genitive suffix, which became the oblique marker in bicasual New Western Iranian languages albeit "de-bonded" from its original stem (following Haig, 2019a, §.3.1). The oblique II forms are an innovative feature of Turkmen Baluchi that separates it from Western Baluchi varieties. They are built upon the oblique form, with the genitive *-ī* and the objective *-(r)ā* suffixes. The oblique II cooccurs with certain prepositions as a rule and others sporadically. It is used on its own in certain locative functions prompting Axenov (2006) to label it as such.[15] Korn (2008a) proposes that this construction carries some "deictic force," and is similar to phrases like "I'm going to my uncle's" or the french *chez moi* (Korn, 2008a, 94, 96). I build

[14]I am excluding the vocative here because the vocative is extra-syntactic (i.e. discourse-generated) following Trubeckoj (1937, 43), Schenker (1964, 18), etc. In this regard, the forms of the vocative are not expressly relevant to the interaction of case, modification, and definiteness at the core of this chapter.

[15]There is some disagreement between scholars as to which of these count as case marking and what the cases should be called (Haig, 2019a, 126, following Jahani & Korn, 2009).

upon Korn's (2008a) theory concluding that the form of the locative in Turkmen Balochi is the result of nominal ellipsis, where the ellipsed nominal is a relational element in the Baluchi post-positional construction.

	M	F	PL	DEF.M	DEF.F	DEF.PL
DIR	-Ø			-Ø		
OBL				-ā		-ān
OBJ				-ā-rā		-ān-ā
GEN	-ay[16]		-ān-ī	-ay[16]		-ān-ī
OBL.II(LOC)	-ay-ā		-ān-ī-ā	-ay[16]-ā		-ān-ī-ā

Table 2.14: Turkmen Balochi case

The fact that Balochi's case system is agglutinatively structured upon a bicasual system like those observed in other Iranian languages is evidence for the more recent development of the more highly differentiated system. The original system would have been like what is observed in the Kurmancî (Northern Kurdish) variety of Muş, where gender is collapsed throughout the system, and number is collapsed in the direct case forms. The non-definite forms are then built upon what I refer to as the indefinite/inanimate type. Or as Axenov puts it, "The categories of definiteness, indefiniteness, and number are irrelevant for a generic noun" (Axenov, 2006, 62).

Koroshi Baluchi Koroshi Baluchi, spoken in Iran, appears to have innovated a definite article *-ok* from evaluative morphology (discussed in §.2.3.2) following Nourzaei (2020). Additionally, there is an innovative plural suffix *-obār* which has displaced the inherited plural marker *-ān* everywhere except in pronominal forms (Nourzaei et al., 2015; Nourzaei, 2020). This form has disrupted the nominal system of Koroshi Baluchi which now features plural marking throughout the system regardless of case or definiteness.[17] Additionally,

[16]The genitive singular suffix is a contracted form of the oblique singular *-ā* and the genitive suffix *-ī* *-ay*(Axenov, 2006, 72).

[17]Following Haig (2019a), there is as of yet no convincing etymology for the formative *-obār*. However, there are idiosyncracies of its distribution that point to a collective noun much like what is observed in Southern Kurdish *geł* 'flock' (Fattah, 2000), Shabaki *gał* 'idem.' (MacKenzie, 1999a), and Āštīāni: *gal* 'idem.' (Central Plateau) (Windfuhr, 1991). For instance, (1) the plural suffix *-obār* governs singular

Koroshi has collapsed the oblique and object cases together. I assume that the collapse is an innovation in Koroshi, because the distinction is otherwise widespread among Baluchi varieties, and the *-rā* object case form is still found in pronominal forms (e.g. *mārā* '[1PL.OBJ] (Nourzaei et al., 2015).

	M	F	PL	DEF.M	DEF.F	DEF.PL
DIR	*-Ø*		*-obār*	*(-ok)-Ø*		*obār*
OBL				*(-ok)-ā*		*-obār-ā*
OBJ						
GEN	*-ay*		*-obār-ay*	*(-ok)-ay*		*-obār-ay*

Table 2.15: Koroshi Balochi case

The resulting system in Koroshi Baluchi is like what Stilo (2008a) describes as the Persian-type system. Except for the genitive, gender and case are collapsed in the non-definite contexts. In the definite, gender is collapsed, but case is retained. This system is identical to what is observed in colloquial New Persian except for the addition of the genitive case. Colloquial New Persian employs the ezafe construction in places where Koroshi Baluchi uses the genitive.

2.2.2 a Typology of New Iranian case

Based on the forms presented here, it is clear that the issue of case in New Iranian languages is a bit more complicated than Stilo's (2008a) typology suggests. In addition to the eight case systems he presents, I have added nine patterns. These patterns maximally include two cases, two genders, and two numbers, although all of them collapse gender in the plural. I present these forms in table 2.16; I have reordered the table to reflect six cells maximally distinguished to a single form undifferentiated in descending order. Vafsi distinguishes feminine singular oblique (*kærg-é* 'chicken [OBL]') and direct plural (*kǽrg-e* 'chickens [DIR]') only by stress placement (Stilo, 2008a, 703, 2004, 263).

agreement; (2) the suffix is incompatible with definiteness marking like the colloquial New Persian *-ha*; (3) the suffix is inflected with singular endings for oblique/object and genitive. The origin of this formative is the subject of forthcoming work. One possible candidate for this formative is *bār 'load.'*

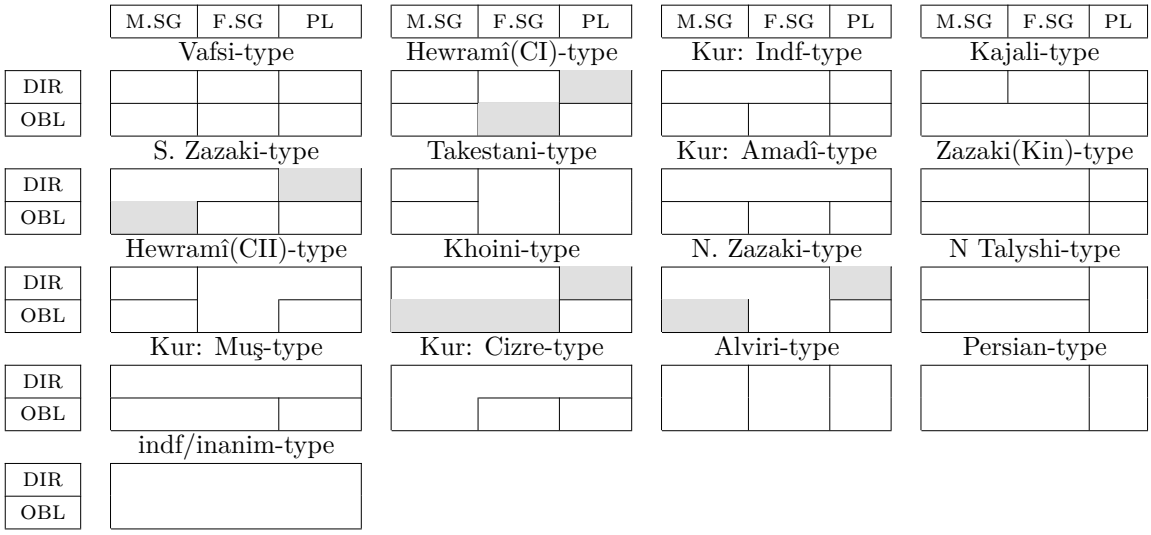


Table 2.16: Stilo's (2008a) typology of Iranian case systems (updated)

Perhaps the most significant addition to Stilo's (2008a) typology is the fact that any language can have multiple systems when definiteness is considered. There are several ways that a system restructures, assuming a Vafsi-type system as a starting point. The stress-based distinction between feminine singular oblique and direct plural can be lost. Gender or case can be lost categorically as well as other mergers and reductions that are not as well understood such as the loss of number distinction in the direct case as seen in Kurmancî and Baluchi. A collective suffix can enter the system reviving a singular-plural distinction that collapses a definite-non-definite distinction, as it does in Koroshi Balochi. A collective suffix can enter the system, creating a singular-plural distinction only in the indefinite, as occurs in all Kurmancî varieties. Additionally, cases can be innovated through the addition of adpositional elements such as the enclitic *-rā* marker for definite direct objects in New Persian and Baluchi and other languages throughout the Iranian world.

This typology looks at the number of cases and the extent to which they feature syncretism in the paradigm. I have only tangentially paid attention to functionality thus far. Stilo (2008a) explains that the bicasual systems of Iranian languages can be examined not just based on the case-marking terminations but their polyfunctionality: subject, agent,

object, recipient, possessor, experiencer, goal, temporal, and location. There is some question as to whether these categories are salient in the pan-Iranian context. For many of the languages, goals, and locations are all expressed through adpositional phrases. As for temporal phrases, they tend to either be expressed by adpositional phrases or in the manner of adverbials (oblique if available). Perhaps a better conceptualization of the typology of nominal systems within Iranian would look at categories subject (S), present-tense agent (A_{PRS}), present-tense object (O_{PRS}), past-tense agent (A_{PST}), past-tense object (O_{PST}), possessor (G), prepositional complement (PC), indirect object (IO), Adverbial (Adv). This system features the categories that are salient in most of the relevant languages. However, not all of these are diagnostic features. For instance, nominal subjects are always in the unmarked form (i.e. direct) in all of these languages, and adpositional constructions in many languages have replaced non-adpositional indirect object constructions.

Furthermore, the apposition of present-tense vs. past tense object (or even agent) makes a foundational assumption that the default is an ergative system. We know that diachronically speaking, there was a split ergative phase that all of these languages went through at some point in their history. However, there are differential object marking (or identified object marking following Bashir, 2008b) systems like the one Stilo (2008a) describes for Vafsi, where the direct object is marked by the oblique when definite or specific and direct in all other contexts regardless of tense (i.e. alignment). Comparison between these forms is hindered by the fact that when available, grammars of New Iranian languages often neglect to analyze nominal marking in the context of definiteness and attribution.

Other studies document very well the existence of IOM but fail in other ways. For instance Paul (1998b) shows the complex system of IOM and animacy effects but fails to explicitly describe the differential marking of indirect objects and adpositional complements. This gap is particularly interesting because much of what is observed for direct objects is equally true of indirect ones. A similar gap exists in Haig (2004) and 2008a which detail the evolution of ergative alignment in Iranian and Kurdish more specifically, while missing

DOM. He clearly shows that the strange transitive alignment (A = O) and A = S with verbs of perception and speech is ubiquitous in Kurmancî and may be reconstructible for a common Kurmancî. However, the question of what triggers the object to be in the oblique case was ignored. Dorleijn (1996), in writing about the loss of ergative alignment in Kurmancî, notes that DOM is at least partially responsible for the shift away from a purely ergative system. In both of these cases, the apparent strangeness of the system distracted from developing an understanding of that system. This issue only gets worse with less well-studied varieties. For instance, MacKenzie (1966) describes the Hewramî (Luhon) language. In his study, he mentions that impersonal agents can condition oblique agent marking (i.e. ergative alignment). This is illustrated in example (4), where the agent *tažnày* 'thirst' is in the oblique case and the typical subject clitic is missing. Later research has shown that it is topicalization that conditions the ergative alignment and not impersonal (Rasekh-Mahand & Naghshbandi, 2013). The sentence in example (4) must be understood as emphasizing the agent, not a marker of its impersonality. This is illustrated in example (5), which shows the agent *að-i* 'he [OBL].' Example (5) does not feature an impersonal agent, and it alternates with examples like (6), which show the unmarked construction, a direct agent with an agent clitic.

(4) tažnà-y kór-e kàrḑ-e
thirst-OBL.SG.M blind-DIR.PL make.PST-3PL
Hewramî (Luhon): 'they were blinded by thirst.' (MacKenzie, 1966, 51)

(5) sipâl-aka að-i šit-Ø
clothsDEF 3SG-OBL.M.SG wash.PST-3SG.M
Hewramî (Pawe): '<u>He</u> washed the cloths.' (Rasekh-Mahand & Naghshbandi, 2013)

(6) âð-Ø sipâl-aka=š šit-Ø
3SG-DIR.M.SG clothsDEF=3SG:A wash.PST-3SG.M
Hewramî (Pawe): 'He washed the cloths.' (Rasekh-Mahand & Naghshbandi, 2013)

A complete typological account of the polyfunctionality of cases in the nominal systems of

Iranian languages is likely to be a career's worth of work for several scholars. It is a task that will require better documentation and better linguistic analysis of those documented languages. Here, I deal with each of these languages on an individual basis as case-marking, definiteness, and attribution interact. Ideally, the issue of function will be revisited when more is know about the relevant languages.

2.3 Definiteness marking strategies

There are various strategies employed for marking definiteness and attribution in the Western Iranian languages. In these languages, definiteness can not be seen as a binary category consisting of definite and indefinite. Rather, they have a tripartite distinction of definite, indefinite, and absolute (Thackston, 2006b; Rastorgueva et al., 2012). The categories of definite and indefinite generally behave as they do in other languages, but the absolute has a general sense or is ambiguous. The definite is employed when the referent is either contextually defined or if it has previously been mentioned. However, new research by Haig & Mohammadirad (2019); Haig (2019a) and Nourzaei (2020) show that the category definiteness may be more complicated than previously understood. It is theorized that the strange distribution of definiteness marking in Kurdish and Balochi may have its roots in the etymological source of the definiteness marker (i.e. from "evaluative morphology" following Nourzaei, 2020, see §.2.3.2 for more on this argument). The indefinite can have a specific interpretation (i.e. "a certain person") as well as an indefinite interpretation. The absolute form of the noun has a more complex distribution. It can contextually be understood as definite or indefinite/generic. Compare the Kurmancî examples (7), and (8).

(7) vêrê hirç-Ø nî-ne
here bear-DIR NEG.COP-3PL
Kurmancî: "there are no bears here (generic)" (Haig & Öpengin, 2018, 16)

(8) hirç-Ø hat-Ø
bear-DIR come.PST-3SG

Kurmancî: 'the bear came' (Haig & Öpengin, 2018, 16)

All the Western Iranian languages examined here have, minimally, an overt marker for the indefinite. However, the distinction between absolute and definite is not always overtly marked. Several of these languages have an overt definite marker, but others still distinguish definiteness only by the occurrence of case endings (e.g. Zazaki). I have outlined these strategies in §.2.3.1 (indefinite marking) and §.2.3.2 (definite marking).

2.3.1 The indefinite markers

There are two types of indefinite markers in the North-Western Iranian languages: YAK (< *aika) and EW[18] (< *aiwa). The languages which have the indefinite article YAK are Soranî (N-*êk*), Kurmancî (N-*ek*) and Central Zazaki (N-*yek*). Standard New Persian (N-*i*), Hewramî (N-*êw(æ)*), Southern Zazaki (N-*ê(n)*), Gilaki (*i* NP) and Chali Tati (*i* NP-(*i*)) have the indefinite article EW. Şirvan Tat (ye NP-*i*) and colloquial New Persian ((*ye*) NP(-*i*)) have developed a hybrid system consisting of both YAK and EW.[19]

The numeral YAK is used in Soranî, Kurmancî, New Persian (standard and colloquial), Hewramî, and Şirvan Tat. The rest of the languages have the numeral EW (e.g. Zazaki (Southern and Central), Gilaki and Chali Tati). The systems of these languages are presented in table 2.17. The two Kurdish languages Soranî and Kurmancî have YAK-type

		So.	Ku.	ŞT.	NP. (col.)	He.	NP.	SZ.	Gi.	CT.	CZ.
YAK	Article	✓	✓	✓	(✓)						✓
	Numeral	✓	✓	✓	✓	✓	✓				
EW	Article			✓	✓	✓	✓	✓	✓	✓	
	Numeral							✓	✓	✓	✓

Table 2.17: The relationship between the numeral "one" and the indefinite article

[18] I use small caps here to signify the numeral and article variants YAK and EW regardless of their realization in the individual languages. This is in contrast with YAK-type, a system with both the numeral and article from *aike, EW-type, both the numeral and article from *aiwa and YAK/EW-type, with the numeral from *aika and the article from *aiwa.

[19] In Şirvan Tat, the indefinite construction without the preposed element exists but seems to be marginal (Suleymanov, 2020a); in colloquial New Persian the distribution is more complicated.

systems; the languages Southern Zazaki, Gilaki, and Chali Tati have EW-type systems and New Persian (standard and colloquial), Hewramî and Şirvan Tat have YAK/EW-type systems, or a system with the numeral YAK and indefinite article EW; Şirvan Tat and colloquial New Persian have a bound form of YAK which accompanies the indefinite article EW. The opposite, a system with a numeral EW and an indefinite article YAK, is only found in Central Zazaki (Werner, 2018). This is odd if it is assumed that both the article EW and the number EW are the inherited forms in Zazaki. In contact with Kurmancî, Central Zazaki borrowed the article YAK (Werner, 2018). One might expect the number to be borrowed from neighboring Kurmancî and to be recruited as an indefinite article assuming that it is easier to borrow words than morphemes. However, the inherited numeral EW has been retained but with the form *yew*.[20]

Unfortunately, the attested Old Iranian languages do not aid in the identification of what is innovation and what is retention. In Old Persian, cardinal numerals are not well attested with the exception of *aiva-* 'one' in the phrase *aivam paruvnām X* "one of many Xs" (Kent, 1950, §204, §250). In all other instances, the cardinal number "one" is expressed not phonetically but by the orthographic numeral. For this reason, it is unclear if there was an *aika variant. Such a variant does not occur in Avestan either. Because of this lack of clear attestation, it is impossible to say for certain if the *aika variant existed in the Old Iranian period and if it did whether or not there was a semantic distinction. Both *aika and *aiwa are reconstructed for Proto-Indo-European. Both are attested on the Aryan branch of the Indo-Iranian language family, as represented by Sanskrit *eka* "one" and *eva* "only," "even," etc.

In Middle West Iranian, the picture is closer to what can be observed today. Parthian (Northwestern) does not show any reflex of the numeral YAK (Durkin-Meisterernst, 2004, 373), and it always uses EW as the cardinal number "one" (Skjærvø, 2009a, 211). This

[20]Other Zazaki varieties have *jew* "one". Based on available data (Paul, 1998b; Werner, 2018), whether or not the *y : j* correspondence is a regular one is unclear. If it is irregular, one possible explanation could be contamination from Kurmancî *yek*.

EW-type system is observed in the New Northwestern Iranian languages Southern Zazaki, Gilaki, and Chali Tati. In contrast, Middle Persian (Southwestern) has both the numeral YAK and the indefinite article EW, which is the YAK/EW-type system as observed in the New Southwestern Iranian languages standard New Persian (southwestern) and Hewramî, which is typically thought of as Northwestern (Durkin-Meisterernst, 2004, 100, 373). In fact, the function of the Middle Persian indefinite marker *-ew* is to mark a specific indefinite noun just as in New Persian (Skjærvø, 2009a, 211, 205).

It is not currently possible to decide whether the existence of a split YAK/EW-type system in New Persian, Hewramî and Şirvan Tat (and Middle Persian) shows these languages to have preserved inherited reflexes of the two lexemes. Another possibility is that the YAK variant was borrowed into these languages from a related language between Old and Middle Iranian. In Kurdish, both the numeral and the indefinite article are from YAK. This fact taken alone suggests that the YAK variant is reconstructible for Proto-Kurdish. The distribution of systems suggests that the YAK-type is the innovative form as it only occurs in Kurdish (see §.2.5.2 for a more detailed discussion of this possibility). The generally accepted sequence of events was that the numeral became the article and not the reverse. Assuming that in pre-Proto-Kurdish, there was a system that better reflected old Iranian with both EW and YAK variants, the numeral YAK may have become an indefinite article filling the gap created by the loss/repurposing of the indefinite article EW as the indefinite ezafe.[21] The fact that the only language to feature the numeral EW and the article YAK (Central Zazaki) did so through direct morphological borrowing of the article YAK into a system that preserved the numeral EW suggests that the numeral can be recruited as an indefinite article. However, an article is unlikely to break away and become a numeral.

[21] This was suggested in Karim (2021c) (forthcoming) based on the distribution of the Central Kurdish ezafe in the context of what is observed in New Persian and Northern Kurdish. However, my most current thinking on the subject is that its "indefinite" properties are epiphenomenal of certain diachronic retentions that set it in contrast to the more innovative Northern Kurdish. The claim that Northern Kurdish is innovative disagrees with Mackenzie (1961) and others, who have long regarded Northern Kurdish as more conservative due to the retention of categories such as gender and case. This is discussed in detail in section 2.4.1 chapter 4.

2.3.2 The definite markers

There is no clear etymological source for many of the various definite markers in the Northwestern Iranian languages. This is likely because there is no record of most of these languages predating the modern era. Several definiteness marking strategies are employed: (1) the marker RĀ (New Persian (standard and colloquial) and Şirvan Tat); (2) case marking (Zazaki and Chali Tati); (3) the article *-(h)e* (New Persian (colloquial); and (4) the article *-eke* (Soranî and Hewramî) and *-ok* in (Koroshi Baluchi). The historical origins of these particles are not a trivial matter. Their etyma are particularly important for understanding the idiosyncrasies of the way they combine with attributes.

The marker rā

The marker RĀ is employed in several languages to mark a definite direct object. It has a clear etymological path (in Persian) from Old Persian *rādiy* "on account of)" to Middle Persian *ray* "for/because," apparently by regular sound change, on to its various forms in New Iranian languages (Korn, 2017b, 613). Additionally, in languages which employ the marker RĀ and an overt definite article (e.g. colloquial New Persian), they can occur together (see ex. (9)).

(9) ketab-æ-ro xund-æm
book-DEF-DEF.DO[22] read.PST-1SG
NP. (col.) "I read the book."

The same is true of many Baluchi varieties, which feature *-ra*-marking in addition to the underlying, read inherited, deferential-case-marking system. Example (10) shows just this stacking, where the nominal object KOHĀRĀ 'mountain' shows both the oblique suffix *-ā* which only occurs on definite/specific oblique nouns and the objective suffix *-(r)a*. The differential case marking strategy is described more fully below.

[22] All glosses have been altered from their original publications to reflect the features discussed in this study and to match Leipzig glossing rules when possible.

(10) man am-ā koh-ā-rā [=a] gınd-in
I EMPH-that mountain-OBL.SG-OBJ [=IPFV] see.PRS-1SG:A
R Balochi: 'I see that very mountain.' (Barker & Mengal, 2014, 142)

Case marking alone

Another strategy is to use differential case marking to mark definiteness. For instance, Zazaki employs oblique to mark definite direct objects and not indefinite/generic ones (Paul, 1998b, 21). Compare examples (11) and (12) where the direct object surfaces as *mırıçıkı* with no oblique marker when used in a generic sense and *mırıçıkerı* with the feminine singular oblique marker *-er(i)* when definite.

(11) seni kes mırıçıkı tepş-en-o
when person bird catch.PRS-IPFV-3SG.M
SZ. "when one catches a bird" (Berz & Malmîsanij (1951), 113.22 apud Paul (1998b))

(12) mırıçık-erı xo mıyan=dı nımn-en-ê
bird-F.SG.OBL self middle=LOC hide.PRS-IPFV-3PL
SZ. "(they) hide the bird among themselves" (Berz & Malmîsanij (1951), 198.4 apud Paul (1998b))

Some Kurmancî dialects also show the pattern where the presence of case marking (when applicable) only occurs on definite nouns; see examples (13) and (14) which differ only by the oblique marker on the direct object *pirtûkê* which marks it as definite. In the direct case where there is no distinction, there is ambiguity about whether the noun is definite or indefinite.

(13) Ez pirtûk di-xwîn-im.
1SG.DIR book IPFV-read.PRS-1SG
Ku. "I do read books (generic)"

(14) Ez pirtûk-ê di-xwîn-im.
1SG.DIR book-M.SG.OBL IPFV-read.PRS-1SG

Ku. “I read/am reading the book” (definite)

It was my first instinct to dismiss this pattern as it does not reflect “standard Kurmancî.” However, standard Kurmancî is plagued by prescriptions, and existing grammars have various flaws. For instance, Aygen (2007) “contains numerous factual and analytical errors” and Thackston’s (2006a) “undue reliance on the model of Persian has led to some distortions” (Haig & Öpengin, 2014, 113). In light of the acknowledgment that there are numerous errors in Kurmancî grammars and Dorleijn’s (1996) observation that differential object marking is at least responsible for the phenomenon known as double oblique marking (transitive alignment), a reevaluation of the Kurmancî data is warranted. This is especially true as the phenomenon in question and the associated disruption of ergativity are pervasive across Kurmancî varieties and have been the focus of many studies (e.g. Dorleijn, 1996; Gündoğdu, 2017b; Haig, 2004, 2008a), none of which specifically target the question of DOM. I take up the issue of DOM in chapter 4 as part of a deeper discussion of the source of differential object marking in Kurmancî.

Colloquial New Persian *-(h)e*

The colloquial New Persian article *-(h)e* is perhaps the most puzzling as it was not attested before the modern period. Its form is similar to the ezafe except that it is stress bearing. The ezafe is thought to have descended from the Old Iranian relative pronoun (Avestan: *yat*), perhaps acting as an article. However, this cannot be the etymon of *-(h)e*, which in addition to being stressed occurs with an epenthetic *h* when it follows a vowel. In contrast, the ezafe employs an epenthetic glide *y*. The fact that an *h* is used to break hiatus implies an etymological (i.e. not a purely phonetic) source. In Persian, the strategy employed to resolve vowel hiatus is the insertion of a glide *-y-*. This is exemplified by ezafe after vowel final nouns and by the personal endings after vowel final verb stems: *xune-ye* and *miy-a-yæm* not *xune-e* and *miya-æm* (there is a colloquial contracted form of *miy-a-yæm*

that is realized as *miy-a-m* but, crucially, not *miy-a-hæm*). However, there is a possible analogical source for the epenthetic *h* a hiatus resolution strategy. The New Persian plural ending *-(h)a* occurs with the *h* after vowel-final stems and without the *h* after consonant-final stems (e.g. *xune-ha* 'houses' but *ketab-a* 'books'). Note that on the surface there is no demonstrable difference between an etymological consonant that is lost after a consonant and a consonant that is epenthesized intervocalically. This etymological difference is only learned later when speakers become literate; the *h* is always prescriptively written in the plural suffix <hā>. Literacy need not affect a speakers ability to analogically extend the *h* in *-ha* to *-é*.

A possible etymon for this article is the Old Iranian demonstrative pronoun *ha* (< PIIr. *sa(s)). This would account for the phonology and the semantics.[23] It cannot, however, account for the absence of such a marker in historical texts. In Old Persian, the demonstrative *ha-* and the relative pronoun *ya-* only occur in the univerbated form *ha-ya-* which was employed as both a definite article and relative pronoun (Kent, 1944). This may suggest a hybrid solution; i.e. the ezafe *-e* from Old Iranian *ya-* and the definite article *-(h)é* from Old Iranian *(ha)ya-* (Karim, 2021c, forthcoming). The proposal of this etymon resolves the issue of the /h/. However, as noted in the previous paragraph, the /h/ may have a valid etymological source, e.g. analogy to the plural marker *-(h)a*. Jahani (2015) has suggested that the definite suffix *-(h)e* is the reflex of the diminutive suffix Middle Persian: *-ak/-ag*, cf. New Persian *bačče* 'child' where the final /e/ is a diminutive suffix. Of course, this form never occurs after a non-etymological /h/. However, this solution has the benefit of offering a uniform etymon for the definite markers in Central and Southern Kurdish, Gorani (next §), and beyond.

Another feature of the colloquial New Persian definite article that is noteworthy is the fact that it is incompatible with plural marking. The form *ketab* 'book' alternates with

[23] The tendency for demonstratives to become definite articles can be observed in many languages, including English: *the* and Syriac: *hu*. This is "the most frequent way in which definite articles evolve" (Heine & Kuteva, 2002, 110).

ketab-i 'a book' and *ketab-e* 'the book,' while in the plural, there are only *ketab-a* 'books, some books, the books' and *ketab-a-i* 'some (particular) books. However, this pattern is more likely to result from the innovative plural marker *-(h)a* descending from a collective suffix that would have been incompatible with definiteness marking. This would be further supported if it could be shown that Old Persian *haya* was indeed the source of colloquial New Persian *-he*, which would have had a different form in the plural (i.e. from OP *taya*).

It is widely accepted that the New Persian plural marker has descended from a collective suffix Middle Persian *ihā* (Durkin-Meisterernst, 2014, following). However, there are certain aspects of this etymology that are less than satisfactory. For instance, Durkin-Meisterernst (2014) states that this morpheme likely consisted of the abstract noun suffix *-īh* and the plural suffix *-ān*. The shortening and eventual loss of the vowel *ī* is unexplained, along with the loss of the final *n*. The final *n* of this formative is lost in various contexts in varieties spoken across the Iranian world (e.g. Northern Zazaki, Rakhshani Balochi, Eastern Kurmancî, etc.), including some colloquial Persian dialects. However, it does not happen in Standard New Persian, and the *-ha* suffix occurs in all dialects. Regularity of sound change suggests that this etymology should be amended.

Assuming that the plural morpheme is indeed from a collective noun-forming suffix, it can be said that the definite suffix does not cause the clash between plural and definite. A similar phenomenon is observed in Koroshi Balochi (Nourzaei, 2020) where the definite suffix *-ok* is not compatible with the plural suffix *-obār*, which is innovative in Koroshi Balochi and exhibits signs of being derived from a collective suffix.

Soranî *-eke* and Koroshi *-ok*

The article *-eke* found in Soranî and Hewramî is also problematic. It has been suggested that this definite article has its origin in the diminutive suffix *-ek*.[24] Another possibility is

[24]The diminutive source theory was proposed as early as Soane (1913). His assertion appears to be based on superficial similarity and a fundamental misunderstanding of the suffix *-eke*, which he does not describe as a definite article at all. Mackenzie (1961) has diminutive markers *-çe, -çke, -ek, -ik, -ke, -eke, -ōk, -ōke,*

that of a hybrid origin. Haig & Mohammadirad (2019) proposed that the diminutive marker *-ek* might in fact be augmented by the demonstrative circumposition *=e*; compare (Soranî) *ew=e* "that" and *ew ktêw=e* "that book". One merit of this explanation is that in addition to the definite article *-eke*, the article *-e* occurs without the *-ek*. The demonstrative clitic *-e* is known as a marker of definiteness in the demonstrative construction (perhaps akin to double determination in Greek). However, it was not described as an article independent of the demonstrative in any Soranî grammar (e.g. Thackston (2006b); McCarus (1956); Mackenzie (1961)) until Öpengin (2016), a study of the Mukrî dialect (cf. (15)).

(15) kuř-e name=yî bird
boy-DEF[25] letter=3SG take.PST
Mukrî: "The boy took the letter." (Öpengin, 2016, 60)

The plausibility of the diminutive origin has largely gone unchallenged in the literature. This issue is now significantly less likely to be challenged now that the Koroshi Balochi definite article *-ok* has been shown to have developed from "evaluative" morphology. This formative certainly bears an affinity to the diminutive marker and the definite article in Kurdish (Nourzaei et al., 2015; Nourzaei, 2020). Additionally, there is a precedent for a diminutive becoming a referential marker similar to a definite article (Pakendorf & Krivoshapkina, 2014). One potential problem with using the Ėven data from Pakendorf & Krivoshapkina (2014) is that the diminutive marker in Ėven is actually an evaluative marker implying some kind of specification (though this is often diminutive). Moreover, "[t]hey are

-kele, -le, -leke, -île, -ûle, -ûlke, -ɬe, -eɬe, -ōɬe, and *-ōɬke*. Each of these is highly lexicalized in its use, and it is not necessarily clear that these are acting as diminutives synchronically or diachronically. For instance, Mackenzie (1961) gives *pûɬeke* "fish scale" and *tûreke* "small bag" as examples of the diminutive suffix *-eke*. However, pûɬ "coin, pittance" from which *pûɬeke* is derived already has a small connotation; the suffix *-eke* narrows the meaning to a specific kind of small round thing. The example *tûreke* "small bag" (Mackenzie, 1961) is not clearly smaller that a *tûr* "small bag." However, this suffix appears to be used derivationally in this form. Öpengin (2016) has *tûreke* "a special type of bag" which conforms more to the definite interpretation. However, he categorizes this *-eke* as separate from the definite article albeit homophonous (Öpengin, 2016, 54). This is justified by the fact that these nouns can take the definite article (e.g. *tûrekeke* "the bag").

[25] Öpengin (2016) glosses this as DEM.CL acknowledging its affinity to the postpositional element of the demonstratives.

optional, and the speaker's intentions play a large role in whether or not to use them (Pakendorf & Krivoshapkina, 2014, 328). Of course, it is challenging to establish the function of these formatives in Middle Iranian texts, and Middle Persian features a large amount of its vocabulary with doublets marked with the "diminutive suffix" (p.c. with Geoffrey Haig). It appears that we may be able to dismiss our understanding of these formatives in Old Indo-Iranian following Whitney (1993) that the Indo-Iranian diminutive marker *-(a)ka-* attached to a variety of bases to create adjectives of appurtenance, diminutives, and to impart no discernible meaning (Whitney, 1993, §1222). Rather the evaluative function of Balochi may have deeper roots. This discussion has been evaluated synchronically in detail for Koroshi Balochi (Nourzaei, 2020). She shows how the function of the *-ok* suffix in some Balochi varieties is reminiscent of what Pakendorf & Krivoshapkina (2014) describes for Ėven evaluative morphology. In the closely related varieties Coastal and Sistani Balochi, the *-ok* suffix does not occur with the "anaphoric definite" function (Nourzaei, 2020, 32); this fact points to the shift from evaluative to definite being a more recent innovation after a common Balochi.

Of the three functions of the *-ok* suffix (1) evaluative usage, (2) deictic and recognitional contexts, and (3) anaphoric definite, only the first is compatible with the plural marker *-obar*. Additionally, Nourzaei2020 assumes that "[the suffixes in question *-ok/ak/ek/lok/o*] are all reflexes of a middle-western Iranian suffix involving a final-K, but with differing vowel values according to the nature of the nominal stem to which it attached in Middle Iranian" (Nourzaei, 2020, 6). Despite this claim, Nourzaei (2020) does not show any examples of formatives other than *-ok* occurring in a definite context. There is no explanation as to why one diminutive/evaluative suffix would be chosen over the others if definite *-ok* is to be understood as the same as the evaluative/diminutive *-ok, -ak, -ek, -lok,* and *-o*. This issue may be equally problematic for such a solution in Kurdish.

The features indicative of the development of a definite marker from evaluative morphology are not necessarily features of the Soranî definite suffix. Although, some of the

functional aspects of the Soranî forms are similar to the Balochi ones. There are several other problems with/open questions about the argument for a diminutive origin of the Soranî definite article *-eke*. Of these concerns, there are several directly relevant to this inquiry:

Why does this article condition the definite form of the ezafe (see §.2.4.1)?

One aspect of the combinatorics of definite forms is that they condition what I call the definite ezafe *-e*.[26] If the definite ezafe is seen as a compounding marker (following Samvelian, 2005), we may assume that the compound nominal can be inflected like any other nominal, thereby receiving the bound morphology *-eke* originally from a diminutive. However, from the perspective that the ezafe is a (morphologically-marked) syntactic phenomenon, one must explain why the nominal morphology is carried by the modifier and not the head noun. This process, known as "debonding" (following Haig & Mohammadirad, 2019), must already be proposed for independent reasons (e.g. in coordination). This points to a separate syntactic entity as the etymon. Soranî *-eke* is a definite article and has a more clitic like attachment; compare *kuř-eke* "the boy", *kuř-e baş-eke* "the good boy", *kuř-e baş-e drêj-eke* "the good tall boy," *kuř-e şwan-eke* "the shepherd's boy" How can the shift from derivational affix to inflectional clitic be explained?[27] In coordination, the same issue occurs: *kiç û kuř-ek-an* 'the boy(s) and girl(s)' shows a plurality, where only the second member is marked for definiteness or plurality. The total in the group is more than one, hence plural, and they are known from context, hence definite.

An additional problem is that in Hewramî, where both the definite article *ækǽ* and the

[26] In section 2.4.1, I discuss the relevant formative, and in chapter 3, I give an analysis of these forms; I propose that the definite ezafe is derivational morphology that converts the noun into an entity looking for a modifier on its right to form a noun phrase.

[27] The example *kuř-e baş=eke* 'boy[-EZ-DEF] good[=DEF]' features the definite suffix *=eke* on the adjective not the head noun as perhaps expected assuming the diminutive/evaluative suffix as its etymon. If this distribution really supports the conclusion that *-eke* is actually a clitic *=eke*, then this is yet another counter-example to claims of "unidirectionality" in grammatical change, which is supposedly always a shift from less- to more-grammatical. I do not take this view here as there is no demonstrable difference between nouns and adjectives in most Western Iranian languages. Therefore, it makes no difference which element takes the formative. See chapter 3 for an in depth account of these construction that favors a morphological (i.e. affixal) understanding of the definite suffix.

demonstrative clitic *ǽ* exist along side case marking, both particles have different declensions (MacKenzie, 1966); compare masculine singular oblique *-ækáy* and *-i=ǽ*. If *-ækáy* is composed of *-æk* and *=ǽ*. It is unclear why the latter occurs outside of inflection, and why, if the definite article is from the diminutive suffix *-æk* plus the suffix *=ǽ*, both suffixes belong to different declensions.

My intent here is not to necessarily argue against current hypotheses but to express the need for a more thorough investigation of this possibility. Haig (2019a) has shown that the diminutive explanation is possible for Soranî. However, he stops short of being able to confirm this cline.[28] Despite some of the problems with the diminutive origin theory, and my initial skepticism, it is currently the best-supported theory. The ideal proposal, going forward, for the origin of this marker should consider not just the semantic functions of the definite article but its morpho-syntactic distribution. The way it interacts with ezafe marking may be an important clue to its diachrony.

2.4 Attribution marking strategies

Western Iranian languages tend to employ some sort of ezafe construction. The ezafe is most simply a linker binding a noun to either an attributive adjective or a genitival possessor. In New Persian, Gilaki, Soranî and Hewramî, the ezafe is also used to link secondary (de-nominal) prepositions to their complements. The canonical ezafe construction consists of a morphologically-marked head noun followed by a modifier (a left-headed construction). This construction may repeat recursively (see (16)). Some languages in the Caspian region

[28] Haig (2019a) states that there is no other likely etymon for the article *-eke*. However, this is not necessarily the case. Another currently unexplored possible etymological source for this marker is the relativizer *ke*. Relationships between relativizers and definite articles are well attested; e.g. English: *the* ∼ *that*, German: *der, das,* etc. "(definite article)" ∼ *der, das,* etc. "(relative pronoun)". Perhaps the best evidence for the shift from relative pronoun to definite article is Iranian-internal. In Old Persian, the relative pronoun *haya* was also used as a definite article (Kent, 1950, §260.IV). The ambiguity between relative pronoun and definite article can be reanalyzed in either direction. Starting in Classical Greek poetry, the definite article of Classical Greek came to be used as a relative pronoun (van Emde Boas et al., 2019, §28.31). This likely began with the so-called "polydefinite" construction (i.e. *the man the good* = "the good man"). The phrase *the man the good* could then be reanalyzed as *the man [that] (is) good* with a zero copula.

and also Balochi by some accounts (e.g. Haghkerdar, 2009) have a so-called reverse-ezafe construction (see (17)).[29] The term "reverse ezafe" (following Stilo, 2018b), refers to the ezafe-like particle that attaches to an attributive adjective that precedes its nominal complement (a right-headed construction). This particle is sometimes referred to as an attribution marker (e.g. in Suleymanov (2020a) for Şirvan Tat and Axenov (2006) for Turkmen Balochi, etc.).

(16) æsp-î zil-î syaw-î xas
horse-EZ big-EZ black-EZ good
He. "(the) good big black horse" Holmberg & Odden (2008)

(17) an rə́st-ə gə́b-ə
that true-REZ word-COP.3SG
Gi. "It is true? (lit. Is that a true word?)" (Rastorgueva et al., 2012)

In addition to the three uses of the ezafe, ad-attribute[30], ad-genitive, and with prepositions, the ezafe makes several morphological distinctions specific to each of the languages. These distinctions are of four types: (1) interactions between ezafe and definiteness, (2) interactions between ezafe and gender, (3) interactions between ezafe and case, and (4) type of modification. The combination of these distinctions has been of particular interest to linguists as gender is a feature of the head noun, definiteness is semantically conditioned, case is governed by a verb, preposition or the ezafe itself, and the type of modifier.[31]

2.4.1 The interaction of (in)definiteness and ezafe

At the core of this study is the interaction between definiteness marking and the ezafe. However, many factors influence this interaction. For instance, in most Kurmancî (Northern

[29] There are varieties of Balochi which have developed canonical ezafe marking under the influence of Persian (cf. Koroshi (Nourzaei et al., 2015)); see Korn (2005) for more on the origin of these particles (gen: p.108, adj.suff: p. 151).

[30] The ad-attributive ezafe may also link to a prepositional phrase.

[31] Samiian (1994) describes the ezafe as a case marker with multiple governors. This description only looks at the interaction of case and type of modifier and not definiteness.

Kurdish) varieties, the ezafe inflects for number, gender, and definiteness. In neighboring Zazaki, the ezafe inflects additionally for case and type of attribution. Like in the discussion of bicasual systems (§.2.2), these categories collapse in particular contexts. In Zazaki, complete distinction of case, number, animacy, and modifier type only holds for grammatically masculine nouns. The innovative case systems in Iranian can be viewed as a system within a system or on top of a system. For instance, in the history of New Persian case, number and gender distinctions were eliminated. The introduction of the *-rā* marking strategy resulted in a caseless system in absolute and non-specific indefinite contexts and a bicasual system in definite contexts. An idiosyncrasy of *-rā* marking is that the system has a morphologically marked form for (definite) accusative and an unmarked form for nominative and oblique function. The same is observed in Balochi, where there is a caseless system overlaid by a definite-oblique system overlaid by a *-rā*-marked system. These last two layers are definite layers and are therefore merged into a complex case system. These layered case systems are further complexified by the introduction of modification in the canonical ezafe construction, which competes for realization with case, number, and gender marking in some of the Iranian languages.

Gilaki (Rastorgueva et al., 2012), Chali Tati (Yarshater, 1969) and Şirvan Tat (Suleymanov, 2020a) employ the so-called reverse ezafe when a noun is modified by an attributive adjective.[32] The reverse ezafe is different from the canonical ezafe in that the adjectival modifier precedes the head noun and the ezafe attaches to the adjective and not the head noun. Another feature of the reverse ezafe is that it does not have a different form in indefinite, definite, and absolute contexts. Compare (18) and (19).

[32]Şirvan Tat additionally has an inherited canonical ezafe construction which has been largely replaced by the reverse ezafe. This can be observed in example (i), where the noun *xuna* house appears as *xune* with its final vowel raised the last vestige of the inherited ezafe marker.

(i) xune gɨlxɨs=a ataʃ gɨrɨft-e
house.EZ PN=OBL fire seize-PRF.3SG
ŞT. "Gilxis's house has caught fire." (Suleymanov, 2020a, ex. 335)

(18) ziyorat-a ǰö
visit-REZ place
ŞT. “a pilgrimage site” (Suleymanov, 2020a, ex. 323)

(19) ye gözäl-ä imorät=i
one beautiful-REZ mansion=INDF
ŞT. “a beautiful mansion” (Suleymanov, 2020a, ex. 749)

Indefinite ezafe

Except for the reverse ezafe languages, where the head noun hosts the indefinite marking but crucially not the ezafe, each of the Western Iranian languages has some idiosyncrasies in the way the ezafe may combine with an indefinite marker. These idiosyncrasies may have their origins in a historical incompatibility based on the fact that the ezafe is the modern descendant of *yat*, which may have acted as a definite article in the Old Iranian period and therefore could not cooccur with the indefinite article *aiwa*. However, the collapse of case in non-definite contexts has undoubtedly affected the ezafe construction.

Various strategies are employed to work around this incompatibility. One such strategy is to have no overt ezafe marker on indefinite nouns marked for indefiniteness, as shown in the juxtaposition of (20) and (21). In addition to this construction in New Persian, the same meaning expressed in (21) can occur as *ketab-e besyar bozorg-i* with the ezafe on the head noun and the indefinite article attaching at the end of the noun phrase. This suggests that synchronically it is not an incompatibility between indefiniteness and the ezafe but rather between the marker *-i* and the marker *-e*.

(20) xune-ye besyar bozorg
house-EZ very big
NP. “(a/the) big book” (Thackston, 1993)

(21) xune-i besyar bozorg
house-INDF very big
NP. “a very big house” (Thackston, 1993)

Examples (22) and (23) show the forms of the ezafe in the Hewramî dialect of Luhon (MacKenzie, 1966). These examples do not show the incompatibility between indefinite and ezafe seen in the New Persian examples. According to Holmberg & Odden (2008), the occurrence of the ezafe with the indefinite article is optional; they deem both (24) and (25) acceptable. Ezafe marking on indefinite nouns may be intrusive in Hewramî. However, the examples from Holmberg & Odden (2008) and MacKenzie (1966) are somewhat ambiguous. For instance, the sequence *-ew-i* can be parsed [IND-OBL]. The oblique marker *-i* blocks the appearance of the adjectival ezafe. Without a full context sentence, it is impossible to tell if the optionality described by Holmberg & Odden (2008) is actually oblique marking. When I presented the ezafe examples to a native speaker from Pawe city, they were deemed ungrammatical (p.c. with Hishmat Shiany).

(22) kitêb-î sîaw
book-EZ black
He. "(a/the) black book" (MacKenzie, 1966)

(23) kitêb-êw-î sîaw
book-INDF-EZ black
He. "a black book" (MacKenzie, 1966)

(24) mar-êwæ zil-î reş
snake-INDF big-EZ black
He. "(a/the) big black snake" (Holmberg & Odden, 2008)

(25) mar-êw-î zil-î reş
snake-INDF-EZ big-EZ black
He. "a big black snake" (Holmberg & Odden, 2008)

Even if the forms are genuinely optional as claimed by MacKenzie (1966) and Holmberg & Odden (2008), there is a clear difference between nouns specifically marked for indefiniteness and those that are unmarked; for the latter, there is most assuredly no optionality. More research is necessary to confirm the inherited incompatibility between indefinite marking and ezafe marking. I tentatively place Hewramî in the group, which shows the Persian-like

incompatibility, although this disagrees with MacKenzie's (1966) description.[33]

Another strategy for dealing with the incompatibility of ezafe and indefinite is the Kurdish pattern. In Soranî and Kurmancî, there is a distinct difference between the ezafe employed in definite and indefinite contexts. Compare the Kurmancî ezafe in (26) and (27) where there is a different form of the ezafe when paired with the indefinite article. This form occurs only when attached to the head noun in most dialects. In other words, additional attributes are marked by the absolute ezafe[34] as can be observed in (28).

(26) welat-ê mezin
country-EZ big
Ku. "(a/the) great country" (Ekici, 2007)

(27) welat-ek-i mezin
country-INDF-EZ big
Ku. "a great country"

(28) şal-ek-î rind-ê erzan
pants-INDF-EZ nice-EZ cheap
Ku. "a nice cheap pair of pants" (Ekici, 2007)

At first glance, the Soranî ezafe does not show the same type of variation (cf. examples (30) and (31)).

(29) ktew-e gewre-ke
book-EZ big-DEF
So. "the big book" (Thackston, 2006b)

[33] Ultimately, MacKenzie (1966)'s analysis may have to be reevaluated in light of the data from Mahand & Naghshbandi (2014), which shows that the oblique is used in topicalization and to show emphasis or contrast. Added to genitive marking, object marking, and as a prepositional complement functioning as a core verbal argument, this may help to disambiguate many unclear examples from MacKenzie (1966).

[34] I use the term "absolute ezafe" to refer to the ezafe as it would attach to a noun not explicitly marked for definiteness or indefiniteness (following Rastorgueva et al. (2012)'s tripartite classification of definiteness, cf. Thackston (2006b)). This should not be confused with the independent form of the ezafe used without a head noun variably termed "free" (Thackston, 2006a), "demonstrative" (MacKenzie, 1995b), "absolute" (Paul, 1998b), or "elliptic" ezafe (Paul, 2009). these forms are phonetically identical to the secondary construct following Thackston (2006a); in ch. 3, I group the secondary (clitic) form of the ezafe with the demonstrative "anaphoric" function, which I believe to be nominal ellipsis.

(30) ktêw-î gewre
book-EZ big
So. "(a/the) big book" (Thackston, 2006b)

(31) ktêw-êk-î gewre
book-INDF-EZ big
So. "a big book" (Thackston, 2006b)

It is unclear from these examples that the same type of alternation occurs in Soranî because the absolute ezafe is the same as the variant used in indefinite contexts. When juxtaposed with the definite ezafe (29) *-e*, it is clear that there is, in fact, a non-definite ezafe that is allowed on bare nouns and those marked for indefinite. According to Kurmancî grammars (etc. Thackston, 2006a; Bedirxan & Lescot, 1986), Kurmancî does not allow the indefinite form of the ezafe with absolute nouns. However, there are examples where absolute nouns take the indefinite ezafe on the bare noun stem, the absolute form. This is illustrated in example (32), where *nîşan*, the nominal complement of the light verb *da* (*nîşan dan* 'to show'), is marked by the indefinite/absolute form of the ezafe *-î*. Of course, the nominal complement of this verb, in particular, could never be interpreted as a definite noun (lit. 'he gave *the showing').

(32) Mêrik ... nivîn-ê ... nîşan-î min da.
man ... and bed-OBL.F.SG ... show-EZ 1SG.OBL LV.PST
Kurmancî: "The fellow showed me ... the bed ...." (Thackston, 2006a, 76)

Likewise, in some Kurmancî varieties, the form of the secondary (enclitic) construct takes the non-definite form of the ezafe. According to Thackston (2006a) this is "An optional—and fairly rare—alternative masc. sing. construct extender;" e.g. (33) and (34).

(33) nivîskar-ek-î din =î zîrek
writer-IND-EZ.M.SG.IND other =EZ clever
Kurmancî: 'another clever writer' (Thackston, 2006a, 16)

(34) bi w-î dengê xwe =î bilind
with that-M.SG.OBL voice-EZ.M.SG self =EZ loud
Kurmancî: 'in that loud voice of his' (Thackston, 2006a, 16)

Kurmancî grammars generally show the indefinite variant only in the restricted context of indefinite (and not as an absolute ezafe). This fact, along with some superficial similarities with Persian forms, led Karim (2021c) (forthcoming) to link the indefinite article EW as the indefinite ezafe with Soranî generalizing it as the absolute ezafe. This possibility seems unlikely given the wider distribution in Kurmancî. However, it is reasonable to assume that if the ezafe is sensitive to definiteness, then its etymon was as well.

Additionally, Samvelian (2005) shows that there are varieties on Kurmancî spoken at the edge of the Kurmancî-Soranî frontier that have borrowed the definite article *-eke* from Soranî which use the absolute form of the ezafe with the definite article. This is clear in the Surç variety described by Mackenzie (1961) (ex. (35)). Here, *y* is the postvocalic orthographic variant of the indefinite/absolute ezafe *-î*.

(35) mirow-aka-y xwārê
man-DEF-EZ lower
Surçî (NK) 'the lower man' (Mackenzie, 1961, 160, apud Samvelian, 2005, 51)

Perhaps, the best way to deal with these data comprehensively is to say that there is a definite ezafe in most Kurmancî varieties, which inflects for number and gender (also case in some varieties), and there is another ezafe which is used in all other contexts. The former is to be understood as a univerbation of the original definite marker and the latter simplex ezafe (first proposed in ch. 4 of this dissertation).

Southern Zazaki alone has developed the distinctive strategy of using a specialized ezafe marker fundamentally different from the definite and the absolute ezafe. This ezafe is the d-form ezafe which also occurs in a restricted set of contexts on nouns that are not specifically marked for indefiniteness (cf. Paul (1998b)'s oblique IIa). These forms can be understood as

different albeit related phenomena.[35] Compare the ezafe in examples (36) and (37); when the indefinite article *-(y)ê(n)* occurs, it blocks the absolute ezafe *-o*, and the specialized form *-do* takes its place.

(36) meřdım-o gırd
man-EZ big
SZ. "(a/the) big man" (Paul, 1998b)

(37) meřdım-ê-do gırd
man-INDF-EZ big
SZ. "a big man" (Paul, 1998b)

Definite ezafe

In addition to the interaction between the indefinite article and the ezafe, each of the Western Iranian languages has some idiosyncrasies in how the ezafe may combine with a definite marker. These idiosyncrasies cannot necessarily be understood in terms of historical semantics as the ezafe was itself likely a definiteness marker in the Old Iranian period. Additionally, the definite articles in the languages which have them have disputed etymologies (cf. 2.3.2). There is no discernible difference between the absolute ezafe and the ezafe employed on a definite noun for standard New Persian, which lacks any definite article. Things are more complicated for languages like Kurmancî, where the definite-absolute distinction is not clearly defined, and varieties vary as to how they mark these forms. Zazaki, Vafsi, likely Kurmanĉ, and others use case-marking alone to signify definiteness (i.e. nouns which are not definite are uninflected). Zazaki masculine singular nouns are only overtly marked in the oblique case. In Zazaki, the definite (ex. (38)) and absolute (ex. (39)) ezafat

[35] The distribution of the d-form ezafe is not well-understood. Paul (1998b) and Todd (2002) disagree over the specific conditioning environments and the realization of the d-forms. The d-forms that surface with variants identical to those used with the indefinite (e.g. with the *-do* variant) only seem to occur with inanimate nouns. The conditioning environments for the d-forms may, like the indefinite, have originated in an inherent incompatibility between the proto-ezafe and the relevant context (see ch. 5).

(PL) only differ in the oblique. This phenomenon points to a complex (possibly phonological[36]) interaction between the case marker and the ezafe and not necessarily a syntactic or semantic incompatibility.

(38) embaz-ê bin-i vin-en-o
friend-EZ.DEF.OBL other-OBL see.PRS-IPFV-3SG
SZ. "he sees the other friend" (Berz & Malmîsanij (1951), 47.23 apud Paul (1998b))

(39) embaz-o bin vin-en-o
friend-EZ other see.PRS-IPFV-3SG
SZ. "he sees other friends" (adapted)

Another strategy for dealing with the interaction between definiteness and the ezafe can be observed in Soranî and Hewramî. These two groups have the *-eke* type definite article. Compare the Soranî examples (40) and (41) as well as the Hewramî counterparts (42) and (43). On the noun phrases unmarked for definiteness, (40) and (42), the absolute form of the ezafe is used. On the noun phrases marked for definiteness, (41) and (43), a specialized definite ezafe is used.

(40) ktêw-î gewre
book-EZ big
So. "(a/the) big book" (Thackston, 2006b)

(41) ktêw-e gewre-ke
book-EZ big-DEF
So. "the big book" (Thackston, 2006b)

(42) æsp-î zil
horse-EZ big
He. "(a/the) big horse" (Holmberg & Odden, 2008)

(43) æsp-æ zil-ækæ
horse-EZ big-DEF
He. "the big horse" (Holmberg & Odden, 2008)

[36] MacKenzie (1966) describes a blocking of the absolute ezafe after the oblique marker *-ê* in the Hewramî dialect of Luhon. The same phonological blocking of the ezafe might be reconstructible for early Zazaki.

In addition the Soranî-Hewramî type definite marker, there is a definite article *-(h)e* in New Persian (colloquial). Samvelian (2007a) shows that the New Persian (colloquial) definite article can cooccur with the ezafe (44). She also shows an alternative construction (45) where the definite article follows the entire noun phrase neutralizing ezafe marking on each internal modifier.[37] Assuming that both the strategies for combining definite and ezafe shown in Samvelian (2007a) exist, there is a parallel in Soranî. Soranî allows the form *ktêw-eke-i gewre*, but with a slightly different meaning than the canonical *ktêw-e gewre-ke*. *ktêw-e gewre-ke* must be read as the big book where the book's size is an intrinsic part of its identity; *ktêw-eke-î* gewre, in contrast, should be read as the book which happens to be big. More work on the New Persian (colloquial) definite article is necessary to decide whether these are parallel constructions. Based on these data, I group typologically colloquial New Persian with the languages that have a separate definite ezafe Hewramî and Soranî (i.e. ∅).[38]

(44) pesær-e-ye æhmæq
boy-DEF-EZ silly
NP. (col.) "the silly boy" (Samvelian, 2007a)

(45) lebas qermez bi astin-e
dress red without.sleeves-DEF
NP. (col.) "the sleeveless red dress" (Samvelian, 2007a)

2.4.2 Summary of forms

Suppose the colloquial New Persian definite article is, in fact, similar to what is observed in Soranî and Hewramî. In that case, there only seems to be one strategy of dealing with the combination of ezafe and definite article. This strategy involves using a specialized definite ezafe on the head noun and a definite article at the rightmost edge of the noun phrase. This

[37] There is a great deal of diversity in the regional varieties of Persian including additional definite articles, Bandari *-ü*, Shirazi *-u*, etc.

[38] Samvelian (2005) has proposed that the lack of ezafe marking on definite nouns as observed in (45) is actually an example of compounding. The merits of this possibility are not discussed here.

definite ezafe occurs as *-e* in Soranî, *-æ* Hewramî and as -∅ in New Persian (colloquial). I present a summary of these forms in table 2.18.[39]

	"one"	Indf. Art.	Def. Art.	Indf. Ez.	Def. Ez.	Abs. Ez.
S. Zazaki	jew	N-ê(n)	N-∅/-i[case]	N-IND-do Adj	N-o/-ê Adj	N-o Adj
Hewramî	yæk	N-êw	N-ækæ	N-IND-∅ Adj	N-æ Adj-DEF	N-î Adj
Soranî	yek	N-êk/-yek	N-eke	N-IND-î Adj	N-e Adj-DEF	N-î Adj
N Persian (col.)	yek	(ye) N(-i)	N-(h)e/-(h)æ-(r)o	N-e Adj-IND	N-∅ Adj-DEF	N-e Adj
C. Zazaki	yew	N-yek	N-∅/-i[case]	N-INDF-o Adj	N-o/-i Adj	N-o Adj
Kurmancî	yek	N-ek	N-∅/-[case]	N-IND-î Adj	N-ê Adj	N-ê Adj
N Persian	yek	N-i	N-∅/-ra	N-IND-∅ Adj	N-e Adj	N-e Adj
Şirvan Tat	yek	ye N-i	N-∅/-ra	IND Adj-a N-IND	Adj-a N	Adj-a N
Gilaki	i N	i(ta) N-i	a N	IND Adj-ə N-IND	Adj-ə N	Adj-ə N
Chali Tati	i N	i N	N-∅/-e[case]	IND Adj-a N	Adj-a N	Adj-a N

Table 2.18: Strategies for Marking Definiteness and Adjectival Attribution

There are two strategies for dealing with the incompatibility of the ezafe and the indefinite article: (1) ezafe-blocking (New Persian and Hewramî), where the ezafe cannot cooccur on a noun which hosts the indefinite article; and (2) the use of a specialized indefinite ezafe (Soranî, Kurmancî and Southern Zazaki). If the (1) strategy is conceptualized as zero-marking, both strategies can be categorized according to which ezafe allomorphs are syncretic, definite, indefinite, or absolute. Hewramî and Southern Zazaki have three separate ezafe markers (counting the zero-marked indefinite). Standard New Persian and Kurmancî employ the same marker for the definite and absolute ezafat but have a unique marker for indefinite. Soranî, colloquial New Persian, and Central Zazaki employ the same marker for indefinite and absolute ezafat but have a unique definite ezafe. Gilaki, Chali Tati, and Şirvan Tat have the reverse ezafe, one invariable form for all three (see table 2.19 for a visual summary).

[39] In addition to the forms in this table, New Persian has an indefinite construction with the absolute ezafe and a phrase-final indefinite marker (N-e Adj-IND) as seen in colloquial New Persian. Additionally, Soranî, Hewramî, and colloquial New Persian each have the option of stacking the definite article and ezafe on the head noun instead of using the definite ezafe albeit with different semantics (see §.2.4.1; Soranî/Hewramî: N-DEF-i Adj New Persian (col): N-DEF-e Adj. These forms have been omitted from the chart, so only comparable forms are presented.

DEF.EZ = IND.EZ = ABS.EZ	:	Gilaki; Chali Tati; Şirvan Tat
DEF.EZ = ABS.EZ ≠ IND.EZ	:	Kurmancî; New Persian
IND.EZ = ABS.EZ ≠ DEF.EZ	:	Soranî; New Persian (col.); C Zazaki
DEF.EZ ≠ IND.EZ ≠ ABS.EZ	:	S Zazaki; Hewramî

Table 2.19: Ezafe Syncretism

2.4.3 A typology of case, definiteness and attribution

The typologies of case marking proposed by Stilo (2008a) and updated in section 2.2.2 are affected additionally by the interaction of ezafe with definiteness and case marking. In a language like Soranî, which does not have case marking but includes definiteness, there is an interesting distribution of forms. In the non-definite contexts, the ezafe is uniform. In definite contexts, there is the definite ezafe -*e*. However, this form is restricted to attributive contexts creating a de facto distinction between the ad-attributive and ad-genitival ezafat. This distinction is codified in the Hewramî and Zazaki with overt morphology. In Soranî, the distinction only exists in the form of this syntactic gap. It is notable that in Hewramî, which has both a separate morpheme for the ad-genitival ezafe -*u* and the definite suffix -*ækæ*, the same distribution is observed; the definite ezafe -*æ* is not compatible with possessors.

<table>
<tr><th></th><th colspan="2">M</th><th>F</th><th>PL</th><th>DEF.M</th><th>DEF.F</th><th>DEF.PL</th></tr>
<tr><td>DIR./ATT</td><td colspan="3" rowspan="4">-î</td><td colspan="3" rowspan="2">-e (Adj-eke)</td><td rowspan="2">-e (Adj-ekan)</td></tr>
<tr><td>OBL./ATT</td></tr>
<tr><td>DIR./GEN</td><td rowspan="2">-an-î</td><td colspan="3" rowspan="2">-eke-î (N)</td><td rowspan="2">-ekan-î (N)</td></tr>
<tr><td>OBL./GEN</td></tr>
</table>

Table 2.20: Soranî definiteness and attribution

In the arguably most complex of the bicasual languages Southern Zazaki, the introduction of ezafe marking rearranges the system, adding some distinctions while collapsing others. Compare the ezafe marked forms at the top of table 2.21 with the unattributed forms at the bottom.

In the indefinite contexts, all distinctions are collapsed. However, when attribution is introduced, the gender distinction is restored even in the plural, where there is no gen-

<table>
<tr><th></th><th>M.INDF</th><th>F.INDF</th><th>PL.INDF</th><th>DEF.M</th><th>DEF.F</th><th>DEF.PL</th></tr>
<tr><td>DIR.EZ:/ATT</td><td>-ê-do</td><td rowspan="6">-ê-da</td><td rowspan="6">= SG</td><td>-o</td><td rowspan="2">-a</td><td rowspan="2">-ê</td></tr>
<tr><td>DIR.EZ:/GEN</td><td>-ê-dê</td><td rowspan="3">-ê</td></tr>
<tr><td>OBL.EZ:/ATT</td><td rowspan="4">= DIR</td><td rowspan="2">(-er)-a</td><td rowspan="2">-an-ê</td></tr>
<tr><td>OBL.EZ:/GEN</td></tr>
<tr><td>OBL(GEN).EZ:/ATT</td><td rowspan="2">-dê</td><td rowspan="2">(-er)-da</td><td rowspan="2">-an-dê</td></tr>
<tr><td>OBL(GEN).EZ:/GEN</td></tr>
<tr><td>DIR</td><td colspan="3" rowspan="2">-ê</td><td colspan="2">-Ø</td><td>-i</td></tr>
<tr><td>OBL</td><td>-i</td><td>-(er)</td><td>-an</td></tr>
</table>

Table 2.21: Zazaki case, definiteness, and attribution

der distinction in the rest of the Zazaki nominal system (and beyond). Additionally, the distinction of modifier type is part of the masculine alone. In the definite contexts, the syncretism between masculine and feminine direct is neutralized, resulting in a morphologically marked gender distinction. Except in the not-so-well-understood feminine singular oblique *-er* forms, there is no case distinction in the feminine singular with or without ezafe marking. The plural remains unchanged. However, case distinctions are eliminated when there is an ad-genitival masculine singular noun. The syncretism between direct plural and masculine singular oblique is maintained and extended to include masculine singular direct when ad-genitival. Additionally, the oblique realm is further subdivided, separating genitival possessors[40] from other oblique functions. In this sense, there is not only a preservation but also an expansion of the case system in the context of modification.

The types of expansions and contractions that occur in the case systems in the context of definiteness and attribution are widespread in the canonical-ezafe languages. However, they do not occur in reverse-ezafe languages. This is likely due to the fact that the reverse ezafe is a morphological property of the modifier and not the head noun that hosts case and definiteness. Because the various formatives involved never occurred on the same entity, there was never competition for realization, univerbation, or fusion (unimorphization). This

[40]I use the term genitival here as shorthand for the forms Paul (1998b) refers to as oblique I and oblique II. I do not follow Paul's (1998b) division of the oblique case into two separate cases because the oblique I forms' deviation from oblique II is explainable due to several factors, including animacy, and is illuminated through comparative evidence. This is elaborated upon in section 5.

process is discussed for Zazaki in chapter 5.

2.5 Thoughts and conclusions

The ezafe phenomenon has been described as being sensitive to case, number, gender, definiteness, and attribution type. The types of systems and patterns created by these sensitivities have not as yet been systematically explored. There have been discussions of etymology which are fairly certain for the New Persian ezafe *-e* and definite direct object marker *rā*. However, for many markers, Soranî *-eke*, colloquial New Persian *-(h)e* the origins are still debatable. The point of this inquiry is not to settle these issues once and for all but rather to identify some of the patterns that will guide future inquiry into the ezafe phenomenon. One of the guiding principles here is that the etymological source of a particular morpheme must be able to explain its various sensitivities. In other words, an ezafe that marks gender most likely has descended from a gender-bearing formative (e.g. a pronoun). Likewise, since some ezafe markers have a definiteness distinction, they most likely have descended from a unit bearing definiteness (e.g. an article, demonstrative, etc.). There has been no attestation for most of these languages either in the Middle or Old Iranian periods. As such, everything we know about the history of these languages is a hypothesis that will constantly be tested as more data is considered. In light of the data presented here, I have identified a few issues that require further study.

2.5.1 The definite ezafe

There are two types of definite ezafat or different ezafe allomorphs for definite contexts. One type is what is observed in both Central and Southern Zazaki. These languages feature differential case marking with the oblique maker only on definite nouns occurring in contexts that call for oblique marking. The distinction between definite and indefinite ezafat is therefore only morphologically marked in a restricted set of contexts. When the noun is

in the direct case, there is no distinction; when the noun is oblique, there is a definiteness distinction. The distinction has been created by the interaction between case marking and the ezafe marker. In Kurmancî, by contrast, the ezafe ending supersedes oblique case marking, yielding a system with no definiteness distinction for the ezafe. This assessment of the Kurmancî ezafe is only valid for the standard language. There are, in fact, several dialects of Kurmancî that make distinctions for case and definiteness (see Haig & Öpengin (2018)). Any exploration of the evolution of ezafic systems must look into all regional language varieties. Furthermore, the distinctions made in this paper may transcend the genealogical relationships between these languages.

The other type of definite ezafe is that observed in Soranî, Hewramî, and colloquial New Persian. These systems are characterized by the existence of a separate definite article (So. *-eke*, He. *-ækæ* and NP. (col.) *-(h)e*) and a reduced form of the ezafe (So. *-e*, He. *-æ* and NP. (col.) ∅). In each of these languages, the definite article shows a clitic-like[41] behavior where the definite article attaches to the end of a noun phrase with any number of intervening adjectives N-DEF.EZ Adj(-DEF.EZ Adj)-DEF. Alternatively, the definite article can attach to the head noun followed by the absolute ezafe N-DEF-ABS.EZ Adj(-ABS.EZ Adj). The similarities of these systems are striking because they represent three different groupings within Western Iranian that have converged on this point. An additional problem is that the current theory with the most traction regarding the origin of the definite article *-eke* is the diminutive origin theory. Is there a comparable possible etymon for *-(h)e*? It is not clear how this etymology can explain the development of a special ezafe for these definite contexts.

[41] I describe this the definite article as clitic-like here in a general sense. My exact interpretation of the article and the definite ezafe is that they are inflectional and derivational affixes, respectively. This assessment is described in chapter 3.

2.5.2 The indefinite ezafe

When it comes to the indefinite ezafe, the general rule is that when the head noun is the host of the indefinite article EW, that same noun cannot host an ezafe. This is true of Southern Zazaki, which has innovated an indefinite ezafe, and standard New Persian, which does not allow an ezafe in those contexts. According to MacKenzie (1966) and Holmberg & Odden (2008), Hewramî optionally allows the coexistence of these forms. Suppose the reason for the idiosyncrasies of the indefinite ezafe in Southern Zazaki and standard New Persian are indeed the result of a historical incompatibility between the indefinite article and the ezafe (as an article). In that case, Hewramî seems to be innovating ezafe marking on indefinite nouns. Another possibility is that this optionality is actually a misinterpretation by linguists of the *-i* in *-êw-i* as [-INDF-EZ]. The correct interpretation would be *-êw-i-∅* as [IND-OBL-EZ] the blocking of the ezafe after the oblique suffix being a well-attested feature of Hewramî (MacKenzie, 1966, §9.i). Unknown at the time of MacKenzie (1966) is the fact that when a Hewramî noun is topicalized or emphasized, it takes oblique marking (Mahand & Naghshbandi, 2014). Although colloquial New Persian has an indefinite article EW, it has developed a clitic-like placement where it attaches to the end of a noun phrase, not necessarily falling on the head noun. Therefore, there is no interaction or conflict between indefinite marking and ezafe marking. Additionally, in the three languages which feature the reverse ezafe, Şirvan Tat, Gilaki, and Chali Tati, there is no interaction or conflict between indefinite marking and ezafe marking.

Soranî, Kurmancî, and Central Zazaki all feature the postposed definite article YAK that must attach to the head noun. Although their placement is akin to the placement of EW in standard New Persian, Hewramî, and Southern Zazaki, the ezafe and the indefinite article can cooccur; there is no incompatibility. In Central Zazaki, it is clear that the indefinite article YAK is a recent morphological borrowing from Kurmancî. There are, therefore, no diachronic idiosyncrasies in their combination. For Soranî and Kurmancî, it is tempting to reconstruct a numeral YAK and indefinite article YAK as there is no obvious reflex of *aiwa

in either language. If the idiosyncrasies of the ezafe in standard New Persian, Southern Zazaki, and at least partially in Hewramî are due to a historical incompatibility between indefinite and ezafe, then there would be a similar system regardless of the etymon of the indefinite article. The co-occurrence of both the ezafe and indefinite article suggests that the form of the indefinite article in Soranî and Kurmancî are more recent innovations albeit not as recent as in Central Zazaki. This innovative form can cooccur with the ezafe because it no longer conveyed definiteness at the time of its recruitment. Another possibility is that the former indefinite article EW could cooccur with YAK (not dissimilar to Şirvan Tat: *ye* N-*i*), and the indefinite article EW was reinterpreted as the indefinite ezafe.

2.5.3 Resolution of the historical questions

Ultimately, the origins of the indefinite, definite, and absolute ezafat are deeply entangled. Soranî and Kurmancî have the same marker for the indefinite ezafe, but Kurmancî shows syncretism between definite and absolute while Soranî shows syncretism between indefinite and absolute. Is the Soranî indefinite ezafe an innovative form that is generalized to absolute contexts, or is the Kurmancî definite ezafe the innovation? Finding the origin of one of these will likely inform conclusions about all of them.

There is rich variation in the ways that attribution and definiteness marking interact within the Western Iranian languages. This study presents a list of some of the ways these interactions take place. However, it is not an exhaustive list. There seems to be quite a bit of variation within each language. Many of the dialects containing this variation are under-studied in general, and more work is necessary to explore these phenomena fully. More needs to be said about these forms as future work brings additional complexity to light.

Chapter 3

An HTLCG account of the Ezafe phenomenon

3.1 Introduction

A distinguishing feature of the nominal morphology of Iranian languages is the ezafe or attribution marker. The term ezafe is used to describe a range of phenomenon. The core function of the ezafe is to mark possession and attribution. Ezafe formatives can fall on the head noun, the modifier, or neither the head noun nor the modifier, the construct, anti-construct, and floating construct states, respectively (following Rießler's (2016) typology of attribution marking). My goal here is to define the ezafe phenomenon, including the precise function of all ezafat (PL of ezafe). I approach this task from the theoretical formalism of Hybrid-Type Logical Categorial Grammar (hereby HTLCG). I have chosen this particular framework specifically (and CG more generally) because of several foundational assumptions of categorial grammars: (1) syntactic structure is not a universal underlying (arboreal) entity that languages can deviate from or not. (2) Words in syntactic arrangement do not exist in relations such as head/complement/adjunct etc. that are defined in phrase-structure theoretic frameworks; (3) instead, some words are atomic categories (e.g. N, NP, S, etc.), and other words correspond to functors; e.g. $NP\backslash S$ an entity that given a noun phrase to its left will return a sentence. The details of this formalism are described more thoroughly in section 3.4. (4) The syntactic category (i.e. the combinatorics) of a particular lexical entry is stored in the lexicon alongside its semantics and prosody. In this regard, CG theories are considered lexicalist, and they are favored by linguists who value a transparent interface between syntax, semantics, and prosody.

In addition to the CG approach to the syntax of the ezafe, I assume an inferential-realizational approach to morphology.[1] The morphology-syntax interface can be described as the system by which morphological or lexical material is inserted into syntactic structures. The core divide between different morphological theories is precisely what material is inserted. Some favor morpheme-based approaches, which allow morphemes (as things) to be inserted directly into syntactic structures, while others favor word-based approaches that maintain a clear divide between morphology and syntax.[2] The latter is favored by Stump (2001), who motivates the need for an inferential-realizational approach based largely on the fact that there is no direct correspondence between formatives and the meanings associated with a particular inflected form. Stump (2001) has developed the framework of Paradigm Function Morphology which accounts for how the syntactic paradigm of a language generates the prosodic form of a word through realizational rules. Morphology is the organizing system of the lexicon. At the core of Stump's (2001), is the assumption of syntactic dependencies rejected by CG.

Although no theory of morpho-syntax has merged the foundational assumptions of CG and inferential-realizational morphology, I believe that they are necessarily compatible. CG assumes that syntactic combinatorics are stored in the lexicon. The combinatorics are inextricably linked to the semantics of the lexical entry. There is likely no system that assumes that meaning is stored outside of the lexicon. It follows from this linkage that if morphology is the organizing system of the lexicon, that meaning and the associated syntactic types are projected by the morphology as well. The need for realizationalism is motivated by the existence of "difficult" cases, where there is not a clear correspondence between morphemes and meaning. This idea that syntax is projected by the morphological paradigm is likewise motivated by difficult cases, where the syntax is performing operations that are alternately conducted by morphology in other parts of the grammar. For instance, the

[1] See Stump (2001) for a discussion of merits of IR theories of morphology

[2] See Neill (2016), *Lexicalism, the Principle of Morphology-free Syntax and the Principle of Syntax-free Morphology*, for a discussion of the motivations of this separation.

Sanskrit perfect tense is formed synthetically for many verbs, e.g. $\sqrt{}$ṛś 'see' gives dadarśa 'see[.3SG.PRF].' However, for a minority of verbs, the perfect is formed periphrastically (i.e. using syntax), e.g. $\sqrt{}$ās 'sit' gives āsām čakāra sit[-ACC do.3SG.PRF].' It is clear that the morphological paradigm for $\sqrt{}$ṛś contains cells for the perfect tense. An approach that values uniformity across the lexicon suggests that the verb $\sqrt{}$ās must also have cells for the perfect. However, these cell are filled by the syntactic, periphrastic perfect, construction. This is essentially Periphrastic Realization Hypothesis (Ackerman & Stump, 2004; Bonami & Samvelian, 2015, etc.), "Inflectional rules that deduce the realizations of a morphological paradigm' s cells include rules defining periphrastic combinations as well as rules defining synthetic forms." A unified theory of morpho-syntax that carries the foundational assumptions of inferential-realizational morphology and categorial grammar must assume that syntax is always projected by the morphological paradigm. This premise is not only motivated by the difficult cases, where any theory must make some accommodation, but also by common derivational morphemes like causatives and passives, which are valence increasing and valence decreasing, respectively.

The assumption underpinning Stump's (2001) theory and phrase structure grammars in general is that syntax feeds morphology (SYN ==> MORPH). Syntactic dependencies condition the lexical insertion of a particular inflected form. I do not assume any such relationship. Rather, there is no ordering of inputs and outputs among components of the grammar. Syntactic, semantic, and prosodic combinatorics are all stored in the lexicon, and morphology is the paradigmatic organization of the lexicon. This assumption bears strongly on my analysis here. However, the specific mechanics of such a system is beyond the scope of the current study (a future book-length endeavor). There are two additional points behind my motivation for employing these theoretical assumptions: (1) Psycholinguistic Plausibility: if the syntax, semantics, and prosody are indeed stored in the the lexicon, then all three of these aspects are subject to cognitive processes, the lexicon being a type

of memory following Pufahl & Samuel (2014).[3] Due to this assumption, I believe that (2) there are implications for diachronic linguistics. Just as changes in semantics, and prosody (phonetic, phonological and analogical) can be reconstructed, syntax may be reconstructible through the comparative method as well. Until now, the field of historical syntax has relied on the interpretation of morphology as preserving a prior state's syntactic arrangement. It is often said that morphology is "yesterday's syntax" (following Weir, 1987, etc.). I hope that an accurate description of a language's syntactic combinatorics could lead to creating a new sub-field of historical syntax that looks at "pure syntax" in the same way that morphemes and phonemes can be studied as discreet units subject to variability and change.

This chapter is organized into an introduction to the ezafe phenomenon (§.3.2), an introduction to Hybrid-Type Logical Categorial Grammar (§.3.3), an exploration of the ezafe withing the HTLCG framework (§.3.4), a new typology of ezafe marking based on this new perspective (§.3.5, and a brief exploration of the implications this account has on the diachronic study of the ezafe (§.3.6). Throughout this chapter, I use the term ezafe to refer to a great variety of attribution strategies, some of which have different terminology in different intellectual traditions. In section 3.5, I make a point of connecting these phenomena to the typology of attribution described by Rießler (2016).

3.2 What is an ezafe

The ezafe is a morpheme present in some (Northwestern and Southwestern) Iranian languages, including Kurdish, Persian, and Zaza-Gorani. In its simplest form, it has been

[3] Pufahl & Samuel (2014) hypothesized that the "Fundamentally, the mental lexicon is a memory system: It is the place where language and memory meet" (Pufahl & Samuel, 2014, 27). They conducted a series of psycholinguistic experiments to show that non-linguistic information was also stored in the lexicon. Ultimately, they concluded that there is no clear demarcation between the lexicon and memory; rather, lexical "representations appear to reflect more episodic traces of words and co-occurring auditory events, even from unrelated sources like background sounds" (Pufahl & Samuel, 2014, 28).

described as an adnominal linker, marking a noun as having a following adjectival or nominal modifier. This description works nicely for Standard New Persian, where there is very little allomorphy. The Persian ezafe surfaces as *-e* after consonants (ex. (1)),

(1) ketab-e naw
book-EZ new
New Persian (S): 'the new book' (Thackston, 1993, 12)

-ye after vowels (ex. (2)),

(2) dâneshju-ye zerang
student-EZ clever
New Persian (S): 'the clever student' (Thackston, 1993, 12)

And *-Ø*[4] when its morphological slot is filled by the indefinite marker *-i* (ex. (3)).

(3) Ali, dâneshju(-Ø)-i zerang =ast
Ali, student(-EZ)-IND clever =COP.3SG
New Persian (S): 'Ali is a clever student' (Thackston, 1993, 15)

However, the ezafe can be more complicated in other languages. For instance, in Zazaki, the ezafe exhibits allomorphy sensitive to several morpho-syntactic features: the modifier type (nominal/adjectival): compare (4), where the ezafe *-yo* marks the masculine singular direct noun *bıra* 'brother' as being followed by an attributive adjective, with (5), where the ezafe *-ê* marks the masculine singular direct noun *beg* 'chief' as being followed by a genitival possessor;

(4) bıra-yo gırd
brother-EZ.M.SG./ATT bigM.SG
Zazaki: 'the big brother' (Paul (1998b) apud Berz & Malmîsanij (1951), 60.6)

[4] I use zero, Ø, throughout this chapter as a way of expressing a unit of meaning that does not correspond to a unit of morphological form. It is not my intention to suggest that there are indeed silent (invisible) units that are parsable by speakers of the language, which would be counter to a lexical-realization approach to morphology.

(5) beg-ê diyarbekır-i
chief-EZ.M.SG./GEN DiyarBekir-SG.OBL
Zazaki: 'the chief of Diyarbekir' (Paul (1998b) apud Berz & Malmîsanij (1951), 51.24)

case: compare (4), where the ezafe *-yo* marks the direct noun *bıra* 'brother' as being followed by an attributive adjective, with (6), where the ezafe *-ê* marks the oblique noun *embaz* 'friend' as being followed by an attributive adjective;

(6) embaz-ê bini vin-en-o
friend-EZ.M.SG other see.PRS-IPFV-3SG.M
Zazaki: 'he sees the other friend' (Paul (1998b) apud Berz & Malmîsanij (1951), 47.23)

number: compare (4), where the ezafe *-yo* marks the singular noun *bıra* 'brother' as being followed by an attributive adjective, with (7), where the ezafe *-ê* marks the plural noun *seran* 'heads' as being followed by an attributive adjective;

(7) ser-an-ê bin-an k-en-o dırbetın
head-OBL.PL-EZ.PL other-OBL.PL make.PRS-IPFV-3SG.M injured
Zazaki: 'he injures the other heads' (Paul (1998b) apud Berz & Malmîsanij (1951), 140.18)

gender: compare (4), where the ezafe *-yo* marks the masculine singular noun *bıra* 'brother' as being followed by an attributive adjective, with (8), where the ezafe *-a* marks the feminine singular direct noun *keynek* 'girl' as being followed by an attributive adjective;

(8) keynek-a werd-i
girl-EZ.F.SG small-F.SG
Zazaki: 'the small girl' (Paul (1998b) apud Berz & Malmîsanij (1951), 143.14)

verbal argument structure: compare (5), where the ezafe *-ê* marks the masculine singular

direct noun *beg* 'chief' as being followed by a genitival possessor, with (9), where the ezafe -*dê* marks the noun *laj* 'boy' as a core (non-accusative), in this case dative/allative argument of the VP and as being followed by a genitival possessor;

(9) arwêş-i ... misn-en-o laj-dê xo
rabbit-OBL.SG ... show.PRS-IPFV-3SG.M boy-EZ.M.SG.OBL.CORE REFL
Zazaki: 'he shows the rabbit to his son' (Paul (1998b) apud Berz & Malmîsanij (1951), 209.18)

animacy: compare (9), where the ezafe -*dê* marks the noun *laj* 'boy' as a core dative/allative argument of the VP and as being followed by a genitival possessor, with (10), where the ezafe -*dô* marks the noun *ca* 'place' as a core dative/allative argument of the VP that is inanimate and as being followed by an attributive adjective;

(10) ez =do keyn-er-a xo bi-da
I.1SG.DIR =FUT daughter-F.OBL-EZ.F.SG REFL SBJ-give.PRS-1SG
ca-do zengin
place-EZ.M.SG.OBL.CORE.INANIM./ATT
Zazaki: 'I will give my daughter to a rich place' (Paul (1998b) apud Berz & Malmîsanij (1951), 90.3)

and topicalization: the noun phrase *ê legleg-an-dê bin-an* 'those other storks' is acting as the oblique subject of the verb *nê-kışt-ê* 'would not have killed'. The expected for in this construction, according to Paul (1998b), should not have the d-form ezafe marker -*dê* (e.g. *ê legleg-an-ê bin-an*). One claim that I make in this chapter is that this form results from a confluence of factors, including topicalization. This distribution is in line with what is observed in the closely related Hewramî languages described in Rasekh-Mahand & Naghshbandi (2013).

(11) ê legleg-an-dê bin-an a legleg-ı
those.PL stork-OBL.PL-EZ.PL.OBL.CORE other-OBL.PL that.F.SG stork
nê-kış-tê
NEG-kill.PST
Zazaki: 'Those other storks would not have killed that stork.' (Paul (1998b) apud Berz & Malmîsanij (1951), 34.18)

In addition to the ezafat that mark a following nominal or adjectival modifier, some ezafe languages employ the ezafe to connect the noun to PP modifiers as well as relative clauses. When linking to a relative clause, the presence of an ezafe can signify a restrictive clause in Zazaki. Persian, in contrast, uses a marker identical to the indefinite marker *-i* to signify that a relative clause is restrictive.

Despite these morphological differences, the syntax of the ezafe is generally homogeneous (i.e. Noun-EZ Modifier). However, there is a subset of the North-Western Iranian languages that employ the so-called reverse ezafe construction. The Gilaki example (12) shows the ezafe ə attached to the modifier *pəsər* 'boy('s),' and the head noun *kitāb* 'book' follows it.

(12) pəsər-ə kitāb
boy-REZ book
Gilaki: 'the boy's book' (Haghkerdar, 2009, 12)

This phenomenon is generally limited to the Caspian languages. However, there is a reverse ezafe construction in (Caucasian) Tat, Southwestern Iranian closely related to Persian, alongside the inherited ezafe construction. Compare example (13), where the ezafe *a* attaches to the modifier *ziyorat* 'pilgrimage,' and the head noun *ǰö* 'place' follows it, with example (14), where the head noun *xuna* house hosts the ezafe which only surfaces as ablaut (*xune* < **xuna-i*), and it is followed by the modifier *Gılxıs* 'G.('s).'

(13) ziyorat-a ǰö
visit-REZ place
Şirvan Tat: 'a pilgrimage site' (Suleymanov, 2020a, 323)

(14) xune gɨlxɨs=a ataʃ gɨrɨft-e
house.EZ PN=OBL fire seize-PRF.3SG
Şirvan Tat: 'Gilxis's house has caught fire.' (Suleymanov, 2020a, ex. 335)

The morphologically distinct forms of inherited ezafe construction are generally restricted to nouns that end in *-a*. However, the ezafe construction can occur as simple juxtaposition with nouns that lost the ezafe without a trace. Şirvan Tat is a Southwestern Iranian language spoken in the Caspian zone close to reverse ezafe languages. This fact points to contact as the motivating factor for the shift from ezafe to reverse ezafe.

Additionally Baluchi possesses a construction that has been equated to the reverse ezafe (Haghkerdar, 2009). In Baluchi, the genitive marker *-ī, -ē, -e, -a, or -ay* attaches to the possessor, and it is followed by its possessum; see example (15), where the genitive marker *-ay* attaches to the noun *mardum* 'man,' and it is followed by the possessum *kōṭ-ā* 'coat[OBL].' Unlike the reverse ezafe construction in some of the Caspian languages and (Caucasian) Tat, the construction in Baluchi is sensitive to the type of modifier. For instance, the adjective *ɣarīb* 'poor,' in example (15), carries the attribution marker *-ēn* and is followed by the noun that it modifies *mardum* 'man.'

(15) yakk ɣarīb-ēn mardum-ay kōṭ-ā
one poor-ATT man-GEN coat-OBL
'a poor man's coat' (Axenov, 2006, 86)

The reverse ezafe in Baluchi has allomorphs that are sensitive to the type of modifier. This distinction occurs in other languages; for instance, Takestani Tati (Caspian) makes the same distinction but only on masculine singular nouns. Additionally, this distinction is not limited to the reverse ezafe languages. At least two languages in the Kurdish zone, Zazaki and Hewramî, have separate ezafe allomorphs for attribution and possession (e.g. Zazaki: -(y)o[EZ:M.SG./ATT]/-(y)ê[EZ:M.SG./GEN]; Hewramî: -î[EZ:/ATT]/-û[EZ:/GEN]). Like Baluchi, Hewramî makes the distinction across the board, and like T Tati, Zazaki only

makes the distinction on masculine singular nouns. Scholars may disagree about whether the Baluchi forms can be considered reverse ezafe. There is no agreement between scholars about how many cases Baluchi has or what they should be called (Jahani & Korn, 2009, 651). If a criterion in deciding which morphemes qualify as ezafat is that they are a reflex of the Old Iranian relative pronoun *yat*, then there are many ezafat among these languages that would not qualify and many more still with unknown etyma. I take the view that there is a range of phenomena that qualify as ezafat, including the Baluchi attributive and genitive markers, that have a range of etyma (see §.3.6).

This chapter aims to produce a working definition of the ezafe that encompasses all ezafat in the Iranian languages (and possibly beyond) firmly grounded in HTLCG. Wherever possible, I use this to reevaluate previous assumptions about ezafat and their etyma.

3.3 Introduction to HTLCG

Following Kubota & Levine (2020), "*Hybrid Type-Logical Categorial Grammar* is a variant of *categorial grammar* (CG; Ajdukiewicz, 1935; Bar-Hillel, 1953; Lambek, 1958) that belongs to the tradition of Type-Logical Categorial Grammar (Morrill, 1994; Moortgat, 1997).[5]" Categorial grammars are in a sense an elaboration of a simple context-free grammar (e.g. an AB grammar, following Ajdukiewicz, 1935; Bar-Hillel, 1953). In an AB grammar, narrowly, and CG, broadly, there are a few foundational features (mechanics). Linguistic representations are written as prosodic representation (π), semantic interpretation (σ), and syntactic category (γ), e.g. $\langle \pi; \sigma; \gamma \rangle$.

The syntactic category (γ) can occur as an atomic category like N, NP or S in natural language or as the variables over types A and B in an AB grammar. Atomic categories in natural languages are a finite set. However, that set may vary from language to language (as I propose for Iranian in §.3.4.3). Additionally, the syntactic category can be a complex

[5]For more studies in the frame work of Hybrid Type-Logical CG see Kubota (2010, 2014, 2015); Kubota & Levine (2012, 2013a,b, 2014, 2016, 2020).

category such as A/B, which should be read as an entity that given a B to its right will render an A, or $B\backslash A$, which should be read as an entity that given a B to its left will render an A. "The set of syntactic categories is infinite and is recursively defined" (Kubota & Levine, 2020, 23). In HTLCG, there are three type constructors the forward-slash (/), the backslash (\), and the vertical slash ($\uparrow$). The first two are spacial encoding linear order of the constituents, and the last encodes a missing argument without spacial specification. There are no other type constructors or categories in the syntactic portion of this formalism.

Summary:

1. N, NP, and S are categories.
2. If A and B are categories, then so are A/B and B\A.
3. Nothing else is a category.

(Kubota & Levine, 2020, 23)

The semantic interpretation (σ) can, in principle, utilize any available theory of semantics. In this dissertation, I have chosen to follow Partee (1976) formulation of Montague grammar as a basis for the semantic representation. In HTLCG, there is a direct, transparent connection between syntax and semantics. We can, therefore, assume that a semantic functor of type $\langle e, t\rangle$, a function from entities to truth-values, corresponds to the syntactic types $(NP\backslash S), (S/NP), (S \uparrow NP)$, etc., functors that that given a phrase corresponding to an entity render a sentence. Likewise, the syntactic category NP corresponds to a semantic entity e, and the syntactic category S corresponds to a semantic truth value t.

Summary:

1. e and t are semantic types.
2. If α and β are semantic types, then so is $\alpha \longrightarrow \beta$.
3. Nothing else is a semantic type.

(Kubota & Levine, 2020, 25)

The prosodic representation (π) corresponds to words in natural language. There are several operators associated with π: there are lambda expressions, which are essential for operations such as gapping and ellipsis; the symbol (ϵ) corresponds to an empty string; there is a concatenation operator ($\circ$) that represents linear order of prosodic words. To this set, I have added ($=$) to mark the concatenation of a syntactic word that forms a single prosodic unit with another or a clitic. I use the term clitic narrowly to refer to an atypical word (following Zwicky's (1994) description of clitics as either atypical words or as atypical affixes). These elements are characterized by different syntactic and phonological hosts (following Klavans, 2017). If the syntactic and phonological hosts are the same, these elements are treated as morphology (e.g. as atypical affixes following Zwicky, 1994).

HTLCG is a proof-theoretic approach to syntax. Following the tradition of Lambek (1958), natural language utterances are proven in the same sense as a logical proof. Here I introduce slash elimination rules in the labeled deduction format[6] following Oehrle (1996); Morrill (1994); Kubota & Levine (2020). As already noted, the distinction between the forward slash (/) and the backward slash (\) corresponds to the surface word order in which functors look for arguments. That is, A/B is a functor looking for a B to its right, and B\A is a functor looking for a B to its left, to become an A. Following Kubota & Levine (2020), I adopt the Lambek-style notation for syntactic categories, in which arguments are always written "under the slash." Likewise, the Functor A/B correspond to the semantic functor $\mathcal{F}$, which combines with the the semantic portion of the premise corresponding to B, $\mathcal{G}$ yielding $\mathcal{F}(\mathcal{G})$. The elimination rules are given in proofs (1a) and (1b). The *premises* (inputs) are written above the line, and the *conclusions* (output) are written below the line. The linear order between the two premises has only mnemonic significance. Only the order

[6] "The labeled deduction presentation is so called since, in addition to the syntactic categories of the premises and conclusions, the rules are also annotated (or labeled) with how the semantics and prosody of the conclusion are computed given the semantics and prosody of the premises" (Kubota & Levine, 2020, 24).

of a and b[7], which corresponds to the prosody of the expression obtained as the conclusion, is significant.

a. $$\frac{\text{a; } \mathcal{F}; A/B \quad \text{b; } \mathcal{G}; B}{\text{a} \circ \text{b; } \mathcal{F}(\mathcal{G}); A} /\text{E}$$

b. $$\frac{\text{a; } \mathcal{F}; B\backslash A \quad \text{b; } \mathcal{G}; B}{\text{b} \circ \text{a; } \mathcal{F}(\mathcal{G}); A} \backslash\text{E}$$

Proof 1: Forward slash and backward slash elimination

Likewise, there are slash introduction rules. These allow hypothetical arguments to satisfy the combinatorics of a particular functor, only to be abstracted upon and resolved in another part of the proof. This is not to be understood as chronological. These rules are exemplified in the proofs (2a and b). Here, there is an argument of the syntactic type B, a meta variable corresponding to the syntactic type sought by the functor A/B. it corresponds to the free variable x, which is to be understood as referring to a hitherto unknown entity. Its prosody is the variable ϕ, which corresponds to some natural language string. I have labeled this triple n because the introduction rules may only target a variable introduced in the proof. The slash introduction rules (e.g. $/\text{I}^n$) then target this triple for abstraction.

a. $$\frac{\dfrac{[\phi;\, x;\, B]^n \quad \text{a; } \mathcal{F};\, A/B}{\text{a} \circ \phi;\, \mathcal{F}(x);\, A} /\text{E}}{\lambda\phi.\text{a} \circ \phi;\, \lambda x.\mathcal{F}(x);\, A/B} /\text{I}^n$$

b. $$\frac{\dfrac{[\phi;\, x;\, B]^n \quad \text{a; } \mathcal{F};\, B\backslash A}{\phi \circ \text{a};\, \mathcal{F}(x);\, A} \backslash\text{E}}{\lambda\phi.\phi \circ \text{a};\, \lambda x.\mathcal{F}(x);\, B\backslash A} \backslash\text{I}^n$$

Proof 2: Forward slash and backward slash introduction

There is much more to say about HTLCG and categorial grammars more generally. However, this brief introduction is sufficient for the discussion of the phenomenon discussed in the following sections.

[7] a and b are prosodic metavariables; they represent arbitrary strings of prosodic units. They may represent both actual prosodic variables (e.g. ϕ_1, ϕ_2, etc.) or prosodic constants (i.e. words in natural languages).

3.4 A HTLCG approach to the ezafe

To my knowledge, before now, there has not been an analysis of the ezafe phenomenon in the framework of categorial grammar (CG). Here I present a novel analysis of both the ezafe and reverse ezafe within the theoretical framework of Hybrid Type-Logical Categorial Grammar. The basic premise of CG is that there is no "arboreal" hierarchy that, hidden behind utterances, dictates the dependencies between words and phrases. Instead, only actual words and how they combine with other words to become well-formed utterances are responsible for the syntactic structure. HTLCG's direct interface between syntax, semantics, and prosodic combinatorics provides the foundation for an account of the ezafe phenomenon that informs the greater discussion on the nature of attribution and modification. It does so by allowing a more nuanced understanding of syntactic categories. For instance, the term Verb can refer to many types with differing syntactic combinatorics:

- Intransitive: $run; NP\backslash S; run$

  The intransitive verb *run* has the prosodic form /run/. It has the syntactic type $NP\backslash S$; i.e. it is an entity that given a noun phrase on its left returns a sentence. Its semantics are $\lambda x.run(x)$. In the string *John runs* (ex. 3), *runs* is looking for an NP on its left. It finds *John* and prosodic string, and the semantics combine as specified by the syntactic type.

$$\frac{runs; \lambda x.run(x); NP\backslash S \quad John; j; NP}{John \circ runs; run(j); S}\ \backslash\text{E}$$

Proof 3: English: *John runs.*

- Transitive: $throw; (NP\backslash S)/NP; throw$

  The transitive verb throw is $(NP\backslash S)/NP$, an entity that takes a noun phrase on its right and yields $NP\backslash S$, an entity that given a noun phrase on its left will yield a sentence. Note that an intransitive verb is nested within a transitive one.

- Ditransitive$_1$: $give; ((NP\backslash S)/NP)/NP; give$

  Likewise, a transitive verb is nested within a ditransitive verb. However, in English, the term ditransitive refers to two (or more) different syntactic types. The type represented here has two noun phrase arguments, a recipient and a direct object, and the type represented by the following bullet.

- Ditransitive$_2$: $give; ((NP\backslash S)/PP_{to})/NP; give$

  Here, give is $((NP\backslash S)/PP_{to})/NP$, an entity that is looking for an NP (DO) to its right and yields $(NP\backslash S)/PP_{to}$ an entity that is looking for a prepositional phrase marked by the preposition *to* to yield an $NP\backslash S$.

  In the string *John gives the book to Mary* (Proof 4), *gives* is looking for an NP (DO) on its right. It finds *the book* and prosodic string, and the semantics combine as specified by the syntactic type. Next, *gives the book* is looking for a prepositional phrase on its right. It finds and combines with *to Mary*, and so on.

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}gives;\\ \lambda x\lambda y\lambda z.give(x)(y)(z);\\ (NP\backslash S)/PP_{to})/NP\end{array} \quad \begin{array}{c}the\circ book;\\ \iota(book);\\ NP\end{array}}{\begin{array}{c}gives\circ the\circ book;\\ \lambda y\lambda z.give(\iota(book))(y)(z);\\ (NP\backslash S)/PP_{to}\end{array}}/\text{E} \quad \begin{array}{c}to\circ Mary;\\ m;\\ PP_{to}\end{array}}{\begin{array}{c}gives\circ the\circ book\circ to\circ Mary;\\ \lambda z.give(\iota(book))(m)(z);\\ NP\backslash S\end{array}}/\text{E} \quad \begin{array}{c}John;\\ j;\\ NP\end{array}}{\begin{array}{c}John\circ gives\circ the\circ book\circ to\circ Mary;\ give(\iota(book))(m)(j);\ S\end{array}}\backslash\text{E}$$

Proof 4: English: *John gives the book to Mary.*

Just as this theoretical formalism can facilitate a more nuanced description of English verbs, it can do so for verbs in other languages. The syntax, semantics, and prosody are fully integrated, and language-specific facts, like constituent order, are, by design, integrated into lexical entries.

3.4.1 The type adjective

To understand the ezafe phenomenon, one must first look at the status of the adjectives and possessors with which ezafe-marked nouns combine. In a CG approach, there is a more nuanced understanding of the concept of syntactic category. However, there is still a need for atomic categories. In this framework, an atomic category is an entity that does not require another element to be well-formed. Instead, it is a constituent with which other entities seek to combine. In English, such categories are nouns (N), noun phrases (NP), and sentence (S). The word book, a noun, does not select for anything, but it can be selected by a determiner (NP/N) yielding a well-formed noun phrase; see example 5, where the determiner *the* is searching for and finds a noun (book) on its right to form the noun phrase *the book*.

$$\frac{book;\ book;\ N \quad the;\ \lambda P.\iota P;\ NP/N}{the \circ book;\ \iota(book);\ NP}\ /\text{E}$$

Proof 5: English: *the book*

A problem arises when examining a traditional category like adjective that may or may not be an atomic category. When deciding the category of adjectives in English, there are two possibilities: (1) the noun is not a primitive and it selects for an adjective, or (2) the noun is a primitive and the adjective selects for it. If the noun were of type $Adj\backslash N$ (possibility 1), then the proof in example 5 would not be possible. In the erroneous example 6, *the* and *book* cannot combine because they are not compatible categories.

$$*\ \frac{book;\ book;\ Adj\backslash N \quad the;\ \lambda P.\iota P;\ NP/N}{!}$$

Proof 6: English: Ø

However, if nouns are to be taken as primitives, then the proof in example 5 still works, and adjective, of type N/N, can be added. In example 7, the adjective *red* (N/N), combines

with the noun *book* (N), which can then be selected by the determiner *the* (NP/N) creating a well-formed noun phrase.

$$\dfrac{\dfrac{book;\ book;\ N \quad red;\ \lambda P\lambda x.red(x) \wedge P(x);\ N/N}{red \circ book;\ red \wedge book;\ N}\ /\mathrm{E} \quad the;\ \lambda P.\iota P;\ NP/N}{the \circ red \circ book;\ \iota(\lambda x.red(x) \wedge book(x));\ NP}\ /\mathrm{E}$$

Proof 7: English: *the red book*

This category adjective, N/N, works nicely for these English examples. However, when the category adjective is examined crosslinguistically, it becomes clear that adjective must be defined language-specifically. For instance, in Japanese, they are two different types of adjectives 形容詞 /keiyōshi/ 'verbal adjectives' and 形容動詞 /keiyō-dōshi/ 'adjectival nouns.' Both groups contain words equivalent to adjectives in English (e.g. きれい /kirei/ 'beautiful,' 'an adjectival noun' and いい /ii/ 'good,' a verbal adjective). The adjectival nouns can combine with the copula to be used predicatively. Only when inflected with the particle *-na* do they have the syntactic type N/N. These combinatorics are illustrated in example 8 where the adjective きれい /kirei/ 'beautiful' combines with the noun 女の子 /onnanoko/ 'girl' to become the (complex) noun きれいな女の子 /kirei=na onnanoko/ 'beautiful girl.[8]'

$$\dfrac{onnanoko;\ girl;\ N \quad \dfrac{kirei;\ beautiful;\ N \quad = na;\ \lambda P\lambda Q\lambda x.P(x) \wedge Q(x);\ N\backslash(N/N)}{kirei = na;\ \lambda Q\lambda x.beautiful(x) \wedge Q(x);\ N}\ \backslash\mathrm{E}}{kireina \circ onnanoko;\ \lambda x.beautiful(x) \wedge girl(x);\ N}\ /\mathrm{E}$$

Proof 8: Japanese: きれいな女の子　/kireina onnanoko/ 'beautiful girl'

On the other hand, the Japanese verbal adjectives are, as a default, of the same type as transitive verbs, $NP\backslash S$. These entities want a noun phrase on their left to form a well-formed sentence (ex. 9). For verbal adjectives to be used attributively (not as predicates), the exact mechanism used to create relative verbal clauses must be employed.

[8]The existence of categories below the phrase level in Japanese is not discussed here. This discussion would theoretically require the status of Japanese particles as either syntactic or inflectional. Such a discussion is beyond what is possible in this format.

$$\frac{onnanoko = wa;\ \iota(\lambda x.girl(x));\ NP \quad ii;\ \lambda x.good(x);\ NP\backslash S}{onnanokowa \circ ii;\ good(\iota(\lambda x.girl(x)));\ S}\ \backslash\text{E}$$

Proof 9: Japanese: 女の子はいい /onnanokowa ii/ '(the) girl is good'

Based only on the facts of English and Japanese, there are at least three syntactic types associated with the traditional classification of adjective: English: N/N; Japanese adjectival noun: N, and Japanese verbal adjective: $NP\backslash S$. At the heart of this investigation is whether or not ezafe-marked nouns, adjectives, and other entities with which ezafe-marked nouns combine fit into established syntactic categories in other languages. Additionally, the exploration of this question informs the application of CG to typology and historical linguistics.

3.4.2 The ezafe and the Iranian adjective

The function of the ezafe (not the reverse ezafe) is to mark a noun as having a following modifier. This modifier can be an adjective, a noun, a prepositional phrase, or a relative clause. These other entities are addressed here alongside adjectives for a complete (unified) picture of the ezafe. Here, I briefly establish my hypothesis that adjectives in ezafe languages are different from the possible types established in section 3.4.1 based on English and Japanese. Then, the hypothesized category is applied to other Iranian languages to establish whether the ezafe is one syntactic phenomenon or multiple idiosyncratic phenomena across the Iranian languages.

I start by looking at Soranî (Central Kurdish) as an essentially random example. As nouns can occur without adjectives, the first logical hypothesis is that adjectives in Soranî are of type $N\backslash N$ just like English adjective.[9] However, there are four main problems with such a solution.

[9]English adjectives are of type N/N signifying that they are looking for a noun on their right. Soranî adjectives follow the nouns they modify hence the backslash $N\backslash N$.

Problem 3.4.1. The first problem is the ezafe itself. It is the noun that is marked for ezafe. If the adjective is the entity that selects the noun that it modifies, then that noun must be sub-typed for ezafe marking. Adjectives would be of type $N_{EZ}\backslash N$; an entity that takes an ezafe-marked noun on its left to produce a noun. This assumes that the ezafe is a case marker, and the adjective governs a noun's inflection for the ezafe case. This approach has been proposed by Samiian (1994) and supported by Larson & Yamakido (2008). However, a drawback of such a theory is that it requires that we propose that adjectives can be governors of case. In the Iranian languages with both case and ezafe marking (e.g. Zazaki, Hewramî, Kurmancî, etc.), the nominal terminations would be akin to a fusional reverse suffixaufnahme. Suffixaufnahme (case stacking), when a noun is both marked for the genitive case and the case of the head noun that it modifies, is a feature most often associated with highly agglutinative languages. The noun would carry properties of its modifier with the ezafe along with its case, number, definiteness, animacy, gender, etc. I am hesitant to accept the case marking proposal, although it is not impossible. The Soranî example 10 shows that if we assume the ezafe case is a legitimate sub-type, the entities have no difficulty combining. However, the proposal of an ezafe case may be unnecessary as the second problem is more problematic.

$$\frac{\textit{kuřî};\ \textit{boy};\ N_{EZ} \qquad \textit{baş};\ \lambda P\lambda x.\textit{good}(x) \wedge P(x);\ N_{EZ}\backslash N}{\textit{kurî} \circ \textit{baş};\ \lambda x.\textit{good}(x) \wedge \textit{boy}(x);\ N}\ \backslash\text{E}$$

Proof 10: Soranî: *kuřî baş* 'good boy

Problem 3.4.2. The second problem is that adjectives are not the only modifiers that can attach to ezafe-marked nouns. The ezafe is also used to link to nominal (but not pronominal) possessors and prepositional phrases. If an adjective is of type $N_{EZ}\backslash N$, an entity looking for an ezafe-marked noun on its left to produce a noun, then a nominal possessor must also be of the same category. Nominal possessors are not marked differently from other nouns in Soranî and New Persian, where there is no case marking. In languages like Hewamî, Mukriyanî, and Zazaki, nominal possessors are marked in the oblique case when acting as

possessors. However, the oblique case is not reserved for genitive nouns alone; they can act as direct objects of present-tense transitive clauses, as agents of past-tense transitive clauses, and nominal elements of prepositional phrases. If possessors and adjectives are unified as nominal modifiers and of type $N_{EZ}\backslash N$, then all nouns would have two morphologically-unmarked syntactic types N, an atomic category, and $N_{EZ}\backslash N$ an entity that is looking for an ezafe-marked noun on its left to render a noun.

Problem 3.4.3. The third problem is that the category noun N as it is known from English does not necessarily work for Ezafe languages. Most declined forms of a noun (e.g. Hewramî: *kuř* 'boy (generic),' *kuřæ* 'boy (demonstrative),' *kuřækæ* 'boy (definite),' etc) do not require the support of determiners to yield well formed noun phrases. In this regard, the atomic category N does not exist in any real sense. Rather all nouns are full well-formed noun phrases on their own. The only nominal forms that cannot exist on their own are the ezafe-marked nouns (e.g. Hewramî[10]: *kuřî*, *kuřæî*, *kuřækæî*, etc.). As these forms cannot combine with an entity that requires a noun phrase, they must not be noun phrases. In example 11, the verb *miwînû* 'I see' cannot combine with *kuř-ækæ-î* 'boy[-DEF-EZ/ATT].'

$$\frac{\textit{kuřækæî};\ \gamma_?;\ \sigma_? \qquad \textit{miwînû};\ \lambda x.see(x)(y_{1SG});\ NP\backslash S}{!}$$

Proof 11: Hewramî: Ø

Problem 3.4.4. The fourth problem is that there is no clearly demarcated line between nouns and adjectives. Just as an adjective like *qeɬew* 'fat' can be used adjectivally in *mirawîyî qeɬew* 'fat duck,' it can also be used substantively declined regularly as a noun (e.g. *qeɬew* 'fat (one(s)),' *qeɬeweke* 'the fat (one),' *qeɬewêk* 'a fat (one),' etc.).

It should be clear that it is not the case that the categories noun and adjective merely overlap. Rather, there is but a single category as exemplified by the following examples.

[10]I switch to Hewramî for these examples to avoid confusion that might arise because of homophony between the Soranî ezafe *-î* marked noun and a noun carrying the third-person singular possessive clitic *=î*.

Both adjectives and nouns can be the arguments of verbal functors whether definite, indefinite, or generic singular or plural:

Generic:

(16) drêj muşqîle =ye
tall trouble =COP.PRS.3SG
Soranî: 'tall is trouble' or 'tall ones are trouble'

(17) kuř muşqîle =ye
shepherd trouble =COP.PRS.3SG
Soranî: 'shepherds are trouble'

Indefinite Singular:

(18) drêj-êk muşqîle =ye
tall-INDF trouble =COP.PRS.3SG
Soranî: 'a tall one is trouble'

(19) şwan-êk muşqîle =ye
shepherd-INDF trouble =COP.PRS.3SG
Soranî: 'a shepherd is trouble'

Definite Singular:

(20) drêj-eke muşqîle =ye
tall-DEF trouble =COP.PRS.3SG
Soranî: 'the tall one is trouble'

(21) şwan-eke muşqîle =ye
shepherd-DEF trouble =COP.PRS.3SG
Soranî: 'the shepherd is trouble'

Indefinite Plural:

(22) drêj-an muşqîle =yin
tall-PL trouble =COP.PRS.3PL
Soranî: 'some tall ones are trouble'

(23) şwan-an muşqîle =yin
shepherd-PL trouble =COP.PRS.3PL
Soranî: 'some shepherds are trouble'

Definite Plural:

(24) drêj-ek-an muşqîle =yin
tall-DEF-PL trouble =COP.PRS.3SG
Soranî: 'the tall ones are trouble'

(25) şwan-ek-an muşqîle =yin
shepherd-DEF-PL trouble =COP.PRS.3SG
Soranî: 'the shepherds are trouble'

Likewise, both adjectives and nouns may host the ezafe morpheme, e.g. *drêj-î, drêjêk-î, drêjeke-î, drêjan-î, drêjekan-î, şwan-î, şwanêk-î, şwaneke-î, şwanan-î,* and *şwanekan-î*. Each of these forms has the same interpretation of their simplex noun counterparts (in examples (16)-(25)) except that they are not well formed phrases without a following modifier. The simplex forms of both nouns and adjectives require a copula to become predicates. The only thing that sets an attributive adjective apart from a noun is that it must be in its generic form to be interpreted as attributive. For example, the phrase *kuřêkî drêj* has two interpretation. The first and most natural interpretation is that there is a boy and he is tall; in the second the adjective *drêj* 'tall' is interpreted as a kind, e.g. *tall ones*.

(26) kuř-êk-î drêj
boy-INDF-EZ tall
Soranî: 'a tall boy' or 'a boy of tall (ones)'

In contrast, the phrase in example (27) has but one interpretation. There is a boy belonging to a specific tall (person/one).

(27) kuř-êk-î drêj-eke
boy-INDF-EZ tall-DEF
Soranî: 'a boy of the tall (one)'

However, the ambiguity of the generic form is not limited to prototypical adjectives. Nominal have the same distribution. In example, (28) *şwan* has two interpretations: either it is a generic noun 'shepherds,' or it is an attributive adjective modifying *kuřêk* 'a boy.'

(28) kuř-êk-î şwan
boy-INDF-EZ shepherd
Soranî: 'a shepherd boy' or 'a boy of shepherds'

The lack of an independant category adjective (N/N or NP/NP) has consequences similar to problem 3.4.2. Problem 3.4.2 must be slightly amended in light of problem 3.4.3: there is no category N, so $N_{EZ}\backslash N$ would be $NP_{EZ}\backslash NP$. Problem 3.4.2 states that if adjectives are of type $NP_{EZ}\backslash NP$, then all other modifiers that require an ezafe-marked noun must be of the same category. This entails that nouns acting as possessors have a different morphologically unmarked syntactic category (e.g. $NP_{EZ}\backslash NP$) than nouns in all other roles (NP). As there is no clear divide between nouns and adjectives, this would entail that the same is true of adjectives.

Problems 3.4.1-3.4.4 are only problems given a proposed category adjective based on the English one $NP_{EZ}\backslash NP$. My solution is to propose a unified category which is not an N, Adj, NP or $AdjP$ but rather an XP[11] an unspecified phrase which is itself an atomic category, and XP and S are the only atomic categories in the languages. The category XP includes what are traditionally known as noun phrases (NP), adjective phrases (AP), and prepositional phrases (PP). The category of a modifier (M) would therefore be XP/XP. This syntactic type reflects the semantics of modifiers which require the modified entity to be well-formed. Ezafe marking changes the syntactic type of a noun from XP to $XP/(XP/XP)$, an entity that wants a Modifier on its right to produce a phrase.

This proposal works nicely for the Hewramî sentence in example 12. The ezafe-marked definite noun *kuřækæî* is of type $XP/(XP/XP)$, an entity looking for a modifier on its

[11]The orthographic convention XP has been chosen because X is a variable and it is a P phrase-level unit. No connections to X-bar theory (Chomsky, 1970) should be drawn from this orthographic similarity.

right to become a phrase. It combines with the adjective *xas* of type XP to form the XP *kuřækæî xas*, which is then a constituent in the verb phrase/sentence *kuřækæî xas miwînû* 'I see the good boy'.

$$\dfrac{\dfrac{\dfrac{

3.4.3 The Ezafe explored

My definition of morphology is the systematic correspondences between form and meaning. In a traditional morphemes-as-things approach, morphemes can be understood as the smallest unit of correspondence. There are languages, referred to as agglutinative, that prefer a one-to-one form-to-meaning correspondence. In Fusional languages, some morphemes may have multiple units of meaning associated with a single form. Additionally, other correspondences exist, e.g. multiple exponence: multiple units of form to one meaning, etc. It has been suggested that there is a tendency for languages to prefer a one-to-one correspondence (Humboldt's Universal von Humboldt, 1836). Here, I use the word meaning to refer to widely accepted morphological categories (e.g. case, number, gender, etc.), syntactic category, and semantics. At the heart of HTLCG and CG is the assumption that the syntactic combinatorics are stored in the lexicon. When going about the partially descriptive task of identifying the syntactic category of a particular inflected lexeme, it is crucial to establish guidelines for making the determination. In principle, any juxtaposition of two words can be interpreted in two ways: the preceding word as the functor and the following as the argument or vice versa. This is a particularly salient issue when describing and explaining ezafe-marked nouns, which are entities that require a modifier to become a well-formed XP. In contrast, their modifiers are entities that become the semantic functors. Here, I propose the following guideline 3.4.1 for identifying the syntactic functor in any juxtaposition of forms.

Theory 3.4.1 (Maximum Meaning)**.** If there is a distinction in the morphological form of one word and not another syntactically related word, then the morphologically marked form should be the augmented functor. This analysis favors the interpretation that there is a one-form-one-meaning correspondence whenever possible (effectively repurposing Humboldt's Universal as an analytical tool; von Humboldt, 1836).

For example, if we have the noun phrase *kuřekeî şwanêk* 'a shepherd's son,' there are two

options: (1) *kuřekeî* is of type NP/NP, or (2) *şwanêk* is of type $NP\backslash NP$. As expressed in problem 3.4.2, that would mean that *şwanêk* is of type $NP\backslash NP$ sometimes and NP elsewhere. This is not an issue for *kuřekeî* (type NP/NP) as it is not a fully formed NP absent the modifier presaged by the ezafe morpheme *(*-î). I use theory 3.4.1 as foundational in how I evaluate the data below. This theory rejects the interpretation that the ezafe merely sub-types the noun to become the argument of its adjectival functor. Instead, it assigns the complete responsibility for the juxtaposition to the morphologically marked form. This is perhaps an obvious characteristic of and category changing derivational morphology like causatives, responsible for the increase in valence a change in the syntactic category.

XP types in Iranian

In Soranî, the only atomic category is the (X) phrase (XP). I continue to use the words noun, adjective, etc. to refer to lexemes that would traditionally be referred to as such for ease of explanation. However, the unified view of these phrases is reflected in the formalism. Note that this analysis is contingent on a unified view of nouns and adjectives in Western Iranian languages. The justification for this view is based on the fact that there is a syncretism between the attributive adjective and the generic form of the noun. Generic nouns may be used attributively, and attributive adjectives may be used substantively, and the declension of nominals and adjectives is the same. I recognize that another view of these data would be that there are two distinct categories, adjective and noun, and a morphologically unmarked derivation converts the adjective from a functor to a substantive. This latter analysis is not explored in detail here as it conflicts with my foundational principle that a theory should not promote the proliferation of meanings associated with the same form unless absolutely necessary.[13]

[13]Note that the principle that a theory should not promote the proliferation of meanings associated with the same form unless absolutely necessary is essentially an argument for "parsimony," i.e. Occam's razor. What drives the importance of this is the idea that complexity, a fundamental feature of human language, should follow from the data and not be introduced by the linguist observer to fit their theoretical assumptions.

As for nouns (including adjectives used as nominals), each has its definiteness marked morphologically. In this regard, there are no determiners. Three values correspond to the category of definiteness: (1) definite nouns are marked by the definite suffix *-eke* and can be represented by the expression $\iota(\lambda x.P(x))$, where P corresponds to the property denoted by the noun. The unselective quantifier binds all free variables in the expression (following Brasoveanu, 2013; Lewis, 2002; Partee, 2002, etc.). So, $\iota(\lambda x.P(x))$ is equivalent to $\iota x.P(x)$. This distinction will be of importance as multiple layers of modification are introduced by ezafat. (2) Indefinite nouns are marked by the indefinite suffix *-êk* and can be represented by the expression $\exists(\lambda x.P(x))$, where P corresponds to the property denoted by the noun. The unselective quantifier binds all free variables in the expression. (3) Generic nouns are unmarked and can be represented by the expression $^{\cap}(\lambda x.P(x))$, where P corresponds to the property denoted by the noun. The unselective quantifier binds all free variables in the expression. The nominalizer $^{\cap}$ (following Chierchia, 1985; Partee, 2002, etc.) takes the set denoted by $\lambda x.P(x) : \langle e,t\rangle$ and returns the kind $^{\cap}[\lambda x.P(x)] : e$. This form of the noun is used for general statements; e.g. *asman berze* 'the sky is high' or *mina~~l~~ muşkîleye* 'kids are trouble.' Additionally, words traditionally referred to as adjectives can be used substantively without any morphological marking (problem 3.4.4). These forms, their syntactic category, semantics, and translation are enumerated below. I use the syntactic category XP for all entries as they are all nominal.

- *kuřeke*; $\iota(\lambda x.boy(x))$; XP: 'the boy'
- *kuřêk*; $\exists(\lambda x.boy(x))$[14]; XP: 'a boy'
- *kuř*; $^{\cap}(\lambda x.boy(x))$; XP: 'boys (generally)'
- *başeke*; $\iota(\lambda x.good(x))$; XP: 'the good (one)'

[14]The form "*kuřêk*; $\exists(\lambda x.boy(x))$; XP: 'a boy"' is given here as a preliminary. However, the true semantic functor of an indefinite noun would be $\lambda\sigma.\sigma$(*kuřêk*); $\exists(\lambda x.boy(x))$; $XP \uparrow (S \uparrow XP)$ for reasons of scope (following Kubota & Levine, 2020). This issue is addressed in section 3.4.3

- *başêk*; $\exists(\lambda x.good(x))$[14]; XP: 'a good (one)'
- *baş*; $^{\cap}(\lambda x.good(x))$; XP: 'good (one)s (generally)'

The morpho-syntax of ezafe constructions

I use theory 3.4.1, that the morphologically marked form in a juxtaposition is the functor, as a basis for this analysis. The simple ezafe construction consists of a noun and an attributive adjective. The ezafe-marked noun *kuřekeî* is of the syntactic type XP/XP and can be understood as $\lambda y[\iota(boy(x) \wedge^{\cup} y(x))]$. It is an entity looking for an XP on its right y that corresponds to a kind $^{\cap}P$, or the set of entities of which P is true but not necessarily true for all ($\forall y$). It then type-shifts that entity into the corresponding property using the predicativization operator $^{\cup}$ (following Partee (2002)). In review, P is a property of type $\langle e,t\rangle$, and $^{\cap}P$ is a kind of type e. Likewise, s is an entity, type e, and $^{\cup}s$ is a property of type $\langle e,t\rangle$. In the Soranî example 13, the ezafe-marked noun *kuřekeî*, XP/XP, takes the adjective *baş*, XP, satisfying its combinatoric requirements. In this case, the nominalization operator and the predicativization operator cancel each other, allowing the adjective and noun to combine, and the unselective definite quantifier ι binds all free variables.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{cc} \textit{kuřekeî}; & \textit{baş}; \\ \lambda y[\iota(\lambda x[boy(x) \wedge^{\cup} y(x)])]; & {}^{\cap}(\lambda x_2.good(x_2)); \end{array}}{\dfrac{XP/XP \qquad XP}{\textit{kuřekeî} \circ \textit{baş};}\,/\text{E}}}{\lambda y[\iota(\lambda x[boy(x) \wedge^{\cup} y(x)])]({}^{\cap}[\lambda x_2[good(x_2)]]);}\,\lambda\text{-conv.}}{\iota(\lambda x[boy(x) \wedge^{\cup\cap} (\lambda x_2.good(x_2)(x)]);}\,{}^{\cup\cap}\text{-canc.}}{\iota(\lambda x[boy(x) \wedge \lambda x_2.good(x_2)(x)]);}\,\lambda\text{-conv.}}{\iota(\lambda x[boy(x) \wedge good(x)]);}$$
$$XP$$

Proof 13: Soranî: *kuřekeî baş* 'the good boy'

The noun *kuřekeî* carries the meaning $\lambda y[\iota(\lambda x[boy(x) \wedge^{\cup} (y(x))])]$. That is to say that it combines with an entity y representing a kind (i.e. generic noun or adjective $^{\cap}P$). It should be clear that the generic ("absolute" following Thackston (2006b)) form of a noun

in apposition or an attributive adjective are identical. The entity (y) is then converted into its associated property.

Aspects of the attributive ezafe The ezafe is recursive. It can (theoretically) attach to as many subsequent modifiers as possible within the constraints of short-term memory. An ezafe-marked adjective and an ezafe-marked generic noun are of the same syntactic and semantic types XP/XP an entity that wants a phrasal unit to its right returning a phrasal unit. They carry the meaning $\lambda y[\cap(\lambda x[P(x) \wedge^{\cup} y(x)])]$; it is an entity that given a kind y, will return a kind carrying the property P corresponding to the ezafe-marked lexeme and the property denoted by y. In example 14, the ezafe-marked adjective *zîrek* 'smart' combines with the adjective *cwan* 'beautiful' to form *zîrekî cwan* 'beautiful and clever (ones). The definite ezafe-marked noun *kiçekeî* 'the girl' then attaches to it just like any other adjective.

$$\dfrac{\dfrac{\dfrac{

($\lambda x_3[beautiful(x_3)]$), is not necessarily accurate. It can be understood as a coordination of two kinds (type e) smart (ones) and beautiful (ones). This interpretation differs from the semantics of the equivalent phrase with iterative ezafe constructions; i.e. $\cap[\lambda x_2[smart(x_2) \wedge beautiful(x_2)]]$ or smart-and-beautiful (ones).

$$\dfrac{\dfrac{\dfrac{\textit{zîrek};}{\cap[\lambda x_2[smart(x_2)]];}\quad \dfrac{\dfrac{\hat{u};}{\wedge;}\quad \dfrac{\textit{cwan};}{\cap(\lambda x_3[beautiful(x_3)]);}}{(XP\backslash XP)/XP}\quad XP}{\hat{u} \circ \textit{cwan};}\,/\mathrm{E}}{\wedge\cap(\lambda x_3[beautiful(x_3)])}$$
$$\dfrac{XP \qquad XP\backslash XP}{\textit{zîrek} \circ \hat{u} \circ \textit{cwan};}\,\backslash\mathrm{E}$$
$$\cap(\lambda x_2[smart(x_2)]) \wedge\cap\ (\lambda x_3[beautiful(x_3)])$$
$$XP$$

Proof 15: Incorrect interpretation with simple coordination

It seems clear that the functor $\hat{u}$ 'and' must be redefined in light of the type mismatch produced by the functor in proof 15. A hint as to the solution for this problem is present in another idiosyncrasy of conjunction in Iranian languages, group inflection. In a sequence of two or more conjoined nouns, only the final noun is inflected; the final noun's inflection distributes across all conjoined nouns. For example:

- *kiç û kuřeke*; $GROUP(\lambda x.boy(x) \wedge girl(x))$; XP: 'the girl(s) and boy(s)'
- *kiç û kuřan*; $some(\lambda x.boy(x) \wedge girl(x))$; XP: '(some) girl(s) and boy(s)'
- *kiç û kuř*; $\cap(\lambda x.boy(x) \wedge girl(x))$; XP: 'girls and boys (generally)'

Here, *kiç û kuřeke* refers to a specific group of individuals containing any non-zero number of girls and and non-zero number of boys, the whole forming a plurality of two or more individuals. Although, it is the case that there are quantifiers that have been proposed that form a group from multiple conjoined entities (e.g. $LINK$), I use the shorthand $GROUP$ for a quantifier that takes a group with multiple properties and returns a set of

individuals where a non-zero subset has each of the properties of the group. *kiç û kuřan* is the indefinite counterpart of *kiç û kuřeke*. *kiç û kuř* is the generic functor. As such, it refers to a kind containing fewer than all individuals that fit in those two categories.

The way the conjoiner *û* 'and' brings these premises together is not simple conjunction. Rather, the conjunction functor must be able to distribute the its quantification across all conjoined entities. Assuming that a noun phrase is a tuple consisting of a property P and a quantifier $\mathcal{Q}$, I propose the following lexical entry for *û* 'and:'

(29) $\hat{u};\ \lambda x_1[\lambda y_1[let^{15}\langle \mathcal{Q}, P\rangle := x_1\ in\ Q(\lambda x[P(x) \wedge^{\cup} y_1(x)])]];\ (XP_{ABS}\backslash XP)/XP.$

$\hat{U}$ is a functor that takes an XP, consisting of a property P and a quantifier $\mathcal{Q}$, on its right, and it yields a functor looking of an XP in the absolute (generic) state. It combines with the first XP; it distributes the XP's quantifier to the whole conjoined phrase, and its integrates the associated property into the semantic functor.

$$\dfrac{\dfrac{\dfrac{

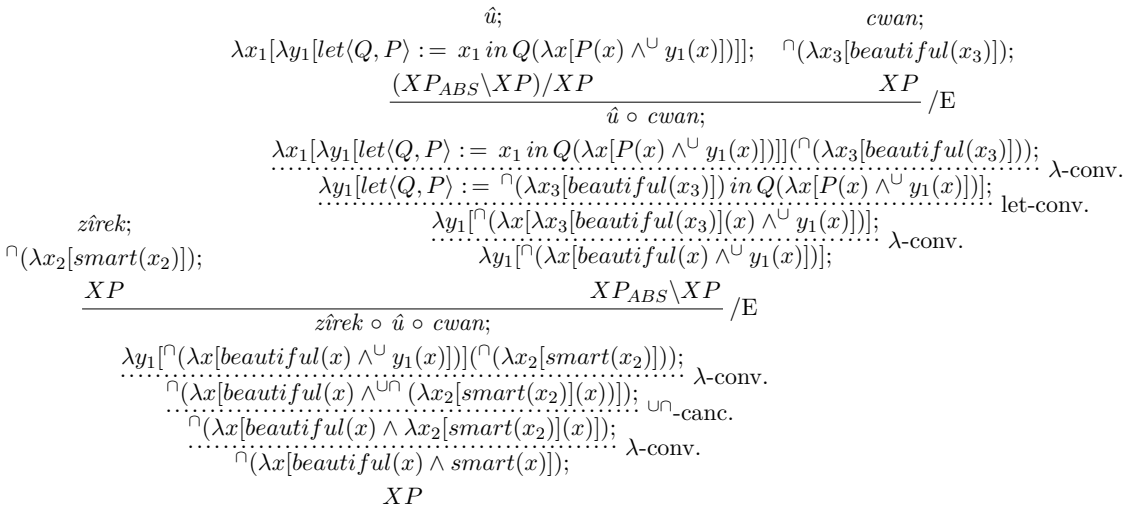
In proof 16, the functor *û* 'and' receives the Property *boy* associated with *kuřekan*, and distributes its quantifier $GROUP$ across the whole conjoined phrase yielding the functor $\lambda y_1[GROUP(\lambda x[boy(x) \wedge^{\cup} y_1(x)])]$. This is significant because it is now similar to the semantic functor of the ezafe-marked noun, which wants a generic noun corresponding to a kind associated with a particular property.

Returning to the issue of conjoined adjectives, this functor for *û*, proposed in example (29), works just as well. All the members of a conjoined string of adjectives must be in the general form. This is true for all but the final conjoined noun, which lends its quantification to the rest. The difference is that an adjective must be generic in order to be received by the semantic functor of the ezafe marked noun.[16] There are two ways to interpret the fact that all conjoined adjectives are generic: (1) the *û* that conjoins adjectives typeraises all the conjoined adjectives converting them into properties, and there is a third functor, e.g. $\lambda z[\lambda y[\lambda x[^{\cup}z(x) \wedge^{\cup} y(x)]]]$. The next interpretation is that (2) the *û* that conjoins adjectives is the same as the *û* that conjoins. Both interpretations produce the same result. However, the interpretation (2) prevents an unnecessary proliferation of meanings. If parsimony is the evaluation metric, then this solution clearly wins. What is certain is that a simple coordination operator $\wedge$ does not produce the correct result for either adjectives or nouns, which need their inflectionally specified quantifiers to be distributed across the coordinated entities. The result of simple coordination in 15, $^{\cap}(\lambda x_2[smart(x_2)]) \wedge ^{\cap}(\lambda x_3[beautiful(x_3)])$, does not capture the fact that there is only one entity.

The formulae in 17 does, however, produce the correct result: *zîrek* $\circ$ *û* $\circ$ *cwan*; $^{\cap}(\lambda x.beautiful(x) \wedge smart(x))$ of type XP. This form can then be the argument of the functor *kiçekeî*; $\lambda y_1[\iota(\lambda x[girl(x) \wedge^{\cup} y_1(x)])]$; XP/XP.

An additional peculiarity of the ezafe in Soranî (as well as Colloquial New Persian, Hewramî, etc.) is that a definite noun phrase can occur with the definiteness marking on

[16] Note that a conjoined string of adjectives can be inflected in any form. However, a different interpretation would be warranted. If the final adjective were definite, the ezafe would be interpreted as ad-genitival, and the adjective would be interpreted as a nominal. See the next section Aspects of the possessive ezafe.



Proof 17: Soranî: *zîrek û cwan* 'smart and beautiful'

the adjective. In contrast, the head noun carries a definite form of the ezafe (Karim, 2021c, forthcoming). The difference between the form with a definite head noun followed by an attributive adjective connected by the general ezafe[17] (e.g. *kuřekeî baş* 'the good boy') and the form with a head noun marked by the definite ezafe followed by an adjective inflected for definiteness (e.g. *kuře başeke* 'the good boy') is that the latter refers to a specific good boy while the former is a specific boy that happens to be good (Thackston, 2006b, §.8). It is not explicitly clear how this shift in meaning should be captured in formal semantics. I leave that discussion and the validity of Thackston's (2006b) analysis as the subject of future study.

The noun marked with the definite ezafe behaves not unlike the operator for nominal coordination. N-EZ.DEF is of type XP/XP_{DEF}, an entity that wants a nominal argument on its right to yield a nominal argument just like all other ezafe-marked nouns. However, the argument must be subtyped for definiteness. The following nominal adjective inflected for definiteness is an XP carrying the meaning $\iota(\lambda x[P(x)])$ where P corresponds to the property denoted by the adjective. This meaning is certain as the definite form of the adjective (e.g.

[17] This form is referred to as the indefinite ezafe in Karim (2021c) (forthcoming). However, the term indefinite is not quite accurate, as this form of the ezafe occurs in all constructions except for when nested within a demonstrative circumposition or other definite construction.

kwêreke 'the blind (one)) must be understood as a nominal(ization). Therefore, the ezafe-marked noun must distribute that definiteness across the whole noun phrase like the nominal coordination operator *û*. N-EZ.DEF carries the following meaning:

(30) $\text{N}_{P1}\text{-}e$; $\lambda x_1[let\langle \mathcal{Q}, P_2\rangle := x_1\ in\ \mathcal{Q}(\lambda x[P_1(x) \wedge P_2(x)])]$; $(XP/XP_{DEF}$.

Here, P_1 is the property corresponding to the ezafe-marked noun. This is a functor that combines with a definite nominal to its right to yield a definite nominal. The nominal that it joins with distributes its quantifier across the entire phrase, and the property it denotes modifies the head noun. Note that this is different from the non-definite ezafat, which feature the quantification on the head noun not the modifier. This is demonstrated by the proof in example 18. The ezafe-marked noun *kuře* selects the following definite nominal adjective *başeke* 'the good one' yielding the XP *kuře başeke* 'the good son' with the definiteness marking scoping over the whole XP.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{cc} \textit{kuře}; & \textit{başeke}; \\ \lambda x_1[let\langle \mathcal{Q}, P\rangle := x_1\ in\ \mathcal{Q}(\lambda x[boy(x) \wedge P(x)])]; & \iota(\lambda x_1[good(x)_1]); \end{array}}{\dfrac{XP_{DEF}/XP_{DEF} \qquad \text{XP}}{\textit{kuře} \circ \textit{başeke};}\ /\text{E}}}{\lambda x_1[let\langle \mathcal{Q}, P\rangle := x_1\ in\ \mathcal{Q}(\lambda x[boy(x) \wedge P(x)])](\iota(\lambda x_1[good(x_1)]));}\ \lambda\text{-conv.}}{let\langle \mathcal{Q}, P\rangle := \iota(\lambda x_1[good(x_1)])\ in\ \mathcal{Q}(\lambda x[boy(x) \wedge P(x)]);}\ \text{let-conv.}}{\iota(\lambda x[boy(x) \wedge \lambda x_1[good(x_1)](x)]);}\ \lambda\text{-conv.}}{\iota(\lambda x[boy(x) \wedge good(x)]);}$$

$$XP_{DEF}$$

Proof 18: Definite Ezafe: Soranî: *kuře başeke* 'the good boy'

Both the conjoiner *û* and the definite ezafe *-e* require that the quantifier of the final item in a string scopes over the whole XP. In addition to this, there is another idiosyncrasy of both formatives; they are both used to form compounds. I do not believe this to be a coincidence (to be discussed in §.3.6).

Another possible interpretation for these data is that the definite suffix *-eke* is actually

an example of "edge inflection" (see Zwicky, 1987). Essentially, the marker *-eke* would be seen as an inflectional element of the head noun that realizes itself on the rightmost element of the NP. Two analyses are possible for most Western Iranian languages: one with the definite suffix followed by the ezafe and one with the definite ezafe followed by an adjective with the definite suffix. The latter is subject to this edge inflection analysis. Assuming such an analysis would be beneficial as it would eliminate the need to separate the quantifier from the associated property in the semantics. Rather the head noun carries the property definite, and it is looking for a modifier on its right that is marked for that property. As the purpose of the current study is to explore the ezafe phenomenon, I leave the precise mechanics of the edge-feature interpretation in a CG framework for future study.

Aspects of the possessive ezafe The ezafe can also be used to connect a noun to a nominal possessor. However, the formula for ezafe-marked nouns established for adjectives (e.g. $\lambda y[\iota(\lambda x[P(x) \wedge^{\cup} y(x)])]$) does not work for possessors. The proof in example 19 shows the result $\iota(\lambda x[boy(x) \wedge^{\cup} s(x)])$ which can be understood as a specific entity x that has the property boy' and the property ${}^{\cup}s$. The latter is a type-shifted form converting the proper noun *şwan* 'Shepard [a name]' of type e to a property of type $\langle e, t\rangle$. The form lacks the possessive meaning, and it is generally difficult to parse.

$$\dfrac{\dfrac{\lambda y[\iota(\lambda x[boy(x) \wedge^{\cup} y(x)])]; \quad \begin{matrix} kuřekeî; \\ XP/XP \end{matrix} \qquad \begin{matrix} şwan; \\ s; \\ XP \end{matrix}}{\begin{matrix} kuřekeî \circ şwan; \\ \lambda y[\iota(\lambda x[boy(x) \wedge^{\cup} y(x)])](s); \end{matrix}}\,/\text{E}}{\begin{matrix} \iota(\lambda x[boy(x) \wedge^{\cup} s(x)]); \\ XP \end{matrix}}\,\lambda\text{-conv.}$$

Proof 19: The ad-genitival ezafe requires a different functor

In the previous section, I reject the proliferation of meanings associated with a single inflected form. Adjectives should not be understood as being both of type XP and $XP\backslash XP$. Likewise, building the possessive semantics into the noun would cause a proliferation of

possible meanings for each noun; e.g. attributive $\lambda x.P(x)$, possessive $\lambda x.P(y) \wedge \mathcal{R}$[18]$(x)(y)$, substantive $^{\cap}P$. A better solution is to allow an additional (2nd) parsing to the ezafe-marked noun rather than three meanings for all nouns.

The second parsing for a definite ezafe-marked noun is N--EZ; $\lambda y[\iota(\lambda x[P(x) \wedge \mathcal{R}(x)(y)])]$; XP/XP. This should be understood as an entity looking for an XP possessor on its right to yield an XP. The possessor y has a relationship $\mathcal{R}$ with the head noun x. This formula represents a wide variety of interpretations that are available for possession. This is exemplified by 20, which is the corrected version of 19.

$$\dfrac{\dfrac{\lambda y[\iota(\lambda x[boy(x) \wedge \mathcal{R}(x)(y)])];}{XP/XP} \quad \dfrac{\textit{kuřekeî};}{\textit{şwan};\ s;}}{\textit{kuřekeî} \circ \textit{şwan};\ XP}\ /\text{E}$$
$$\dfrac{\lambda y[\iota(\lambda x[boy(x) \wedge \mathcal{R}(x)(y)])](s);}{\iota(\lambda x[boy(x) \wedge \mathcal{R}(x)(s)]);\ XP}\ \lambda\text{-conv.}$$

Proof 20: *kuřekeî şwan* 'Şwan's son'

Although the proliferation of meanings associated with single forms is discouraged here (theory 3.4.1), it is necessary for the syntax-semantics interface to produce the correct result. It should be noted that this is required for Kurdish and Persian but not for some other Iranian languages (e.g. Zazaki and Hewramî). These languages have different ezafe allomorphs for ad-nominal and ad-adjectival ezafat; e.g. Zazaki: *laj-o gırd* '(the) big boy [boy-M.SG.EZ:/ATT big]' ∼ *laj-ê şwane-y* '(the) shepherd's boy [boy-M.SG.EZ:/GEN shepherd-M.SG.OBL]' and Hewramî: *kuř-eke-î xas* '(the) good boy [boy-DEF-EZ:/ATT good]' ∼ *kuř-eke-û şwanî* '(the) shepherd's boy [boy-DEF-EZ:/GEN shepherd-OBL].' For these languages there is in fact a one-form-to-one-meaning correspondence. The three Hewramî forms are summarized below where N is the inflected head noun, P_N corresponds to the

[18]Here, $\mathcal{R}$ it represents a contextually salient possessing relation; it is not a free variable or a constant.

property denoted by the noun, and $\mathcal{R}$ is the relationship between the possessor y and the possessum x.

(1) N; $Q(\lambda x.P_N(x))$; XP.

(2) N-*î*; $\lambda y[Q(\lambda x[P_N(x) \wedge^{\cup} y(x)])]$; XP/XP_{GEN}.

(3) N-*û*; $\lambda y[Q(\lambda x[P_N(x) \wedge \mathcal{R}(x)(y)])]$; XP/XP.

Whenever theory 3.4.1 must be violated to produce a coherent form, there is at least one Iranian language that has innovated an ezafe allomorph to serve that function (see §.3.6).

Indefinite ezafat Soranî nouns are inflected for definiteness. There are three values that this feature can have absolute (general) (N: $^{\cup}(\lambda x.P(x))$), definite (N-*eke*: $\iota(\lambda x.P(x))$), and indefinite (N-*êk*: $\exists(\lambda x.P(x))$). I have addressed, through examples, the general behavior of definite and absolute nouns. When marked for ezafe, indefinite nouns work in approximately the same way as other nouns. In the phrase *kuř-êk-î baş* 'a good boy [boy-IND-EZ good],' *kuř* carries the morpheme marking indefiniteness which is followed by the ezafe marker. This is the same as the first attribution strategy with definite nouns (e.g. *kuř-eke-î baş* 'the good boy [boy-DEF-EZ good]). However, definite nouns had an alternative strategy where the head noun augmented by the definite ezafe attached to a definite adjective (e.g. *kuř-e baş-eke* 'the good boy [boy-DEF.EZ good-DEF]). There is no parallel strategy with indefinite nouns.

Another difference between indefinite nouns and their definite and absolute counterparts is that in instances of coordination, the (in)definiteness value of the second member is not distributed across the whole group. This is only tangential to the combinatorics of ezafe-marked nouns. However, it is important to establish the fact that the two functions of *û* 'and,' that for (definite and absolute) nouns and adjectives and that for functors (verbs and indefinite nouns) are independently motivated. When conjoining verbs, for instance, *û* is

not distributive. In example (31), the two conjoined segments *nanim drûst kird* 'I cooked food' and *xwardim* 'I ate it' are of the same type S.

(31) nan =im drûst= kird û xward-im-Ø
bread =1SG:A build= LV.PST and eat.PST-1SG:A-3SG:O
'I cooked food and ate it.'

Simple coordination is all that is necessary to connect them. Therefore, *û* is of syntactic type $(X\backslash X)/X$ and carries the meaning $\wedge$. Here X represents any syntactic type except XPs, which require distributive determination. This begs the question of what type are indefinite nouns if not XPs.

The answer to this question is rooted in the semantics of existential quantification. The existential quantifier is not like the generic quantifier $^{\cup}$, which takes a property $\langle e,t\rangle$ and returns a kind e or the definite quantifier ι, which takes a property $\langle e,t\rangle$ and returns an individual e. Instead, the existential quantifier is a function over properties $\langle\langle e,t\rangle,t\rangle$. Likewise, it bears the syntactic type $S \uparrow (S \uparrow NP)$. In other words, it is an entity that wants a sentence missing an argument to render a sentence as illustrated by ex. 21. Here, the VP had been lying is missing a subject argument which is abstracted upon. The existential quantifier can then scope over the whole sentence as opposed to occurring merely as an argument. In other words, the parsing $\exists(person)(\lambda x.lie(x))$; is more accurate than $lie(\exists x.person(x))$.

This type of analysis becomes important when considering issues of scope (see Kubota & Levine (2020)). Without going too far afield, it is sufficient to say that the *û* 'and' must have two types: $(X\backslash X)/X$ and $(XP\backslash XP)/XP$, the former marking simple coordination and the latter marking group inflection including coordination of adjectives. This distinction is independently motivated and necessary to understand compounding strategies in Soranî (§.3.4.5).

The type of an indefinite ezafe-marked noun (N-*êk*) is therefore $(S \uparrow (S \uparrow XP)) \uparrow XP$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\varphi\,; \quad x \quad NP}{\varphi \circ \text{had} \circ \text{been} \circ \text{lying};}\,\backslash E \qquad \dfrac{\text{had} \circ \text{been} \circ \text{lying};}{\textit{lie}(x);}}{S}}{\lambda\varphi\,[\varphi \circ \text{had} \circ \text{been} \circ \text{lying}];}\,\uparrow\text{I} \qquad \dfrac{\lambda\sigma.\sigma(\text{someone}); \quad \exists(\textit{person}); \quad S \uparrow (S \uparrow NP)}{\lambda\sigma.\sigma(\text{someone})(\lambda\varphi\,[\varphi \circ \text{had} \circ \text{been} \circ \text{lying}])}\,\uparrow E}{\lambda\varphi\,[\varphi \circ \text{had} \circ \text{been} \circ \text{lying}](\text{someone})}\,\lambda\text{-conv.}}{\text{someone} \circ \text{had} \circ \text{been} \circ \text{lying};}\,\lambda\text{-conv.}$$

$$\exists(\textit{person})(\lambda x.\textit{lie}(x));$$

$$S$$

Proof 21: *someone had been lying*

and carries the meaning $\lambda y.\exists(\lambda x.P(x) \wedge^{\cup} y(x))$, an entity that is looking for a restrictor on its right in the form of an XP corresponding to a generalized attribute to render a functor over properties. In example 22, the ezafe-marked indefinite noun *kuřêkî* combines with the adjective *xrap* 'bad' to become *kuřêkî xrap* 'a bad boy.' This form then selects for the property $\textit{lie}'$ represented by the lightverb construction *droî kird*, which is technically a complete sentence. This form is abstracted upon becoming the type that that the indefinite noun is looking for.

Strings of ezafat Ezafat are recursive morphemes that can occur in strings. When nouns have multiple modifiers, each occurs sequentially with all forms except for the ultimate marked by the ezafe. This is unremarkable as each ezafe construction renders an XP which can then be the argument of another ezafe-marked XP (ex. 23). This process felicitously combines with a theoretically infinite number of modifiers in practice limited by the constraints of short-term memory.

One potential issue with ezafe chains is apparent when combining with multiple types of ezafat ad-attributive and ad-genitive. The fundamental problem is that of scope; if there is

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\lambda\phi_1[\lambda\sigma.\sigma(\textit{kuřêkî} \circ \phi_1)];\quad \textit{xrap};}{\lambda y.\exists(\lambda x.\textit{boy}(x) \wedge^{\cup} y(x));\quad {}^{\cap}(\lambda x_1.\textit{bad}(x_1));}}{(S \uparrow (S \uparrow XP_{3SG})) \uparrow XP}\ \text{XP}}{\lambda\phi_1[\lambda\sigma.\sigma(\textit{kuřêkî} \circ \phi_1)](\textit{xrap});}\ /\text{E}}{\lambda\sigma.\sigma(\textit{kuřêkî} \circ \textit{xrap});}\ \lambda\text{-conv.}}{\lambda y.\exists(\lambda x.\textit{boy}(x) \wedge^{\cup} y(x))({}^{\cap}(\lambda x_1.\textit{bad}(x_1)))}\ \lambda\text{-conv.}}{\exists(\lambda x.\textit{boy}(x) \wedge^{\cup\cap} (\lambda x_1.\textit{bad}(x_1))(x))}\ \cup\cap\text{-canc.}}{\exists(\lambda x.\textit{boy}(x) \wedge \lambda x_1.\textit{bad}(x_1))(x)}\ \lambda\text{-conv.}}{\exists(\lambda x.\textit{boy}(x) \wedge \textit{bad}(x))}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\textit{droî} \circ \textit{kird};\quad \lambda x[\textit{lie}(x)];\quad XP_{3SG}\backslash S;\quad \left[\begin{array}{c}\phi;\\ x_1;\\ XP_{3SG}\end{array}\right]^1}{\phi \circ \textit{droî} \circ \textit{kird};}\ \backslash\text{E}}{\lambda x[\textit{lie}(x)](x_1);}\ \lambda\text{-conv.}}{\textit{lie}(x_1);}\ \lambda\text{-conv.}}{\dfrac{S}{\lambda\phi[\phi \circ \textit{droî} \circ \textit{kird}];}\ \uparrow\text{I}^1}$$

$$\lambda x_1[\textit{lie}(x_1)];$$

$$\dfrac{\dfrac{S \uparrow (S \uparrow XP_{3SG}) \qquad S \uparrow XP_{3SG}}{\lambda\sigma.\sigma(\textit{kuřêkî} \circ \textit{xrap})\ (\lambda\phi[\phi \circ \textit{droî} \circ \textit{kird}])}\ \lambda\text{-conv.}}{\lambda\phi[\phi \circ \textit{droî} \circ \textit{kird}](\textit{kuřêkî} \circ \textit{xrap})}\ \lambda\text{-conv.}$$

$$\textit{kuřêkî} \circ \textit{xrap} \circ \textit{droî} \circ \textit{kird};$$

$$\exists(\lambda x.\textit{boy}(x) \wedge \textit{bad}(x))(\lambda x_{3SG}.\textit{lie}(x_{3SG}));$$

$$S$$

Proof 22: *kuřêkî xrap dro kird* 'a bad boy lied'

$$\dfrac{XP/XP \quad \dfrac{XP/XP \quad \dfrac{XP/XP \quad \dfrac{XP/XP \quad \dfrac{XP/XP \quad XP}{XP}\ /\text{E}}{XP}\ /\text{E}}{XP}\ /\text{E}}{XP}\ /\text{E}}{XP}\ /\text{E}$$

Proof 23: Ezafe Recursivity

a combination of N-EZ < N-EZ < Adj, it is not necessarily clear which noun is modified by the ezafe. This is not a problem in Soranî and New Persian as the semantic combinatorics of the ezafe only allow an adjective to combine with the proximate noun.

In example 24, the adjective *baş* 'good' attaches to the possessor *şwanekeî* 'the shepherd (EZ)' through the ezafe connection. The semantic functor that I have proposed for combining with attributes, where there absolute form of a noun or adjective combines recursively with the preceding modifier until it is stopped by a the head noun, naturally terminates at this point as *şwaneke* is definite ($\lambda y[\iota(\lambda x_2[\textit{shepherd}(x_2) \wedge^{\cup} [y(x_2)]])]$). When *kuřekeî* 'the boy (EZ)' combines with *şwanekeî bas* 'the good shepherd,' it does so with the ad-genitival

ezafe $\lambda y[\iota(\lambda x_1[boy(x_1) \wedge \mathcal{R}(x_1)(y)])]$.[19] There is no mechanism by which the adjective *baş* can be understood as referring to the boy in this phrase. This assumption is predicted by the proposed semantic functor and it is upheld by native speaker judgements. The phrase here is felicitous and pragmatically called for in certain contexts. However, the most common way of expressing this is to nest the adjective under the definite article using the definite ezafe (see §.3.4.3); e.g. *kuřekeî şwane başeke* 'the [good shepherd's] son' or *kuře başekeî şwaneke* 'the shepherd's [good son]'.

$$\dfrac{\dfrac{\dfrac{

[good shepherd's] son' or 'the shepherd's [good son].' A hint to the proper treatment of this ambiguity is hidden in the fact that the latter is alliteratively written *kurê şivan ê baş*, which distinguishes the two readings.

(32) kur-ê şivan-ê baş
boy-M.SG.EZ shepherd-M.SG.EZ good
Kurmancî: 'the [good shepherd's] son' or 'the shepherd's [good son].'

The writing of the ezafe as a separate particle reflects the interpretation that the Kurmancî secondary[20] ezafe is a clitic (i.e. not part of morphologically complex (inflected) forms). This is also supported by the fact that it agrees in ϕ-features with the noun it modifies. This becomes clear when the head noun and the possessor have different genders/numbers cf. examples (33) and (34), which correspond to 'the [good shepherd]'s daughter' and 'the shepherd's [good daughter]' respectively. In the former, *şivanê* 'the shepherd' is inflected for masculine singular ezafe *-ê*; in the latter, the ezafe, generally written separately, is inflected for feminine singular *=a*. The orthographic convention is not necessarily indicative of a true linguistic difference. However, the discord between the ezafe's ϕ-features and its host (i.e. *şivan* [M.SG] and *=a* [F.SG]) and the independence of syntactic and phonological hosts (i.e. following Klavans, 2017) points to syntactic combination rather than morphological inflection.

(33) keç-a şivan=ê baş
girl-F.SG.EZ shepherd=M.SG.EZ good
Kurmancî: 'the [good shepherd]'s daughter'

(34) keç-a şivan=a baş
girl-F.SG.EZ shepherd=F.SG.EZ good
Kurmancî: 'the shepherd's [good daughter].'

Of course, introducing an ezafe clitic as a solution to this challenging case suggests that

[20]The term secondary is employed here in reference to these ezafat, markers of what Thackston (2006a) terms the secondary construct state.

it may be a better solution for parsimony's sake to look at all ezafat as clitics and, therefore, syntactic elements. This would, however, be the wrong solution. First of all, syntax is a much more powerful tool than is required for all ezafe constructions in most ezafe-languages, including Soranî, Hewramî, and new Persian. The reason this is an issue is that Kurmacî and Zazaki have lost the definite ezafe (see chapter 4 for more on this development). The only remaining possession strategy forces the possessor into a space that breaks all other ezafe chains. Therefore, any solution to this issue would be language-specific and not a property of all ezafat. Another fact that points to different treatment of the secondary ezafat is that they are not in competition with other morphological features.

In Kurmancî, the ezafe supersedes all other morphological endings. Compare example (35), which shows *rojê* 'day [-F.SG.OBL]' and example (36), which shows *roja* 'day [-F.SG.EZ].' Both contexts require an oblique (genitive) noun, and were it not for the demonstrative which carries same set of ϕ-features, the noun would be ambiguously marked in respect to case.

(35) behs-a girîngiy-a w-ê roj-ê
talk-F.SG.EZ important-F.SG.EZ that-F.SG.OBL day-F.SG.OBL
Kurmancî: 'discussion of the importance of that day' (Thackston, 2006a, 14)

(36) behs-a girîngiy-a w-ê roj-a pîroz
talk-F.SG.EZ important-F.SG.EZ that-F.SG.OBL day-F.SG.EZ celebrated
Kurmancî: 'discussion of the importance of that celebrated day' (Thackston, 2006a, 14)

There is no such competition for realizations with the secondary ezafe. This is exemplified by (37), which shows that the ezafe *yê* with the glide for hiatus resolution after the vowel. This ezafe cooccurs with the oblique case ending -*î*. This amounts to further proof that we are not looking at morphology, or at the very least, not the same morpheme.

(37) nav-ê w-î mirov-î =yê rastîn
name-M.SG.EZ that-M.SG.OBL man-M.SG.OBL =M.SG.EZ$_2$ real

Kurmancî: ‘that man’s real name’ (Thackston, 2006a, 15)

The same is true of Zazaki, where the ezafe has allomorphs that combine the values of case and type of attribution as well as number and gender. This can be observed in example (38), where the noun *dêwiyê* ‘demon [M.SG.OBL PL:EZ] carries both the masculine singular oblique inflection *-i* and the secondary ezafe *-(y)ê* inflected with the ϕ-features of the head noun *seran* ‘heads.’ In this case, the adjective *dırbetınan* ‘wounded [PL],’ which also agrees in ϕ-features with *seran*, can only be interpreted as describing *seran* ‘heads’ (‘both [wounded heads] of the demon’).

(38) wırna ser-an-ê dêw-i-yê dırbetın-an
both head-PL.OBL-PL.EZ demon-M.SG.OBL-PL.EZ wounded-PL.OBL
cı-k-en-o
PV-kill-PRS.IPFV-3SG.M
Zazaki: ‘he kills both wounded heads of the demon’ (Berz & Malmîsanij (1951), 142.5 apud Paul (1998b))

If the adjective wounded were to describe *dêw* ‘demon,’ the phrase would read *dêw-dê dırbetın-i* (ex. (39)). The noun *dêwdê* is inflected for ezafe in the context of being an oblique possessor, but there are not independant agglutinating morphs for both values as there are when the adjective describes the head noun. Additionally, the adjective *dırbetını* ‘wounded [M.SG.OBL] agrees with *dêwdê* in ϕ-features.

(39) wırna ser-an-ê dêw-dê dırbetın-i
both head-PL.OBL-PL.EZ demon-M.SG.GEN.EZ wounded-M.SG.OBL
cı-k-en-o
PV-kill-PRS.IPFV-3SG.M
Zazaki: ‘he kills both heads of the wounded demon’

Based on these data, we must re-evaluate the recursive nature of the ezafe as far as Zazaki and Kurmancî are concerned. While it is true that there can be recursive strings of ezafat like example 23, each subsequent modifier must modify the previous XP. When a

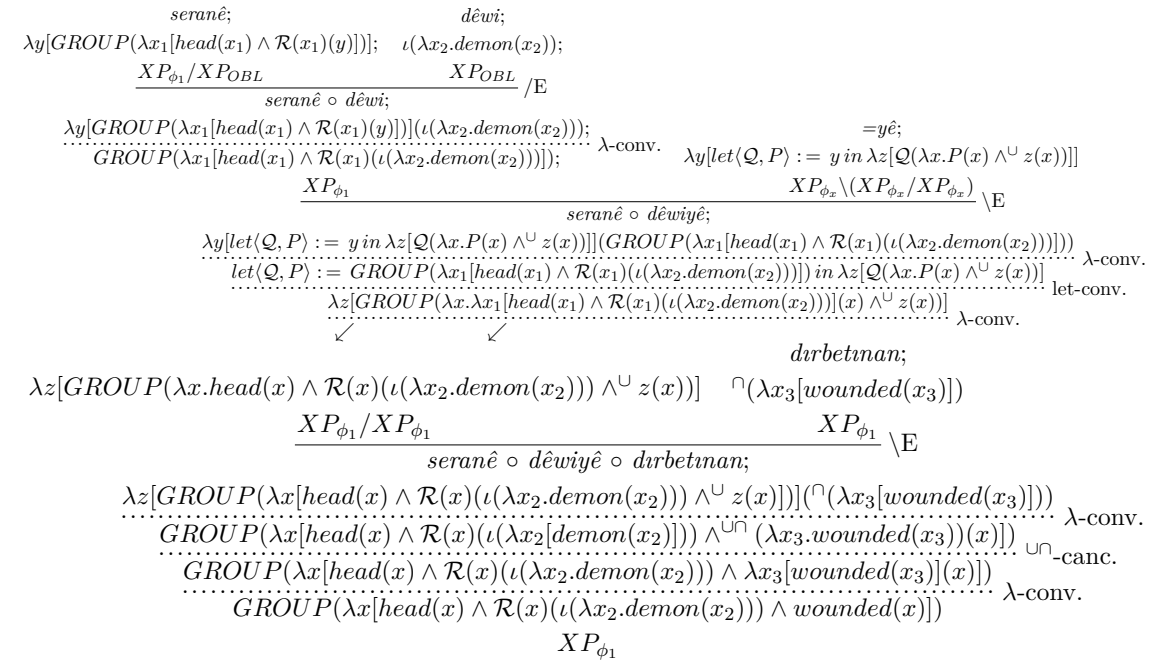
modifier must skip the adjacent XP to modify the head noun, a secondary ezafe (clitic) must be employed. In example 25, the initial ezafe-marked noun, represented by XP/XP, combines with a modifier XP to form an XP. The secondary ezafe $XP\backslash(XP/XP)$ is then the recursive functor (in alternating red and black), which could theoretically repeat ad infinitum. However, it is limited in practice both by the constraints of memory and by the unlikelihood of multiple possessors, each with multiple attributes.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{XP/XP \quad XP}{XP}\,/\mathrm{E} \quad XP\backslash(XP/XP)}{XP/XP}\,\backslash\mathrm{E} \quad XP}{XP}\,/\mathrm{E} \quad XP\backslash(XP/XP)}{XP/XP}\,\backslash\mathrm{E} \quad XP}{XP}\,/\mathrm{E}$$

Proof 25: Secondary Ezafe Recursivity

The practical application of the secondary ezafe is illustrated in proof 26. Here, the ezafe-marked head noun *seranê* 'heads' is of type XP_{ϕ_1}[21]$/XP_{OBL}$ with the same semantic functor proposed for the ad-genitival ezafe $\lambda y[GROUP(\lambda x_1[head(x_1) \wedge \mathcal{R}(x_1)(y)])]$. It is followed by its possessor *dêwi* 'the demon,' which must be in the oblique case, the case for possessors. The secondary ezafe selects for a noun on its left and returns a functor looking for a modifier on its right. I write the syntactic functor as $XP_{\phi_x}\backslash(XP_{\phi_x}/XP_{\phi_x})$ with the XPs subtyped ϕ_x, which is coded by the first noun to which it attaches i.e. combining with XP_{ϕ_1} and becoming XP_{ϕ_1}/XP_{ϕ_1}. This is essentially a shorthand as the secondary ezafe has allomorphs that agree in ϕ-features with the noun it modifies. The more accurate description of this functor *=yê* would, therefore, be $XP_{PL.OBL}\backslash(XP_{PL.OBL}/XP_{PL.OBL})$. This is a functor that wants an oblique plural noun on its left to render a functor looking for an oblique plural adjective on its right to form an oblique plural XP.

[21] In the formula XP_{ϕ_1}/XP_{OBL}, ϕ_1 signifies the case, number, and gender (but not ezafe) of the head noun. In the case of *seranê* those features are [PL.OB]. I use the ϕ_1 in this case to emphasize the relationship between the ϕ-features of the head noun and its attributive adjective.



Proof 26: S Zazaki: *seranê dêwiyê dırbetınan* 'the demon's [wounded heads]'

In Zazaki, just as with the primary (morphological) ezafe, there is a separate functor for ad-attributive ezafe and ad-genitival ezafe. The result is that there are four semantic functors: (1) the ad-attributive ezafe-marked noun, (2) the ad-genitival ezafe-marked noun, (3) the ad-attributive secondary ezafe, and (4) the ad-genitival secondary ezafe. These functors are enumerated below.

1. N-EZ:/ATT; $\lambda y[\mathcal{Q}(\lambda x[P(x) \wedge^{\cup} [y(x)]])]$; XP/XP

2. N-EZ:/GEN; $\lambda y[\mathcal{Q}(\lambda x[P(x) \wedge \mathcal{R}(x)(y)])]$; XP/XP

3. =EZ:/ATT; $\lambda y[let\langle \mathcal{Q}, P\rangle := y\, in\, \lambda z[\mathcal{Q}(\lambda x.P(x) \wedge^{\cup} z(x))]]$; $XP\backslash(XP/XP)$

4. =EZ:/GEN; $\lambda y[let\langle \mathcal{Q}, P\rangle := y\, in\, \lambda z[\mathcal{Q}(\lambda x.P(x) \wedge \mathcal{R}(x)(z))]]$; $XP\backslash(XP/XP)$

The first two are represented in all ezafe languages. The final two exist in Kurmancî and Zazaki alongside the primary ezafat. However, there is no Iranian language, to my

knowledge, that has the secondary ezafat (3 and 4) but not the primary (morphological ezafat (1 and 2). Although, it may be a reasonable diachronic outcome.

The ezafe with adpositional phrases In addition to the ad-attributive and ad-genitival ezafat, the ezafe can connect the head noun to an adpositional phrase. This is exemplified simply by the Persian and Soranî examples (40) and (41). In each of these the ezafe-marked nouns *ketabe* and *ktebi* 'book' are followed by prepositional phrases *rūye mīz* and *le_ser mêz=a* 'on the table.'

(40) ketāb-e rūye mīz
book-EZ on table
Persian: 'the book on the table'

(41) ktêb-î le_ser mêz=a
book-EZ on table=on.POST
Soranî: 'the book on the table'

There are two possibilities for how these phrases must combine: either there is a third possible meaning for the ezafe-marked noun XP/PP, or prepositional phrases are also of type XP. Based on theory 3.4.1, a preference is given for the latter explanation that does not require a proliferation of meanings. Furthermore, the data support the interpretation of prepositional phrases as XPs just like nouns and adjectives. Just like other XPs, prepositional phrases can be arguments of verbs. For example *wergirt*; $\lambda z.\lambda x.\lambda y.take(x)from(z)(y)$; $XP_{le}\backslash(XP_A\backslash S)$. *Wergirt* is looking for a prepositional phrase to its left to become a functor looking for a noun corresponding to the direct object subtyped for agent marking. Ignoring the complexities of the Kurdish verb phrase, it is clear here that the prepositional phrase can be an argument of the verb just like any other XP. Of course, this only proves phrasehood and not that a prepositional phrase does not constitute a unique atomic category. The important question is whether a different semantic functor is necessary for combination with prepositional phrases.

Clearly, the ad-genitival functor $\lambda y[\mathcal{Q}(\lambda x[head(x) \wedge \mathcal{R}(x)(y)])]$ does not produce the correct result (**[on the table]'s book*). I assume that the preposition *on*'s semantics is $\lambda x.\lambda y.on(x)(y) : \langle e, \langle e, t\rangle\rangle$; it combines with a noun $x : e; XP$ to become a a functor looking for the noun $y : e$ that is "on noun_{x}." However, $\lambda y.on(\iota)(y)$ is a functor of type $\langle e, t\rangle; XP \uparrow S$, which cannot be the argument of a functor looking for an XP(e.g. a transitive verb $XP \uparrow S$). This disagrees with what is observed in the languages. Just like adjectives used substantively, prepositional phrases are type-lowered in their default function (i.e. for the semantics to match what is observed in the languages studied here). The functor for *li ser* is *li ser*; $\lambda x.^{\cap}on(x)$; XP/XP. The ad-attributive functor $\lambda y[\mathcal{Q}(\lambda x[P(x) \wedge^{\cup} [y(x)]])]$ works perfectly. In proof 27, the Kurmancî ezafe-marked noun *pirtûka* 'the book' combines with the prepositional phrase *li ser masê* 'on the table.' The result is a an entity bearing both the properties *book* and $on(\iota(table))$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{cc} li \circ ser; & masê; \\ \lambda x[^{\cap}on(x)]; & \iota(\lambda x_2[table(x_2)]); \end{array}}{\dfrac{XP/XP_{OBL} \qquad XP_{OBL}}{li \circ ser \circ masê;}\,/\text{E}}}{\begin{array}{c} pirtûka; \\ \lambda y[\iota(\lambda x_1[book(x_1) \wedge^{\cup} y(x_1)])]; \end{array} \quad \dfrac{\lambda x[^{\cap}on(x)](\iota(\lambda x_2[table(x_2)]))}{^{\cap}on(\iota(\lambda x_2[table(x_2)]));}\,\lambda\text{-conv.}}}{\dfrac{XP/XP \qquad XP}{pirtûka \circ li \circ ser \circ masê;}\,/\text{E}}}{\dfrac{\dfrac{\lambda y[\iota(\lambda x_1[book(x_1) \wedge^{\cup} y(x_1)])](^{\cap}on(\iota(\lambda x_2[table(x_2)])))}{\iota(\lambda x_1[book(x_1) \wedge^{\cup^{\cap}} on(\iota(\lambda x_2[table(x_2)]))(x_1)]);}\,\lambda\text{-conv.}}{\iota(\lambda x_1[book(x_1) \wedge on(\iota(\lambda x_2[table(x_2)]))(x_1)]);}\,{}^{\cup}{}_{\cap}\text{-canc.}}$$
$$XP$$

Proof 27: Kurmancî: *pirtûka li ser masê* 'the book on the table'

These combinatorics show that there is no difference between prepositional phrases and adjectives in the sense that they are properties that can be arguments of functors as well as attributes without distinguishing the two functions morphologically. My view here is supported by the languages that specify the difference between ad-adjectival and ad-genitival ezafat. For instance, in Zazaki, the noun is marked by the ad-attributive ezafe.

(42) pehliwan-ê-do zey to
hero-IND-IND.EZ:/ATT like you.OBL
Zazaki: 'a hero like you' (Berz & Malmîsanij (1951), 163.14 apud Paul (1998b))

See example (42), where the prepositional phrase *zey to* 'like you' is the argument of the functor *pahliwanêdo* 'a hero' marked with the ad-attributive ezafe *-do*. Zazaki considers prepositional phrases to be the same as attributive adjectives in this regard.

3.4.4 Reverse ezafe

The reverse ezafe is, in a sense, an ezafe, and in another sense, not an ezafe. It is a morphological element that that acts as a linker between attributive adjectives and, in some languages, also genitival possessors. The critical difference is that the modifier precedes the noun and hosts the (reverse) ezafe, while the canonical ezafe marks the noun and is followed by the modifier. Interestingly, such a phenomenon would develop within Iranian when both ezafe and reverse ezafe are seemingly rare cross-linguistically.[22] As discussed in section 3.4.1, the typical understanding of adjective is an entity that given a noun renders a noun $N(P)/N(P)$. In languages with (syntactic) determiners, this happens below the phrase level. The adjective is, therefore, the functor, and the noun is the argument.

There is no apparent difference between nouns and adjectives in ezafe languages, and the nouns are morphologically marked for modification. In these languages, paradoxically, nouns are functors, and adjectives (and other nouns) are their arguments. Reverse ezafe languages are more like non-ezafe languages than ezafe languages in this regard. Reverse-ezafe-marked adjectives (and nominal possessors) are functors with nominal arguments. What makes these languages unique is that there is a morphologically-marked form when used as a modifier, and the simplex form is employed for use as a substantive. It is common

[22]The assertion that either of these types of formatives is rare is very difficult to confirm. The term ezafe is Iranian specific (not related to its Arabic namesake iḍāfat). The term reverse ezafe is even more problematic as it competes with terms like attribution marker and genitive marker. Similar constructions in other languages have idiosyncratic terminology (e.g. Albanian nyjë 'knot' particle). No search could quickly produce a list of languages with ezafe and ezafe-like constructions.

for simplex nouns to be substantive with a marked form for the corresponding possessor (i.e. genitive case). However, the idea that the marked form of an adjective is attributive conflicts with preconceived notions about the essence of adjectives (e.g. based on English).

The questions that this section aims to address are (1) what nominal types exist in reverse-ezafe languages; (2) what are the various syntactic functors of reverse-ezafe-marked nouns and adjectives; and (3) how do these forms compare with what is observed in ezafe languages. In reverse-ezafe languages, somewhat unremarkably, attributive adjectives select noun (phrases) as their arguments, and the quantifier of the selected noun distributes over the whole phrase. The entries for adjectives would be as follows:

- Adj/N; $\mathcal{Q}(\lambda x.P_{Adj/N}(x))$; XP

  The default reading for a simplex adjective is substantive. Unlike in Central Kurdish, the quantifier is specified by context in most situations (ι or $\cap$); the simplex form can refer to a definite noun that corresponds to the property P denoted by the adjective or noun, or it can refer to a kind, the group of entities that share the property denoted by the adjective or noun (but not necessarily the whole group). The indefinite substantive would necessarily be morphologically marked.

- Adj-REZ; $\lambda y[let\langle \mathcal{Q}, P_N\rangle := y\ in\ \mathcal{Q}(\lambda x.P_{Adj}(x) \wedge P_N(x))]$; XP/XP

  The reverse-ezafe-marked adjective is then the functor that selects a following nominal. That nominal is a quantified expression that can either be general, definite, or indefinite, although the first two are identical in most reverse-ezafe languages in most functions. The quantifier $\mathcal{Q}$ must distribute over the properties denoted by both the noun and the adjective. It is just this distribution that separates the revere-ezafe-marked adjective from adjective in English (i.e. $\lambda P_N[\lambda x.P_{Adj}(x) \wedge P_N(x)]$; N/N).

  This is illustrated in proof 28, where the Gilaki adjective *rǻst* 'straight, correct' morphologically-marked with the reverse ezafe -ə must take as its complement a fol-

lowing nominal (XP: a category containing both nouns and adjectives). It combines with *divár* 'wall' to form the phrase *rə́stə divár* 'sheer wall.' The phrase is quantified by the determiner $\mathcal{Q}$ which is unspecified here.

$$\dfrac{\dfrac{\dfrac{

which employs a type-lowering functor for attributive constructions and not for possessive ones.

$$\dfrac{\dfrac{\dfrac{\lambda y[let\langle \mathcal{Q}_{N2}, P_{N2}\rangle := y\ in\ \mathcal{Q}_2(\lambda x_2[\mathcal{R}(x_2)(\mathcal{Q}_{GEN}(niece)) \wedge P_N(x_2)])]; \quad \mathcal{Q}_1(\lambda x_1[birthday(x_1)]);}{\lambda y[let\langle \mathcal{Q}_{N2}, P_{N2}\rangle := y\ in\ \mathcal{Q}_2(\lambda x_2[\mathcal{R}(x_2)(\mathcal{Q}_{GEN}(niece)) \wedge P_N(x_2)])](\mathcal{Q}_1(\lambda x_1[birthday(x_1)]));}\ \lambda\text{-conv.}}{let\langle \mathcal{Q}_{N2}, P_{N2}\rangle := \mathcal{Q}_1(\lambda x_1[birthday(x_1)])\ in\ \mathcal{Q}_2(\lambda x_2[\mathcal{R}(x_2)(\mathcal{Q}_{GEN}(niece)) \wedge P_N(x_2)]);}\ \text{let-conv.}}{\mathcal{Q}_1(\lambda x_2[\mathcal{R}(x_2)(\mathcal{Q}_{GEN}(niece)) \wedge \lambda x_1[birthday(x_1)](x_2)]);}\ \lambda\text{-conv.}}{\mathcal{Q}_1(\lambda x_2[\mathcal{R}(x_2)(\mathcal{Q}_{GEN}(niece)) \wedge birthday(x_2)]);}$$

$$XP$$

xaxurzáyə; *təvəllud*;
XP/XP XP /E
xaxurzáyə ∘ *təvəllud*;

Proof 29: Gilaki: *xaxurzáyə təvəllud* 'niece's birthday' (Rastorgueva et al., 2012, ex. 113b)

One thing that sets the reverse ezafe apart from canonical ezafe construction is that they cannot combine with prepositional phrases. The order of the reverse ezafe, modifier-noun, words for noun-noun (possessive), and adjective-noun (attributive) constructions but does not work for PP-noun construction. This problem makes sense in terms of the combinatorics. For instance, the proofs in 30 show how a preposition P is of type XP_{prep}/XP; it combines with a noun XP to form a prepositional phrase XP_{prep} corresponding to a particular preposition. The XP with which it combines can be simplex or complex, but it must be of type XP. This dictates that it be the last member of an ezafe chain, as the ezafe-marked noun is of type XP/XP and cannot felicitously combine. In canonical ezafe languages, a prepositional phrase breaks the ezafe chain. Likewise, in reverse-ezafe languages, a preposition needs a type XP to combine with and could never host an ezafe.

a. $\dfrac{\dfrac{\text{N-EZ; } XP/XP \quad \text{N; } XP}{\text{N; } XP} \quad \text{P; } XP_{prep}/XP}{\text{PP; } XP_{prep}}$ b. $\dfrac{\text{N-EZ; } XP/XP \quad \text{P; } XP_{prep}/XP}{!}$

Proof 30: Broken ezafe chain

Differentiation of reverse-ezafat

Some languages, like Zazaki and Hewramî, for the canonical ezafe construction, have developed unique allomorphs or the reverse ezafe. Two of these languages I examine here are Şirvan Tat (Suleymanov, 2020a) and Baluchi (Barker & Mengal, 2014). The former is a Southwestern Iranian language spoken at the far Northeast of the Iranian zone. It has been in longstanding contact with the Caspian languages, which are known for the reverse ezafe and Azeri (Turkic). The latter is a Northwestern Iranian language spoken in the far southeast of the Iranian zone and has gone through waves of contact with different linguistic groups, most prominently with Indic languages and Brahui (Dravidian) but also including the Caspian languages (Korn, 2019b).

It might reasonably be suggested that the reverse-ezafe constructions in these languages are contact-induced innovations. However, they seem to have recruited native material for the purpose. Additionally, it is not necessarily clear that these constructions represent the same phenomenon. I address these topics in this section.

Possessive constructions in Şirvan Tat In Şirvan Tat there is a reverse-ezafe construction (ex. (43)) by which the adjective is marked by the morpheme *-yä*. This marks the adjective as attributive as opposed to substantive. Suleymanov (2020a) refers to this marker as an attribution marker, which is in line with basic terminology in reverse-ezafe languages. Based on its function, perhaps attribution marker is a better term than reverse ezafe, which is in its nature Irano-centric.

(43) dürgüčü-yä čubon
lying-ATTR(REZ) shepherd
Şirvan Tat: 'lying shepherd' (Suleymanov, 2020a, ex. 324)

Şirvan Tat differs from Gilaki, which has a single reverse ezafe morpheme that can mark adjectives as attributive and nouns as genitive (both being modifiers with similar combina-

torics). Instead, Şirvan Tat possesses a separate possessive construction. Like the reverse ezafe, there is a postpositive morpheme that attaches to the possessor, and it is followed by its possessum. This can be observed in example (44), where *sora* 'Sara' is marked by the morpheme *-ra* signifying, in this case, that it is the possessor. However, this construction differs from the reverse ezafe construction as it was described in the previous section (§.3.4.4) in that the possessum must also be marked with a pronominal clitic indexing the possessor, *=i* '[=3SG]' in example (44).

(44) sora=ra birör=i
Sara=OBL brother=POSS:3
Şirvan Tat: 'Sara's brother' (Suleymanov, 2020b, ex. 29a)

The phrase *birör=i* 'his/her brother' is an XP that can be an argument of a functor (i.e. a verb or adposition) on its own. The phrase *sora=ra*, in this case 'Sara's,' has multiple interpretations. The *ra*-marked form of a noun can be used to mark a direct or indirect object, an experiencer, a possessor, or as a placeholder, a construction where the *ra*-marked noun is indexed on a preposition with a resumptive pronoun (Suleymanov, 2020a). The placeholder construction can be observed in example (45). The proper noun *minǰivon* 'Minǰivon' is marked by the morpheme *-(r)a*, and it is followed by a phrase containing the adpositional phrase *bö* (< bä + ü) 'on them(sg).'

(45) minǰivon=ä bö pö-üst-und
PN=OBL LOC.3SG stand-PST-3PL
Şirvan Tat: 'They stopped at Mincivan.' (Suleymanov, 2020b, ex. 57)

When the *ra*-marked noun is acting as the direct or indirect object, or as the experiencer, they are phrase-level arguments of the type XP subtyped for oblique case (i.e XP_{OBL}). However, when they are acting as possessors or "placeholders," their combinatorics are more complicated. The issue is that without the *ra*-marked argument, both the possessive and placeholder constructions are complete. The possessive in example (44) and the sentence

in example (45) are complete without the *ra*-marked arguments *sora* 'Sara' and *minǰivon* 'Minǰivon;' e.g. *biröri* 'his/her brother' and *bö pö-üst-und* 'They stopped at it.' The implication is that what occurs in both these examples is more akin to argument cross-indexing, where there the possessor is indexed on the noun in example (44). The prepositional complement is indexed on the preposition in example (45).[23] There are two possible ways to deal with these facts. The first and most parsimonious solution is that the sentence or phrase containing the pronominal index can abstract upon the pronominal variable x_{3SG} converting it into a functor $\lambda x_{3SG}.x_{3SG}$. In both the placeholder and possessive constructions, the result is an entity looking for and oblique nominal to its left, $XP_{OBL}\backslash S$ and $XP_{OBL}\backslash XP$, respectively.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\substack{\textit{sorara};\\ s}}\quad \dfrac{}{\substack{\textit{biröri};\\ \mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})}}}{XP_{OBL}\quad XP_{OBL}\backslash XP}\backslash\text{E}}{\textit{sorara} \circ \textit{biröri};}}{\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})](s)}\lambda\text{-conv.}}{\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(s)}\quad \dfrac{\dfrac{}{\substack{\textit{biröri};\\ XP}}}{\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})]}\lambda\text{-abs.}$$

XP

Proof 31: Şirvan Tat: *sorara biröri* 'Sara's brother'

This is exemplified by proof 31, which shows the fully-formed XP *biröri* 'his brother abstracted upon. It then takes the oblique-marked noun (XP_{OBL}) *sorara*. This is parsimonious because the oblique marked noun has the same entry for all of its functions; e.g. N-OBL; $\iota(P)$; XP_{OBL}. Additionally, abstraction on indexed variables is already necessary to account for cross-indexing in other parts of the lexicon.

[23]There is an affinity between the placeholder constructions and Central Kurdish absolute prepositions, which cause the indexing of the adpositional complement. These forms have been analyzed as applicatives because of the way they incorporate adjunct and peripheral arguments into the core argument structure of the verb (Karim, 2021a; Salehi, 2020). To my knowledge, no one has examined Şirvan Tat from a similar perspective.

However, two factors encourage a different solution: (1) if this is a variety of reverse-ezafe construction, one would expect the possessor to be the functor; and (2) the Şirvan Tat possessive construction is thought to be a pattern borrowing of the Azeri possessive construction (Suleymanov, 2020b, 286). This in Azeri (ex. (46)) construction is similar to the Şirvan Tat equivalent (ex. (44)) in that the possessum is marked by a third-person singular (pronominal) possessive.

(46) Sara-nın qardaş-ı
Sara=OBL brother=POSS:3
Azeri (Turkic): 'Sara's brother' (Suleymanov, 2020b, ex. 29')

However, the marker on the possessor is unifunctional, and cannot stand alone. This implies that the "genitive" noun *Saranın* 'Sara's' is the functor, and the XP is its argument. This becomes problematic in light of the fact that *qardaşı* 'his/her brother already contains a possessor; e.g. $\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})]$. This issue requires a hybrid solution; the noun *qardaşı* can be (λ-)abstracted upon, and the functor *Saranın* is a functor looking for a functor (i.e. type-raised). In proof (46), the functor *saranın* ($\lambda P.P(s)$; $X/(X \uparrow XP)$) combines with the λ-abstracted *qardaşı* ($\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})]$; $XP \uparrow XP$). This solution uses only functions that must exist for independently-verified reasons (i.e. λ-abstraction and type-raising), and it is necessary to account for the facts of Azaeri. However, such a solution for Şirvan Tat would unnecessarily violate theory 3.4.1.

The differences between the possessive constructions in Şirvan Tat and Azeri do not preclude the former from being a pattern borrowing from the former. Their structural similarity and close geographic proximity suggest that Suleymanov's (2020b) assertion is correct. This is perhaps a reminder that diachronic facts should not influence synchronic analysis, although they may confirm or provide insight. I address these developments more thoroughly in section 3.6.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\begin{array}{c}\textit{saranın};\\ \lambda P.P(s)\end{array}}}{X \uparrow (X \uparrow XP)} \quad \dfrac{\dfrac{}{\begin{array}{c}\textit{qardaşı};\\ \mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})\end{array}}}{XP}\,\lambda\text{-abs.}}{\begin{array}{c}\textit{qardaşı};\\ \lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})]\end{array}}}{XP \uparrow XP}}{\begin{array}{c}\textit{saranın} \circ \textit{qardaşı};\\ \lambda P[P(s)](\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})])\end{array}}\,\lambda\text{-conv.}}{\lambda x_{3SG}[\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(x_{3SG})](s)}\,\lambda\text{-conv.}}{\mathcal{Q}(\lambda x.brother(x)) \wedge \mathcal{R}(x)(s)}$$

XP

Proof 32: Azeri: *saranın qardaşı* 'Sara's brother'

Possessive constructions in Takestani Tati Like Gilaki and Şirvan Tat, Takestani Tati[24] has a reverse ezafe construction. The attributive construction is a morphological variant of an attributive adjective, where the unmarked form is substantive. Unlike Gilaki (Rastorgueva et al., 2012), the possessive construction is not identical to to the attributive construction (Adj-ɛ 'Adj-ATTR' ∼ N-*e* 'N-GEN'). There is a separate morpheme to express the possessive and attributive functions of the reverse ezafe, but the syntactic combinatorics are the same. This is essentially parallel with the difference between Central Kurdish and Hewramî, the former having a single ezafe representing both the possessed and attributed interpretations and the latter having distinct allomorphs for the two functions. The T Tati possessive construction is illustrated in proof 33, where the noun *mardak* 'man' is marked by the possessive reverse ezafe (i.e. genitive), and it combines with the head noun *das* 'hand.'

Baluchi reverse ezafat Like Gilaki and Tati, Baluchi has variably been described as having reverse ezafat (Haghkerdar, 2009). However, the traditional literature refers to the possessive construction as "genitive" and the attributive construction as "attributive" (following Barker & Mengal, 2014; Axenov, 2006, etc.). This terminology is in line with what has been proposed for Tati as well. Like T Tati, Baluchi has two forms of the reverse ezafe.

[24]The Takestani Tati data I use here has been provided by Neda Taherkhani Stoney Brook University. She is developing a grammar of Takestani Tati.

$$\dfrac{\dfrac{\lambda y[let\langle \mathcal{Q}, P\rangle := \; y \; in \; \mathcal{Q}(\lambda x[P(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(man))])]; \quad \mathcal{Q}_1(\lambda x_1[hand(x_1)); \quad \textit{mardake}; \quad \textit{das};}{XP/XP \qquad\qquad XP}\,/\text{E}}{\dfrac{\dfrac{\dfrac{\lambda y[let\langle \mathcal{Q}, P\rangle := \; y \; in \; \mathcal{Q}(\lambda x[P(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(man))])](\mathcal{Q}_1(\lambda x_1[hand(x_1)))); \quad \textit{mardake} \circ \textit{das};}{let\langle \mathcal{Q}, P\rangle := \; \mathcal{Q}_1(\lambda x_1[hand(x_1))) \; in \; \mathcal{Q}(\lambda x[P(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(man))])];}\,\lambda\text{-conv.}}{\mathcal{Q}_1(\lambda x[\lambda x_1[hand(x_1)(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(man))])];}\,\text{let-conv.}}{\mathcal{Q}_1(\lambda x[hand(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(man))]);}\,\lambda\text{-conv.}$$

$$XP$$

Proof 33: Takestani Tati: *mardake das* 'the man's hand

There is a genitive marker *-ay*[25] [GEN.SG] and *-ī* [GEN.PL], and there is an attributive marker *-ēn* [-ATTR]. However, there is one difference between the Baluchi reverse ezafat and the T Tati ones; in Baluchi, the attributive marker is polyfunctional. In T Tati and Ş Tat, just as in all of the canonical ezafat described thus far, the bare form of an adjective must be interpreted as a substantive $^{\cap}P$. However, in Baluchi, the data is more complicated. Like other Iranian languages, the bare form of the adjective is used in predicative constructions and substantively. However, the "attributive" form is used both attributively and substantively. There are three ways to deal with these facts: (1) the attributive *-ēn* form is substantive, and attribution is done through juxtaposition alone; (2) there are two homophonous suffixes *-ēn* -ATTR and *-ēn* -NMLZ; or (3) the attributive form is, in fact, attributive, and the nominalized function represents something akin to nominal ellipsis, albeit at the morphological level.

It is the third strategy that I favor here. The bare forms behave exactly as they do in other Iranian languages. In example (47), the adjective *ɣarīb* 'poor' occurs as the bare stem and it is clear that it refers to a nominal. Recall that in most Iranian languages, there is no independent marker for definite; I employ the quantifier $\mathcal{Q}$ here to represent the zero-marked form in Baluchi, which can be interpreted as either $^{\cap}$ or ι. In example (47)

[25]The genitive marker *-ay* can be understood as a combination of what Barker & Mengal (2014) call the definite singular marker *-ā* and the normal genitive marker *-ī* (following Axenov, 2006). However, Korn (2005) suggests that this may be the regular outcome of Old Iranian *ahi̯a.

ɣ*arīb* must be understood as ι(*poor*).

(47) y-ē ɣarīb y-ē bēadabī-ā dīst
HI-DEM poor man HI-DEM rudeness-OBJ see.PST.3SG
T Baluchi: 'This poor man saw this rudeness.' (Axenov, 2006, ex. 142)

As expected the bare form can occur declined in all forms that other nominals can. In example (48), ɣ*arīb* is marked for object case -*ā*.

(48) kull-i ɣarīb-ā nazz-ī āwurt u mazan-ēn xayrāt-ē
all-EZ poor-OBJ gathered-ENC.3SG bring.PST.3SG and big-ATTR charity-INDF
dāt
givePST.3SG
T Baluchi: 'He brought together all the poor and made a big charitable offering.'
(Axenov, 2006, ex. 387)

With the attributive suffix -*ēn*, the substantive ɣ*arīb* becomes the attributive adjective ɣ*arībēn* which must combine with a nominal like *zag* 'boy' in example (49). This form has the syntactic combinatorics XP/XP just as the reverse ezafat observed in Gilaki, T Tati, and Şirvan Tat.

(49) nasrō yakk ɣarīb-ēn zāg-ē at.
Nasro one poor-ATTR boy-IND COP.PST.3SG
T Baluchi: 'Nasro was a poor boy'. (Axenov, 2006, ex. 14)

If this were the whole story, the interpretation of Baluchi as a reverse ezafe language would be secure as far as "reverse ezafe" is a salient term. However, the attributive form also occurs in substantive use. In example (50), the numeral *du* 'two' is attributivized *dukēn*, as it normally is when preceding a nominal. However, in this context, it occurs by itself without the modified head noun.

(50) ā du-k-ēn bi-m-ē abar-ay sar-ā at-ant ki bāz
DEM two-HI-ATTR to-EMPH-DEM word-GEN on COP.PST-3PL SUB falcon
āt
come.PST.3SG
T Baluchi: 'Both of them were discussing this problem when the falcon came.' (Axenov, 2006, ex. 607)

Additionally, the attributive adjective can occur in predicate position just like a bare adjective. In example (51), *pruštag* 'broken' is attributized *pruštagēn*, yet it functions as a predicate followed by the copula *atant* 'they were.'

(51) du prušt-ag-ēn at-ant
two break.PST-PP-ATTR COP.PST-3PL
T Baluchi: 'Two were broken.' (Axenov, 2006, ex. 177)

A clue to the correct understanding of these forms is found in the following example (52). Here the adjective *ṭū* 'big (elder)' is used twice with two different contextually identified referents. Both *ṭū-ēn-ā* and *ṭū-ēn-ayā* correspond to 'the elder (brother)' and 'the elder (sister)' respectively, and the referents are contextually retrievable and marked for the relevant cases.

(52) gis dāt ṭū-ēn-ā ṭū-ēn-ayā
house give.PST.3SG big-ATTR-OBJ big-ATTR-LOC
T Baluchi: 'He married the elder (brother) to the elder (sister).' (Axenov, 2006, ex. 146)

Examples (50)-(52) all refer to contextually specified referents. From these data, I propose that this distribution is similar to the English "Adj *one*" construction. Each of these forms has an anaphoric function. This is not problematic because these forms most often distinguish themselves morphologically as they are case marked, unlike attributive adjectives.

In table 3.1, I show a partial adjectival declension. The bare stem of the adjective (i.e. without the attributive marker) can occur declined in all cases. This includes the

Substantive adjective				
mazan	-Ø;		XP_{DIR};	$\mathcal{Q}(\lambda x.big(x))$
mazan	-ā;		XP_{OBJ};	$\mathcal{Q}(\lambda x.big(x))$
mazan	-ay;		XP_{ϕ}/XP_{ϕ};	$\lambda y[let\langle\mathcal{Q},P\rangle := y\, in\, \mathcal{Q}(\lambda x.P(x) \wedge \mathcal{R}(x)(\mathcal{Q}_2(\lambda x.big(x))))]$
etc.				
Attributive adjective				
mazan	-ēn;		XP_{ϕ}/XP_{ϕ};	$\lambda y[let\langle\mathcal{Q},P\rangle := y\, in\, \mathcal{Q}(\lambda x[big(x) \wedge P(x)])]$
Attributive adjective (with anaphorically retrievable referent)				
mazan	-ēn	(ɛ)-Ø;	XP_{DIR};	$\mathcal{Q}(\lambda x[big(x) \wedge P_{\epsilon}(x)])$
mazan	-ēn	(ɛ)-ā;	XP_{OBJ};	$\mathcal{Q}(\lambda x[big(x) \wedge P_{\epsilon}(x)])$
mazan	-ēn	(ɛ)-ay;	XP_{ϕ}/XP_{ϕ};	$\lambda y[let\langle\mathcal{Q},P\rangle := y\, in\, \mathcal{Q}(\lambda x.P(x) \wedge \mathcal{R}(x)(\mathcal{Q}(\lambda x[big(x) \wedge P_{\epsilon}(x)])))]$
etc.				

Table 3.1: Baluchi adjectival inflection *mazan* 'big'

genitive, where it becomes a functor looking for a nominal possessum. The attributive marker converts the adjective into a functor looking for the noun that it modifies, bringing a quantifier and a property. In this function, the adjective is never declined for case or number, but the noun that it modifies must be. When the attributive form occurs on its own, it must be marked for case and number. However, there is a referent that is missing from the morphological form that must be resolved. There is something akin to NP ellipsis built into the inflected forms. (53) shows the principle by which this anaphora should be resolved.

(53) Principle of anaphora resolution on the XP with ellipsis built into the morphological paradigm:

a. If there is a syntactic constituent with category XP in the antecedent clause, then the value of P_{ϵ} is identified with the denotation of that constituent.

b. If there is no such syntactic constituent, then the value of P_{ϵ} is anaphorically identified with some salient property in the discourse that is the syntactic category XP.

Given the difference in the analyses of parallel forms like *mazanā*; XP_{OBJ}; $\mathcal{Q}(\lambda x.big(x))$ and *mazanēnā*; XP_{OBJ}; $\mathcal{Q}(\lambda x.big(x) \wedge P_{\epsilon}(x))$, it is reasonable to propose that the attribu-

tive form marked by -*ēn* is in fact attributive, and the nominalized function represents nominal ellipsis at the morphological level. The phonetically identical forms *mazanēn*; $\lambda y[\lambda\mathcal{Q}[\lambda P[\mathcal{Q}(\lambda x[big(x) \wedge P(x)])]]y]$; XP_{ϕ}/XP_{ϕ} and *mazanēn*; $\mathcal{Q}(\lambda x.big(x) \wedge P_{\epsilon}(x))$; XP_{DIR} are, therefore, a case of accidental synonymy based on the fact that the direct case ending in Baluchi is -$\emptyset$. Of course, syncretism between paradigm cells is common crosslinguistically. Despite the additional complexity created by the two types of adjectival substantives, the bare adjectives and the forms with integrated nominal ellipsis, Baluchi is a typical reverse ezafe language. It has both reverse ezafe types, an attributive suffix (*Adj-ēn*; XP/XP; $\lambda y[\lambda\mathcal{Q}[\lambda P[\mathcal{Q}(\lambda x[P_{Adj}(x) \wedge P(x)])]]y]$) and a genitive case (*N-ay*; XP/XP; $\lambda y[\lambda\mathcal{Q}[\lambda P[\mathcal{Q}(\lambda x[\mathcal{R}(x)(\iota(P_N)) \wedge P(x)])]]y]$).[26]

The moniker reverse ezafe I think the term reverse ezafe adds nothing to our understanding of attributive constructions in Iranian languages. The term is born of a desire for a unified understanding of modification in Iranian languages. The term ezafe (< Arabic *iḍāfat* 'the construct state') is deeply integrated into the Persian grammatical tradition and has been extended to refer to various constructions among the Iranian languages. In the sense that all of these are, generally speaking, "linkers," they can be seen as a continuum of morpho-syntactic phenomena. This is the general approach of Haghkerdar (2009); Larson & Samian (2020); etc. However, these phenomena are different enough to be considered separate entities.

Moreover, there are similar/identical phenomena in other languages and other established terminology that should take precedence over a coinage like reverse ezafe e.g. attributive state, or following Rießler's (2016) exploration of attribution strategies crosslinguistically, anti-construct. Anti-construct is perhaps just as reasonable as reverse ezafe when considering attributive adjectives. However, for nominal possessors, genitive case or possessive state are well established, and their meaning is transparent.

[26]Baluchi features different allomorphs for genitive case in the context of plural and with pronouns.

3.4.5 Notes on compounding strategies

In the languages with definite ezafat, including Soranî and colloquial New Persian, the definite ezafe allomorph is used to make nominal compounds. For example, Soranî *gu-le sûr* 'rose' is different from *gu-lî sûr* 'red flower' (Thackston, 2006b, 13). The difference is only that the former uses the definite ezafe allomorph *-e* to connect *gul* 'flower' and *sûr* 'red.' The latter uses the non-definite ezafe *-î* to connect them. This fact has led Samvelian (2005) to reject the view that the definite ezafe is a syntactic construction at all but rather a form of compounding (Samvelian, 2005, 18). According to Rießler (2016), compounding is a strategy that some languages use as a method of nominal attribution. In principle, this conclusion is not an issue. However, when the construction is truly productive, and there are no phonological consequences of the merger of the two lexemes; relegating the function to morphology is perhaps reductionist.

I believe that this solution misses an important generalization. The function that I have proposed for the definite ezafe in section 3.4.3 requires quantification to be distributed across the nominal elements. This distributive property of the definite ezafe is part of the coordination operator *û* 'and' which is used in so-called group inflection, or grammatical marking on only the final member of a conjoined phrase. The conjoined phrase must be interpreted as a single set consisting of various entities, some of each of the various conjoined types. Based on these facts, it is perhaps no surprise that In Soranî, the *û* operator is also used for compounding. For instance, *teng-û-çeleme* 'trouble [lit. narrow and mucous],' *cwan-û-esp* 'a male foal [lit. a male youth and a horse],' *řaw-û-řawjgar* 'a hunting trip [lit. hunting and spending-a-day].' It might reasonably be theorized that languages that only allow simple coordination $\wedge$; $(X)/x$ (i.e. that don't feature group inflection) would not develop compounds from the coordination operator. These types of compounds are pervasive across the Iranian world (e.g. in Soranî, colloquial New Persian Balochi, etc.), all of which feature group inflection.

3.5 Typology of ezafat

I describe many diverse constructions, some of which have been described as ezafat in the literature some of the time and others all of the time. I present a brief typology of ezafat employing terminology from Rießler's (2016) typology of attribution marking. The construct refers to attribution marking on the head (ezafe); anti-construct refers to attribution marking on the dependant (reverse ezafe); and floating construct refers to a situation where the attribution marker is neither closer to the head nor the dependant. The forms in table 3.2 represent the types of ezafat based on how they appear in the languages referenced in this study. Two unifying features characterize these languages (Soranî, Kurmancî, Hewramî, Southern Zazaki, Gilaki, Takestani Tati, Şirvan Tat, and Turkmen Baluchi):

- **N = NP (XP here):** The inflected stem of a noun carries as part of its morphology a determiner. In some of these languages, there is a morphological marked three-way distinction; e.g. Soranî: definite N-*eke*, indefinite N-*êk*, or generic N-Ø. In others, there is a two-way distinction; e.g. Farsi: definite/generic N-Ø, indefinite *ye* NP-*i*. However, many of the languages that collapse definite and generic only collapse the distinction in certain cases; e.g. Zazaki: nominative: definite = generic N-Ø but accusative: definite N-*ê* $\neq$ generic N-Ø. In these languages, nouns not marked for attribution should be understood as type $\mathcal{Q}(P)$ where $\mathcal{Q}$ is a quantifier $\iota, \exists$, or $^{\cap}$, and P is a property corresponding to the noun.

- **Adj = NP (XP here):** The bare stem of the adjective is substantive. It is converted into a modifier either by an attributive suffix (in reverse-ezafe languages) or by juxtaposition with a noun marked for adjectival modification (in ezafe languages). In these languages, adjectives not marked as attributive should be understood as type $\mathcal{Q}(P)$ where $\mathcal{Q}$ is a quantifier $\iota, \exists$, or $^{\cap}$, and P is a property corresponding to the noun.

In attributive constructs, they appear only as with the quantifier $\cap$, and in definite attributive constructs, they appear only with the quantifiers ι.

Canonical Ezafat	Prosody	Syntax	Semantics
Possessive Construct	N-EZ;	XP/XP;	$\lambda y[\mathcal{Q}(\lambda x[P_N(x) \wedge \mathcal{R}(x)(y)])]$
Attributive Construct	N-EZ;	XP/XP;	$\lambda y[\mathcal{Q}(\lambda x[P_N(x) \wedge^{\cup} y(x)])]$
Definite Ezafat			
Definite Att. Construct	N-EZ;	XP/XP;	$\lambda y[let\langle \mathcal{Q}, P_{Adj}\rangle := y\, in\, \mathcal{Q}(\lambda x[P_N(x) \wedge P_{Adj}(x)])]$
Reverse Ezafat			
Att. Anti-construct	Adj-ATTR;	XP/XP;	$\lambda y[let\langle \mathcal{Q}, P_N\rangle := y\, in\, \mathcal{Q}(\lambda x[P_N(x) \wedge P_{Adj}(x)])]$
Possessive State (GEN)	N-GEN;	XP/XP;	$\lambda y[let\langle \mathcal{Q}, P\rangle := y\, in\, \mathcal{Q}(\lambda x[P(x) \wedge \mathcal{R}(x)(\iota(P_N))])]$
Secondary Ezafat			
Att. Floating Construct	(=)EZ;	$XP\backslash(XP/XP)$;	$\lambda y[let\langle \mathcal{Q}, P\rangle := y\, in\, \lambda z[\mathcal{Q}(\lambda x[P \wedge^{\cup} z)]]]$
Pos. Floating Construct	(=)EZ;	$XP\backslash(XP/XP)$;	$\lambda y[let\langle \mathcal{Q}, P\rangle := y\, in\, \lambda z[\mathcal{Q}(\lambda x[P \wedge \mathcal{R}(x)(z))]]]$
Not Ezafat			
Possessor Cross-indexing	N-POSS:ϕ;	$XP_{OBL}\backslash XP$	$\lambda y_{\phi}[\mathcal{Q}(\lambda x[P_N(x) \wedge \mathcal{R}(x)(y_{\phi})])]$

Table 3.2: Typology of ezafat

Table 3.2 can be understood in light of these two features. Here, I continue the primitive XP to refer to the category containing adjectives, nouns, prepositional phrases. Canonical ezafat are marked on the head noun and recursively on each subsequent modifier. These forms can be attributive or possessive. Additionally, the canonical ezafat can be definite or unspecified. The unspecified ezafat are the most commonly studied type as is known from standard New Persian. The head noun is marked for modification and is followed by a nominal possessor (possessive construct). In the attributive construct, the following nominal must be in the generic form, and it is type-raised by the ezafe functor becoming a property. The generic quantifier $\cap$ takes a property P and returns a kind $^{\cap}P$, which is to be understood as some subset of the set of entities denoted by P (type e) (following Partee, 2002). The attributive ezafe functor contains the predication operator ($^{\cup}$), which converts the generic substantive into an associated property ($^{\cup\cap}P = P$). It is this type-raising function–the conversion of a substantive into an attribute–that characterizes the canonical attributive construct, the ezafe. Furthermore, this type of ezafe is derivational

(not necessarily inflectional[27]); it converts an XP into an XP/XP. In some languages (e.g. Kurmancî, Zazaki, etc.), the possessive ezafe is necessarily followed by a noun in the oblique case. I have taken this to be a sub-typing relationship; The possessive construct is of type XP/XP_{OBL}. This analysis is justified because the oblique form has other functions within the language (e.g. past-tense transitive agents, complements of prepositions, etc.).

The reverse ezafat are, in a sense, not ezafat. The possessive construction is identical to a genitive construction in any other language with the caveat that the entity that is its complement is a phrase-level entity. It must, therefore, distribute its quantifier over the whole expression; this is not rare cross-linguistically as many languages lack syntactic determiners. The attributive anti-construct works in the same way as the possessive state except that it represents a property P_{Adj} of the following head noun and not a relation $\mathcal{R}$.

The definite ezafat are characterized by the word order of the canonical construct $NP > Adj$ and the semantic functor of the reverse ezafe. They take a following substantive and distribute the corresponding properties over the whole noun phrase. Still, it is the ezafe-marked head noun that precedes the attributive adjectival substantive marked for definiteness. Although these constructions differ according to which entity carries the quantifier distributed across the whole phrase, they are otherwise identical. Given that there is no discernible difference between adjectives and substantives, it may be beneficial to see the definite ezafat as reverse ezafat within a canonical ezafe language.

The secondary ezafat are purely syntactic. That is to say that they are independent words that attach to an XP putting it into the construct state. The secondary construct is an independent word, although it is considered, at least orthographically, enclitic in some languages (e.g. Zazaki). I use the term "enclitic" here to mean an independent word that forms a prosodic unit with the preceding word but has its own combinatorics both independent from and not projected by its phonological host.[28] The secondary ezafat

[27] In some languages (e.g. Zazaki, Kurmancî, Hewramî, etc.), the ezafe morpheme has univerbated with other inflectional affixes yielding single morphs that correspond to both inflectional and derivational features.

[28] This definition of (en)clitic is my own, mainly based on the combinatorics of the secondary ezafat. However, this definition does not add anything to what is know about clitics. Instead, my definition of

are only different from the primary (canonical) construct/ezafat according to the syntactic functor. It is necessary for Kurmancî and Zazaki varieties, which do not have a way to nest modifiers within a single phrase. It can take a word modified by the possessive ezafe and also modify it with an attributive one or vice versa.

Possessor cross-indexing is a strategy that is wholly different from the ezafe. Although it superficially operates like the possessive state, neither the nominal possessor nor the morphologically marked pronominally-possessed noun is a functor. Instead, the nominal occurs cross-indexed on the noun and is akin to the optional cross-indexing seen in verbal systems.

Language	Att. Floating Construct	Poss. Floating Construct	Possessive construct	Attributive construct	Definite Att. Construct	Attributive Anticonstruct	Possesive State	Possessor Crossindexing
Kurmancî	✓	✓	✓	✓				
Southern Zazaki	✓	✓	✓	✓				
Soranî			✓	✓	✓			
Hewramî			✓	✓	✓			
Colloquial New Persian			✓	✓	✓			
Standard New Persian			✓	✓				
Şirvan Tat			✓	✓		✓		✓
Gilaki						✓	✓	
Takestani Tati						✓	✓	
Turkmen Baluchi						✓	✓	

Table 3.3: Ezafe types by language

The languages considered in this study are Kurmancî, Southern Zazaki, Soranî, Hewramî,

clitic can be understood as an amalgamation of Klavans's (2017) typology based on the independence of phonology and syntax. I also follow Zwicky's (1994) assertion that rather than three categories word, clitic and affix, there is a two-way distinction between words and affixes, clitics being either atypical words or atypical affixes, and Zwicky's (1987) assertion that clitic be "prosodically dependent." Note that Zwicky's (1994) notion of "typicallity is needed independently; one need not propose any new notions to implement his two-way distintion between word and affix.

New Persian (colloquial and standard), Şirvan Tat, Takestani Tati, and Gilaki. Table 3.3 shows each of these languages and the ezafe types they possess. The table does not refer to whether or not there are distinct allomorphs for each ezafe type within the system. Generally speaking, if a language employs a particular strategy for modification marking, it must have both an attributive and possessive version of the ezafe functor. This is true for all languages that employ and floating construct and an attributive construct and a definite construct. However, Şirvan Tat breaks this trend; it has an attributive anti-construct (reverse ezafe) state but not the corresponding possessive state. In place of the possessive state construction, it employs possessor cross-indexing. As suggested by Suleymanov (2020a), this is the likely result of longstanding contact with Azeri (Turkic), which employs both possessor cross-indexing and the possessive state. Şirvan Tat is most closely related to standard New Persian, which has only the canonical ezafat (i.e. the attributive and possessive construct states). It has innovated the attributive anti-construct in contact with the Caspian languages (e.g. Tati and Gilaki) and possessor cross-indexing in contact with Azeri. The two languages missing the definite construct only employ the canonical construct forms in non-definite contexts. The theoretical implications of these facts are discussed in the following section 3.6, which focuses on the diachrony of the interface between the syntax and semantics of the ezafe.

3.6 Diachronic patterns in the development of ezafat

Very little is known about most of these languages going back into history. Our knowledge of Iranian is much like an hourglass: due to the nature of genealogy, as we go further back into time, diversity decreases. We have a reasonably detailed picture and rich understanding of Proto-Indo-Iranian and Proto-Indo-European through comparative historical linguistics. This has been greatly augmented by the fact that Avestan on the Iranian side and Sanskrit on the Indic provide attested forms going back more than three millennia. Going into

Middle Iranian and early New Iranian, we know almost nothing. Today, we have a wealth of living New Iranian languages that we continue to know more about each day. However, the diversity is so rich that it is not necessarily clear what took place on the diachronic path from Old Iranian to any given New Iranian language. With the possible exception of Middle and New Persian, it is not a stretch to say that none of the known New Iranian languages are the direct descendants of any of the extant Old and Middle Iranian languages. Ideally, the Middle Iranian languages that were the parents of Kurdish, Zazaki, or Central Iranian Kermanic languages could be reconstructed through the comparative method just as PIE was reconstructed (and continues to be) from comparing the extant old Indo-European languages. In other words, New Iranian evidence can help us understand what earlier iterations of the known languages were like. Based on the synchronic syntactic analysis I have proposed in section 3.4, several patterns have emerged. This section on the diachrony of the ezafat is primarily focused on observing these patterns and proposing a narrative for what significance they might bear on the development of ezafe marking systems. Each of the following sections continues a problem or issue that arose as part of the discussions in sections 3.4.3 and 3.4.4. These issues are presented in no particular order.

3.6.1 Directionality

Several patterns have emerged from the distribution of ezafe types I have outlined in table 3.3, the proposed ezafe functors in table 3.2, and the overview of New (Western) Iranian nominal morphology (ch. 2): (1) languages which have floating ezafat (Kurmancî and Zazaki) also have a canonical ezafe construction but not a definite ezafe construction. (2) Languages with a definite ezafe construction correspond to the languages with a definite suffix (see ch. 2). (3) the definite ezafat and the reverse ezafat have the same semantic functors attributive: $\lambda y[\lambda\mathcal{Q}[\lambda P_1[\mathcal{Q}(\lambda x[P_N(x) \wedge P_2(x)])]]y]$; XP/XP and genitive: $\lambda y[\lambda\mathcal{Q}[\lambda P_1[\mathcal{Q}(\lambda x[P_N(x) \wedge \mathcal{R}(x)(\iota(P_2))])]]y]$; XP/XP. An additional fact that seems relevant here is that all Iranian languages, regardless of the attribution system they have

developed, have an adjective-noun word order in at least some constructions. For instance, Soranî, which is a canonical ezafe language (i.e. N-EZ Adj), also has some adjectives (e.g. superlatives) which precede the head noun (ex. (54)).

(54) cwan-tirîn kiç-êk
beautiful-SUPL girl-INDF
Soranî: 'the most beautiful girl'

This word order is also attested more broadly in Western Middle Iranian languages. It occurs extensively in Parthian and sporadically in Middle Persian, which had already come to rely on the Ezafe construction. Skjærvø (2009b) refers to the simple juxtaposition of an adjective preceding a noun (e.g. (55)) in Western Middle Iranian as the "older and rarer construction."

(55) weh dīn ∼ yazd-ān nām
good religion *sim* god-GEN.PL name
Middle Persian: 'the good religion' ∼ 'the gods' name' (Skjærvø, 2009b, 222)

Here, I propose that the reverse-ezafe, a genitive, and an attributive construction, is probably a retention from the inherited Old Iranian system of adjectival and genitival modification. Now that more is known about the spread of the definite article in Iranian, it is possible to say that it is reconstructible for both common Northwestern Iranian and common Southwestern Iranian, albeit from different sources. The definite ezafe construction, the one nested inside of a definite article, employs the same syntactic and semantic functor as the reverse ezafe construction and is, in my view, a natural bridging context for the shift from reverse ezafe to canonical ezafe. If the reverse ezafe construction is Adj-EZ NP-*def* but Adj = NP (i.e. XP), then it matters not which property comes first in such a construction XP-EZ XP-*def*. The relative pronoun was inducted into the morphological system when integrating multiple modifiers and modifiers of different types following the model of the definite ezafe. Some of these languages subsequently lost the definite article

and, therefore, the definite ezafe construction. For languages like Gilaki and Takestani Tati, this loss occurred before the development of a canonical ezafe. For languages like Standard (not colloquial), New Persian, Kurmancî, and Zazaki, this happened after the development of the ezafe. Of course, there is no sign of this definite construction in Middle Persian or Parthian. However, its survival in colloquial New Persian leads me to believe that it was once present in a common New Persian.

It is my assertion here that there is a cline from reverse ezafe marking, actually, just genitive and attributive marking, to a definite construct, to a canonical ezafe, and in the case of Kurmancî and Zazaki, the loss of definiteness marking lead to the development of a floating ezafe. This is motivated by retentions of older systems in subsets of the grammar, the lexical entries for these forms (in the CG sense of the term), and the distribution of these formatives across the Iranian languages. Additionally, the nominalizing function with "deictic force," or the nominal use of the attributive forms, only occurs when the ezafe marker is close to the adjectival form. The reverse ezafe and the floating ezafe have the capacity for such usage. Still, I do not predict such an outcome in languages with only a canonical ezafe without a dummy noun to host the ezafe such as Soranî *hîn* 'thing' in *hînî baş* 'a good thing.'

The ad-genitival ezafe, possessive construct, is of type XP/XP_{OBL} an entity looking for a nominal in the oblique case on its right to yield a nominal. This subtyping relationship is justified because the oblique form has other functions within the languages that have preserved case marking (e.g. past-tense transitive agents, complements of prepositions, etc.). However, the oblique forms began their life as the inherited genitive case (oblique feminine *-ê* < *-āyā, oblique plural *-an* < *ānām). This suggests that the possessive construct forms went through a stage where the oblique form was a genitive case (reverse ezafe/anti-construct/possessive state) and the pre-ezafe was looking for this type: N-GEN; XP/XP and N-EZ; $XP/(XP/XP)$. Rießler (2016) has deemed this type of construction, the double construct, a noun in the construct state, and a modifier in the anti-construct. I assert that

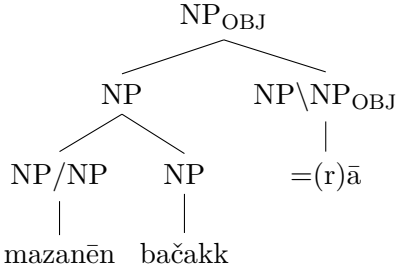
this is not a good synchronic analysis for Kurmancî but that it was a likely stop on its way to the current system. This could be a clue to another way such a reanalysis could take place. The functor XP/XP gets reinterpreted as XP_{OBL} as it becomes the argument of another functor.

3.6.2 The development of nominalizers from the construct state

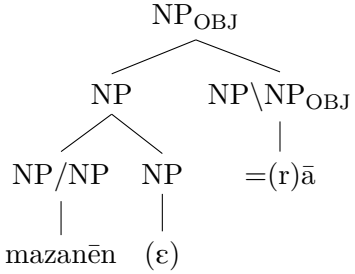
The first issue that I address here is based on the fact that there are nominalizers that appear to be related to the various ezafe constructions in these Iranian languages. I assert that the bare form of adjectives is substantive is a key factor in the development of ezafe constructions. However, in Baluchi, attributive marking could additionally signify a nominalized form albeit with "deictic force" (following Korn, 2008a). A form like *mazanēn* is, strictly speaking ambiguous between the two readings XP_{ϕ}/XP_{ϕ}; $\lambda y[\lambda\mathcal{Q}[\lambda P[\mathcal{Q}(\lambda x[big(x) \wedge P(x)])]]y]$ and XP_{DIR}; $\mathcal{Q}(\lambda x[big(x) \wedge P_{\epsilon}(x)])$. The former is the expected attributive functor, and the latter is the same functor with ellipsis of its nominal complement. This is superficially ambiguous only in the unmarked direct case. However, the syntactic combinatorics disambiguate the forms. The other case endings are thought to have originally been clitic forms that would take a preceding noun phrase and return a noun phrase marked for case (ex. (56)a). When the NP consisted of an attributive adjective and a nominal, the nominal element could be omitted through nominal ellipsis (ex. (56)b). The clitic element then fell on the attributive adjective, eventually univerbating. This univerbation created a form that was paradoxically a nominalized adjectivalized nominal.[29]

[29] The use of tree diagrams here is simply an illustrative tool and should not be interpreted as a description of structure other than what is epiphenomenal of the syntactic combinatorics.

(56) a.



b.



The locative case in Turkmen Baluchi

Given this type of change with the anti-construct in Baluchi, it stands to reason that the same would be true of the possessive state (e.g. *šwānag-ay-ā 'the shepherd's (ε)'). There are no such forms in Axenov's (2006) thesis (Turkmen Balochi). However, there are several in Barker & Mengal (2014) like example (57)[30], where the genitive marker is followed by the oblique marker implying that the noun *olakã* 'cattle' modified by the possessor *vat* 'self' is missing but retrievable from context.

(57) man bānd-ā vat-ī-ā[n=a] bar-īn
1SG tomorrow-OBL REFL-GEN-OBL.PL[=IPFV] bring.PRS-1SG:A
R Balochi: 'I will bring mine own (ε) tomorrow' (Barker & Mengal, 2014, 152)

There is a construction in Turkmen Balochi that appears to be made up of the same formatives. The locative case in Turkmen Baluchi is built from the genitive suffix *ay/ī* and the object case ending -*ā* yielding *ayā/īā*. However, the term locative is perhaps an oversimplification of the actual function of this suffix. Its meaning can be locative, dative, or allative. Within the locative realm, it can be essive, superessive, inessive, etc. The question becomes how did these functions develop from a genitive and object case. This question was solved by Korn (2008a), who first proposed that the form of the locative carried a "deictic force" much like the English "at my uncle's" (Korn, 2008a, 94). This solution works quite

[30]Examples from Barker & Mengal (2014) have been altered to reflect the orthography employed by Axenov (2006). Additionally, examples have been altered to include the imperfective markers that Barker & Mengal (2014) systematically removed from his manuscript (See Barker & Mengal, 2014, page 149 for their motivations for removing this particle.)

well with my assertion of the nominalized attributives. However, there is one more thing to be said about these locative constructions in addition to what Korn (2008a) proposed. In Baluchi (including the variety of Turkmenistan), adpositional phrases consist of the nominal head in the genitive case possessing a nominal element with a relational denotation. For instance example (58) shows the typical superessive construction literally 'on the house's head with *gis* 'house' in the genitive case and *sar* 'head' in the object case.

(58) gis-ay sar-ā
house-GEN head-OBJ
R Balochi: 'on the house'

The superlative construction (ex. (59)) signifying movement onto something is signified by the same construction as the superessive in example (58) with the simple preposition *bi* 'to.'

(59) bi gis-ay sar-ā
to house-GEN head-OBJ
R Balochi: 'on the house' (Axenov, 2006, ex. 62)

In Korn (2008a), she examines the distribution of the locative case in Afghan and Turkmen Baluchi. She shows that certain prepositions govern the locative case, some that never do, and some that sometimes do. One of these forms is *čō* 'like' glossed by Barker & Mengal (2014) as 'so' for Rakhshani Baluchi (Pakistan). Assuming nominal ellipsis as the source for the locative construction, a viable bipartite construction–or many bipartite constructions–could explain which prepositions govern the locative case. For *čō* such a possibility is *čō* N-*ay ḍawl-=a* 'in such a manner.' This form occurs in Rakhshani Baluchi as illustrated by example (60). Where 'like fire' is expressed with a bipartite construction containing the prepositional element *čō* and the nominal element *ḍawl* 'kind, type, manner.' With nominal ellipsis, the role of the postposition is subsumed by the prepositional element alone.[31]

[31]There is a tendency among the western Iranian languages to lose one element of a bipartite (circumpositional) construction. This is certainly true of Kurdish varieties that employ circumpositions in formal

(60) ā šāir-ay gal čō ās-ay ḍawl-ā soč-ant
that poet-GEN words as fire-GEN manner-OBJ burn.PRS-3PL
R. Baluchi: 'the words of that poet burn like fire.' (Barker & Mengal, 2014, 461)

Additionally, the locative case occurs rarely with the preposition *gō* 'with,' which is surprising as the preposition does not govern the object case and does not appear from Axenov (2006) or Barker & Mengal (2014) to occur together with postpositions. Korn (2008a) gives example (61) which appears to show the locative case in combination with the preposition *gō*.

(61) ē zarr-ānā mn-ī wa watī-ānā gōn gō kirē-y-ā
DEM money-OBJ.PL I-GEN and own-OBJ.PL together with rent-GL-OBL
kōṭ-y-ayā dāt-Ø
room-GL-LOC give.PST-3SG
"[she] gave this money for me and [her]self toghether with the rent for the room"
(Buddruss, 1988, §133 apud Korn 2008a, ex. 10)

However, there is intervening matter *kirēyā* 'the rent' in the oblique case which is modified by *koṭā* room seemingly in the locative case. This particular construction has parallels in the Baluchi system concerning attributive adjectives. When an attribute is emphasized, it is repeated after the noun it modifies, and the ellipses form of the adjective is employed. This is illustrated in example (62), where the noun *ǰwab-ā* 'answer' is followed by the modifier *mardī* 'manly' which is marked for attribution and the case of the ellipsis form. According to Axenov (2006), this is the form of an adjective when the modifier is emphasized.

(62) Nadiršā ki tāǰir-ay ǰwab-ā mardī-ēn-ā uškit, ā-ī
PN *sub* merchant-GEN answer-OBJ manly-ATTR-OBJ hear.PST.3SG EM-GEN
mal-u-dunyā-ā padā dat
property-and-world-OBJ back give.PST.3SG

language but loose wither the preposition or the postposition in normal speech; compare the circumpositional Basselhâya: *di mal da* 'in the house,' the prepositional Hewlêr: *le maɬ* 'id.,' and the postpositional Bingöl [Çewlîg]: *malê da* 'id.' (Matras et al., 2016).

'When Nadir Shah heard the brave (manly) answer of the merchant, he gave back his property.' (Axenov, 2006, ex. 147)

Korn (2008a) shows that regardless of the case of the head noun, the locative case can be used when a modifier is postposed (e.g. *emē aspā bēzēnayā* 'this horse without-a-saddle'). Korn (2008a) parses *bē-zēn-ayā* as [without-saddle-LOC]. However, I believe that in this case, as in example (61), that the correct parsing should be *-ay-ā* as [-GEN-OBJ]. The reason this attributive compound appears with a genitive ending and not an attributing is likely sue to the nominal character of *zēn* 'saddle.'

I assert that both the substantivized attributive forms *-ēn-ā* etc. and the substantivized possessive forms *-ay-ā* should be treated as if they are cases of nominal ellipsis. In some cases, these forms truly have a deictic character referring back to an entity already introduced or otherwise available from context. Additionally, each of these forms can serve to emphasize a preceding (adjacent or otherwise) noun. There are a select few de-nominal constructions involving either a relational noun in the object case preceded by the head noun in the genitive case. There is another set of constructions where a preposition precedes the de-nominal postpositional construction. In both of these contexts, the relational noun may be ellipsed, rendering the so-called locative case ending *ayā*. This form can occur with prepositions like *gō* 'with' only when emphasized hence the rarity. However, the locative case ending occurs alone or along with prepositions like *bi* 'to,' which historically occurred with a now ellipsed postpositional element.

nominalizers in Kurmancî

Just like in Baluchi, there are a series of ways to represent an adjectival substantive. The bare stem of the noun is naturally substantive, and it can be converted into a property (attribute) by a preceding ezafe functor. The secondary construct form can also attach to a following modifier without a head noun to form a demonstrative form (e.g. *yê/ya/yên*

baş 'good one(s) [M/F/PL]' or *yê/ya/yên wî şivanî* 'that shepherd's one(s) [M/F/PL]'). This form is referred to by Samvelian (2008) (following Mackenzie, 1961) as the demonstrative ezafe. However, Samvelian (2008) and Mackenzie (1961) alike group the secondary ezafe and the demonstrative ezafe together. I assert that this grouping is not exactly unwarranted, but it misses a few crucial points that should be clear in light of the Baluchi data.

The three functions of the demonstrative and secondary ezafat mentioned by Samvelian (2008) show a parallel to the attributive markers in Baluchi. The "demonstrative ezafe" occurs when " (1) preceding a noun, a pronoun or an adjective modifier and giving a possessive or a substantive sense to the whole group, (2) in all cases when the modifier is separated from the head it modifies, and (3) when the head noun is followed by more than one modifier" (Samvelian, 2008, 355). The first criterion introduces the nominalizing function of the ezafe, which is paradoxical as simplex adjectives are already nominals. The second and third are what set the secondary ezafe functions apart from the canonical ezafe. The secondary ezafe is a floating construct state, which is to say that it is neither bound to the head nor the modifier. This construction is employed when changing modifier types. It is the second function that is the "demonstrative" function. It can be used like substantivization albeit with a deictic character much like in Baluchi; compare Baluchi: *faqīrēn* '(the) poor (one)' with Kurmancî *yê faqîr* 'id.' Additionally, when the noun is dislocated from its modifier, the demonstrative ezafe must be used just like the substantivized attributive form in Baluchi; see example (63), where the (demonstrative) substantivized adjective *granbā* 'expensive' modifies the predicate *zarr-ē* 'money.' It is separated by the copula from the noun that it modifies.

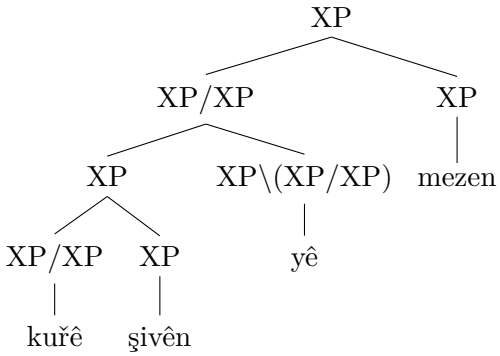
(63) tī ǰind zarr-ē w-ay granbā-ēn
you.SG.GEN self money-IND HI-COP.PRS.2SG heavy.price-*attr*
Baluchi: 'You yourself are a precious treasure.' (Axenov, 2006, ex. 494)

Similarly, in Kurmancî, The demonstrative ezafe is used when the modifier is extraposed (ex. (64)). This sentence can be literally translated as 'he is a man, the strong (one).'

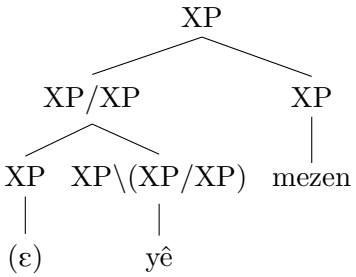
(64) ew mirov =e yê bi-qiwet
PN.3.DIR man.DIR =COP.3SG EZ.M.SG with strength
Kurmancî: 'he is a STRONG man.'

I assume that as the base form of an adjective is already substantive in Kurmancî that the demonstrative ezafe forms can be regarded as the result of a kind of nominal ellipsis like I reconstruct for a stage of the Baluchi forms. Although the scenario which conditioned the secondary ezafe is perhaps more complicated than the attributive marker in Baluchi, its interpretation is the same. The construction must refer to some entity retrievable from context (see ex. (65)). The significant difference between the Kurmancî and the Baluchi forms is that the Kurmancî construction remains purely syntactic likely since the secondary ezafe is neither completely bonded to the head or the modifier.

(65) a.



b.



The Baluchi construction is bonded to the modifier, allowing the head noun to be ellipsed. The Kurmancî form is neither bonded to the head nor the modifier allowing the head noun to be ellipsed. However, other (canonical) ezafe constructions are bonded to the head noun, which precludes the possibility of being ellipsed, leaving behind a new seemingly substantivized construction. Because these types of constructions have not developed in canonical ezafe languages that haven't also developed a floating ezafe, I propose that this

is an innovative construction in Kurmancî and Zazaki. This goes against the assumptions of Samvelian (2008), Franco-Rita & Savoia (2012) and others, implied and overt that the demonstrative ezafe is a conservative feature descended from the ezafe's original relative function. This is discussed further in the context of Rießler's (2016) diachrony of attribution markers in section 3.6.3 below. Additionally, chapter 4 expands upon this discussion in the context of the development of Kurmancî and Zazaki's innovative–albeit with an "archaic" quality–nominal inflection systems.

3.6.3 Against Rießler's (2016) diachrony (DEM > NMLZ > ATTR)

Rießler (2016) discusses the ways that several attribution-marking strategies developed diachronically. Regarding the development of the anti-construct and construct states (reverse ezafe and ezafe), he introduces two possible scenarios: (1) a demonstrative becomes a relative, and subsequently, an attribution marker, and (2) a demonstrative becomes a nominalizer and subsequently an attribution marker. Rießler's (2016) discussion of the development focuses on the anti-construct in Proto-Balto/Slavic. This aspect is relevant in the Iranian context as the Balto-Slavic anti-construct marker *jь* has descended from PIE *H̯io and is therefore cognate with some but not all ezafe allomorphs.

Rießler (2016) argues against scenario one where a demonstrative becomes a relative first before becoming an attribution marker, a theory that he attributes to Koch (1999). According to Rießler (2016), Koch (1999) supports this theory primarily due to Iranian evidence. The Old Persian proto-ezafe construction *haya/taya* had the attributive function but could also be used demonstratively in a type of nominalization. Rießler's (2016) issue with Koch's (1999) analysis is that "[he] does not disprove the assumption that the relative function of the pronoun derives from the deictic-anaphorical marking utilizing a demonstrative" (Rießler, 2016, 187). This issue seems to assume several facts, which may not be supported. First, Rießler assumes that PIE *H̯io is primarily a demonstrative pronoun and only developed relative functions secondarily. This is, in principle, a possibility as the

PIE root associated with the Indic, Iranian, and Slavic relative pronoun (*e-3, ei-, i-, fem. ī-) is associated with demonstratives as well. However, there are enough examples of the form of the relative pronoun to reconstruct this function for Indo-European. With that in mind, the question becomes is the Slavic anti-construct in *jь* actually a reflex of PIE *H̯io. Creissels (2008) seems to suggest that this is not the case despite the existence of the Old Bulgarian (OCS?) relative pronoun *jь-že*.

The question of cognacy extends to the Iranian data as well. According to Samvelian (2008), Old Persian *hya (tya-)* becomes *-i* in Middle Persian and progressively looses its demonstrative value to end up as a simple linker. However, the jump from *hya (tya-)* to *-i* is not predictable based on regular sound changes. Furthermore, it is now thought (following Skjærvø, 2009b, 2017b, 2007, etc.) that the Old Persian orthographic <hya> and <tya> actually represented *ha-ya* and *ta-ya*, which were bipartite. They consisted of a univerbated form of the demonstrative *h/ta-* and the relative *ya-* pronouns. These forms had both a relative and a demonstrative function in Old Persian. The Middle Persian *-i* (New Persian *-e*) is more likely the descendent of the relative pronoun *ya-* that the univerbated forms *ha-ya* or *ta-ya*.[32] In light of this, Samvelian's (2008) comment about the ezafe eventually losing its demonstrative value may not be valid. It should not be assumed that it began with any demonstrative character at all. As for the canonical ezafe as the reflex of the Proto-Iranian relative pronoun, it seems that the cline was Relative > attribution marker > nominalizer, which does not fit either scenario proposed by Rießler (2016). As for the reverse ezafe (anti-construct), it is fairly clear, at least as Balochi is concerned, that the attribution marker has descended from the Old Iranian adjectivizing formative (Avestan:

[32]The isolated form of the relative pronoun *ya-* is conspicuously missing from the small corpus Old Persian. In these texts the univerbated forms *ha-ya* and *ta-ya* take the function of both relative and demonstrative. However, it is perhaps a stretch to consider the dialect (or rather idiolects) of several elite Persian kings in the 5th century BCE to be representative of the speech of all of the ancestors of all of the Middle and New Persian speakers. I have myself proposed that there might be both reflexes of the relative pronoun *ya-* in the Persian ezafe and the univerbated form observed in Old Persian *haya-* in the Persian definite article *-(h)e*. The definite article is a part of the colloquial language but is paradoxically missing from over a millennium of attestation through the Middle and New Iranian periods; It does not exist in the standard language (Karim, 2021c, forthcoming).

-aēna- Korn, 2003, 151). This suffix started its life as an attributive marker and became a nominalizer later, and in specific circumstances.

3.6.4 The role of syntactic functors in morpho-syntactic change

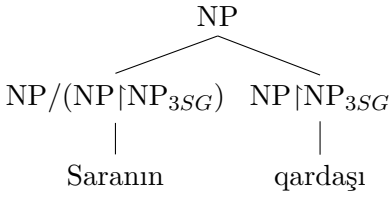
One of my stated goals of analyzing the ezafe phenomenon from a CG perspective is to develop an understanding of the ezafe grounded in psycholinguistic plausibility. As CG and HTLCG more specifically take a lexicalist approach, one might be able to do syntactic reconstruction based on what is stored in the lexicon. This is an endeavor to look at historical syntax the same way we look at historical phonology (and morphology). A complete analysis of the viability of such an approach is likely to be a lifetime project. However, there are a few observations that can be made based solely on the forms presented here. There are several historical developments I have mentioned. Among them are the development of possessor cross-indexing in Şirvan Tat as a compromise system between Azeri (Turkic) and the reverse ezafe of the Caspian languages, and the shift from right-headed attributive and genitive constructions to the ezafe construction. Both of these constructions have some commonalities, which may be salient in developing an understanding of the role of syntactic functors in morpho-syntactic change.

Possessor cross-indexing as a compromise system

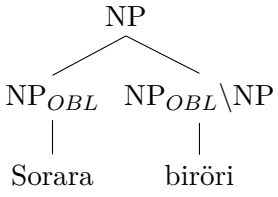
As suggested by Suleymanov (2020b), the possessor cross-indexing in Şirvan Tat is a hybrid system between the Azeri (Turkic) double construct and the Caspian reverse ezafe (i.e. genitive case). The Azeri system features a nominal possessor in the genitive case followed by its possessum morphologically marked for possession. The Caspian system features a nominal possessor in the genitive case followed by its possessum, unmarked with respect to possession. The Şirvan Tat system features a nominal possessor in the object case followed by its possessum morphologically marked for possession. The object case in Ş Tat is marked by the *-ra* suffix, which originally marked a dative relationship and later came

to be a marker of the direct object, experiencer, possessor, and applied object.[33] Based on the understanding of the forms in Azeri and Şirvan Tat I proposed in section 3.4.4 (proofs 31 and 32), I have constructed the diagrams in examples (66) a (Azeri), b (Ş Tat), c (a hypothetical system). In the Azeri form *qardaşı* 'his/her brother' on its own would have an anaphorically retrievable possessor. With an overt possessor within the phrase, the variable associated with the pronoun must be abstracted over, and the possessor is looking for an entity that is looking for a noun to become a noun.

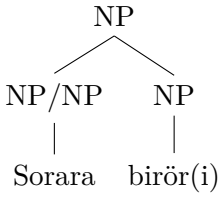
(66) a.



b.



c.



When the syntactic functor features a complex category under the slash, the combinatorics of argument become in a certain sense irrelevant. I propose that as long as such a form never has the opportunity to combine with the type that it is looking for, the complex category can reduce to a corresponding simplex category. This is illustrated by the hypothetical form in example (66) c. Here, $NP \uparrow NP$, an entity looking for an NP to yield an NP reduces to the atomic category NP as it never has the chance to act as the functor. This is a hypothetical form as it does not occur as such in Şirvan Tat. However, according to Suleymanov (2020b), two Caucasian Tat varieties have possessors marked with *-ra* and bare possessa (Literary Judeo-Tat and Abşeron Tat). One might speculate that the form of the bare possessum of type NP could be a morphological reflection of this principle. On the other hand, (66) b looks like a simple reanalysis of which nominal in the linear order is the functor. Based on these data, I vary tentatively propose two theories:

[33] What I refer to as the applied object, is referred to as the placeholder construction in Suleymanov (2020b). This construction is similar to the Absolute prepositions of Middle Persian (MacKenzie, 1964), Soranî (Rawlinson, 1839; Mackenzie, 1961), and other Iranian languages (Mohammadirad, 2020), a type of applicative marker (following Karim, 2021a; Karim & Salehi, 2020).

Theory 3.6.1 (input reduction)**.** Complex categories that occur under the slash may become a corresponding atomic category, perhaps innovating a new sub-type.

Theory 3.6.1 is motivated by the change observed in Judeo-Tat and Abşeron Tat but also in the following section about the ad-genitival ezafe.

Theory 3.6.2 (direction shifting)**.** For any two felicitously combining premises, the functor and its argument may switch, the former argument becoming the functor and vice versa. This shift likely takes place in two steps: (1) the slash switches directions, and (2) the complex category under the slash is reduced (i.e. 3.6.1: input reduction).

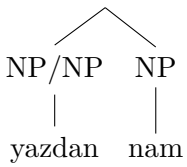
Theory 3.6.2 likely works with theory 3.6.1 as part of its mechanics. Two premises such as a genitive construction XP/XP and its nominal argument XP can be reinterpreted as having the latter as the functor. The former would initially retain its syntactic category: $XP/XP : XP > XP/XP : (XP/XP)\backslash XP$; the results, the category above the slash remains unchanged. As the usage of the the original functor becomes circumscribed or limited to contexts where it is the argument, the premise under the slash is reduced: $XP/XP : (XP/XP)\backslash XP > XP_{XP/XP} : XP_{XP/XP}\backslash XP$; the resulting reduced category may form a new sub-typing relationship like the oblique case in Iranian languages. The functions of that case marker may have an idiosyncratic distribution, i.e. not corresponding to well-known categories. Theory 3.6.2 is motivated by the Şirvan Tat form in (66) b. Undoubtedly, further research into the typology of grammatical categories from a CG perspective will illuminate both the validity and constraints surrounding these changes.

The genesis of the ezafe from an anti-construct system

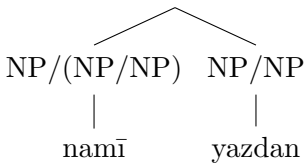
In section 3.6.1, I showed that in Western Middle Iranian, there was a hybrid system where an inherited genitive case and attributive adjectives preceded the nouns they modified (Skjærvø, 2009b, 221). Over time, the modifier-head word order became less frequent,

losing ground to the ezafe construction. These forms are represented by examples (67) a and b both 'gods' name(s) (ex. (55)).

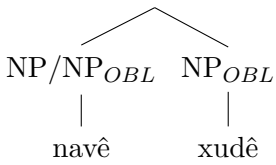
(67) a.



b.



c.



In the Middle Persian system, the original genitive represented by *yazdan* 'gods" precedes the noun and is of type NP/NP, the head noun can therefore be in any case. Eventually, an alternative construction came along by which the nominal became the marked functor, yet the original genitive maintained its status as a functor; the genitival possessor was both the functor NP/NP and the argument of the functor $NP/(NP/NP)$. As the original construction lost ground to the innovative construction only surviving in perhaps a few frozen forms, the functor under the slash was reanalyzed as a new atomic category NP_{OBL}. The Kurmancî form in (67) c 'god's name' shows just this type of reanalysis where the oblique form *xudê* 'god' which no longer has the canonical genitive function. This seems to be further evidence for theory 3.6.1. However, it is unclear what bearing this change has had, if any, on the establishment of the category oblique from the original genitive case.

3.7 Reflections

Using a theoretical framework under the umbrella of categorial grammar, it is now possible to develop a more nuanced understanding of syntactic categories. It is clear that Adjective, as a category, is something that is idiosyncratically defined and language-specific. In the Western Iranian world, it is clear that the category is in many ways non-existent. Attributive adjectives are by default substantive and are inflected as nominals. As for the languages which have the so-called reverse ezafe constructions, there is no demonstrable reason to have

developed new idiosyncratic terminology. The possessive reverse ezafe already has well-established terminology, the possessive state or genitive case. Likewise, the anti-construct is perhaps more accurately classified as attributive marking or adjectivizing. However, a typological line should be drawn between the languages with attributive adjectives and genitival possessors that take a noun and yield a noun like English (N/N) and ones that operate over noun phrases (NP/NP), the reverse ezafe languages.

As for the canonical ezafe, it has been variably called a "linker, relativizer, nominalizer [and] tense-particle" (the title of Haig, 2011). It has been analyzed syntactically and morphologically (Franco-Rita & Savoia, 2012; Ghomeshi, 1997; Larson & Yamakido, 2008; Larson & Samian, 2020; Samiian, 1994; Samvelian, 2005, 2007a, 2008, 2018). Here, I will update the definition to fit the theoretical framework of the current study and the current best knowledge of the ezafe.

Definition 3.7.1 (ezafe)**.** The ezafe is a type of derivational morphology that changes the syntactic category of a nominal from XP, a phrase level unit, to XP/XP, a functor looking for a nominal on its right to become a nominal.

The definition of ezafe 3.7 is the simplest conception possible. Note that this definition works for reverse ezafe languages by default (despite the poor terminology). It may be said that ezafe marking (by any name) is likely to develop in languages which inflect for (in)definiteness and do not distinguish the category adjective. Despite the typological diversity among these languages, they agree on each of these points. Although it has not been covered here, there are likely typological parallels that confirm the facts. For instance, the Albanian *nyhë* 'knot' particle, a type of ezafe marking (following Franco-Rita & Savoia, 2012), occurs as a "linker" within a nominal system that inflects for definiteness. Of course, most inherited adjectives only occur with the corresponding ezafe when being used substantively. This would be an extension that has yet to occur in any of the Iranian languages.

Another point that has become clear throughout this study is that a major area of inquiry going forward is the mundane. Syntacticians are generally interested in solving complex syntactic problems such as gapping, pseudo-gapping, and ellipsis (see Kubota & Levine, 2020). However, many aspects of language are considered simple and mundane, such as attribution and case relations (a subtyping relationship in CG). However, many of these mundane aspects of language offer complex problems. Truly atomic categories in any given language vary greatly. A verb that requires a direct object in the accusative case and a subject in the nominative case may be as simple subtyping relationship synchronically, e.g. a transitive verb of type π; γ; $\mathrm{NP_{ACC}}\backslash(\mathrm{NP_{NOM}}\backslash\mathrm{S})$. However, we know from attestation that certain cases have developed from others that originally caused a syntactic derivation, e.g. the New Indo-Aryan ergative case π; γ; $\mathrm{NP_{erg}}$. This case is the reflex of the Old Indo-Aryan instrumental case, which introduced a verbal adjunct, π; $\lambda P[P \wedge with(x)]$; $(\mathrm{NP_{NOM}}\backslash\mathrm{S})/(\mathrm{NP_{NOM}}\backslash\mathrm{S})$. When the stative adjectival form of the verb (Past Passive Participle) was recruited for a perfect construction, the instrumental could be used as an optional agent (by) phrase. As this construction took on grammatical status and adpositional constructions displaced the instrumental, it was reanalyzed as a verbal argument and not an adjunct. Its status as a functor changed. The search for a nuanced theoretically driven approach to this type of change may be possible due to the work that I have begun in the current study.

An additional virtually unstudied avenue of research that has come to light due to the current study is the morphosyntax interface. It is clear from the ezafe and other derivational morphemes that there is a direct connection between morphological marking and syntactic category. From a CG (lexicalist) perspective, we assume that the syntactic category is an unalienable part of the lexical entry, like its prosodic realization and semantics. As a morphologist who works in inferential-realizational approaches, I see morphology as the organizing principle of the lexicon. Given these two assumptions, syntactic combinatorics are a function of the morphological paradigm. In the future, inquiry into the specifics of such

a theoretical framework will lead to a unified field theory of linguistics. In other words, it is abundantly clear that syntax, semantics, prosody, and morphology are so fundamentally bound that the study of one is necessarily the study of the others. A future linguistic theory should reflect this fact.

Chapter 4

Why Kurmancî has no Definite Article

4.1 Introduction

The question at the core of this chapter is why Kurmancî has no definite suffix when other Kurdish varieties do. However, this question could be conceptualized more broadly as why are there four types of definiteness marking systems in Iranian languages: those with a K-form definite suffix, those with a K-less definite suffix, those with case marking only when definite, and those with no definiteness marker. Here I focus narrowly on Kurmancî, with data from other languages to support my hypothesis: the original evaluative marker Old Iranian *-(a)ka- became a definiteness marker. Due to regular sound changes and analogy, these formatives changed in the various Iranian languages. The nominal suffixes and case markers of all the languages that have some type of definiteness marking have descended from OIr. *-(a)ka-.

Within the geographic area where the majority of the population consider themselves to be ethnic Kurds (hereafter the Kurdish zone or Kurdistan), there are two natively-spoken sub-groups of the Iranian branch of the Indo-European language family. These are the Kurdish (Northern, Central, and Southern) and the Zaza-Gorani language sub-families. Although the Kurdish varieties are closer to one another than they are to the Zaza-Gorani varieties, Kurmancî (Northern Kurdish) and Zazaki have come to be more alike, and Hewramî (Gorani) and Soranî (Central Kurdish) have also come to share many features. MacKenzie (1961), citing Professor K. Barr, attributes some differences within Kurdish to Gorani influence on the [Central and Southern] dialects. He further argues that

"there is no avoiding the conclusion that [Central and Southern] dialects of Kurdish have overlaid a Gorani substratum[1], while the Northern dialects have to a much greater extent preserved their purity" (MacKenzie, 1961, 86). Of course, there is no question that there has been interaction between Central Kurdish and Gorani. The Gorani-speaking area has decreased since Rawlinson's (1839) account of his travels in the region.

In contrast with MacKenzie (1961), Leezenberg (2015)[2] rejected this claim asserting that in addition to Gorani contact, the convergences between Central and Southern Kurdish and Gorani could also be explained as common inheritance, "parallel innovations of a Sprachbund-like nature, as prestige borrowings, or as innovations specific to Kurmancî." I take this assertion a step further here, adding that these Kurmancî innovations, including the loss of the definiteness marker, can be attributed to Zazaki contact and their mutual participation in shared innovations with a Northern belt of languages extending across Armenia and Azerbaijan, and through the Caspian region.

The convergences shared between Kurmancî and Zazaki are summarized in table 4.1. They include the following features: the use of a future auxiliary in Wackernagel's position, case marking (bicasual), grammatical gender, object indexing, the definite suffix *-eke*, a synthetic passive, enclitic pronouns, and the use of an enclitic pronoun to mark agents. Object indexing is described as agreement by Jügel (2014). However, object indexing in Kurmancî is dependent on several factors such as animacy and definiteness. Generally, past-tense transitive verbs are marked for the direct object. An overt conominal may occur with a verb inflected for object agreement (i.e. a cross-index following Haspelmath, 2013).

[1]There is no way of knowing precisely what MacKenzie (1961) meant by substratum. It is unlikely that in 1961 the term carried much of the theoretical weight that it does today.

[2]Leezenberg (2015) provides a more theoretically-driven approach rooted in the tradition of Thomason & Kaufman (1988). His goal was to analyze the type of contact that resulted in the borrowing from Gorani found in Central and Southern Kurdish. Essentially, he challenged the narrative, conjured up by (MacKenzie, 1961) use of the term substratum, of a Gorani-speaking population shifting to Kurdish and bringing along aspects of their language as a result. His ultimate conclusion was that the borrowings attested in Kurdish were of the type that could be prestige borrowing from an elevated literary Gorani. The so-called Gorani koiné was the official language of the Erdelan court. The Erdelan dynasty was a time and place where the Gorani language flourished, and many poets composed in Gorani despite being speakers of other varieties.

	Sprachbunde I		Sprachbunde II			
	Kurmancî	Zazaki]	Hewramî	M.CK	S. Soranî	SK
future aux	✓	✓				
case	✓	✓	✓	✓	Ø	Ø
gender	✓	✓	✓	Ø	Ø	Ø
object index	✓	✓	✓	(✓)	(✓)	Ø
synth. passive	Ø	✓	✓	✓	✓	✓
DEF (*-eke*)	Ø	Ø	✓	✓	✓	✓
encl. PrN	Ø	Ø	✓	✓	✓	✓
agent CL	Ø	Ø	✓	✓	✓	Ø
ve= > =ewe			✓	✓	✓	✓

Table 4.1: Grammatical convergence between Kurdish-zone languages (edited and expanded from Jügel, 2014, 136)

However, definiteness also comes into play as a verb never agrees with a noun in the oblique case, which can be conditioned by specificity (as suggested by Dorleijn, 1996). In contrast, Central Kurdish (M.CK and S.Soranî) does not allow a conominal (i.e. a pro-index following Haspelmath, 2013). I show innovations[3] in red. Here the development of a future auxiliary and the loss of the agent clitic, enclitic pronouns, and the definite suffix can be attributed to innovation between Kurmancî and Zazaki. The renovation of the synthetic passive is thought to be due to contact with Hewramî (MacKenzie, 1961, 84). However, the form of the Central Kurdish passive morpheme *-ra* is unlike the one in Hewramî *-i*, which is likely the inherited form, well attested in Avestan, Old Persian, and Sanskrit (Leezenberg, 2015). In other words, the Kurdish form could be explained by a combination of mutual inheritance and the Central Kurdish internal innovation of the *-ra* variant. It should be clear from this chart that there are many shared features between Central and Southern Kurdish and Hewramî (Gorani), which are attributable to parallel innovation. The only shared feature that appears to be a parallel innovation is the occurrence of the preverb *(e)ve=* (Kurmancî) attached as a post-verbal element, e.g. *=ewe* (Soranî) (Leezenberg,

[3]I use the term innovation here to refer to features of these languages that are thought to be developments since a Proto-Kurdish or a Proto-Zaza-Gorani. However, many of these developments could be inherited as very little is known about these languages in the Middle Iranian period. Only data from comparative historical linguistics can broaden that understanding, and thus, these assertions are subject to updates in the future.

2015).

It was also MacKenzie's (1961) assertion that the definite suffix *-eke*, occurring in Gorani and Zazaki[4], must also be borrowing from Gorani, as it is notably absent from Kurmancî (MacKenzie, 1961, 81). Now that more is known about various western Iranian languages, it is clear that this article is a more widespread phenomenon that may even be reconstructible for Proto-(North)western Iranian.[5] In this chapter, I take a narrow view and explore definiteness marking in Kurmancî (Northern Kurdish) and its origin.

4.1.1 The development of the K-form in Kurmancî

I assert that the K-type definite suffix was a part of Proto-Kurdish and probably Proto-Northwestern Iranian. Although they probably were not definiteness markers at the time, the building blocks of these formatives were part of Old Iranian going back to Proto-Indo-European. In this chapter, I take the diminutive theory of the origin of these formatives to be fact with the stipulation that the term "evaluative" is probably better than diminutive (following Nourzaei, 2020). A brief outline of the narrative of this shift is as follows.

1. The Proto-Iranian *-(V)ka- marker was inherited by the daughter languages, including Kurdish (and Zaza-Gorani).

2. The reflex of *-(V)ka-, used to express an evaluative meaning, gives way to deictic and recognitional functions and eventually an anaphoric function.

3. The Iranian languages rearrange themselves into multiple sprachbünde, dividing Kurdish and Zaza-Gorani into a northern and a southern sprachbünde.

[4] MacKenzie (1961) cites Hadank (1932) for the article *-eke* in Zazaki. However, this is to my knowledge absent from any contemporary variety (following Paul, 1998b; Todd, 2002; Aygen, 2007, etc.). Hadank (1932) *ǧeniäkä* 'diese Frau' (p.65), which could in principle be a determinate suffix *-ä* (p.65) and the diminutive suffix *-(ä)k* (p.62).

[5] Although they are only distantly related, there are parallels in colloquial New Persian that may be connected (e.g. the definite suffix *-he*).

4. The northern sprachbund undergoes a subsequent change, where intervocalic *k is lenited and lost.

5. Further reductions and syncretism result in case disappearing from most Kurdish varieties. It is paradoxically retained in the varieties in which the loss of *k obscured the definite article.

Several facts support the case for the loss of the is a feature of K-less systems but not K-ful ones.K-type definiteness marker–as opposed to its being a borrowing into Central Kurdish–:

- K-forms are widespread all over the Iranian world.
- IOM, inherited case marking only on definite nouns,
- A hypothetical system with K-forms would have differential object marking.
- The reflexes of -aka are comparable to Kurmancî inflectional suffixes.
- The Kurmancî ezafe reflects the development of a combination of K-type definite suffix and the inherited absolute/indefinite ezafe.

These facts and the general developments I Kurmancî are addressed in the following section 4.2.

4.2 The loss of K-forms in Kurmancî

Kurmancî does not have a definite marker, which sets it apart from Soranî (Central) Kurdish and Southern Kurdish as well as Hewramî. The definite article shared by Hewramî and Central and Southern Kurdish coupled with its absence from Kurmancî appeared to MacKenzie (1961) as a clear example of convergence between Hewramî and Central and Southern Kurdish. In this chapter, I assert that the lack of such an article is an example

of convergence between Kurmancî and Zazaki and retention of the article in Soranî and Hewramî. There are many divergences between Kurmancî and Soranî, as well as features shared between them and Zaza-Gorani that have led some to suggest that there are perhaps two separate sprachbünde which each have different properties. I follow Jügel's (2014) "Hypothetical development (4b, 128)." This theory assumes a Proto-Kurdish, which split into several dialect groups that subsequently evolved in two separate sprachbünde one with Gorani varieties including Hewramî and the other with Zazaki (and likely Armenian, Asia Minor Greek, and Caucasian languages). The first sprachbund retained the K-type suffixes, which allowed some of the languages to lose case and gender marking. The second sprachbund lost the K-suffixes, unimorphating the definite suffix, the case-gender marking, and sometimes the ezafe morphemes.

4.2.1 The Proto-Iranian *-(V)ka- marker was inherited by the daughter languages

I assume that the K-form definite suffix was inherited from Old Iranian into all of its daughter languages. This mutual inheritance is supported by its existence in Old Iranian, cf. Old Persian *ba^{n}da-ka-* 'bondsman' (Skjærvø, 2007, 903) and Avestan: *maṣyā-ka-* (Kanga, 2003, 403). Additionally, it has reflexes in Middle Western Iranian, cf. Middle Persian: *xān-ag* 'house' (Skjærvø, 1989b, 262), and in other Middle Iranian, e.g. the Sogdian vocalic declension (described in §.4.2.5). As for the New Iranian languages, some have a K-type definite suffix, those that have a definite suffix without /k/, and those with only case-marking suffixes.

The K-type markers occur in a variety of languages, e.g. The dialect of Emāmzāda Esmā'īlī (Fars): *doft-ak-ō* 'the girls [girl-DEF-PL]' (Windfuhr, 2012), Bušehrī (Fars): *ī havā-y-akū* 'this weather' (Windfuhr, 2012), Gīonī (Lor): *asp-Ø gap-eka* "the big horse [horse-DEF.EZ big-DEF]" (McKinnon, 2001), Northern Lori *-(e)ka* (McKinnon, 2011), Dezfuli and Šuštari (S Lori): *-aka* (McKinnon, 2011), Bakhtiāri (S Lori): *-ekū* (McKinnon, 2011), *-(e)ke*

(Anonby & Taheri-Ardali, 2019, 452).[6] This is in addition to the widespread occurence of K-form definiteness markers in Gorani and Central and Southern Kurdish, e.g. Central Kurdish: *-eke* (Mackenzie, 1961), Southern Kurdish: *-aka -aga* and *aǧa* (Fattah, 2000, 245), Hewramî (Lihon): *-aka* (MacKenzie, 1966), Paweyane: *-ækæ* (Holmberg & Odden, 2008), Zerdeyane: *-aka* (Mahmoudveysi & Bailey, 2013), Gewrecûî: *-aka* (Mahmoudveysi et al., 2012).

In addition to the languages that have a K-type suffix, there are others which have definite suffixes that have the same morpho-syntactic properties as the K-type suffixes albeit with different forms. Sīvandi has *-u* [M.SG.DEF] and *-e* [F.SG.DEF] (Windfuhr, 1991); Judeo Isfahanî has *-e* [SG.DEF] (Windfuhr, 1991); Khuri has *-u* [SG.DEF] (Windfuhr, 1991); Kermani languages have *-u* (Borjian, 2017a); the Median dialects (Kašan) have *-a/-e* (Borjian, 2012b); Kešaʻi has *-é* (Borjian, 2017c); Kumzari has *-ō* (Anonby, 2019, 631), etc. Like the languages and language varieties which have the K-type suffixes, these languages share syncretism between the definiteness marker and a diminutive/evaluative suffix.[7] It is in principle possible that all of these suffixes have the same etymon or related etyma. There is some certainty to this evaluation for at least some of the languages e.g. the Median dialects of Kašan. According to Borjian (2012b), "a whole class of words in each dialect carries the ending vowels /-a/, /-e/, or both, derived from the Old Iranian suffix *-aka," being the likely etymon of the K-type suffix (following Haig, 2019a; Haig & Mohammadirad, 2019; Nourzaei, 2017, etc.). This group likely includes colloquial New Persian which features the definite suffix *-(h)e* (proposed by Jahani, 2015).[8] Unfortunately, not enough is known about

[6] Examples are given where available.

[7] See Jurafsky (1996) for a discussion of how the diminutive can become evaluative. Additionally, see Pakendorf & Krivoshapkina (2014) for a specific case study involving Ėvev (Siberia), which develops a definiteness marker from an evaluative morpheme. In Jurafsky's (1996) broader study of diminutives, he does not specifically mention the step of becoming a definite article, although he shows an example from Khasi (Mon-Khmer), which features a definite article "for diminutive nouns and to mark affection or respect for 'members of one's family or to persons of superior social position such as teachers or employers' (Ravel, 1961, 95)" (Jurafsky, 1996, 571).

[8] Jahani (2015) links the colloquial New Persian definiteness marker *-(h)e* to the diminutive suffix *-aka as well. However, that account does not take into the consideration the origin of the /h/, which sets the form apart from the diminutive marker *-e*. Of course, this could be explained away by analogy with the

all of these languages to confirm (or deny) the possibility of a common etymon for all these formatives.

In addition to the languages that have a definite article, some languages use case-marking alone to distinguish definiteness (described in ch.2). Some of these languages have an innovative definite direct-object marker cognate with (or borrowed from) New Persian *-rā*. Some languages feature this marker superimposed on top of an otherwise caseless system like New Persian (standard and colloquial). Other languages feature the *-rā* marker superimposed on top of a bicasual system like Eastern Balochi (e.g. Rakhshani Barker & Mengal, 2014. Turkmen Axenov, 2006, etc.). The underlying system of Eastern Balochi features differential case marking (Korn, 2003, 332). Additionally, the languages have differential object marking in the form of the inherited case marker on definite or specific direct objects and no marking on absolute (general) and non-specific indefinite. A preliminary observation of these varieties is that the languages with a K-type definite article (or one that is the likely reflex of Old Iranian *-aka) and the languages that have inherited case marking and differential object marking constitute two disjoint sets.[9] Note that there are languages that have preserved the inherited case marking and the K-form article but not DOM (MacKenzie, 1966, e.g. Hewramî).

I assert that all inherited definiteness marking systems in the Iranian languages have a similar origin except for the secondary *rā*-type, which has a well-known history (< Old Persian *rādiy*). The systems that seem to lack a definite article had inherited one that was subsequently lost in some contexts. I take the attestation of these forms across time and space as proof that these are the expected inherited forms. This attestation is just one piece of evidence that culminates in my conclusion that the definite article was lost in Kurmancî.

plural marker *-(h)a*, which occurs with an /h/ after vowels just like the definite marker. See section 2.4.1 for a deeper discussion of the possible etyma of *-(h)e*.

[9]One possible counter-example to this trend is that of the Northern Kurdish variety of Surči, which has the definite article *-aka*. However, this variety is spoken at the North-Central Kurdish transition zone. The occurrence can be understood as dialect borrowing. Note that in the same region there are Turkic and Semitic languages that have borrowed the Central Kurdish definite marker *-eke* as well, e.g. Iraqi Turkmani: *o lan-akâ* [boy.SPEC] (Bulut, 2019, 368), Arbel (Jewish): *belă-ke* [house-DEF] (Khan, 2019, 322), etc.

4.2.2 The reflex of *-(V)ka-, used to express an evaluative meaning, gives way to deictic and recognitional functions and eventually to anaphoric function

The Kurdish definite article *-eke* and its cognates across the Iranian world are difficult to study. The definite marker and its cognates' functions in each of the Iranian languages differ not only from definite articles in English, for instance, but also from each other. Following Nourzaei (2020), the contexts where definiteness markers are used include:

- Anaphoric definiteness: The referent has an antecedent in the preceding textual context: *A man and a woman entered.* ***The man*** *sat down.*
- Bridging definiteness: The referent has not been previously mentioned in the discourse context, but its existence can be inferred from associated expressions: *We bought a new car, but* ***the brakes*** *were faulty.*
- Proper nouns: The noun is conventionally associated with a specific entity: *Sweden, Angela Merkel, Mount Kilimanjaro*
- Possessed nouns: The noun is accompanied by a grammatical possessor, often syntactically fulfilling the determiner function: *my house, their child, Henry's birthday*
- Deictically modified nouns: Nouns accompanied by demonstrative elements: *this article, that place*
- Unique referents: Entities which are assumed to be uniquely identifiable by all members of a given speech community, hence require no preceding or inferable mention: *the sun, the river (in a given community), the President, etc.*
- Situational definiteness: Identifiability is achieved through the immediate speech context, possibly aided through additional gestures and adverbial expressions: *the man over there (pointing).*

(Nourzaei, 2020, 5)

In Central Kurdish, the definite suffix *-eke* is employed in anaphoric, bridging, and possessed contexts (Haig & Mohammadirad, 2019). In contrast, the Koroshi Balochi cognate is used with deictically modified nouns and unique nouns but not with possessed nouns in addition to the anaphoric and bridging functions it shares with Kurdish. These articles have only recently been analyzed from the perspective of their function (e.g. Haig & Mohammadirad, 2019; Haig, 2019a; Nourzaei, 2017, 2020, etc.). However, it is already clear that there is a great deal of variability between languages. Additionally, examination of other Balochi varieties shows different degrees of grammaticalization in the usage of these particles (Nourzaei, 2017). Presumably, this suggests that although these particles have a shared etymological origin that can be reconstructed for the family, the definite function may be a later development. The path from "evaluative marker" to definite suffix is just one of the possible grammaticalization paths that might have been selected or encouraged by contact with languages that had already gone down such a path (See Nourzaei, 2020, for more on the development from "evaluative" morphology to definiteness marking.).

If the variability in function of the K-type markers in Central Kurdish and Koroshi Balochi is representative of the cognate morphemes in other Iranian languages, it makes sense that there has been no systematic study of these forms or recognition of their common function. Likely due to this comparative lack of scholarly study and the variability, they are referred to differently by different scholars. The monikers 'definite' (Mackenzie, 1961; MacKenzie, 1966; Mahmoudveysi et al., 2012; Mahmoudveysi & Bailey, 2013; Öpengin, 2016, etc.), demarcative (McKinnon, 2011), determinative (Windfuhr, 2012), or deictic (Windfuhr, 1991) appear in the literature. A large-scale typological study is necessary to understand the full distribution of these suffixes, their forms, and their functions.

4.2.3 The Iranian languages rearrange themselves into multiple sprachbünde dividing Kurdish and Zaza-Gorani into a northern and a southern sprachbund.

According to Jügel (2014), based on shared grammatical features, "we may erroneously conclude that Zazaki and Kurmanji, on the one hand, and Sorani, Gorani, and Hawrami on the other, are more closely related. This conclusion is incorrect as it ignores other features which separate Zazaki, Gorani, and Hawrami from Kurmanji and Sorani." This problem is exacerbated based on Korn's (2019a) assertion (citing Clackson, 2007, 5) that "it is now generally agreed among linguists that the most certain subgroups are constructed based on unique shared morphological innovations." Korn (2017b) mentions that several known New Iranian languages have their "phonemic systems ... adjusted to those of neighboring languages" in reference to Ossetic and Balochi, which have become more like Caucasian and Indic languages, respectively. However, there is no reason why similar phenomena would affect genetically related languages differently.

Assume that Korn's (2019a) assertion that significant shared "morphological" innovations are the most reliable and that phonological developments should be less reliable when deciding genealogical relationships.[10] Based on that assumption, there are shared features of Kurmancî and Zazaki that are relevant to this study.

According to Gippert (2009), "In älterer Zeit dürfte es jedoch vom Nordwesten bis in den Nordosten Irans ein Kontinuum nahe verwandter nordwestiranischer Dialekte gegeben haben, das sich deutlich von den südwestiranischen Dialekten weiter im Süden abgehoben haben dürfte [in older times, however, there must have been a continuum of closely related north-west Iranian dialects from the north-west to the north-east of Iran, which is likely to

[10] I believe that both morphological and phonological innovations can be borrowed. However, borrowings do not have to come into every sub-variety, as is the case with shared innovation. Likewise, Gippert (2009) states that "Will man die Position einer Sprache wie des Zazaki bestimmen, so erfordert dies die Betrachtung nicht einzelner Isoglos- sen, sondern von Isoglossenbündeln auf allen Ebenen der Grammatik (Phonologie, Morphologie, Wortbildung, Syntax, Lexikon) [If one wants to determine the position of a language such as the Zazaki, this requires the consideration not of individual isoglosses, but of isogloss bundles at all levels of grammar (phonology, morphology, word formation, syntax, lexicon)]."

have stood out clearly from the south-west Iranian dialects further south][11]" (Gippert, 2009, 80). He speculated that looking at imperfective marking alone might suggest that there is a Northern-belt of languages with Semnānī in the Center and the Gorani languages at the far left periphery (Gippert, 2009, 79). It should be noted that Kurdish does not partake in this isogloss even though Northern Kurdish and Zazaki share many others not shared by the rest of Kurdish. The majority of the verbal and nominal morphology of Kurmancî puts it clearly in the same category as Central and Southern Kurdish. However, its case system puts it in the same group with Gippert's (2009) northern-belt languages. Some of the Kurdish features absent from Zaza-Gorani are listed in table 4.2.

	Zazaki	Hewramî	Kurmancî (NK)	S.Soranî (CK)	SK
LOC > IPFV[12]			✓	✓	✓
1PL:S *-(î)n*			✓	✓	✓
2PL:S *-n*			✓	✓	✓
3SG:COP *-(y)e(s/t)*			✓	✓	✓
PRS.PRF < *(i/u/a)g			✓	✓	✓
PST.COND *-ba*			✓	✓	✓
PST.PRF *-bû*			✓	✓	✓
etc.					
IPFV *-ēn(e)*	✓	✓			
1PL:S *-(m)(e)*	✓	✓			
etc.					

Table 4.2: Kurdish morphological features (not shared by ZG)

Zazaki and Kurmancî join Gippert's (2009) northern-belt languages as far as nominal morphology is concerned. The languages in this group have, for the most part, preserved (1) bicasual inflection systems and show (2) identified object marking (IOM) with the inherited case markers. Although, they do not all feature IOM in both the past and present tense. Additionally, several language sub-families count among their members at least one language

[11] English translation provided by Google Translate.

[12] The shift from an original locative to an imperfective marker is based on a new proposal following Karim (2021b). This proposal unites all the various Kurdish imperfective prefixes and suffixes *de-, d(i)-, e-, =y e-, -a, -ya, -ẅa* and the negative-imperfective allomorphs *na-, neye-, nê-, ney-, nede-,* and *neʔe-* as cognates of a single construction the locative circumposition *de-...=da*. The variation is accounted for by only long-established historical sound changes.

that does not have these two properties. For example, among the Tati languages, there are not only languages like Vafsi, fully inflected for case, number, gender with IOM, but also languages like Alviri with no case marking and Vidari with a definite direct-object marker *-i* (Yarshater, 1964).

	Kurmancî (N Kurdish)	Zazaki (Zaza-Gorani)	Vafsi (Tatic)	Chāli (Tatic)	Leriki (Talyshi)	Lāhijāni (Gilaki)
IPFV *-ēn(e)*		✓	✓	✓	✓	✓
numeral EW		✓	✓	✓	✓	✓
marked possessor	✓	✓	✓	✓	✓	✓
bicasual	✓	✓	✓	✓	✓	
IOM (PRS)	✓	✓	✓	✓	✓	(✓)
IOM (PST)	✓	✓	✓	✓		(✓)
animacy		✓	✓			

Table 4.3: Northern-belt features shared by Kurmancî

There is a great amount of diversity within the Tatic group described in Stilo (2018b). Regardless of family, the northern-belt features important to this study are case marking and identified object marking (IOM). In table 4.3, I show just a few of these languages. They are arranged from northwest to northeast. The important points are that Kurmancî (Northern Kurdish) does not partake in the shared lexical innovations like the imperfective-/present-tense stem with the *-en(e)* formative or the numeral 'one' of the EW-type (i.e. from *aiu̯a). However, it does continue possessor marking, a bicausal system, and it has IOM in the present- and past- tenses (double oblique). These are shared by most of the Tatic and Talyshi languages. As one moves further south and east, this system breaks down as languages shift to the Persian-type without case marking as such. Rather, as one moves further south and east, IOM re-emerges as *rā*-marking. *Rā*-marking is represented in the table in parentheses for Lāhijāni (Gilaki). It is not clear where the animacy feature fits into this group. It is a feature of both the Vafsi (Tatic) and Zazaki (Zaza-Gorani) nominal systems. However, in Zazaki, for instance, it is only a property of masculine nouns. Whatever the reason, it is not clear whether this is an inherited feature of the family or represents independent innovations.

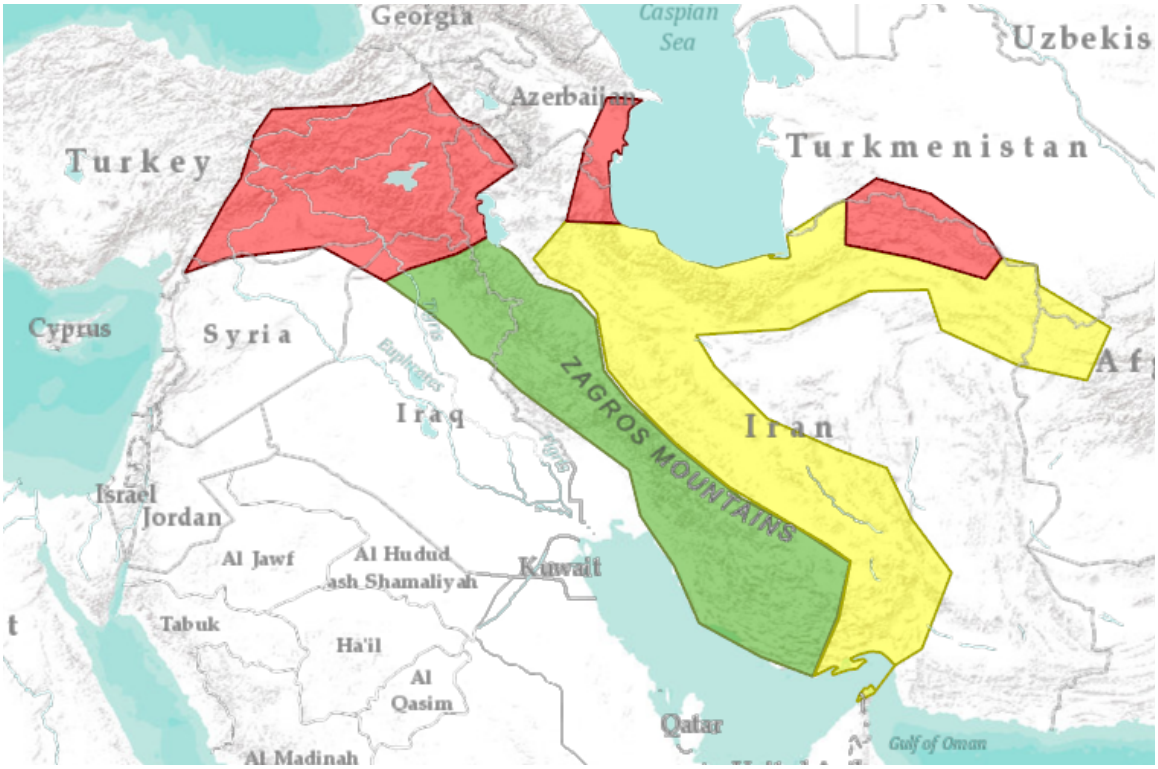


Figure 4.1: K-form definite articles, reduced forms, and reduced forms with IOM

I roughly sketch the geographic locations of these types in the map in figure 4.1. It shows the three different types of languages based on the form of the definiteness suffix. The green zone marks the languages, including Central and Southern Kurdish, Lori, and many of the so-called Fars dialects, that have a K-form definite article. The yellow zone represents the varieties, including the Kermanic languages, Bandari, Kumzari, the so-called Median dialects of Kašan, Some Tatic varieties, and nearly all colloquial Persian varieties, that have a definite article without a velar formative. I assume these formatives based on Jahani's (2015) hypothesis for colloquial New Persian but also based on Windfuhr's (1991) acknowledgment that the reflex of the *-aka suffix is either -e or -a in the Median dialects of Kašan. This assertion is equally supported for other regional languages. The Northernmost section in red represents Kurmancî, Zazaki, and many Caspian languages including Talyshi, and Southern Tati varieties. These are the languages that have retained case marking in

the context of definiteness resulting in the system of IOM. There are gaps in the contiguity of this northern belt caused by the displacement of Iranian varieties by Turkic (e.g. Azeri) and, of course, Persian incursion. Additionally, a pocket of Western Kurmancî migrated to the area northeast of Mazandaran on the Iran-Turkmenistan border. This patch of red appears out of place as not what is expected further east, where colloquial New Persian varieties dominate with the reduced definite article.

It should be noted that not all of the languages belonging to any particular group are required to fit into the same category as its sisters. Additionally, internal migration puts some languages in sections with unlike languages. For instance, the K-form Koroshi Balochi is spoken within the yellow shading, and Sivandi has preserved a grammatical gender on the definite ezafe and the definite article. These are typically red features on an otherwise yellow language spoken in the green zone. These varieties are not represented in this map, which has been provided as a general illustration and does not consider regional variability within varieties, displacement by Persian, or the migration of small communities that make this map somewhat inaccurate. This illustration should be understood as an imperfect sketch of the general regions where these features are located.

4.2.4 The northern sprachbund undergoes a subsequent change, where intervocalic *k is lenited and lost.

I assert that the case marking suffixes and the forms of the ezafe are the reflexes of the inherited *-aka evaluative suffix. This form subsequently became a definiteness marker in the way described in section 4.2.2. This shift was proposed to be the source of the colloquial New Persian definite suffix *-(h)e* by Jahani (2015). For the Median dialects of Kašan, Borjian (2012b) shows that the reflexes of the suffix *-aka are /a/ and /e/, which correspond to the masculine and feminine definite markers *-a* and *-e*. In Sogdian, it has been shown that nouns in *-aka have developed into a unique declension class with a distinct pattern of polyfunctionality from the strong (bicasual) or weak (six-case) stems. Gippert

(2009) has suggested that in Zazaki, the masculine singular oblique ending has descended from masculine nouns in *-a-ka but crucially not other classes (Gippert, 2009, 90). This development may be responsible for definiteness and animacy (specificity) splits that are only properties of Zazaki masculine singular nouns. As of this point, no other study has proposed such a development for case marking in Kurmancî. However, it is exactly this development that is responsible for some of the peculiarities of the Kurmancî system (see the following sections)

4.2.5 Further reductions and syncretism result in case disappearing from most Kurdish varieties. It is paradoxically retained in the varieties in which the loss of *k obscured the definite article.

IOM, inherited case marking only on definite nouns, is a feature of K-less systems but not K-ful ones. There appears to be a continuum of definiteness-marking systems. In the far (south)east, Central and Southern Kurdish, Lori, and many of the Fars dialects have K-marked systems. There are varieties with reduced systems of the colloquial New Persian-type, which show a definite article that has been reduced, losing the velar element spreading north and east from the K-marked region. As one moves northward, that same reduced definiteness marker may retain gender. Finally, in the far northwest of the region, some retain case in the context of this definiteness marker. The result is IOM. Case that is only marked on nouns would have taken the de-evaluative definiteness marker found in other varieties. Much of the discussion of these languages have overlooked Kurmancî DOM, e.g. Haig (2015), which claimed that DOM was "notably absent in Kurdish and Zazaki." Haig's (2015) note was despite DOM being described for Zazaki in great detail by Paul (1998b) and DOM hinted at by Dorleijn (1996). The details of the DOM system in Kurmancî have never been systematically explored. It would be impossible to conduct such a study within the confines of this dissertation. However, looking at the known developments in related languages and the internal idiosyncrasies of Kurmancî, I believe it is possible to understand

how the paradigms of the *-aka and bare stems combined to become what is observed in Kurmancî today. This hypothetical scenario is described below.

A hypothetical system with K-forms would have differential object marking.

One major issue with the understanding of the development of case in Iranian is how the phonological forms of the direct and oblique map to different thematic roles. As mentioned in chapter 2, there is a problem of comparability. That is to say that not enough is known about the various systems in most Iranian languages to classify the case systems according to syncretism between functions. Preliminary work on this is Stilo (2008a) that classifies languages based on the syncretism between paradigm cells (see ch. 2). He also describes some of these languages based on developments according to the polyfunctionality of those cases, i.e. according to thematic roles. However, much is missing in terms of the coverage of under-documented forms.

The second problem is that of diachrony. As shown in table 2.2, repeated here as 4.4 without the agent and kinship nouns represented for convenience. Here is the result assumed if final codas were lost across the board as they were eventually at some point in the history of all the New Western Iranian languages. For the most part, New Iranian languages have adopted adpositional constructions for locative, instrumental, ablative, and sometimes dative constructions. Excluding the vocative, one needs to account for the nominative, accusative, and genitive.

<table>
<tr><th></th><th>LOC</th><th>ACC</th><th>VOC</th><th>NOM</th><th>INS ABL</th><th>DAT GEN</th></tr>
<tr><td>Avestan (a) [SG]</td><td colspan="4" rowspan="2">yasn</td><td></td><td>yasneh</td></tr>
<tr><td>Avestan (a) [PL]</td><td>yasneš</td><td>yasneb</td><td>yasnān</td></tr>
<tr><td>Avestan (ā) [SG]</td><td>daēnay</td><td colspan="3" rowspan="2">daēn</td><td colspan="2">daēnay</td></tr>
<tr><td>Avestan (ā) [PL]</td><td>daēnāh</td><td>daēnāb</td><td>daēnān</td></tr>
</table>

Table 4.4: Case syncretism due to accidental homophony with DAT-GEN and INST-ABL mergers

I have simplified this chart showing just the nominative, accusative forms of the nominal suffixes with the sound changes assumed for Kurmancî (and other NW Iranian languages) in table 4.5. A couple of things stand out about this development: one is that the plural forms of both the masculine *a*-stems and the feminine *ā*-stems have fully merged, i.e. collapsing gender in the plural. This pattern is pervasive in Western Iranian languages (see ch.2).

	SG			PL		
	NOM	ACC	GEN	NOM	ACC	GEN
masc. (a)	-Ø	-Ø	-ī (< *-eh)	-Ø	-Ø	-ān
fem (ā)	-Ø	-Ø	-ē (< *-ay)			

Table 4.5: Inherited case syncretism

The second is that there is syncretism between the nominative and accusative regardless of number and gender. Based on this development, one would expect the unmarked case (direct) to cover nominative and accusative. In contrast, the marked (oblique) case would be used in the genitive, as complements of de-nominal adpositions (and adpositions that would have governed the genitive) as well as possibly the dative. This system is observed in some Middle Iranian languages; this is true of Khwarezmian (Durkin-Meisterernst, 2009, 342). In Zoroastrian Middle Persian, "nouns as direct object are regularly unmarked" (Skjærvø, 1989b, 233). In Sogdian, the light stem nouns retain a six-case system that includes an accusative. However, the heavy stem nouns, which underwent the phonological reduction described in table 4.4, merge the nominative and accusative into the unmarked direct case (Yoshida, 2009, 306). Based on the assumption that the inherited nominative and accusative are homophonous forms, confirmed by Middle Persian, Khwarezmian, and Sogdian, it is unclear how or why direct objects came to be marked with the oblique case. Analogy between the present nominative-accusative and the past ergative-absolutive system is not a viable solution. The past-tense construction would have had a genitive agent, and the present-tense construction would have had a nominative agent. However, the past-tense nominative object and the present-tense accusative object would have been homophonous.

Sogdian gives us yet another clue as to how to answer this question. In addition to the light-stem nouns, which had stressed final syllables and retained a six-case system, and the heavy-stem nouns, which had stressed stems and lost final syllables yielding a bicasual system, there were vocalic stems. The Sogdian vocalic stems are the reflex of nouns with the suffix *-aka, the same suffix that is thought to be the etymon of the various definiteness markers. The stems became vocalic due to the loss of intervocalic *k. If the Sogdian vocalic stems are juxtaposed with the inherited simplex forms, a new case system emerges.

	SG			PL		
	NOM	ACC	GEN	NOM	ACC	GEN
masc. (a)	-Ø	-Ø	-ī	M.PL = F.PL		
masc. (aka)	-ē	-ē	-(ī?)/ē			
fem (ā)	-Ø	-Ø	-ē	-Ø	-Ø	-ān
fem (ākā)	-ā	-ē	-ē	-ē(t)[13]	-ē(t)	-ētī/-ān

Table 4.6: *a*-stem and *aka*-stem forms

Assuming that the evaluative *aka*-stems became the definite forms, there is now a split case system with a different distribution of formatives in definite and non-definite contexts. Here there would be a definite-indefinite distinction for present-tense objects (ACC) and the same distinction for past-tense objects (NOM).

		M	F	PL
PRS agent	INDF	-Ø	-Ø	-Ø
	DEF	-ē	-ā	-ē(t)
PRS object	INDF	-Ø	-Ø	-Ø
	DEF	-ē	-ē	-ē(t)
PST object	INDF	-Ø	-Ø	-Ø
	DEF	-ē	-ā	-ē(t)
PST agent		-ī/-ē	-ē	-ān
possessor		-ī/-ē	-ē	-ān

Table 4.7: Hypothetical case system

[13] In Sogdian, the plural forms are built from a collective marker in *ta*, e.g. Avestan: *yazata-* 'divinities.' These forms are declined as if feminine singular (see Yoshida, 2009, for the archaic *aka-ān* ending).

There are several striking resemblances between the forms of this hypothetical system and what is observed in Kurmancî and other Western Iranian languages (table 4.7). There is no marking on non-definite forms (labeled INDF). This results in differential case marking of both present- and past-tense objects but not past-tense agents. However, mapping of forms is imprecise; the definite feminine form *-ā* always occurs as *-ē*. Likewise, the *ē*-form plural is missing from Kurmancî, at least in the system of simplex nouns. However, all Central and Southern Kurdish varieties have generalized the oblique plural suffix *-an* as a plural suffix. It is not unreasonable to assume that this process took place in stages, becoming the definite plural suffix (parsing already available in Middle West Iranian).

Despite these similarities, there are quite a few conceptual hurdles to cross before a system like this hypothetical one can render a system like the one observed in Kurmancî. I assert that Kurdish historical syntax and the history of the ezafe hold the key to how this hypothetical system levels out to become the one we observe today in Kurmancî.

The Kurmancî ezafe reflects the development of a combination of K-type definite suffix and the inherited absolute/indefinite ezafe. In chapter 2, I describe the ezafe system of Kurmancî. The main relevant point is that the ezafe has an "indefinite form" and another that is used in definite contexts. The indefinite ezafat *-î* [EZ:M.SG.IND] and *-e* [EZ:F.SG.IND]/[EZ:PL.IND][14] occur with modified nouns featuring the indefinite suffix *-(y)ek*, the indefinite plural suffix *-(h)in*, on the (absolute-state) nominal complements of light verbs, nouns used generically, and on nouns with the definite suffix *-eke* in the Surçî variety. Except for the occurrence in a definite context in the Surçî variety, the indefinite ezafat are limited to non-definite contexts.

In definite and specific contexts, the ezafe variants are *-ê* [EZ:M.SG.DEF], *-a* [EZ:F.SG.DEF], and for [EZ:M.SG.DEF], there are *-ê, -êt, ên, ê ti,* and *ê di* depending on dialect and register. It should stand out that these forms are identical to the nominative definite forms in the

[14]The indefinite plural ezafe is identical to the indefinite feminine singular ezafe. This is likely due to the grammatical gender of *hin* 'stuff,' which is thought to be the source of the indefinite plural suffix *-in*.

hypothetical system (based on the Sogdian vocalic declension. Of course, the similarity of the Sogdian plural in *-ēt* may be coincidental; the existence of the discontinuous forms suggests a separate lexeme as the ultimate etymon. For the discontinuous forms *-ê ti* and *-ê di* it is clear that the parsing should be [EZ:DEF PL].These forms bear a striking similarity to what is observed in the Northern Zazaki indefinite, e.g. *-o dı* [-EZ:M.SG IND], *-a dı* [-EZ:F.SG IND], etc. However, these seem to be independent innovations as the link between plural, and indefinite is not supported in the rest of the language, i.e. by a collective plural marker. As the dental consonant is marking plural, it is tempting to assume that it is the reflex of a collective marker of some type. The discontinuous nature of the formative seems to suggest that it is unlikely to be the reflex of the Old Iranian collective suffix *-ata. However, this point is tangential, and as such, I avoid further speculation.

If the noun with the definite suffix *-eke* was followed by the absolute/indefinite ezafe *-î*, the resulting sequences would contract from *-ek-î*, *-ekâ-î*, and *-ekê-î* to *-ê-î*, *-â-î*, and *-ê-î*, for masculine singular, feminine singular, and common plural, respectively. Although these forms have been hitherto unstudied in regards to Kurmancî, there is a precedent in neighboring Zazaki for the change. For instance, in the variety of Çermik-Severik, Hadank (1932) shows wāy-āy [sister-EZ:F.SG] ǰey [3SG.M.OBL] "his daughter" (Hadank, 1932, 73), which features the feminine singular ezafe *-āy* (-aî in Hawar (Kurmancî) orthography). This form has become *-ā* in the modern variety of Çermik-Severik (Paul, 1998b, 39, §.54(e)). Likewise, in Hewramî MacKenzie (1966) shows that the ezafe is blocked after certain formatives for what he deems to be phonological reasons including stress suffixes *-á,* and *-ê.* It is telling that of the hundreds of paradigmatic permutations possible in Zazaki nearly all of them are built from the formatives *-a,* and *-ê* (see ch. 5 for more on the Zazaki nominal paradigm). *-ê,* [EZ.M.SG] *-a,* [EZ.F.SG] and *-ê* [EZ.PL] are the forms observed in Surçî, Akre, Amadî, Berwar, Zaxo, and Şêxan described in Mackenzie (1961). However, some of these dialects some of the time employ either a dental form in the plural or the nasal (e.g. *-ên*) as is standard in the Cizre-Botan variety. This serves to disambiguate [EZ:M.SG] and [EZ:PL].

Both the plural ending *-an* and indefinite plural ending *-in* are viable analogical sources for the plural ezafe variant *-ên*.

The account that I put forth here entails that the nominative forms of the definite ezafe, the definite article plus the absolute (and in fact only) ezafe, have become the canonical (definite) ezafe in Kurmancî. The effect is that the original definite nominative forms have been reanalysed as being a combination of the direct *-Ø* and the definite ezafe *-ê*, *-a*, or *-ê*. Additional support for this hypothesis comes from two additional idiosyncrasies of the Kurmancî ezafe: (1) there is no "definite construct" as described in chapter 3, and (2) sequential ezafat do not continue the definite ezafe in many varieties.

In languages with K-form definite articles, there is a pattern where the definite head noun carries a definite ezafe (e.g. Soranî: *-e*, (col.) New Persian: *-Ø*, etc.) and the following attributive adjective carries the definite suffix (see ch. 3 for a detailed discussion). This construction is limited to adjectival attribution. There are two ways a definite noun can combine with an adjectival modifier in Soranî (CK) with the definite ezafe as in example (1) and with the definite article followed by the absolute ezafe as in example (2).

(1) ker-e gewre-ke
donkey-EZ:DEF big-DEF
Soranî: 'the big donkey'

(2) ker-eke-î gewre
donkey-DEF-EZ big
Soranî: 'the big donkey'

With the contraction of the Kurmancî definite ezafe forms yielding *ker-ê gewre* for what is observed in example (1), forms like what is observed in example (2) fell out of the language. Note that with possessed nouns, only the second attribution strategy, which is conspicuously missing from Kurmancî, is a possible modification strategy. Example (4) is felicitous, but crucially the definite construct in example (3) is not.

(3) *ker-e Şwan-eke
donkey-EZ:DEF shepherd*def*
Soranî: *

(4) ker-eke-î Şwan
donkey-DEF-EZ PN
Soranî: 'Şwan's donkey'

The consequence of the loss of the definite strategy with nouns modified by an attributive adjective was that there was no morphological or syntactic difference between ad-attributive and ad-genitival ezafat in Kurmancî. In Central and Southern Kurdish and Hewramî, the head noun could be modified by any number of adjectives in a definite ezafe construction followed by the absolute ezafe and a possessor. However, In Kurmancî, with the loss of the definite ezafe construction, a noun could only support a single modifier, which led to the development of the so-called floating construct or secondary ezafe construction. The floating construct is a syntactic solution to a problem exposed by a morphological change.

One issue that needs to be explained with this hypothesis is how there would be a sequence of definite ezafat in Kurmancî if they historically descend from a combination of definite article and ezafe. This historical redundancy is illustrated in example (5), where the contracted definite ezafe forms are expanded in brackets.

(5) [kûçik-eke-î] kûçik-ê spî [=(y)eke-î] =yê Şivan
[dog-M.SG.DEF-EZ] dog-EZ:M.SG.DEF white [=M.SG.DEF-EZ] =EZ:M.SG.DEF PN
*Kurmancî (hyp): Şivan's white dog.'

Although there was likely no historical state where the etymon of the article would have been repeated (i.e. at a past synchronic state), there is a possible analogical source for this seeming repetition. Once the complex ezafe forms are reinterpreted as definite ezafe, they can be used as the definite ezafe both in primary (i.e. connected to the head noun) and secondary (i.e. displaced by a modifier) constructs. Of course, this is not an issue in many/most dialects, which have preserved the absolute/indefinite form of the ezafe -

î in secondary contexts (Mackenzie, 1961, 160). Assuming that this, the most common construction, is the original construction, then the Kurmancî ezafe system mirrors the Central and Southern Kurdish precisely. This is, of course, with the caveat that the definite K-form suffixes reduced phonologically, obscuring their origin, albeit not their function.

Implications on the case system of Kurmancî With the definite nominative forms being repurposed as the definite ezafe, a type of balance was restored to the Kurmancî system. This balanced system is illustrated in table 4.8. Here I show the forms of the nominative repurposed as the ezafe. This form neutralizes all other case-marking as it would not have historically been able to occur with other forms. The present-tense agent, which is always unmarked, now matches the past-tense agent, which is always marked regardless of tense or definiteness. The differential object marking, which is definiteness-sensitive, not tense-sensitive, is preserved. However, there are a few changes that are not accounted for though not necessarily unmotivated. The definite oblique plural levels out in all varieties adopting the indefinite/definite plural possessive/agent endings. Likewise, the indefinite/definite masculine-singular possessive/agent endings tend to level out as the definite masculine-singular ending, either adopting the non-definite form perhaps in analogy with the demonstrative forms. In some dialects (e.g. Muş, etc.), the definite form is selected, collapsing grammatical gender distinctions.

	INDF			DEF		
	M	F	PL	M	F	PL
EZ	-î	-e	-e	-ê	-a	-ê
PRS agent	-Ø	-Ø	-Ø	-Ø	-Ø	-Ø
PRS object	-Ø	-Ø	-Ø	-ê (-î)	-ê	*-ê (-an)
PST object	-Ø	-Ø	-Ø	(-ê/-î)	(-ê)	(-an)
PST agent	-î (-ê)	-ê	-an	-ê (-î)	-ê	-an
possessor	-î (-ê)	-ê	-an	-ê (-î)	-ê	-an

Table 4.8: The reorganized Kurmancî system

There is a not-insignificant amount of variation in this table. However, it reflects the

variation observed in Kurmancî varieties. Some varieties feature differential object marking across the board, while others only in the present tense. However, all varieties have inherited the definite ezafe, which behaves differently from simplex ezafe forms. Paradoxically, this phonological reduction has given way to a fusional form preserving the distinction of grammatical gender eliminated in other varieties.

Note that when the formerly nominative forms were repurposed as the definite ezafat, the present-tense agent and past-tense object forms were lost. The result for the present-tense agent, no marking, is observed in all varieties. However, most varieties show some sort of oblique object marking in the past (i.e. DOM). However, there are both Kurmancî varieties that have DOM only in the present-tense (courtesy of an anonymous reviewer for Karim, 2021c, forthcomin) and in some of the Caspian languages e.g. Leriki (Talyshi) (Stilo, 2018b). The oblique marking of past-tense objects requires an additional analogical innovation with present tense objects, genitival possessors, and prepositional complements as obvious sources. Perhaps the genitival possessor is the most likely of these as there are ambiguous contexts where the possessor appears in object position. For instance the noun *mamoste* 'teacher' in example (6) is marked for feminine singular oblique *-(y)ê*. This marker shows that it is the genitival complement of an ezafe-marked noun *kurê* [son-EZ:M.SG]. However, *kur* 'son' is otherwise in object position and the whole noun phrase *kurê mamosteyê* ends in the oblique suffix *-(y)ê*.

(6) Min kur-ê mamoste-yê dît
M.SG.OBL son-EZ:M.SG teacher-F.SG.OBL see.PST.3SG
Kurmancî: 'I saw the teacher's son.'

The exact details are difficult to establish without reasonable doubt as there is a great deal of variation and many possible analogical sources. This section outlines the case for such a development without committing to whether the ultimate source of the forms is from analogy to one particular function or another. Instead, it is my aim here to establish that

Kurmancî has gone through the same changes proposed for colloquial New Persian by Jahani (2015) or Median of Kašan by Borjian (2012b), etc. The *-aka forms marking definiteness contracted, creating a definite article. In addition to this, the retention of case- and gender-marking along with further analogical changes transformed a relatively straightforward definite suffix into the complex system of IOM observed not just in Kurmancî but also in Talyshi and Tatic languages (see Stilo, 2018b).

The reflexes of -aka are comparable to Kurmancî inflectional suffixes. Assume that the present hypothesis that the Kurmancî nominal paradigm results from the convergence of two stems, the non-definite nominal stem and the definite stem from the noun in *-aka, is correct. In this case, we have three different systems: (1) the Kurmancî-type system with differential case marking, (2) the Soranî-type with the definite article *-eke*, and the colloquial New Persian-type with the contracted definite article *-e*. Assuming Jahani's (2015) hypothesis that the New Persian type article is the reflex of Middle Persian *-ag*, case and gender were lost before the phonological change, which resulted in the loss of the velar obscuring the origins of the formative. In the Kurmancî system, there was no such reduction before the change took place, resulting in a maximally distinguished system with definite nouns but not with non-specific non-definite ones. It stands to reason then that there would be every possible iteration of these changes. That is exactly what is observed. However, there are so many stages in the process proposed here that the results vary across the Iranian languages. For instance Sīvandi, a central Iranian Kermanic language spoken in Fars among K-type definite varieties, shows gender marking in definite ezafe-constructions, e.g., kor-i me 'my son,' žen-a me 'my wife,' and with the "deictic suffix" (definite article) *-ű́* for masculines, e.g., quč-ű́ 'the ram,' and unstressed *-e* for feminines, e.g., *usúr-e* 'the horse' (Windfuhr, 1991). This system is the same as what is observed in Kurmancî only without IOM.

4.3 Impact of this hypothesis/further research

The hypothesis that the Kurmancî case system is the reflex of the de-evaluative definite suffix is important on several levels. For Kurdish studies, it serves to unify the various nominal systems observed in Kurdish varieties Northern, Central and Southern. The variation/diversity observed among these varieties has led me to, at times, question the closeness of their genealogical grouping. However, as more information comes in and diachronic theories are generated, it seems that their closeness is to be continually confirmed. When MacKenzie (1961) claimed that the differences between Northern and Southern Kurdish were attributable to Gorani influence on the Southern dialects, the implication was that without that influence, Northern, and Central and Southern groups would be more alike. This description is undoubtedly confirmed here. Although, I assert that two factors (1) closeness with Zazaki and (2) Zazaki and Kurmancî's mutual participation in a northern belt (sprachbund?) was ultimately responsible for the changes observed. At least the Hewramî core of Gorani[15] has avoided changes that took place in the other local languages.

Assuming that Jahani's (2015) hypothesis that the colloquial New Persian definite article *-(h)e* is indeed the reflex of the Middle Persian evaluative suffix *-ag*, then the current hypothesis unites the systems of perhaps the whole of Western Iranian (not just Central Iranian). The languages with K-type definiteness markers include the Kurdish (Central and Southern), the Lori group, and many of the so-called Fars dialects. The languages that show a contracted form of the definiteness suffix either as *-ú, -é,* or *-ó* include languages within the Central Iranian Kermanic group, the Median dialects of Kašan, Kumzari, the dialect of Bandar Abbas, Judeo Isfahani, Khuri, etc. Among these, Sīvandī is of particular note; it shows both reduced forms *ū* [M.SG.DEF] and *-e* [F.SG.DEF] showing how both formatives might have come into the language with one subsequently being adopted as the sole form. This group also contains colloquial New Persian as suggested by Jahani (2015).

[15]There are Gorani varieties that have gone through changes under the influence of Central and Southern Kurdish such as the Gorani variety of Gewrecû described in Mahmoudveysi et al. (2012)

It is known form Middle Persian that gender was lost early on, and later, case was sacrificed in favor of a number distinction, i.e. the oblique-plural suffix became just a plural suffix.[16] The Middle Persian evaluative suffix *-ag* would not have carried case or gender suffixes, but it may have carried a plural suffix. When the /g/ was eventually lost, one would expect a system with a definite singular *pesær-e* 'the boy [boy-DEF]' and definite plural *pesær-æg-an* 'the boys [boy-DEF-PL].' However, at some point in the history of colloquial New Persian, the Middle Persian collective plural suffix *īhā* came to replace the inherited plural marker *-ān*. There was an inherent incompatibility between the evaluative marker *-ag* and the collective marker *īhā*, the former selecting an individual and the latter a crowd. The result was a system where definiteness was collapsed in the plural. *pesær* 'boy(s),' *pesær-e* 'the boy [boy-DEF.SG],' and plural *pesær-a* 'the boys [boy-PL].'[17]

Stacking of defining and deictic elements. One theory of the composition of the definite suffix *-eke* in Central Kurdish is that it contains the evaluative *-ek* and the deictic *-e*. In Central Kurdish, the suffix *-e* is used when a noun is preceded by a demonstrative, e.g. *ew kuř-e* 'that boy [that boy-DEI].' Additionally, it occurs in a variety of poorly understood conditions (Öpengin, 2016, 60). My original, speculative, solution was that this was something akin to the double determination observed in Greek. However, the nominal system of Koroshi Baluchi may shed light on this development. In section 2.2.1, I describe the case system of Turkmen Baluchi as being a bicasual nominal system overlaid by a *ra*-marked system. In contrast, the Koroshi Balochi system features the innovative *ra* marking only in the pronominal system, e.g. *man-ā* [1SG-OBJ], *ta-rā* [2SG-OBJ], *mā-rā* [1PL-OBJ],

[16] The loss of case in Middle Persian is in line with Kuryłowicz's fifth law of analogy "In order to reestablish a distinction of central significance, the language gives up a distinction of more marginal significance" (Kuryłowicz, 1947 apud Hock, 1991, 227). Although there is a certain aspect of what is considered 'marginal' that is admittedly post-hoc, this tendency seems to generally make sence. As for Middle Persian, the singular oblique case ending had already been lost. The result was that case was only a salient category in the plural. In this sense, case became marginal in more than an abstract sense before the case distinction was lost to reinforce the plural distinction.

[17] I am basing this incompatibility between definite and de-collective plural on the New Persian data as well as languages that have collective plurals with clear etymologies, e.g. Kirindî (Southern Kurdish) -EYL [-DEF.SG] < *gel 'flock.'

šomā-rā [2PL-OBJ], etc. (Nourzaei et al., 2015, 46). These forms occur as expected with the /r/ occurring only in post-vocalic position. In contrast, nominals occur in the oblique case with the suffix -*ā* regardless of whether the noun is vowel- or consonant-final, e.g. *ažd*a*hā-ā* [dragon-OBL] (Nourzaei et al., 2015). There are two things that can be gleaned from the word *ažd*a*hāā*: (1) *ā* can be assumed as the reflex of a middle Iranian *ag* suffix, as *ažd*a*hā* has Middle Persian and Avestan cognates *ažd*a*hāg* and *ažiˉ dahāka-*, respectively; and (2) the repeated /ā/ in *ažd*a*hā-ā* acting as an oblique suffix could be the same formative debonded and repurposed as the oblique suffix.[18]

Nourzaei (2020) clearly shows that among the closely related Koroshi, Coastal, and Sistani Balochi varieties exhibit different stages of the grammaticalization of the definiteness marker -*ok*. Coastal Balochi only employs the marker in the original evaluative use; Sistani Balochi uses -*ok* both in evaluative contexts and deictic/recognitional contexts, and Koroshi uses -*ok* in both deictic/recognitional contexts and anaphoric definite contexts but crucially not the original evaluative contexts. This cline must have taken place fairly recently in the time after the split of a common Balochi. However, if my hypothesis that the case marking systems of Iranian and the definite markers all have their origins in these evaluative suffixes, then what is occurring in many of these languages is a spiral of recruitment. Evaluative morphology becomes definiteness marking with a particular distribution indicative of its origin as evaluative morphology.[19] Over time that function generalizes into any of the patterns observed in the various languages. Other evaluative markers are then recruited for the same function, and a complex system is formulated. In the Koroshi Balochi data given

[18] An additional point that I have yet to address is that at least in one example dar'yā-hā [sea-OBL] the -*ā* suffix occurs with an epenthetic /h/ to break the vowel hiatus(Nourzaei, 2020, 23). However, it is hard to say for certain what can be gleaned from this form. It is most certainly a borrowing from Persian as the native Balochi cognates are *zirā* 'sea' and *zirih* 'spring' (Korn, 2005, 380). As discussed in section 2.3.2, the colloquial New Persian definiteness marker *-(h)e* has an /h/ that is inserted after vowels. It is possible that this phonetic anomaly was borrowed into Balochi along with the lexeme.

[19] This retention of idiosyncratic features indicative of a particular formative's diachronic origins is often refered to as persistence, Hopper's (1991) fourth principle of grammaticalization. He defines Persistence as follows: 'When a form undergoes grammaticization from a lexical to a grammatical function, so long as it is grammatically viable some traces of its original lexical meanings tend to adhere to it, and details of its lexical history may be reflected in constraints on its grammatical distribution."

in Nourzaei (2020), the only marker to be use as an anaphoric definite is *-ok*, while the evaluative markers include *-ok, ak, ek, lok,* and *o*. Likewise, the only marker of definiteness in Central Kurdish is *-eke*, while Mackenzie (1961) gives *-çe, -çke, -ek, -ik, -ke, -eke, -ōk, -ōke, -kele, -le, -leke, -île, -ûle, -ûlke, -ɬe, -eɬe, -ōɬe,* and *-ōɬke* as diminutive morphemes (evaluative here).

It is a telling coincidence that so many of the K-type formatives exist in languages alongside a deictic suffix, e.g. Central Kurdish: *-eké* ∼ *-é*, Luri: *-ekū* ∼ *-ū*, etc. It remains purely speculative at this point, but there are three possibilities for the development of these bipartite formatives: (1) the K-type formatives involved two stages in a spiral of grammaticalization. The first step involved the development of a deictic/definite marker from the *-ag* suffix. The second stage involved the recruitment of another evaluative suffix *-ak/-ok*, etc., which occurs in a subset of the situations that call for the original deictic/definite marker. This distribution is reflected by Northern Luri dialects and Central and Southern Kurdish but not Koroshi Balochi, which can show the definiteness marker without the *-ā* suffix. Of course, Balochi has features of the northern-belt languages where there is only case marking in definite contexts. I assert that definiteness marking was reanalyzed as IOM. Balochi's migratory history has led it to partake in the developments of multiple groups. In a sense, the Koroshi *-ok* suffix does underlie an earlier deictic/definite marking system. It is just that that that system was only active in accusative contexts. *-ok* marking revived that system extending it (back?) to direct contexts as well.

The second possibility (2) is that the K-type formatives involved two stages but not a spiral of grammaticalization. The original deictic/definite marker had some other source. Probably an original definiteness marker from a relative or demonstrative pronoun as proposed in Karim (2018) and discussed in section 2.3.2. The third (least likely given the findings of Nourzaei (2020)) possibility is that there the bipartite origin of this form is from a connector and the relative pronoun *ke, ku,* etc. (discussed in section 2.3.2). However, this solution does not explain the strange usage patterns of the different deictic/definite suffixes,

which are thought to indicate the cline from evaluative to definite (following Haig, 2019a; Haig & Mohammadirad, 2019; Nourzaei, 2020, etc.). It is certain that one of these solutions is necessary because an inherited intervocalic /k/ would have been voiced in all languages (except Balochi following Korn, 2005, 72). In at least Kurdish and Luri it is necessary for the /k/ to have been initially a word-final or word-initial phoneme.

There is much further research that needs to be done on this subject both synchronically and diachronically. This echoes nearly every work that I have cited in this section. The forms and functions of definiteness markers in Iranian languages are truly understudied. Since Haig (2019a) and Nourzaei (2020), there is a clear framework for the synchronic study of the functions and distribution of these formatives. Hopefully, as more information comes in from increased documentation efforts, the exact distribution of these formatives will become more evident. Only then can the historical hypothesis at the core of this chapter be fully evaluated.

Chapter 5

The Origin of the *d*-form Ezafe

5.1 Introduction

One possible way that languages change over time can be described as a reductive process in which regular sound change whittles[1] away at the phonetic content of words. Subsequently, the ability of the linguist as well as the layperson to segment words into their constituent recurring units of form and meaning becomes more difficult or even impossible. The forms of the Kurmancî ezafe exemplify this process. Their realization is that of the nominative singular of the old *-aka stems, e.g. *-a* [нnom.sg.f]. However, their synchronic parsing contains the function of of formatives now lost, e.g. the synchronic *-a* [нsg.f.ez.def] comes ultimately from *-ā-kā-yat, where DEF corresponds to *-āk F.SG to *-ā and EZ to *-ȳat. The resulting units, which feature cumulative exponence, a correspondence of many units of meaning to one form, are described as fusional. The fusional morphological system is just one of several idealized[2] types (e.g agglutinative/isolating/polysynthetic). The Zazaki language, a Northwestern Iranian language spoken in SE Turkey by 1,344,000 people according to ethnologue.com, has a fusional system that seems to have developed said fusionality without the hallmark combination of phonological reduction and reanalysis that

[1] I refer to this process as whittling away because loss and reduction are the effects of change most relevant to this discussion. Other possibilities are fortition and assimilation, which preserve content, and epenthesis, which adds content, etc.

[2] I refer to these types as idealized because most languages occur on a continuum; a fusional language may feature many morphemes with a one-to-one correspondence (though crucially not all). For a description of mixed morphological systems (i.e. not idealized), see Plank (1999). There are additional criteria for distinguishing fusional languages as well (see Igartua (2015)); for this study, I will only consider cumulative exponence as indicative of fusion.

is thought to characterize fusional systems. In other words, there is cumulative exponence without the merging of morphs due to phonological erosion.

In general, Western Iranian languages tend to have relatively little nominal morphology. However, there is a range from Farsi (SW Iranian) at one extreme, which features a one-to-one form-to-meaning correspondence, to Zazaki (NW Iranian) at the other, which features a complex inflection system. Even the medieval languages of the region resemble the Farsi system. According to Paul (1998a), "Parthian (NW Iranian), spoken nearly 2000 years ago [was] in its noun morphology more modern than any of the closely related Northwestern dialects spoken today" (Paul, 1998a, 172). The two attested medieval Western Iranian languages, Middle Persian and Parthian, had already lost case, number, and gender in almost all nominal categories.[3] Paul's comment about the "modern" character of Parthian recognizes the tendency of Iranian languages to shift away from marking case and gender (and sometimes number), as can been seen in Farsi, Luri, Baluchi, Soranî and other languages. Zazaki, in contrast, has retained—or more accurately expanded—gender and case marking. In this chapter, I focus on Southern Zazaki spoken in Çirmek-Siverek as described by Paul (1998b).

At the heart of the S Zazaki nominal declension is the interaction of the ezafe[4] and case marking. The ezafe is most simply described as an adnominal linker that connects a head noun to an attributive adjective or a genitival possessor (Fa.: /pesær-e qævi/ [boy-EZ strong] 'strong boy'). The Farsi example maps the single unit of form /-e/ to the single unit of meaning ezafe, modified. In addition to ezafe, the equivalent phrase in S Zazaki conveys the type of modifier, definiteness, case, gender, and number information (Za. /laʒ-

[3]There is a retention of case in familial terms where there is an alternation between what were formerly the nominative and accusative forms.

Additionally, the genitive plural form of all other nouns was retained as the oblique plural, which stands in contrast with all other forms (i.e. genitive and non-genitive). See Durkin-Meisterernst (2014) for a grammatical description of Western Middle Iranian.

[4]Ezafe (Pl. ezafat) marks a noun that is modified by either a following attributive adjective or a genitival possessor. The distinction between genitive and attributive is in reference to the type of the following modifier. The ezafe additionally is marked for the case, number, and gender of the noun to which it is attached. For a discussion of the ezafe as a case marker with multiple governors, see Samiian (1994).

o bi-quwət/ [boy-M.SG.DEF.EZ:/ATT[5] strong.M.SG] 'strong boy'). Additionally, many of the S Zazaki forms appear with a complex (i.e. with a bipartite origin) ezafe marker featuring a d-form (/laʒ-e-do bi-quwət/ [boy-INDF-M.SG.INDF.EZ:/ATT strong.M.SG] 'a strong boy'). The d-forms are peculiar to S Zazaki; they do not occur in other Iranian languages. This study explores how the interaction between ezafe marking and case marking has led to the development of S Zazaki's fusional nominal morphology.

I examine primarily the synchronic system of S Zazaki described in Paul (1998b), which was the product of a combination of fieldwork and a grammatical description of a book of folktales collected by Berz & Malmîsanij (1951).[6] Through comparison with the Proto-Zaza-Gorani[7] nominal declension, I propose a hypothesis for how the complexity of the S Zazaki developed. Any understanding of the S Zazaki system must focus on two parts, the development of the vocalic formatives, which like Kurmancî, was the result of the loss of the ezafe formative after vocalic case endings, and the development of the d-forms, which I propose were innovated due to syntactic displacement. In other words, S Zazaki's complex nominal inflection has developed from a historical displacement of the proto-ezafe/definite article, which has parallels in S Zazaki's sister languages, Hewramî and Goranî, and the major regional languages, Soranî and Kurmancî. These languages are spoken in the same cities and villages as S Zazaki and its sisters.

[5]Throughout glosses of S Zazaki nouns, I will mark the ezafe type as EZ:/ATT for nouns followed by an attributive adjective and as EZ:/GEN for those followed by a genitival possessor (in the oblique case).

[6]Paul (1998b) was careful to document all the forms he encountered without "correction" to forms that met with prior assumptions about the grammar.

[7]On of the focuses of my current research is to reconstruct the nominal system of the common ancestor of the Zazaki and Gorani languages based on commonalities between the languages augmented by areal trends. The Zaza-Gorani languages included Hewramî, Shabaki, Kandula'i, Paweyane, Zerdeyane, Gorani, S Zazaki, Dimili, Kurdki, Kurmacki, and others. The precise demarcation of these terms and their related dialects is a topic that needs further study.

5.2 Background

Interestingly, the S Zazaki nominal declension is larger, and the syntactic information it conveys is more complex than in any other Western Iranian language. This development appears to be a recent innovation (i.e. not a feature of Proto-Zaza-Gorani (Karim forthcoming a)). The S Zazaki noun inflects for case (DIR and OBL), number (SG and PL), gender (M and F), animacy (animate and inanimate), definiteness (DEF, INDF and unspecified) and modifier type (EZ:/GEN and EZ:/ATT). Additionally, the noun inflects for what Todd (2002) has referred to as subordination and Paul (1998b) has referred to as two additional cases (oblique IIa and oblique IIb); I will refer to this as the prepositional case (PREP) for reasons that are in sections 5.4.4 and 5.4.5. This complex system stands in contrast to S Zazaki's closest sister Hewramî, which has two cases, two genders, two numbers, and two separate particles which mark EZ:/GEN and EZ:/ATT, which is similar to the nominal system that can be reconstructed for Proto-Zaza-Gorani.[8] In keeping with my goal of exploring the interaction between ezafe marking and case marking, I show how the Proto-Zaza-Gorani nominal system (2 cases · 2 numbers · 2 genders = 8 paradigm cells) developed into the S Zazaki nominal system (96-192 cells, depending on whether you posit two cases like Todd (2002), three as I consider here, or four like Paul (1998b)), due to a combination of sound change and analogy.

The majority of languages referenced in this study are Kurdish-Zone languages. The Kurdish Zone should be understood as a contiguous region along the Taurus and Zagros mountain ranges within which the majority of the population are ethnic Kurds. This region contains the three major Kurdish languages: Kurmancî (Northern Kurdish), Soranî (Central Kurdish) and Southern Kurdish, as well as the Zaza-Gorani languages, including S Zazaki and Hewramî, which are, according to Mackenzie (1961), not Kurdish.[9] In addition to

[8]The differences between the nominal declension of Hewramî and that of Proto-Zaza-Gorani are limited to phonological changes in the direct feminine singular and oblique plural, and the loss of gender distinction in the plural, which is robustly attested throughout Western Iranian.

[9]Note that the linguistic divisions do not directly correspond to ideas of ethnic identity in the region. S Zazaki, Goranî and Hewramî speakers within the Kurdistan region both consider themselves Kurds and

the genealogical relationships between these languages, they have converged in terms of phonology, vocabulary, and aspects of morpho-syntax. According to Jügel (2014), this convergence zone or sprachbund should instead be thought of as two sprachbünde. There is a Northern zone consisting of Kurmancî and S Zazaki, and a Southern zone consisting of Soranî and the Goranî languages, including Hewramî. These relationships are relevant because it is along geographic lines that we can observe parallel changes, some of which are at the heart of this study (see section 5.4.5).

I propose a new etymological account of the various formatives involved in creating the S Zazaki nominal declension. This proposal is based on a comparison between the system observed in S Zazaki and the systems of the closely related Zaza-Gorani languages, as well as the primary contact languages Kurmancî and Soranî. For this exploration into the S Zazaki nominal declension, I consider S Zazaki's most adequately described sister languages, Goranî and Hewramî. As an imperfect exemplar of S Zazaki in the Old Iranian period, I consider the ancient Iranian language Avestan. Baluchi (NW) and Pashto (SE) are included for a broader Iranian perspective (e.g. if it exists in Pashto, Baluchi and S Zazaki, it may be reconstructable for Proto-Iranian).[10] This etymological account approaches S Zazaki's development into a fusional system from two directions: comparative phonology and historical syntax. This chapter focuses on the historical syntax that led to the development of S Zazaki d-forms. However, a summary of the phonological developments are included in the following section

Additionally, when a particular feature is not relevant in the context of another, I will omit it to aid in the organization of data; for example, gender distinction is neutralized in

are considered Kurds by the majority Kurdish population. Additionally, there are dialects of North-Eastern Neo-Aramaic (Semitic), Domari (Indo-Aryan), and Eastern Armenian spoken in the region.

[10]Data from the various languages cited in this study come from a variety of grammatical studies and articles, including Avestan (Jügel, 2017; Jackson, 1968), Baluchi (Barker & Mengal, 2014), Farsî (Thackston, 1993), Goranî (Gewreju) (Mahmoudveysi et al., 2012) and Goranî (Zerde) (Mahmoudveysi & Bailey, 2013), Hewramî (MacKenzie, 1966), Kurmancî (Bedirxan & Lescot, 1986; Blau, 1976), Pashto (Tegey & Robson, 1996; David, 2014), Soranî (Mackenzie, 1961; McCarus, 1956; Thackston, 2006b), and S Zazaki (Paul, 1998b; Todd, 2002; Hadank, 1932).

the plural, so I list masculine, feminine, and plural as mutually exclusive categories.[11] I assume a three-case system that combines Paul (1998b)'s third and fourth cases (Obl.IIa and IIb). The differences between these forms are predictable based on noun class membership and conditioning environment (see section 5.4.5).

5.3 Paradigmatic Changes

As mentioned in section 5.2, S Zazaki has developed a nominal declension system maximally distinguished with 144 paradigm cells calculated based on a three-case system. One of the sources of this paradigmatic explosion is phonological reduction which can be observed synchronically in Hewramî (MacKenzie, 1966). A feature of this reduction is that only one of two morphemes, the ezafe particle or the original case ending, is allowed to occur. This selection of a morpheme from separate etymological sources is similar to one way in which suppletive stem allomorphs have been known to develop. The inciting incident was phonological change, which led to the loss of entire morphemes within a particular sequence and an analogical reorganization of the S Zazaki nominal paradigm.

In S Zazaki, only masculine nouns reliably mark the distinction between modifier type: EZ:/ATT and EZ:/GEN. For this reason, the forms of the masculine declension are most salient. According to Gippert (2009), the masculine forms, particularly the oblique ending, descended from the *-(a)ka diminutive/evaluative suffix. The formative *-(a)ka is the etymon of the various definiteness markers in Western Iranian languages (see ch. 4). Like Kurmancî, this is responsible for differential object marking and the ezafe formative -*ê*. As is shown by the languages that have preserved a reflex of the *-(a)ka suffix, a definite noun has two strategies for adjectival attribution, the definite construct (ex. (1)) and the definite noun followed by the ezafe (ex. (2)).

[11] I do not claim that masculine, feminine and plural are three possible values for the same feature; this is merely an expositional device.

(1) hær-æ gewræ-kæ
donkey-EZ:DEF big-DEF
Hewramî: 'the big donkey'

(2) hær-ækæ-î gæwræ
donkey-DEF-EZ big
Hewramî: 'the big donkey'

With possessed nouns, only the second attribution strategy is a possible modification strategy. Example (4) is felicitous, but crucially the definite construct in example (3) is not.

(3) *hær-æ Şwan-ækæ
donkey-EZ:DEF shepherd-DEF
Hewramî: *

(4) hær-ækæ-û Şwan-î
donkey-DEF-EZ shepherd-M.SG.OBL
Hewramî: 'Şwan's donkey'

In S Zazaki, the second strategy has developed as the masculine singular definite form contracted to *-ê*. Then the ad genitival ezafe *-û was lost after the vowel /ê/. This shift is in line with the synchronic phonology of Hewramî, and the diachronic phonology of S Zazaki. This same distinction was not available in the feminine singular or the common plural. MacKenzie (1966) describes this blocking of the ezafe or the case marker due to the regular phonological rules[12] of the language:

- When two vowels come together, hiatus is resolved by deletion:

  a. i → ∅ / _i

  b. i → ∅ / _u

[12]The assumption that these are indeed phonological rules of Hewramî and not the result of morphological competition for the same slot in a rule block is based on MacKenzie (1966)'s description of the language. There is partial evidence of this in the Hewramî verbal system which hosts the morphs /-u/ [1.SG] and /-i/ [2.SG]. In the few vowel-final stems we find blocking of forms in line with what is seen in the verbal system. However, this is a relatively small class of verbs, and there is no root that ends in the vowel [e]. In the end, confirming MacKenzie (1966)'s analysis may not be possible based solely on these morphological forms alone. My argument in this section assumes that MacKenzie's observation was correct.

c. $\text{i} \rightarrow \emptyset \; / \; \left\{ \begin{array}{l} \text{í} _ \\ \text{é} _ \\ \text{e} _ \end{array} \right.$

d. $\text{u} \rightarrow \emptyset \; / \; \left\{ \begin{array}{l} \text{í} _ \\ \text{é} _ \\ \text{e} _ \end{array} \right.$

- The consequences of the phonological rules on the ezafe and case markers explained as morphological rules:

Rule 1. The ending /-i/ blocks only [EZ:/ATT] /-i/ (a).

Rule 2. The [EZ:/GEN] /-u/ blocks the ending /-i/ (b).

Rule 3. The endings /-í/, /-é/ and /-e/ block both ezafat (c and d).

MacKenzie's (1966) description of phonologically motivated ezafe blocking is supported by the fact that one can predict whether the ezafe will surface based on the phonological form to which it attaches. For instance /siní/ 'tray,' which ends in a stressed /í/, blocks both ezafat in the direct and oblique singular. When the ezafe is attached to a Hewramî noun, these rules are applied, and only one of the morphemes may surface. I refer to the rules as they affect the ezafe and case marker, not according to their raw phonological implications. Table 5.1 contains a selection of words with the relevant endings to illustrate the resulting combinations.

word	+ /-i/ [EZ:/ATT]		/-u/ [EZ:/GEN]	
áwi 'water (SG.DIR)'	áwi	(Rule 1)	áwu	(Rule 2)
bedaḑí 'injustice (SG.DIR)'	bedaḑí	(Rule 3)	bedaḑí	(Rule 3)
hǽre 'donkeys (PL.DIR)'	hǽre	(Rule 3)	hǽre	(Rule 3)
kɪmat͡ʃé 'girl (SG.DIR)'	kɪmat͡ʃé	(Rule 3)	kɪmat͡ʃé	(Rule 3)

Table 5.1: Blocking of the ezafe in Hewramî

The precise details of the development of the vocalism of the S Zazaki nominal declension include multiple stems merging and several formatives unimorphating. These formatives

include the *-aka definite suffix, the proto ezafe -*ya(s)/-yat, and inherited case endings. Perhaps the most important clue to this is the fact that of 96-192 paradigm cells, nearly all the formatives that fill those cells are built from either *-a,* or *-ê.*

Underspecification is required to understand the forms of S Zazaki. For example, if the morpheme /-e/ took on all possible units of meaning from the paradigm cell occupied by /laʒ-e/, it would be parsed [boy-(M).PL.DIR.EZ], which would itself clash in case when attached to laʒ-ane [boy-(M).PL.OBL.EZ]. Based on the understanding of /-an-e/ as [-PL.OBL-PL.EZ], there is no clash between morphemes. Crucially, one (or all) of the paradigm cells filled by the morph /-e/ has influenced the form of the [PL.OBL.EZ:/ATT], yielding /-ane/ for expected /-ano/. There appears to be some kind of "neighborhood effect" (Burzio & Tantalou, 2007) which has influenced the form. I assume that the PL form of the ezafe gets extended to oblique plural and that those two categories are cognitively adjacent. This assumption is reinforced by the fact that a separate morpheme marks the obliqueness. However, as expressed by Joseph (2009), this concept of neighborhood is not explanatory, and the group of paradigm cells that make up such a neighborhood can refer to just about anything. What is important for this chapter's underlying goals–to explore the interaction of case and ezafe in S Zazaki and chart the development of S Zazaki from less fusional to more fusional—is that this analogical change indeed took place and not necessarily why it took place.

	(no ezafe)	EZ:/ATT	EZ:/GEN
DIR.SG	-∅	-o	-e
OBL.SG DIR.PL	-i	-e	-e
OBL.PL	-an	-an-e (*-an-o)	-an-e

Table 5.2: Case and Number by Ezafe Type

What MacKenzie (1966) described as phonological rules for Hewramî have caused either the ezafe particle or the case ending to be lost. The resulting unit's phonological matter has one etymological source, the case marker or the ezafe, which carries the meaning of both.

The diachronic processes (described in MacKenzie (1966)) that caused the other morph to be lost may be described as the erosion of phonological material. The remaining piece features cumulative exponence—the hallmark of fusionality—and is absent any phonological bulk from the lost morph. After these phonological processes had their effect, analogical processes caused the plural ezafe to be extended to both direct and oblique cases. This final step was not phonological but analogical extension motivated by the initial phonological change, which was the inciting incident.

The S Zazaki paradigm has undergone leveling and even loss of distinctions, which have paradoxically increased the size of its nominal paradigm. The phenomena responsible for each stage of the evolution of S Zazaki's nominal declension are attested in related Kurdish Zone languages. These phenomena include phonological reduction, which can be observed synchronically in Hewramî, and which in S Zazaki conspired with analogical changes to create a fusional paradigm. These facts account for most of the forms observed in S Zazaki, but there is more to the story.[13]

5.4 Displacement of the Proto-Ezafe

The phonological/analogical changes described in the previous section account for most of the forms of the S Zazaki nominal declension, except that sometimes a /d/ appears as part of the otherwise expected ezafe marker. The d-form ezafat occur with the indefinite as well as some oblique constructions. In S Zazaki, the combination of oblique and ezafe is divided into forms with /d/, which are complements of pre-, post,- and circumpositions, datives,

[13]There is an open question regarding the discrepancy between the S Zazaki and Hewramî forms. Why does S Zazaki have [DIR.M.SG.EZ:/ATT] /-o/ to Hewramî's /-i/ and [DIR.M.SG.EZ:/GEN] /-e/ to Hewramî's /-u/? A solution to this problem is not immediately apparent based on comparison with Hewramî; an additional clue that may hold the answer could be in the distribution of ezafat in Goranî languages, which are generally understudied, some of which have the same diversity observed in Hewramî. Mahmoudveysi & Bailey (2013) shows a similar set of allomorphs in Zerdeyane as are observed in Hewramî, but with a pattern that is not yet discernible. Future research is still necessary to posit a convincing answer for this possible inversion of forms.

allatives, and adverbials and forms without /d/, which are ergative agents and present-tense direct objects. This division reflects a kind of compromise between the systems of S Zazaki's sister Hewramî and its neighbor Kurmancî (Karim forthcoming b). That is to say, S Zazaki has oblique marking in all circumstances where Kurmancî has oblique marking. The circumstances where Hewramî has oblique marking are a proper subset of those of Kurmancî. In S Zazaki, however, when oblique and ezafe marking are combined, d-forms occur only in circumstances that would warrant the oblique in Hewramî. As there is no relationship immediately clear between indefinite and (some but not all) oblique, I examine these forms independently. Although these forms do not have an apparent grammatical connection, they are united by a shared history as words that could displace the proto-ezafe construction.

5.4.1 The d-forms

The d-forms have been described by Todd (2002) as the subordinated ezafe. Simply put, these forms occur when a noun is followed by a modifier (adjective/possessor) and is the complement of a pre-, post-, or circumposition, as shown in (5) where /zey/ conditions the d-form ezafe in /kolide/.

(5) tɪ ... bi-ya zɛy koli-de wiskɪ
you.DIR ... be.PST like firewood-PREP.PL.EZ dry-DIR.PL
Za. 'you've become like dry firewood' (Berz (1951: 114.14) apud Paul (1998b))

Paul (1998b) has shown that the d-forms occur with nouns acting as genitival possessors, recipients (dative), or verbal goals (allative). Additionally, some of the syntactic domains described by Todd (2002) as conditioning d-forms seem to only optionally occur with them, as in (6) where /padiʃa/ 'king' is the complement of the postposition /=re/ 'to' but does not link to /xo/ 'self' with a d-form ezafe. According to Todd (2002)'s account, one would expect /padiʃade/ instead of /padiʃaye/. At a glance, the fact that this form occurs in

Paul (1998b)'s data set—a collection of folktales ultimately drawn from Berz & Malmîsanij (1951)—points to Paul's claim that the d-form is optional (to be refuted in section 5.4.4).

(6) padiʃa-ye xo=re xɛbɛr-i ɾiʃ-ɛn-o
king-OBL.EZ self=to news-OBL.M.SG send-PRS.IMPF-3SG.M
Za. 'he sends word to his king' (Berz (1951: 212.6) apud Paul (1998b))

The combination of Paul (1998b)'s more nuanced description of the ezafic system and the fact that /-o/ [DIR.M.SG.EZ:/ATT] can occur as part of a complex d-form (i.e. /-do/) when employed in all the contexts Paul has deemed optional has led him to propose two new cases, oblique IIa (which has the option of taking d-form ezafat) and oblique IIb (which always takes d-form ezafat). I have rejected Paul (1998b)'s expansion of the case system due to the fact that the variation between the forms of the oblique IIa and IIb is predictable based on noun class membership and a deeper look at the specific lexemes which govern the appearance of the /d-/ (the focus of my ongoing work). I propose that d-form ezafat have their origins in a particular construction that displaced the proto-ezafe/definite article in the Old Iranian period. In other words, the allomorphs of the ezafe have different etyma, which appeared under different syntactic/semantic conditions in the proto-language.

5.4.2 Displacement of the Proto-Ezafe by the Indefinite Article

In S Zazaki, a noun with the indefinite article /-e(n)/ must always take a d-form ezafe (illustrated with the words /laʒ/ 'boy' and /kɛjnɛk/ 'girl' in Table 5.3). The d-forms of the indefinite ezafe distinguish the type of attribution (EZ:/ATT and EZ:/GEN) and gender (M and F), but crucially not case or number, which are only marked on definite nouns in S Zazaki. The forms of the indefinite ezafe are as follows: /-e-do/ [-INDF-M.SG.INDF.EZ:/ATT], /-e-de/ [-INDF-M.SG.INDF.EZ:/GEN], /-e-da/ [-INDF-F.SG.INDF.EZ].[14] Although the d-forms are peculiar to S Zazaki, there is a trend of incompatibility between the ezafe and the

[14]Note that this strategy is only employed in Southern Zazaki. In the Northern varieties, the simple ezafat *-o, -a, -ê* are used followed by the discontinuous formative *dı*, and the modifier.

indefinite article among the Iranian languages. This trend is likely rooted in the etyma of the two morphemes.

	M.INDF	F.INDF	PL.INDF
DIR	laʒ-e	kɛjnɛk-e	laʒ-e/kɛjnɛk-e[15]
DIR.EZ:/ATT	laʒ-e-do	kɛjnɛk-e-da	laʒ-e/kɛjnɛk-e
DIR.EZ:/GEN	laʒ-e-de	kɛjnɛk-e-da	laʒ-e/kɛjnɛk-e
OBL	laʒ-e	kɛjnɛk-e	laʒ-e/kɛjnɛk-e
OBL.EZ:/ATT	laʒ-e-do	kɛjnɛk-e-da	laʒ-e/kɛjnɛk-e
OBL.EZ:/GEN	laʒ-e-de	kɛjnɛk-e-da	laʒ-e/kɛjnɛk-e

Table 5.3: The Indefinite Nominal Declension

The relative pronoun *ya̲t*—the historical predecessor of the ezafe—may have become a definite article by Young Avestan, which is in line with what Kent (1944) proposed for Old Persian *haya*.[16] The indefinite article (< Av. *aēuua*) and definite articles could not occur in the same noun phrases. In young Avestan, there are no instances of *ya̲t* co-occurring within the same noun phrase with the indefinite article *aēuua*. The Avestan indefinite article always precedes the noun phrase (e.g. *aēuua caxra* 'a wheel' not **caxra aēuua* 'wheel a'). However, its descendant in the Western Iranian languages is enclitic (e.g. Persian: /=i/, S Zazaki: /-e(n)/, Hewramî: /-ew(æ)/, etc.). If the noun-phrase second-position definite article *ya̲t* was reanalyzed, becoming the ezafe, it is just as likely that an indefinite article habitually occupying the same place in the linear order in respect to the noun could be reinterpreted as the ezafe in the context of indefinite. The standard New Persian system, in which the ezafe is blocked on a noun carrying an indefinite article, seems to support this hypothesis (as can be observed in (7) and (8); *kitab-i-e is not possible.).

[15] The indefinite plural utilizes the bare noun stem, which takes the plural ezafe. Additionally, the plural indefinite suffix /-ɪn/, which does not condition the d-form ezafe and has no known etymon, can be used to specify indefiniteness (ex. /laʒ-ɪn-e gɪrdi/ 'some big boys') (Paul, 1998b). This does not cause a problem to my analysis for two reasons: (1) This form may be innovative and is not regularly applied, and (2) it is the form of the indefinite plural found in neighboring Kurmancî with which S Zazaki has grammatically merged in many ways.

[16] Avestan, along with Old Persian, represents the Old Iranian period. Although it is not the direct ancestor of S Zazaki, Avestan can be assumed to be an imperfect exemplar of what S Zazaki might have been like at that time. At the time in which Avestan was attested, its divergence from Proto-Zaza-Gorani was fresh.

(7) ketab-e xub
book-EZ good
Fa. 'the good book'

(8) ketab-i xub
book-INDF good
Fa. 'a good book'

S Zazaki's close relative Hewramî allows the ezafe to occur with the indefinite article in the dialect of Luhon (MacKenzie, 1966), but it is strongly disfavored. In conversations with a Hewramî speaker from Pawe city, he deemed the use of the ezafe with the indefinite article ungrammatical (p.c. with Hişmet Şiyenî). Additionally, it seems that in cases where the indefinite article seems to host the ad-attributive ezafe -*î*, it may be the oblique (topical) suffix -*î*.

I propose three outcomes of the Old Iranian definite and indefinite articles, as the ezafe developed in various Western Iranian languages: (1) an ezafe has developed from the definite article, which is incompatible with the indefinite article; note this is not the same as the *-aka definite article; (2) an indefinite and a definite ezafe have developed (Soranî); and (3) an ezafe has developed from the definite article which is incompatible with the indefinite article, and a new ezafe is created to perform its linking function in indefinite noun phrases (S Zazaki). The S Zazaki d-form ezafe is always employed in the context of indefinite.[17] A noun phrase cannot carry more than one of the semantic values definite or indefinite. This is exemplified by the Old Iranian evidence where only one of these articles may surface in the linear order of a noun phrase. This differs from the following examples (5.4.3 and 5.4.4) in that there is a semantic motivation for which lexeme was ultimately selected as the ezafe. The following sections are dedicated to exploring other occurrences of d-forms in S Zazaki, or the words that displaced the proto-ezafe in S Zazaki's ancestor.

[17]The S Zazaki d-form ezafe has not descended from the historic definite article. Its precise etymon will be discussed further in section 5.4.3.

5.4.3 Displacement of the Proto-Ezafe by the Head Noun in Genitival Constructions

Like indefinite nouns, nouns serving as genitival possessors always take d-form ezafat. When a genitival possessor modifies the head noun, the possessor is in the oblique case; when that oblique noun is itself modified by an attributive adjective or genitival possessor, it takes a d-form ezafe (as in (9)).

(9) meydān-ē ǰāmī-dā siyā
plaza-M.SG.EZ:/GEN mosque-PREP.F.SG.EZ black
Za. 'the plaza of the black mosque' (adapted from Paul (1998b): 208.32)

The word order in S Zazaki is fixed, with the head noun always in the left-most position and subsequent modifiers following it directly. In the case of (9), the d-form, which agrees in case, number and gender with ǰāmī 'mosque,' allows only the reading 'black mosque' (not 'black plaza').[18] Although we have no record of S Zazaki in the Old Iranian period, one might assume a situation similar to what is observed in Avestan. In (10), the noun *asp-* 'horse' and the adjective *sām-* 'black' agree in case, number and gender [GEN.M.SG], but they are separated by the head noun *kəhrp-* 'body,' which the black horse possesses. The occupation of noun phrase second position by the possessed head noun precludes the presence of the definite article *yaṯ* in that position.

(10) asp-ahe kəhrp-a sām-ahe
horse-GEN.M.SG body-INS.F.SG black-GEN.M.SG
Av. 'in the form of a black horse' (Yt.8.21)

(11) [kəhrp-əm] sām-ahe kərəsāsp-ahe yaṯ gaēs-āuʃ
body-ACC.F.SG B.-GEN.M.SG K.-GEN.M.SG DEF curly-haired-GEN.M.SG
Av. 'the body of the curly-haired K. B.' (Yt.13.61)

(12) sām-ahe.kərəsāsp-ahe kəhrp-əm taṯ gaēs-āuʃ
PN-GEN.M.SG.PN-GEN.M.SG body-ACC.F.SG DEF curly-haired-GEN.M.SG

[18] cf. /meydān-ē ǰāmī-yē siyā/ 'the mosque's black plaza'

'the body of the curly-haired K. Black' (Adapted to reflect proposal for PZG)

In Avestan, there are no examples of *ya̲t* marking a definite possessor that maintains the canonical word order in a genitival construction, possessor > possessum > adjective[modifying the possessor] (seen in (10)). Instead, the repair word order, possessum > possessor > *ya̲t* > adjective[modifying the possessor], is employed (seen in (11)). I propose that, unlike the Avestan system, when an attribute of a genitival possessor was separated from the head noun by its possessum displacing the second-position definite article, the far-demonstrative would take up the role of the definite article (**ta̲t*), maintaining the canonical order word order. This is illustrated in example (12) which takes the form from example (11) and returns the possessum to the canonical order found broadly in Avestan; then, the distance between the definite article and its noun conditions the use of *ta̲t* for expected *ya̲t*.

If accepted, this displacement of the definite article by the head noun in genitival constructions could be responsible for d-form ezafat not just with genitives but with secondary postpositions as well. These are the denominal postpositions /bin/ 'under N < back,' /dimi/ 'behind N < mouth,' /gore/ 'according to,' /het(i)/ 'to N < towards, compared to N' /miyān/ 'in N < inside,' /pey/ 'behind N,' /ser/ 'on N < over,' /ver/ 'before N < outside,' and /zeře/ 'in N < inside.' These have their historical origin in full nouns, but are used idiomatically as postpositions. In (13), the head noun /lingāndē/ is marked as [OBL.PL-PREP.PL.EZ], where the d-form is conditioned by the postposition SER 'on < head').

(13) ling-ān-dē peyēn-ān **ser**
legs-OBL.PL-PREP.PL.EZ rear-OBL.PL on('head')
Za. 'on its hind legs' (Berz (1951: 97.3) apud Paul (1998b))

The genitival relationship is not clear from this use, but upon examination of other Iranian languages, the genitive origin of the construction becomes clear. In Baluchi (14), the postposition is clearly the head noun, and its complement is in the genitive case. The PP in (14) translated as 'on the table' can be literally translated as 'on the table's head.' Likewise, in

the Pashto (SE Iranian) example (15), 'from above the village' can be literally translated as 'from the village's top.' Each example shows the postposition acting as the head noun.

(14) mez-əy sər-a
table-GEN head-LOC
Ba. 'on the table' (Barker & Mengal (2014))

(15) də-kil-i lə-pās-a
GEN-village-OBL.M ABL-top-ABL.M
Pa. 'from above the village' (David (2014): 9.17)

Based on Avestan evidence, there is reason to believe that, in addition to indefinite contexts, genitive contexts displaced the second-position definite article.

5.4.4 Displacement of the Proto-Ezafe by a Postposition

(16) dest bi lež-dē xo k-en-ē
hand on fight-PREP.M.SG.EZ self do.PRS-IPFV-3PL
Za. 'they begin (lit. 'lay hands on') their fight.' (Berz (1951: 107.8) apud Paul (1998b))

(17) mā do ... ležē xo bi-domn-ē
1PL=FUT ... fight-M.SG.EZ:/GEN self SUBJ-continue.PRS-1PL
Za. "we will continue the fight ...' (Berz (1951: 107.24) apud Paul (1998b))

S Zazaki prepositions condition d-forms, as can be seen in (16) where the preposition BI conditions the occurrence of a d-form ezafe. In contrast, the same phrase can be seen in (17) where /ležē xo/ occurs without the d-form in the absence of the preposition. Except for the denominal postpositions described in section 5.4.3, the prepositions, postpositions, and circumpositions of S Zazaki are thought to have descended from PIE prepositions and postpositions (sometimes acting as preverbs) (p.c. with Agnes Korn). However, their position within a phrase in S Zazaki is not indicative of their position in the Old Iranian period. It is their position in the Old Iranian period that conditioned the displacement of

ya̲t resulting in d-form ezafe in S Zazaki. As can be observed in the Avestan example (18), the postposition *hacā* occupies noun phrase second-position; this caused the article *ya̲t* to occupy third position; N *ya̲t haca* would not be possible. This displacement is fundamentally different from the semantically conditioned total incompatibility of the indefinite article with *ya̲t*. However, it bears a similarity to what is seen in the genitive (ex. (11) and (10)), where either *ya̲t* occurs in sentence second-position or the possessum does but not both. I propose, as I did for the genitive, that as the distance increased between the definite article and the noun it governs, the far-deictic **ta̲t* took over as the definite article, ultimately leading to the d-forms observed in S Zazaki.

(18) aṣ̌āt hacā ya̲t vahištāt
Asha.ABL.SG.N from DEF.SG.N greatest.ABL.SG.N
Av. 'from the greatest Asha'

This concludes the examples that Paul labels oblique II b, which always take d-form ezafat (except for the dative, which is addressed in section 5.4.5). The remaining d-forms were thought to be optional by Paul (1998b), although I reject their optionality. There is no coherent set of categories that condition the remaining d-form ezafat. Instead, specific lexemes always take arguments that host d-form ezafat (e.g. YEN- 'to come'), and membership in this group (d-form conditioners) is not otherwise predictable. Lexemes that do not take arguments that must host d-form ezafat can host d-form ezafat if conditioned by some other structure. The following section shows that these examples are also the likely result of displacement of the proto-ezafe/definite article in the Old Iranian period.

5.4.5 The Remaining Forms (Paul's Oblique IIa)

Firstly, it is essential to point out that Paul's Oblique IIa and Oblique IIb are identical in the feminine singular and common plural. The difference is in the masculine singular, which supports a definiteness distinction as well as an animacy distinction. This is likely

due to the bipartite origin of the masculine stems, e.g. nominal stems in *-a and *-aka. It follows that the root of the distinction between Paul's Oblique IIa and Oblique IIb is also related to this bipartite etymology.

The remaining forms which have yet to be accounted for are the allative (Paul (1998b)'s Obl. IIa) and the dative (Obl. IIb), which appears to be the key to understanding the former, as well as complements of the primary postpositions (Obl. IIa). Paul (1998b) has divided Obl. II (d-form oblique) in two, because one set of functions (Obl. IIa) is characterized by a distinction between EZ:/ATT and EZ:/GEN and "optionality". I reject this description because what appears to be optionality across the category allative is lexical specification. D-forms always occur as complements of certain verbs (e.g. /aməjɪʃ/ 'to come' and /kɛwtɪʃ/ 'to fall/happen (into)'). Likewise, primary postpositions only condition d-forms in combinations with specific verbs or specific prepositions (e.g /vana NP=re/ 'to say to NP'). There is no additional evidence of a distinction between the dative and the allative within the Kurdish Zone languages. Though it is not impossible for the distinction to be innovative, additional S Zazaki-internal evidence shows that there is likely no distinction at all (Karim, forthcoming b). The innovation of the distinction between EZ:/ATT and EZ:/GEN, which should in theory be blocked by /-e/ [M.SG.OBL] (see section 5.3) and distinguishes the allative from the dative, is exactly what is expected of [−ANIM] nouns (Paul, 1998b); they never appear in the oblique case, and therefore do not neutralize the ezafic distinction in the oblique (as described in section 5.3). Paul (1998b)'s categories Obl. IIa and IIb can be collapsed into the prepositional case based on the fact the morpho-phonological distinctions of S Zazaki are predictable based on noun class membership ([+ANIM] or [−ANIM]), and that there is a historical preposition associated with what is now the lexical specification observed with the dative and allative.

There is a connection between the lexical specification, which conditions d-form ezafat and the directional particle (Mahmoudveysi & Bailey, 2013) and demonstrative circumposition. Both the directional particle and demonstrative circumposition occur across Kurdish

Zone languages.[19] With the exception of Hewramî, which uses the preposition PÆY (probably an archaism), Kurdish Zone languages tend to use what is referred to as the directional particle (Mahmoudveysi & Bailey, 2013) or the allative clitic (McCarus, 1956). This clitic attaches to a finite verb, and it behaves just like any other preposition. Mahmoudveysi & Bailey (2013) has suggested that it is actually the preposition /wə/ 'to' which has encliticized to the verb, as in (19) and (20).

(19) hat-m=ə kurdistan.
go.PST-1SG=DRCT kurdistan
So. 'I went to Kurdistan.'

(20) kaft**=a** bimārisān
enter.PST=DRCT hospital
Ze. 'they went to the hospital' (Mahmoudveysi & Bailey (2013): 263)

There is no directional particle phonologically speaking in S Zazaki, but if there were such a particle, it would likely not surface after the vowel-final verbal agreement markers. S Zazaki's directional particle has lexicalized with verbs with which it commonly occurs in other Kurdish Zone languages (Karim forthcoming b). Because the complements of prepositions are already in the category, which always take d-form ezafat (Paul's Oblique IIb), I have merged Paul's categories into the prepositional case. As for the verbs which appear to take d-form ezafat only some of the time (e.g /ʃi/ 'to go,' (21) and (22)), comparative evidence points to another construction which is conspicuously missing from S Zazaki, the demonstrative circumposition.

(21) ma=do ʃɪr-e a dɛw-da binɪ
we=FUT go.SUB-PL this.F.SG village-PREP.F.SG.EZ other
Za. 'we will go to that other village' (Berz (1951: 116.26) apud Paul (1998b))

[19]The directional particle and the allative clitic have not been described for S Zazaki in any grammar to date. Based on Paul (1998b)'s data, there seem to be both syntactic and phonological correlates of the demonstrative circumposition (Karim forthcoming b). I follow this understanding of the S Zazaki system here, although it has not been tested against native speaker intuitions.

(22) cf. ʃɯ-e suk-ā gɪrdɪ
go.PRS.IMPF-PL city-F.SG.EZ big
Za. 'they go to the big city.' (Berz (1951: 168.6) apud Paul (1998b))

	DEM(NP)=DEM.POST			
Soranî	ɛm=ə	'this'	ɛm mez=ə	'this table'
Gewrecuî	in=æ	'this'	i daːr=ič=æ	'this tree too'
Zerdeyane	an=æ	'that'	aː saːɬ=æ	'that year'
Hewramî	in=æ	'this'	i hær=æ	'this donkey'
Kurmancî [DIR]	ev	'this'	ev kur	'this boy'
Kurmancî [OBL]	vi(**=i**)	'this'	vi kur**=i**	'this boy'

Table 5.4: The Demonstrative Circumposition in Kurdish Zone Languages

In the Kurdish Zone languages, demonstratives exhibit a bipartite structure consisting of an initial element DEM and a postpositional element =DEM.POST. The demonstrative pronoun is a single bipartite unit DEM=DEM.POST, and the demonstrative adjective circumposes the entire noun phrase DEM NP=DEM.POST (Table 5.4). This construction exists, albeit slightly differently, in all the languages of the region except S Zazaki. When d-forms occur with verbs that don't themselves condition d-forms, they occur in combination with nouns preceded by demonstrative adjectives. In the discussion of the directional particle, my argument assumes that it (likely /=ə/) was lost word-finally. The same would be true of the demonstrative circumposition, except that it may be the case that it was not lost in all environments. In a relatively small class of masculine singular nouns with a stem ending in /-a/, the DEM.POST, followed by an ezafe connection, surfaces as /-y-/ (example (23)). This /-y/ cannot otherwise be explained (i.e. on phonological grounds).

(23) to ma ne bɛla-y-de
you.2SG.OBL us this.OBL.M.SG misfortune-DEM.POST-PREP.M.SG.EZ
gird-i=ra rɛyn-ay
great-PL=from freed-PAST.PL
Za. 'you have freed us from this great misfortune' (Berz (1951: 146.8) apud Paul (1998b))

There must be an oblique context and an adnominal modifier (the ezafe construction) in order for correlates of the DEM.POST to surface. This is a complex set of circumstances, but it has a parallel in S Zazaki's most prevalent contact language within the Kurdish Zone, Kurmancî. S Zazaki and Kurmancî form the Northern Kurdish Zone sprachbund.[20] These languages have converged on similar innovations, including the behavior of the demonstrative circumposition. In the other Kurdish Zone languages (table 5.4), the circumpositional element is either invariant (Soranî) or inflects for case, number, and gender (Hewramî). In Kurmancî, by contrast, the direct form of the circumposition has no phonological form. In contrast, the oblique takes the form /=i/, as shown in Table 5.4. It is exactly this distribution that was the likely predecessor to the S Zazaki system where the Kurmancî-type form is only preserved when followed by an ezafe connection.

To summarize, I reject Paul (1998b)'s division of the Obl. II into a and b. Instead, I categorize S Zazaki verbs into two groups those which—because of a now lost directional particle—condition the presence of a d-form ezafe and those which do not condition one. The d-form conditioning verbs are then in the same category as prepositions. Despite the retention of its post-nominal position, I put the demonstrative circumposition in the same category.

In summary, all examples which featured optional ezafat according to Paul (1998b) have been conditioned by the prepositional case. The prepositional case conditions d-forms because of the historical displacement of the proto-ezafe by what were historically postpositions (as discussed in section 5.4.4). Additionally, there is a precedent to the lexical specification I propose for S Zazaki in Hewramî. According to MacKenzie, "a noun governed by a simple preposition appears in the oblique case only when it forms an integral part of the clause."

[20] For more on the areal relationships between Kurdish languages and their neighbors see Jügel (2014).

5.4.6 An Alternate Theory

The relevant morpho-phonological pieces that represent the S Zazaki ezafat are both small and highly syncretic. Paul (p.c. 2018) has suggested that one possible origin of S Zazaki's ezafe would be borrowing from Aramaic. The Aramaic languages of the region possess a particle /d(a)-/ which acts as a relativizer and linker between a noun and a possessor (Syriac: /ɛwangɛljon d-luqā/ 'the gospel of Luke'). The possibility of Aramaic /d(a)/ as a source for the S Zazaki d-form ezafe is unlikely for four reasons: (1) Although it is possible that d-forms result from borrowing, there is ample evidence that points toward language-internal sources. (2) The d-forms in S Zazaki are multi-morphemic, consisting of the /d-/ marking either indefinite, genitive, or prepositional case and the corresponding simplex ezafic ending marking number, gender, animacy, and the type of attribution (EZ:/ATT or EZ:/GEN). The corresponding particle in Aramaic is indeclinable and proclitic. A theory that claims borrowing from Aramaic must explain the shift from proclitic to enclitic and potentially its position before other ezafic endings. (3) The particle, as it is known from Syriac, was a relativizer, and it was used in the genitival construction. If it were the source of the S Zazaki form, its extension to adjectival environments would have to be explained. If the distribution of the d-forms were indicative of an EZ:/GEN relationship, that would support borrowing as the source because it would overlap with the function of the Aramaic particle. As that is not the case, it is better to assume a source that can explain its complex distribution. (4) The d-forms in S Zazaki only occur in the context of historical displacement. To entertain the idea of Aramaic borrowing, one would have to explain why the d-forms occur in such an idiosyncratic set of environments, on nouns serving as prepositional complements, genitival possessors, and when marked by an indefinite article. Indeed this set of morpho-syntactic categories does not form a natural class.

My claim that the d-forms are the descendants of the Old Iranian far-demonstrative t-forms is supported by the pronominal system of the other Zaza-Gorani languages (see Table 5.5). Hewramî, Shabaki, and Zerdeyane have d-form personal pronouns, which are

conspicuously missing from S Zazaki. I assert that as the d-forms got pulled into S Zazaki's ezafic system, their pronominal function was lost. In this chapter, I have chosen to reject

	'he'		'she'		'they'		
	DIR	OBL	DIR	OBL	DIR	OBL	
S Zazaki	o	ey	a	ay	e	inan	Paul (1998b)
Hewrami	aḏ[21]	aḏi	aḏæ	aḏe	aḏe	aḏeʃan	MacKenzie (1966)
Zerdeyane	ad				adʃan		Mahmoudveysi et al. (2012)
Shabaki	ed				eʃan		MacKenzie (1999a)

Table 5.5: d-forms in Zaza-Gorani Languages

the possibility of Aramaic borrowing because there is no additional evidence that supports borrowing as a source. Furthermore, there is a plausible language-internal source for the d-forms (the Old Iranian t-series pronoun). Furthermore, the d-form ezafat occur in a subset of the S Zazaki nominal paradigm, which is identical to the Hewramî oblique (see section 5.4). This distribution would require an explanation if borrowing were identified as the source.

5.5 Conclusion

S Zazaki's nominal paradigm is unique among the Kurdish Zone languages and even within Western Iranian as a whole. It has developed a large and complex paradigm characterized by fusional morphological patterns with massive syncretism and morphemes recruited from etymologically different materials. This paradigm preserves many distinctions which can be described as inherited (i.e. case, number, and gender). Still, it does so in a way that is innovative and is characterized by massive paradigmatic restructuring. Phonological reduction allowed only one of two formatives (either the case ending or the ezafe) to surface in some forms, merging their functions. The resulting (complex) case ending is sensitive

[21]The sound represented by /ḏ/ here is part of the phenomenon referred to as Zagros-d. Zagros-d is a blanket term for the outcome of postvocalic /d/ in Kurdish Zone languages. The specific phone in Hewramî has been described as a non-syllabic schwa (MacKenzie, 1966) and is probably a lateralized alveolar approximate. For the purposes of this study, its specific articulation is unimportant; it should be understood as the regular outcome of postvocalic /d/ in Hewramî.

to properties of the head noun, and its modifiers (see Samiian, 1994, for a discussion of the ezafe as a case ending with multiple governors, and ch. 3 (this dissertation) for an alternative theory).

As the proto-ezafe construction univerbated with the noun to form the multi-morphemic inflected noun, a variety of factors conditioned whether the proto-ezafe was from *ya̲t* or *ta̲t*. The result was the genesis of the (y-form and d-form) ezafe allomorphs. This is similar to the likely source of definite and indefinite allomorphs of the ezafe in Kurmancî (e.g. the proto-ezafe/definite article *ya̲t* and the evaluative *-aka). In S Zazaki, when the proto-ezafe/definite article was displaced by the indefinite article, head noun (modified by a genitival possessor), or a postposition, the result was that the proto-ezafe/far-demonstrative *ta̲t* took over, ultimately becoming the d-form ezafat. The reason that this had not been noticed before is precisely because of the complexity of the S Zazaki system. Many of the historical conditioning environments for the d-forms now have no phonological form in most, but crucially not all, environments.[22] It is only through the comparative study of the Kurdish Zone languages that it becomes clear that S Zazaki indeed has forms of the directional particle and the demonstrative circumposition found in virtually every other language in the region.

S Zazaki's system has developed from a less fusional system to a more fusional one (though it was probably not canonically agglutinative at any point). What makes the S Zazaki story unique is that it developed this way by absorbing the meaning of a variety of morphs for disparate reasons. The morphs were selected based on a syntactic distribution, INDF, PREP and GEN, which is now arbitrary in the sense that it is not a natural class. The selection of morphs from different etymological sources, one of the ways suppletion may develop diachronically, infused the case endings with meaning from lexemes that surfaced and

[22]The directional particle has no phonological form. It would have conditioned the d-forms in an earlier period. After its phonological reduction, the verb with which it was formerly encliticized took on the status as the conditioner of the d-form. Therefore, the loss of the phonological form is a historical fact and not a proposed underlying form in a synchronic analysis.

from those that were blocked. In short, fusion can occur without the hallmark phonological reduction that makes the boundaries between morphs unparsable. Instead, phonological reduction can motivate the selection of allomorphs that can take on the meaning of the displaced morphemes, creating fusionality.

Chapter 6

Big Picture: the place of this study in (Iranian) Linguistics

Although my primary focus has been narrow, looking only at the nominal systems of Iranian languages, the material treated here actually covers a lot of ground. Still, there have been many obstacles to this analysis deeply rooted in the state of the field. As outlined in chapter 1, there is very little data available for nearly all Iranian languages. Despite the broad typological work of scholars like Habib Borjian and Don Stilo, who have compiled thorough descriptions of particular languages, and brief overviews of a significant number of others, it is impossible to accurately account for all the phenomena extant in Iranian languages or their precise distributions. In this sense, the job of the diachronic linguist is a particularly difficult one.

Furthermore, it is impossible for each linguist who seeks to do this kind of research to conduct fieldwork on all the relevant languages. The unfortunate compromise is to do the best work possible with imperfect data. As more and better data becomes available, the analyses presented in this study will be tested and refined. It has already come to my attention that current research by Richard Larson (Stony Brook University Larson, 2021) is looking at the ways ezafat are used to link a head noun to different types of relative clauses. In some languages certain relative clauses break an ezafe chain, and in others they do not. As I don't directly address the issue of relative clauses, this necessarily has consequences for my diachronic analysis. I show in chapter 3 that an adpositional phrase can break an ezafe chain. This follows from the combinatorics as the primary ezafe is a morpheme that is part of nominal inflection and could not be at the end of an adpositional phrase. It seems that some languages/varieties have developed secondary ezafat to connect multiple

relative clauses. Note that this does not conflict with my analysis; rather, it builds upon my analysis by suggesting a cause for the development of secondary ezafat.

The problem of sporadic documentation is compounded by the nature of publication in Iranian linguistics. The majority of publication happens in special bound volumes. These volumes are often inaccessible, especially to the speakers of most Iranian languages and scholars living in Iran. I sincerely hope that this trend changes to connect Iranian scholars to the research on their languages and to connect scholars in the west, like myself, to the great work that they are doing behind the scholarly divide.

The majority of my research up to this point has had a clear diachronic focus. However, diachrony is but a series of synchronic states that have been stacked chronologically. It is necessary to have knowledge of parallel well-described synchronic states for several languages with an established genetic relationship to do comparative-historical reconstruction. That precise relationship is still very much uncertain despite significant strides in the field thanks to Korn (2016a) and (2019a). The recognition of Central Iranian together with the fact that most other sub-branches of Iranian do not sub-categorize has forever altered Iranian linguistics.

In this study, I have focused on nominal morphology. To develop a strong foundation for my historical hypotheses, I began with two synchronic analyses. Chapter 2 is an overview of the many types of systems extant in New Iranian languages and what is known about their pre-history. It should be clear from this chapter that for everything that is known, there are many issues with the received knowledge. For instance, it is now well-known that the ezafe in New Persian is the reflex of the Old Iranian relative pronoun *ya- or possibly the Old Persian univerbated form *haya-* (Samvelian, 2007a; Haider & Zwanziger, 1984; Kent, 1944, etc.). However, Persian does not constitute the only Iranian language. Furthermore, the assumption of Persian facts for other Iranian languages is likely misleading. Haider & Zwanziger (1984) show that the ezafe-like construction in Parthian *če* likely comes from the Old Iranian relative pronoun *ka-. Likewise, if my hypotheses in chapters 4 and 5 are

correct, the Kurmancî definite article has a hybrid origin: the evaluative suffix from Old Iranian -(a)ka- and the relative pronoun *ya-.[1] The *d*-form ezafat in Zazaki are likely from OIr. *aita-, and the Kurmancî dental plural ezafe likely have yet another etymological source. Altogether, there are at least five etyma for the Iranian ezafat and perhaps more. This is not surprising when examined cross-linguistically. The Albanian nyje particle, which has been described as a type of ezafe (following Franco-Rita & Savoia, 2012), embodies at least three separate etyma Ø from the *ty- demonstrative, t from the *t- demonstrative, and i and e from the *yo- relative- (p.c. with Brian Joseph, the Ohio State University).

One thing that stands out from the juxtaposition of data in chapter 2 is that the ezafe is not one thing. Instead, it is a polyfunctional morpheme with a different distribution and different sensitivities from language to language. In chapter 3, I show that the constellation of phenomena that fall under the broad category of ezafe-marking represents at least eight separate but closely related constructions: the possessive construct, the attributive construct, the definite construct, the attributive anti-construct, the possessive state (genitive case), the attributive floating construct, the possessive floating construct, and possessor cross-indexing. These phenomena are each distinguished by morphology, with regard to the formatives involved, syntax, with regard to how they combine with modifiers, and semantics, with regard to the meaning of the expression.

It should be clear from this study as a whole that none of the aspects of nominal morphology in Iranian operate independently. Ezafe marking interacts with definiteness, number, gender, case, and animacy. Likewise, definiteness interacts with number in languages like colloquial New Persian, Koroshi Balochi, some Southern Kurdish varieties, and many others where the definite morpheme is incompatible with the plural morpheme. An approach to the diachrony of Iranian nominal morphology must look holistically at all nominal subsystems to recognize the patterns consistent with the persistent (following Hopper,

[1] If the Kurmancî free ezafe, not discussed in this study has a demonstrative origin as suggested by the Badînî form *wê*, we may assume an additional etymon, Old Iranian *awa-.

1991) aspects of their etyma.

My analysis of the ezafe phenomena in chapter 3 is the first of its kind. HTLCG and CG more generally have almost exclusively been used for analyses of syntactic phenomena in English. Perhaps the most striking thing about this account is that I naturally concluded that these eight different types of ezafat exist based on which examples required the proposal of a new functor. Each member of this set of eight ezafat is morphologically distinguished in at least one language. I believe this fact lends credibility to my assertion that there is some psycholinguistic plausibility to the HTLCG approach. In other words, it may be the case that these functors are genuinely stored in the lexicon (i.e. memory) as is foundational in CG. From this perspective, I see the potential for pioneering contributions to a new kind of historical syntax that focuses on how aspects of syntax stored in the lexicon tend to change.

Another implication of this analysis that I must mention here is that syntax is fundamentally linked to morphology in that syntactic combinatorics must be found in morphological paradigms. This premise underlines everything that I have covered in this study but most especially chapter 3. Essentially, the foundational assumption of HTLCG (and CG more broadly) is that the syntactic combinatorics, prosodic form, and semantics are all stored in the lexicon, a function of memory. For those morphologists that work in inferential realizational theories, morphology is seen as the organizing principle of the lexicon. It follows from these two assumptions that syntax is a function of morphology. Of course, this assertion was already necessary to understand derivational and inflectional morphology that changes the syntactic category, e.g. causative, applicative, genitive, etc., and to explain complex phenomena like inflectional periphrasis (see Ackerman & Stump, 2004).

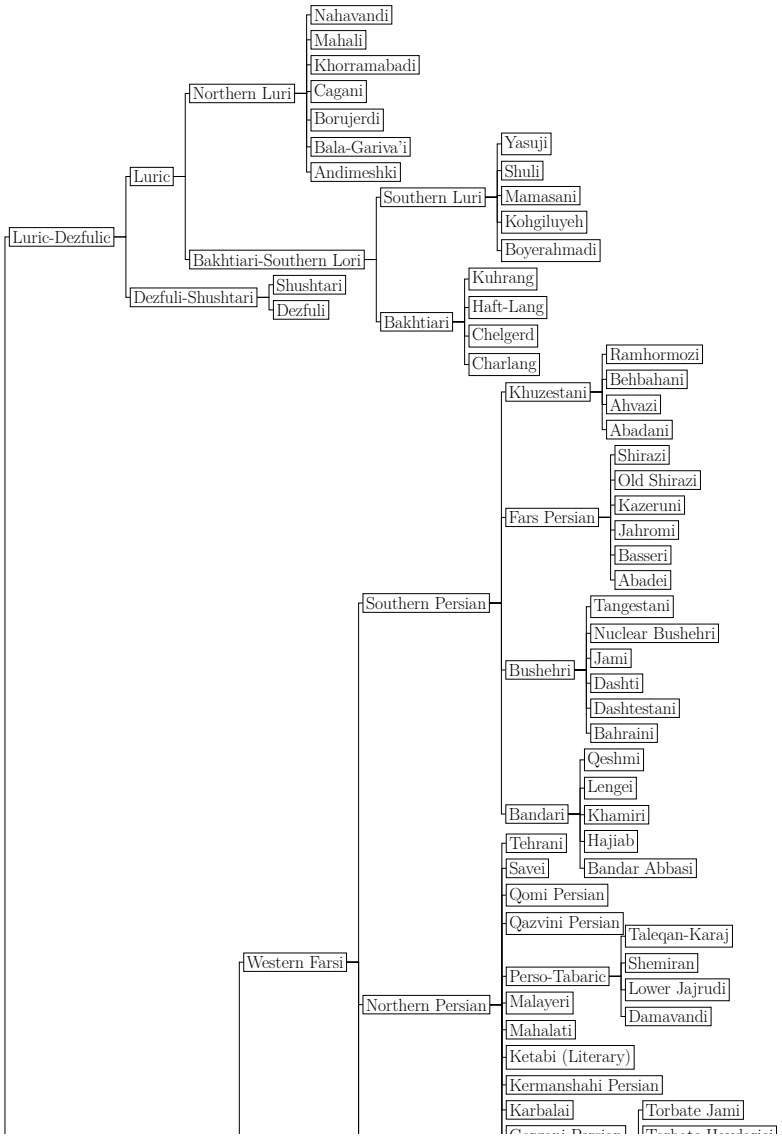
This implication is not something that I devote space to in this study. However, it is something that has come to my attention while conducting this study. There is no doubt that the contents of this dissertation will be the catalyst for exciting future research. In addition to questioning the foundational aspects of morphosyntactic research, it is clear that a new approach to nominal paradigms is necessary based on the Iranian evidence. For

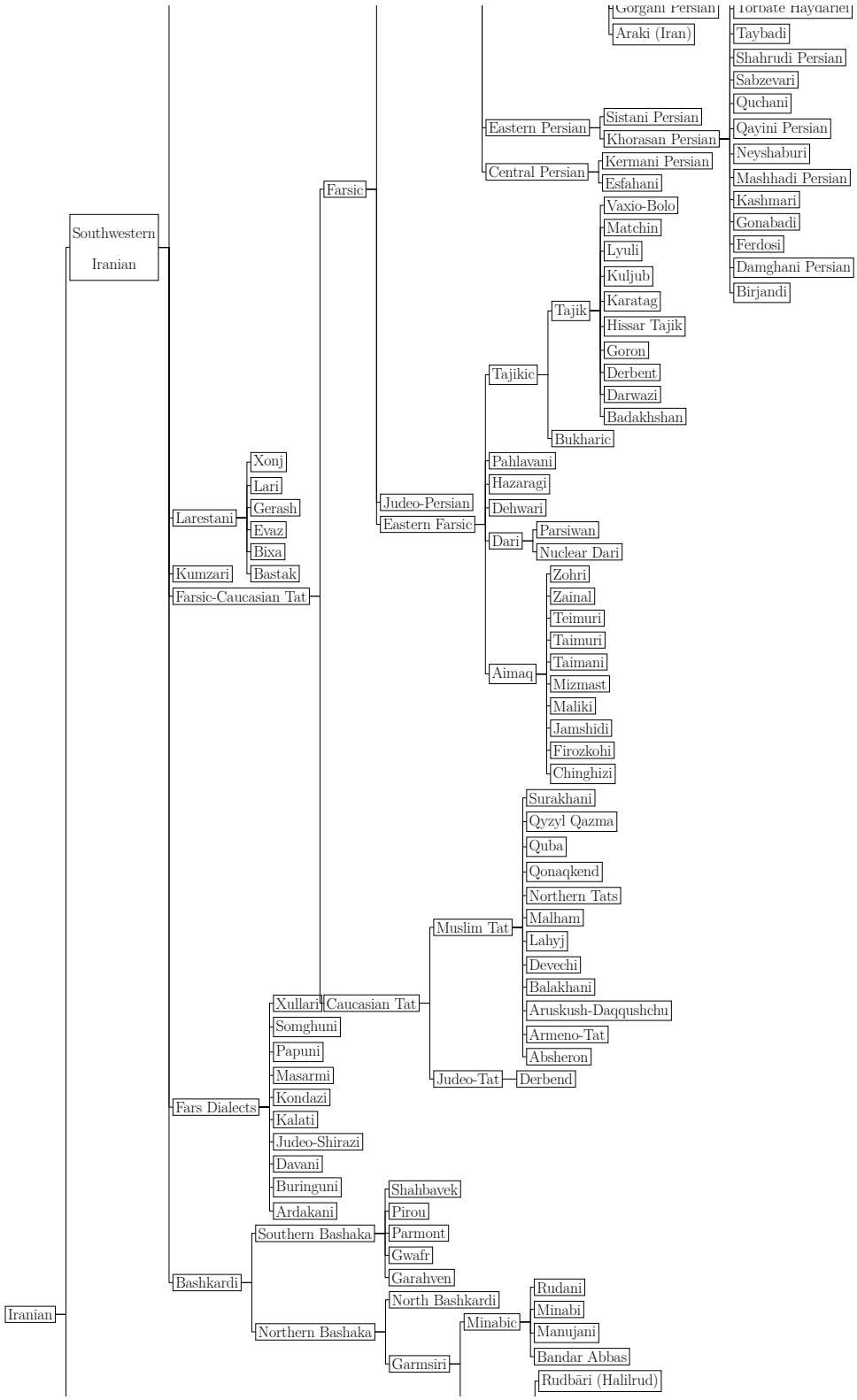
example, in Zazaki (ch. 5), if all combinations of case, number, gender, animacy, definiteness, and attribution are taken into account, there are well over 200 cells in the nominal paradigm. However, there are only ten paradigm cells if you only look at feminine nouns, which don't make distinctions for definiteness, animacy, or modifier type. Furthermore, in chapter 2, I show that many Iranian languages show what might be referred to as stacked systems. These systems can be conceptualized as inflectional systems built upon the bones of others. The new system has hybrid formatives when both the old and new categories apply and simplex formatives when only one applies; see the discussion of Balochi in chapter 2. Speculatively, I highly doubt that child learners of Zazaki are trying to fill 200+ paradigm cells. I foresee that these data may be a reason to reconceptualize the psycholinguistic reality of the Paradigm Cell Filling Problem (following Ackerman et al., 2009).

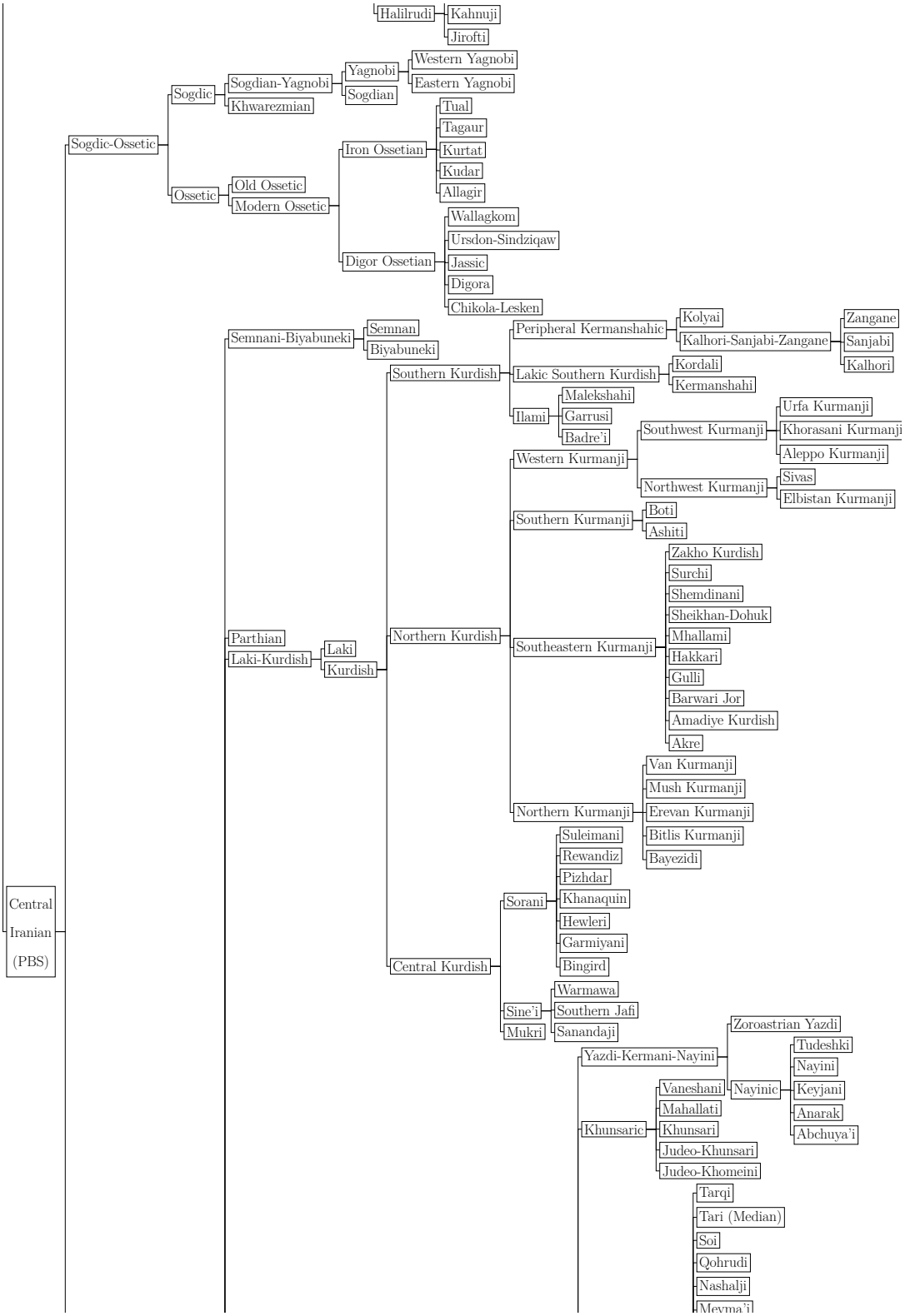
There are no doubt many theoretical and practical implications of this research. I will likely spend the rest of my career exploring these implications. Of particular interest to me is how the same diminutive/evaluative suffix *-(v)ka- became definiteness markers and ezafat, contributed to the development of stacked nominal systems, and preserved gender and case marking in languages that lost these features in most contexts. These changes give specific languages like Zazaki and Balochi an artificial archaic aesthetic. Paradoxically, a series of innovations conspired to make the languages preserve features such as case, number and gender. By exposing these paradoxes, I hope that I have provided the foundation for future scholars to explore these languages and challenge received wisdom based on superficial observations about a given language's complexity.

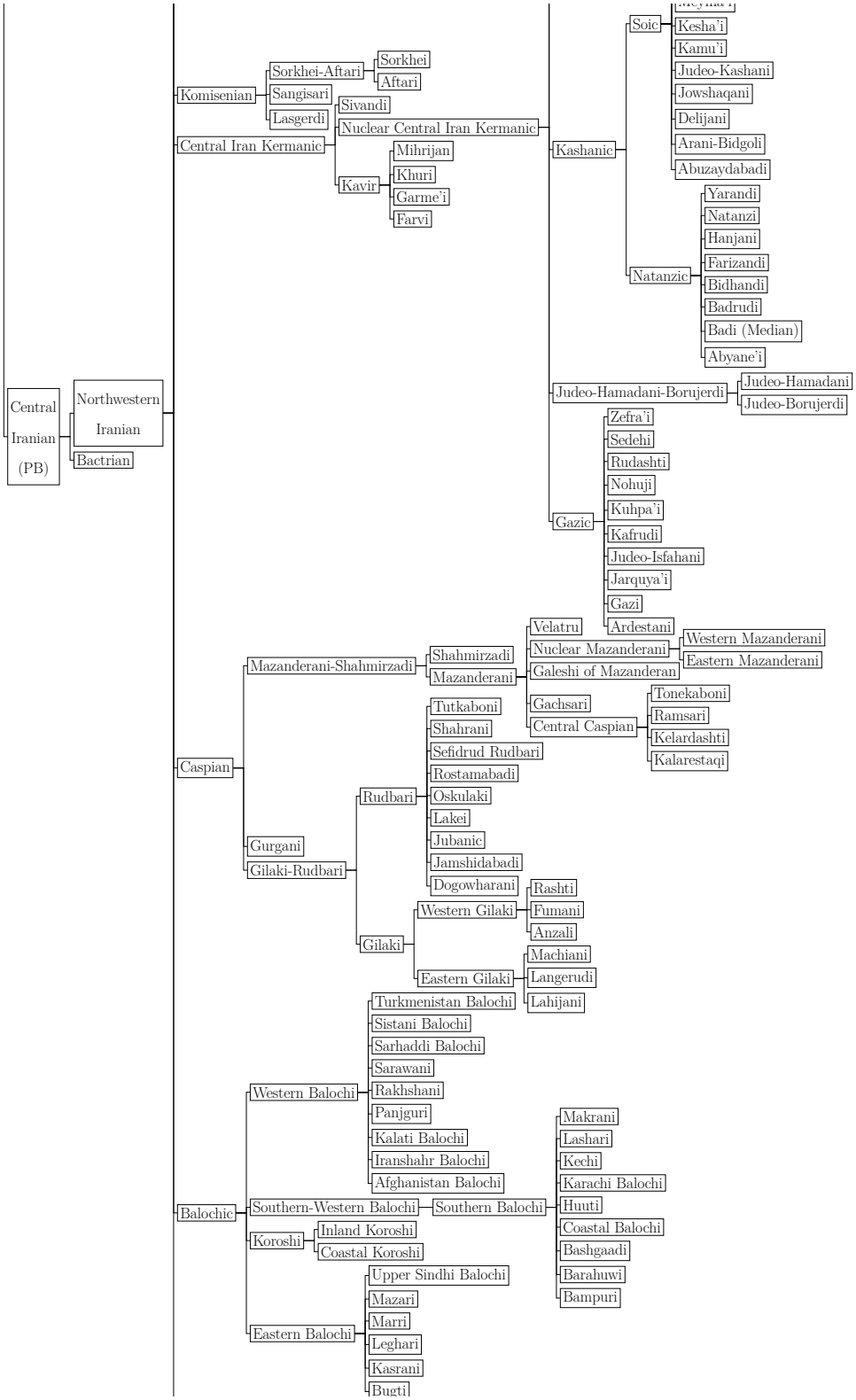
Appendix A

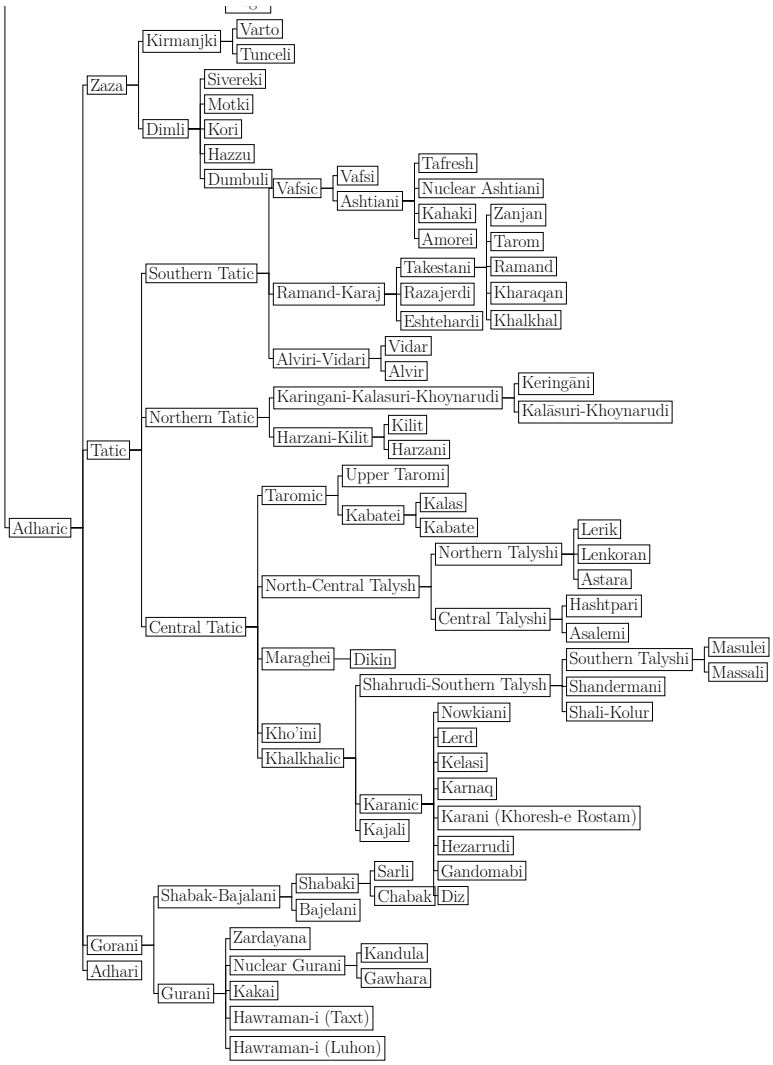
Language Family Tree: Central Iranian and Southwestern Iranian




Iranian
Southwestern Iranian
Larestani
Xonj
Lari
Gerash
Evaz
Bixa
Kumzari
Bastak
Farsic-Caucasian Tat
Fars Dialects
Kondazi
Kalati
Judeo-Shirazi
Davani
Buringuni
Ardakani
Bashkardi
Southern Bashaka
Shahbavek
Pirou
Parmont
Gwafr
Garahven
Northern Bashaka
North Bashkardi
Garmsiri
Minabic
Rudani
Minabi
Manujani
Bandar Abbas
Rudbāri (Halilrud)
Farsic
Judeo-Persian
Eastern Farsic
Tajikic
Tajik
Vaxio-Bolo
Matchin
Lyuli
Kuljub
Karatag
Hissar Tajik
Goron
Derbent
Darwazi
Badakhshan
Bukharic
Pahlavani
Hazaragi
Dehwari
Dari
Parsiwan
Nuclear Dari
Aimaq
Zohri
Zainal
Teimuri
Taimuri
Taimani
Mizmast
Maliki
Jamshidi
Firozkohi
Chinghizi
Caucasian Tat
Xullari
Somghuni
Papuni
Masarmi
Muslim Tat
Surakhani
Qyzyl Qazma
Quba
Qonaqkend
Northern Tats
Malham
Lahyj
Devechi
Balakhani
Aruskush-Daqqushchu
Armeno-Tat
Absheron
Judeo-Tat
Derbend
Eastern Persian
Sistani Persian
Khorasan Persian
Qayini Persian
Neyshaburi
Mashhadi Persian
Kashmari
Gonabadi
Ferdosi
Damghani Persian
Birjandi
Central Persian
Kermani Persian
Esfahani
Gorgani Persian
Araki (Iran)
Torbate Haydarieh
Taybadi
Shahrudi Persian
Sabzevari
Quchani


Central Iranian (PBS)
Sogdic-Ossetic
Sogdic
Sogdian-Yagnobi
Yagnobi
Western Yagnobi
Eastern Yagnobi
Sogdian
Khwarezmian
Ossetic
Old Ossetic
Modern Ossetic
Iron Ossetian
Tual
Tagaur
Kurtat
Kudar
Allagir
Digor Ossetian
Wallagkom
Ursdon-Sindziqaw
Jassic
Digora
Chikola-Lesken
Parthian
Laki-Kurdish
Laki
Kurdish
Halilrudi
Kahnuji
Jirofti
Semnani-Biyabuneki
Semnan
Biyabuneki
Northern Kurdish
Southern Kurdish
Peripheral Kermanshahic
Kolyai
Kalhori-Sanjabi-Zangane
Zangane
Sanjabi
Kalhori
Lakic Southern Kurdish
Kordali
Kermanshahi
Ilami
Malekshahi
Garrusi
Badre'i
Central Kurdish
Western Kurmanji
Southwest Kurmanji
Urfa Kurmanji
Khorasani Kurmanji
Aleppo Kurmanji
Northwest Kurmanji
Sivas
Elbistan Kurmanji
Southern Kurmanji
Boti
Ashiti
Southeastern Kurmanji
Zakho Kurdish
Surchi
Shemdinani
Sheikhan-Dohuk
Mhallami
Hakkari
Gulli
Barwari Jor
Amadiye Kurdish
Akre
Northern Kurmanji
Van Kurmanji
Mush Kurmanji
Erevan Kurmanji
Bitlis Kurmanji
Bayezidi
Sorani
Suleimani
Rewandiz
Pizhdar
Khanaquin
Hewleri
Garmiyani
Bingird
Sine'i
Warmawa
Southern Jafi
Sanandaji
Mukri
Yazdi-Kermani-Nayini
Zoroastrian Yazdi
Nayinic
Tudeshki
Nayini
Keyjani
Anarak
Abchuya'i
Khunsaric
Vaneshani
Mahallati
Khunsari
Judeo-Khunsari
Judeo-Khomeini
Tarqi
Tari (Median)
Soi
Qohrudi
Nashalji
Meyma'i


Central Iranian (PB)
Northwestern Iranian
Bactrian
Komisenian
Sorkhei-Aftari
Sorkhei
Aftari
Sangisari
Lasgerdi
Central Iran Kermanic
Sivandi
Nuclear Central Iran Kermanic
Kavir
Mihrijan
Khuri
Garme'i
Farvi
Kashanic
Soic
Kesha'i
Kamu'i
Judeo-Kashani
Jowshaqani
Delijani
Arani-Bidgoli
Abuzaydabadi
Natanzic
Yarandi
Natanzi
Hanjani
Farizandi
Bidhandi
Badrudi
Badi (Median)
Abyane'i
Judeo-Hamadani-Borujerdi
Judeo-Hamadani
Judeo-Borujerdi
Gazic
Zefra'i
Sedehi
Rudashti
Nohuji
Kuhpa'i
Kafrudi
Judeo-Isfahani
Jarquya'i
Gazi
Ardestani
Caspian
Mazanderani-Shahmirzadi
Shahmirzadi
Mazanderani
Velatru
Nuclear Mazanderani
Western Mazanderani
Eastern Mazanderani
Galeshi of Mazanderan
Gachsari
Central Caspian
Tonekaboni
Ramsari
Kelardashti
Kalarestaqi
Tutkaboni
Shahrani
Sefidrud Rudbari
Rostamabadi
Gurgani
Gilaki-Rudbari
Rudbari
Oskulaki
Lakei
Jubanic
Jamshidabadi
Dogowharani
Gilaki
Western Gilaki
Rashti
Fumani
Anzali
Eastern Gilaki
Machiani
Langerudi
Lahijani
Balochic
Western Balochi
Turkmenistan Balochi
Sistani Balochi
Sarhaddi Balochi
Sarawani
Rakhshani
Panjguri
Kalati Balochi
Iranshahr Balochi
Afghanistan Balochi
Southern-Western Balochi
Southern Balochi
Makrani
Lashari
Kechi
Karachi Balochi
Huuti
Coastal Balochi
Bashgaadi
Barahuwi
Bampuri
Koroshi
Inland Koroshi
Coastal Koroshi
Eastern Balochi
Upper Sindhi Balochi
Mazari
Marri
Leghari
Kasrani
Bugti


Adharic
Zaza
Kirmanjki
Varto
Tunceli
Dimli
Sivereki
Motki
Kori
Hazzu
Dumbuli
Vafsic
Vafsi
Ashtiani
Tafresh
Nuclear Ashtiani
Kahaki
Zanjan
Amorei
Tarom
Alviri-Vidari
Vidar
Alvir
Tatic
Southern Tatic
Ramand-Karaj
Takestani
Ramand
Razajerdi
Kharaqan
Eshtehardi
Khalkhal
Northern Tatic
Karingani-Kalasuri-Khoynarudi
Keringāni
Kalāsuri-Khoynarudi
Harzani-Kilit
Kilit
Harzani
Central Tatic
Taromic
Upper Taromi
Kabatei
Kalas
Kabate
North-Central Talysh
Northern Talyshi
Lerik
Lenkoran
Astara
Central Talyshi
Hashtpari
Asalemi
Maraghei
Dikin
Southern Talyshi
Masulei
Massali
Shahrudi-Southern Talysh
Shandermani
Shali-Kolur
Kho'ini
Khalkhalic
Karanic
Nowkiani
Lerd
Kelasi
Karnaq
Karani (Khoresh-e Rostam)
Kajali
Hezarrudi
Gandomabi
Diz
Gorani
Shabak-Bajalani
Shabaki
Sarli
Chabak
Bajelani
Adhari
Gurani
Zardayana
Nuclear Gurani
Kandula
Gawhara
Kakai
Hawraman-i (Taxt)
Hawraman-i (Luhon)

Appendix B

List of Abbreviations

Abbreviation/ Symbol	Description
languages	
So.	Soranî
Ku.	Kurmancî
NP.	Standard New Persian
NP. (col.)	Colloquial New Persian
He.	Hewramî
ŞT.	Şirvan Tat
SZ.	Southern Zazaki
CZ.	Central Zazaki
Gi.	Gilaki
CT.	Chali Southern Tati
TT.	Takestanî Tati
TB.	Turkmen Balochi
RB.	Rakhshani Balochi
OP.	Old Persian
Av.	Avestan
MP.	Middler Persian
OIr.	Old Iranian

Morphological Glossing	
M	masculine
F	feminine
SG	singular
PL	plural
ANIM	animate
INAN	inanimate
GEN	genitive case/ possessive state
NOM	nominative case
ACC	accusative case
DIR	direct case
OBL	oblique case
INS	instrumental case
EZ	ezafe: modification marking on the head noun (≠ Leipzig)
EZ:/ATT	ad-attributive ezafe: an ezafe anticipating an attributive adjective (≠ Leipzig)
EZ:/GEN	ad-genitival ezafe: an ezafe anticipating a noun in the genitive case (≠ Leipzig)
REZ	reverse ezafe: a category uniting GEN and ATT (≠ Leipzig)
ATT	attributive marker: marking an adjective as attributive
COP	copula
1	first person
2	second person
3	third person
PST	past tense
PRS	present tense

A	agent index
O	object index
R	oblique argument (applied object) index
-	morpheme boundary
=	clitic boundary
DEM	demonstrative
CIRC	circumpositional element (≠ Leipzig)
PV	preverb
LV	light verb
DEF	definite
INDF	indefinite
NEG	negation
DO	direct object
IFPV	imperfective aspect
SUBJ	subjunctive mood

Syntactic categories and CG operators	
NP	Noun Phrase
AdjP	Adjectival Phrase
PP	Prepositional/Postpostional Phrase
XP	X Phrase: a phrase level unit that makes no distinction between NP, AdjP, or PP
VP	Verb Phrase (a stand in for a variety of types $NP\backslash S$, $NP\backslash(NP\backslash S)$, etc.)
/	A functor looking for an argument on its right
\	A functor looking for an argument on its left
↑	A functor looking for an argument regardless of position
E	Elimination: the step of resolving a syntactic functor

I	Introduction: creating a functor by abstracting on a variable
Semantic operators, categories, and variables	
S	subject
A	agent
O	object
x, y, z	variables corresponding to entities $\langle e \rangle$
P, Q, R	variables corresponding to properties $\langle e, t \rangle$
$\mathcal{Q}$	a variable corresponding to a quantifier: in these example $^{\cup}$, or ι
$\mathcal{R}$	a relationship of one entity $\langle e \rangle$ over another
$\exists$	at least one
$^{\cup}$	takes an entity $\langle e \rangle$ and returns a property denoted by that entity
$^{\cap}$	takes a property $\langle e, t \rangle$ and returns a kind $\langle e \rangle$ equivalent to a subset but not a proper subset of the set dented by the property
ι	takes a property $\langle e, t \rangle$ and returns an entity $\langle e \rangle$ equivalent to a particular member of the set dented by the property
$LINK$	takes a property $\langle e, t \rangle$ and returns a specific subset $\langle e \rangle$ of the set denoted by the property
λ	binds a variable in a function
λ-conv.	Lambda-conversion
Scripts	
Hawar: a	IPA: aː ; Orientalist: ā; Zazaki = Hawar
Hawar: (æ)	IPA: æ; Orientalist: a; Zazaki = Hawar
Hawar: b	IPA: b ; Orientalist: b; Zazaki = Hawar
Hawar: c	IPA: d͡ʒ; Orientalist: ǰ; Zazaki = Hawar
Hawar: ç	IPA: t͡ʃ; Orientalist: ç; Zazaki = Hawar
Hawar: d	IPA: d; Orientalist: d; Zazaki = Hawar
Hawar: (ð̩)	IPA: [lateralized alveolar approximate]; Orientalist: ð̩

Hawar: (ḍ)	IPA: ɗ; Orientalist: ḍ; Zazaki = Hawar
Hawar: e	IPA: ə; Orientalist: a; Zazaki = Hawar
Hawar: ê	IPA: eː; Orientalist: ē/e; Zazaki = Hawar
Hawar: f	IPA: f; Orientalist: f; Zazaki = Hawar
Hawar: g	IPA: g; Orientalist: g; Zazaki = Hawar
Hawar: ẍ (ɣ)	IPA: ɣ; Orientalist: gh; Zazaki = Hawar
Hawar: h	IPA: h; Orientalist: h; Zazaki = Hawar
Hawar: ḧ (ħ)	IPA: ħ; Orientalist: ḥ; Zazaki = Hawar
Hawar: i	IPA: i̯; Orientalist: i/ı; Zazaki: ı
Hawar: î	IPA: iː; Orientalist: ī; Zazaki: i
Hawar: j	IPA: ʒ; Orientalist: ž; Zazaki = Hawar
Hawar: k	IPA: k/k^{h}; Orientalist: k; Zazaki = Hawar
Hawar: l	IPA: l; Orientalist: l; Zazaki = Hawar
Hawar: ɬ	IPA: ɬ; Orientalist: ɬ; Zazaki = Hawar
Hawar: m	IPA: m; Orientalist: m; Zazaki = Hawar
Hawar: n	IPA: n; Orientalist: n; Zazaki = Hawar
Hawar: o	IPA: oː; Orientalist: o/ō; Zazaki = Hawar
Hawar: ‘	IPA: ʕ; Orientalist: ‘; Zazaki = Hawar
Hawar: p	IPA: p/p^{h}; Orientalist: p; Zazaki = Hawar
Hawar: q	IPA: q; Orientalist: q; Zazaki = Hawar
Hawar: r	IPA: ɾ; Orientalist: r; Zazaki = Hawar
Hawar: ř	IPA: r; Orientalist: ř; Zazaki = Hawar
Hawar: s	IPA: s; Orientalist: s; Zazaki = Hawar
Hawar: ş	IPA: ʃ; Orientalist: š; Zazaki = Hawar
Hawar: (ṣ)	IPA: ʂ; Orientalist: ṣ; Zazaki = Hawar
Hawar: t	IPA: t/y^{h}; Orientalist: t; Zazaki = Hawar
Hawar: (ṭ)	IPA: ʈ; Orientalist: ṭ; Zazaki = Hawar

Hawar: u	IPA: ʊ; Orientalist: u; Zazaki = Hawar
Hawar: û	IPA: uː; Orientalist: ū; Zazaki = Hawar
Hawar: v	IPA: v; Orientalist: v; Zazaki = Hawar
Hawar: w	IPA: w; Orientalist: w; Zazaki = Hawar
Hawar: x	IPA: x; Orientalist: x; Zazaki = Hawar
Hawar: y	IPA: j; Orientalist: y; Zazaki = Hawar
Hawar: z	IPA: z; Orientalist: z; Zazaki = Hawar

References

Abaev, V.I. 1964. O Dialektax Osetinskogo Yazlyk. In *Indo-iranica: mélanges présentés à georg morgenstierne, à l'occasion de son soixante-dixième anniversaire*, 1–7. Wiesbaden: Otto Harrassowitz.

Abdollahnejad, Elias & Dennis Ryan Storoshenko. 2020. Syntactic and semantic constraints on pronoun and anaphor resolution in Persian. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 2, 15–28. Amsterdam; Philadelphia: John Benjamins.

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford: Oxford University Press.

Ackerman, Farrell & Gregory T Stump. 2004. Paradigms and periphrastic expression: A Study in Realization-based Lexicalism. In Louisa Sadler & Andrew Spencer (eds.), *Projecting morphology*, 111–157. Stanford, CA: CSLI Publications.

Ajdukiewicz, Kazimierz. 1935. Die syntaktische Konnexität. In Storrs McCall (ed.), *Polish logic 1920-1939*, 207–231. Oxford: Oxford University Press. [Translated from Studia Philosophica, Vol. 1: 1-27].

Amān Allāhī, Sikandar & Wheeler McIntosh Thackston. 1986. *Tales from Luristan (Matalyâ Lurissu): tales, fables, and folk poetry from the Lur of Bâlâ-Garîva (Harvard Iranian Series 4)*. Cambridge, Mass.: Harvard University Press.

Angali, Koorosh. 2004. *The Angali Dialect*: University of California at Berkeley Phd.

Anonby, Christina van der Wal. 2019. Kumzari. In Geoffrey Khan & Geoffrey Haig (eds.), *The languages and linguistics of western asia: An areal perspective*, chap. 4.7, 625–658. Berlin, Boston: De Gruyter.

Anonby, Erik, Amos Hayes & Robert Oikle. 2020a. A multi-dimensional approach to classification of Iran's languages. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 3, 29–56. Amsterdam; Philadelphia: John Benjamins.

Anonby, Erik, Masoud Mohammadirad & Jaffer Sheyholislami. 2019. Kordistan Province in the Atlas of the Languages of Iran: Research process, language distribution, and language classification. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 9–38. Bamberg: University of Bamberg Press.

Anonby, Erik & Hassan Mohebbi Bahmani. 2016. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in south-west Iran. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 13–36. Association pour l'avancement des études Iraniennes.

Anonby, Erik, Mortaza Taheri-Ardali, et al. 2020b. Atlas of the Languages of Iran (ALI). http://iranatlas.net/.

Anonby, Erik & Mortaza Taheri-Ardali. 2019. Bakhtiari. In Geoffrey Khan & Geoffrey Haig (eds.), *The languages and linguistics of western asia: An areal perspective*, chap. 4.3, 445–480. Berlin, Boston: De Gruyter.

Arkadiev, Peter. 2007. Poor (two-term) case systems: limits of neutralization. In Andrei Malchukov & Andrew Spencer (eds.), *The oxford handbook of case*, chap. 47, 686–699. Oxford: Oxford University Press.

Axenov, Serge. 2006. *The Balochi language of Turkmenistan a corpus-based grammatical description*. Uppsala: Uppsala University dissertation.

Aygen, Gülşat. 2007. *Kurmanji Kurdish Grammar*. Munchen: Lincom Europa.

Aygen, Gülşat. 2010. *Zazaki/Kirmanckî Kurdish*. Lincom Europa.

Azami, Cheragh Ali. & Gernot L Windfuhr. 1972. *A dictionary of Sangesari with a grammatical outline*. Tehran: Franklin Book Programs.

Baghbidi, Hasan Rezai. 2011. New light on the Middle Persian-Chinese bilingual inscription from Xi'an. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 105–116. Wiesbaden: Reichert.

Bailey, Denise. 2016. *A grammar of Gawraju Gurani*: Georg-August-Universität Göttingen dissertation.

Bailey, T Grahame. 1924. Book Review: The Balochi Language: A Grammar and Manual. *Bulletin of the School of Oriental Studies, University of London* 3(2). 383–384.

Bakaev, C. X. 1962. *Govor Kurdov Turkmenii*. Moskva: Izdatelitstvo.

Bakaev, C. X. 1965. *Yazlyk Azerbaidjanskix Kurdov*. Moskva: Izdatelitstvo.

Baksî, M. 1991. *Gundikî Dono*. İstanbul: Gökyüzü Sanat Ürünleri.

Bar-Hillel, Yehoshua. 1953. A quasi-arithmetic notation for syntactic descriptions. *Language* 29. 47–58.

Baranzehi, Adam Nader. 2003. The Sarawani Dialect of Balochi and Persian Influence on It. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbors: Ethnic and linguistic contact in balochistan in historical and modern times*, 75–112. Wiesbaden: Reichert.

Barjasteh-Delforooz, Behrooz. 2003. The Structure of the Present and Past Stems in Balochi Compared to Old, Middle and New Persian. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbors: Ethnic and linguistic contact in balochistan in historical and modern times*, 19–32. Wiesbaden: Reichert.

Barjasteh-Delforooz, Behrooz. 2007. *Discourse Features in Balochi of Sistan (Oral Narratives)*. Uppsala: Uppsala universitet. doi:10.1017/CBO9781107415324.004. http://uu.diva-portal.org/smash/record.jsf?pid=diva2:169774.

Barjasteh-Delforooz, Behrooz. 2008. A Sociolinguistic Survey among the Jadgal in Iranian Balochistan. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and others: Linguistic, historical and socio-political perspectives on pluralism in balochistan*, 23–44. Wiesbaden: Reichert.

Barker, Muhammad Abd-al-Rahman & Aquil Khan Mengal. 2014. *A course in Baluchi*. Toronto: McGill University Institute of Islamic Studies.

Barry, Daniel. 2019. Pharyngeals in Kurmanji Kurdish: A reanalysis of their source and status. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 39–72. Bamberg: University of Bamberg Press.

Bartholomae, C. 1895. Awestasprache und Altpersisch. In Wilhelm Geiger (ed.), *Grundriss der iranischen philologie*, chap. ii, 152–246. Strassburg: Trübner.

Bashir, Elena. 2008a. Some Transitional Features of Eastern Balochi: An Areal and Diachronic Perspective. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and others: Linguistic, historical and socio-political perspectives on pluralism in balochistan*, 45–82. Wiesbaden: Reichert.

Bashir, Elena. 2008b. Some Transitional Features of Eastern Balochi: An Areal and Diachronic Perspective. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and*

others: Linguistic, historical and socio-political perspectives on pluralism in balochistan, 46–82. Wiesbaden: Reichert.

Bedirxan, Celadet Ali & Roger. Lescot. 1986. *Kurdische Grammatik : Kurmancî-Dialekt*. Bonn: Kurdisches Institut : Verlag Kultur und Wissenschaft.

Belelli, Sara. 2016. *A study on language and folklore in the city of Harsin (in Kermānshāh province of West Iran): Sketch grammar with texts and lexicon*: Naples: L Orientale University Phd.

Belleli, Sara. 2016. *A study on language and folklore in the city of Harsin (Kermānshāh Province, West Iran). Sketch grammar with texts and lexicon*: University of Naples Phd.

Belleli, Sara. 2019. Towards a dialectology of Southern Kurdish: Where to begin? In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 73–92. Bamberg: University of Bamberg Press.

Belyaev, Oleg & Arseniy Vydrin. 2011. Participle-Converbs in Iron Ossetic: Syntactic and Semantic Properties. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 117–134. Wiesbaden: Reichert.

Benveniste, Emile. 1964. La racine yat- en indo-iranien. In *Indo-iranica: mélanges présentés à georg morgenstierne, à l'occasion de son soixante-dixième anniversaire*, 8–12. Wiesbaden: Otto Harrassowitz.

Berz, Koyo & Malmîsanij. 1951. *Na xumxum a ...* Uppsala: Jina Nû Yayınları.

Bielmeier, R. 1993. Das Alanische bei Tzetzes. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 1–28. Leuven: Peeters.

Blau, Joyce. 1976. *Le Kurde de 'Amadiya et de Djabal Sindjar : analyse linguistique, textes folkloriques, glossaires.* Paris: Klincksieck.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382. doi:10.1017/S0022226714000243.

Borjian, Habib. 2004. Mazandaran: Language and People (The State of Research). *Iran and the Caucasus* 8(2). 289–328.

Borjian, Habib. 2005. Talish and the Talishis (The State of Research). *Iran and The Caucasus* 9(1). 43–72.

Borjian, Habib. 2008a. JARQUYA ii. The Dialect. https://iranicaonline.org/articles/jarquya-ii-the-dialect-1.

Borjian, Habib. 2008b. The Komisenian Dialect of Aftar. *Archiv Orientální* 76(3). 379–416.

Borjian, Habib. 2010. KALĀRESTĀQ ii. The Dialect. https://iranicaonline.org/articles/kalarestaq-2-restored.

Borjian, Habib. 2011. The Dialect of Kuhpâya. *Studia Iranica* 40(1). 7–68.

Borjian, Habib. 2012a. Judeo-Kashani: A Central Iranian Plateau Dialect. *Journal of the American Oriental Society* 132(1). 1–22.

Borjian, Habib. 2012b. KASHAN ix. THE MEDIAN DIALECTS OF KASHAN. http://www.iranicaonline.org/articles/kashan-ix-the-median-dialects-of-kashan.

Borjian, Habib. 2012c. *Meyma'i: A Central Iranian Plateau Dialect*. Lincom Europa.

Borjian, Habib. 2012d. Shemiran Dialect Group. In *Enat'mec'nierebis sakit'xebi/issues of linguistics 2011 (jemshid giunashvili festschrift)*, 79–90. Tbilisi, Georgia: Tbilisi State University.

Borjian, Habib. 2012e. The Dialects of Velātru and Gachsar: The Upper Karaj Valley in the Caspian-Persian Transition Zone. *Journal of the Royal Asiatic Society* 22(2). 227–263.

Borjian, Habib. 2013a. FARVI DIALECT. https://iranicaonline.org/articles/farvi-dialect.

Borjian, Habib. 2013b. Perso-Tabaric Dialects in the Language Transition Zone Bordering Mazandaran. *Studia Iranica* 42(2). 195–225.

Borjian, Habib. 2013c. The Caspian Dialect of Kujūr in Central Alborz. *Journal of the British Institute of Persian Studies* 51. 235–246.

Borjian, Habib. 2013d. *The Raji Dialect of Jowshaqan*. Lincom Europa.

Borjian, Habib. 2013e. The Tabaroid Dialects of South-Central Alborz. *Acta Orientalia* 66(4). 427–441.

Borjian, Habib. 2013f. Yushij: A Dialect of Central Alborz. *Persica* 24. 127–153.

Borjian, Habib. 2014. Koroshi Balochi. *Acta Orientalia* 67(4). 453–465.

Borjian, Habib. 2015. Judeo-Iranian Languages. In Lily Khan & Aaron D. Rubin (eds.), *Handbook of jewish languages*, chap. 10, 234–297. Leiden; Boston: Brill.

Borjian, Habib. 2016. NEŠALJ ii. The Dialect. https://iranicaonline.org/articles/neshalj-02-dialect.

Borjian, Habib. 2017a. KERMAN xvi. LANGUAGES. https://iranicaonline.org/articles/kerman-16-languages.

Borjian, Habib. 2017b. KERMANSHAH vii. LANGUAGES AND DIALECTS. https://iranicaonline.org/articles/kermanshah-07-languages.

Borjian, Habib. 2017c. KEŠAᵓI DIALECT. http://www.iranicaonline.org/articles/keshai-dialect.

Borjian, Habib. 2018a. The Caspian Dialect of Māhā. *Studia Iranica* 47(2). 257–292.

Borjian, Habib. 2018b. The Dialect of Khur. In M. De Chiara, Adriano Valerio Rossi & D. Septfonds (eds.), *Mélanges d'ethnographie et de dialectologie irano-aryennes à la mémoire de charles-martin kieffer*, 77–98. Peeters.

Borjian, Habib. 2019a. Judeo-Isfahani: The Iranian Language of the Jews of Isfahan. *Journal of Jewish Languages* 7. 121–189.

Borjian, Habib. 2019b. Mazandarani: A Typological Survey. In A. Korangy & B. Mahmoudi-Bakhtiari (eds.), *Typology of iranian languages*, 79–101. Berlin: De Gruyter Mouton.

Borjian, Habib. 2019c. The Caspian Language of Šahmirzād. *Journal of the American Oriental Society* 139(2). 361–380.

Borjian, Habib. 2019d. The Language of the Kharg Island. *Journal of the Royal Asiatic Society, Series 3* 29(4). 659–682.

Borjian, Habib. 2019e. The Mazandarani Dialect of Kalijān Rostāq. *Journal of Iranian Studies* 52(3-4). 551–573.

Borjian, Habib & Maryam Borjian. 2008. The Last Galesh herdsman: Ethno-linguistic Materials from South Caspian Rainforests. *Iranian Studies* 41(3). 634–644.

Boyce, Mary. 1964. The use of relative particles in Western Middle Iranian. In *Indo-iranica; mélanges présentés à georg morgenstierne, à l'occasion de son soixante-dixième anniversaire*, 28–47. Wiesbaden: Otto Harrassowitz.

Brasoveanu, Adrian. 2013. The grammar of quantification and the fine structure of interpretation contexts. *Synthese* 190(15). 3001–3051.

Buddruss, Georg. 1988. *Aus dem Leben eines jungen Balutschen, von ihm selbst erzählt.* Stuttgart: Franz Steiner Verlag.

Bulut, Christane. 2019. Iraqi-Turkic. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspective*, 354–384. Berlin: De Gruyter.

Burzio, Luigi & Niki Tantalou. 2007. Modern Greek accent and faithfulness constraints in OT. *Lingua* 117(6). 1080–1124.

Cantera, Alberto. 2007. The Accusative of the i- and u- stems with Presuffixal Full or Large Grade in Avestan. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 9–20. Wiesbaden: Harrassowitz.

Cantera, Alberto. 2017. The phonology of Iranian. In Jared S Klein, Brian D Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 33, 481–503. Berlin: De Gruyter Mouton.

Cereti, Carlo G & Ludwig Paul. 1999. *Iranica diversa (Serie orientale Roma 84)*. Roma: Istituto italiano per l'Africa e l'Oriente.

Chafe, Wallace L. 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.

Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb*. Leiden: Brill.

Chierchia, Gennaro. 1985. Formal Semantics and the Grammar of Predication. *Linguistic inquiry* 6(3). 417–443.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A Jacobs & Peter S Rosenbaum (eds.), *Reading in english transformational grammar*, 184–221. Waltham: Ginn.

Christensen, Arthur. 1930. *Dialecte guilāki de Recht, dialectes de Fārizānd, de Yaran et de Natanz. (Historisk-filologiske meddelelser, 17.2.)*. København: B. Lunos.

Christensen, Arthur. 1935. *Contributions à la dialectologie iranienne II: Dialects de la région de Semnan: sourkheī, lāsguerdī, sängesärī et chämerzadī. (Historisk-filologiske meddelelser, 21.3.)*. København: Adr. Fred. Høst & Søn.

Christensen, Arthur & Åge Meyer Benedictsen. 1921. *Les dialectes d'Awroman et de Pawa*. Kobenhavn: B. Lunos.

Christensen, Arthur Emanuel. 1972. *Sîvändî, Yäzdî und Sôî*. Nendeln, Liechtenstein: Kraus Reprint.

Chyet, Michael L. 2003. *Kurdish-English dictionary = Ferhenga Kurmancî-Inglîzî*. New Haven: Yale University Press.

Ciancaglini, Claudia A. 2011. The formation of the periphrastic verbs in Persian and neighbouring languages. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 3–22. Wiesbaden: Reichert.

Clackson, James. 2007. *Indo-European Linguistics. An Introduction*. Cambridge: Cambridge University Press.

Comrie, Bernard. 2016. Ergativity in Iranian languages: a typological perspective. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 37–54. Association pour l'avancement des études Iraniennes.

Condoravdi, Cleo & Ashwini Deo. 2014. Aspect shifts in Indo-Aryan and trajectories of semantic change. In Chiara Gianollo, Agnes Jäger & Doris Penka (eds.), *Language change at the syntax-semantics interface*, chap. 9, 261–291. Berlin/Munich/Boston: Walter De Gruyter. doi:10.1515/9783110352306.261.

Creissels, Denis. 2008. Direct and indirect explanations of typological regularities: the case of alignment variations. Revised and extended and version of a paper presented at SLE 2006, Bremen, August 30 -September 02 2006. *Folia Linguistica* 42(1-2).

Dabir-Moghaddam, Mohammad. 2008. On Agent Clitics in Balochi in Comparison with Other Iranian Languages. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and others: Linguistic, historical and socio-political perspectives on pluralism in balochistan*, 83–100. Wiesbaden: Reichert.

Dabir-Moghaddam, Mohammad. 2018. Non-Canonical Subject Construction in Endangered Iranian Languages: Further Investigation into the Debates on the Genesis of Ergativity. In Saloumeh Gholami (ed.), *Endangered iranian languages*, 9–40. Wiesbaden: Reichert.

David, Anne Boyle. 2014. *Descriptive grammar of Pashto and its dialects*. Boston: De Gruyter. doi:10.1515/9781614512318.

Degener, A. 1993. Zur Syntax des Khotanischen. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 45–56. Leuven: Peeters.

Delforooz, Behrooz Barjasteh & Stephen H. Levinsohn. 2014. The third person singular pronominal clitic in Balochi of Sistan: A progress report. *Studia Iranica* 43(2). 203–220.

der Wissenschaften, Berlin-Brandenburgische Akademie. 2007. *Turfan studies*. Berlin: Berlin-Brandenburg academy of sciences and humanities. http://catalogue.bnf.fr/ark:/12148/cb41139646j.

Deravi, Farzaneh & Jean-Yves Dommergues. 2011. Bilingual Speech of Highly Proficient Persian-French Speakers. In 197-212 (ed.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 197–212. Wiesbaden: Reichert.

Doostan, Gholamhossein Karimi & Fatemeh Daneshpazhouh. 2019. Kurdish -râ as an Anti-Actor marker. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 205–224. Bamberg: University of Bamberg Press.

Dorleijn, Margreet. 1996. *The decay of ergativity in Kurdish*. Tilburg: Tilburg University Press.

Durkin-Meisterernst, Desmond. 2004. *Dictionary of Manichaean Middle Persian and Parthian*. Turnhout, Belgium: Brepols.

Durkin-Meisterernst, Desmond. 2009. Khwarezmian. In Gernot L Windfuhr (ed.), *The iranian languages*, 336–376. New York, NY: Routledge.

Durkin-Meisterernst, Desmond. 2011. The importance of the Middle Persian texts from Turfan. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 97–104. Wiesbaden: Reichert.

Durkin-Meisterernst, Desmond. 2014. *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch)* (Sitzungsberichte / Österreichische Akademie der Wissenschaften. Philosophisch-Historische Klasse, 0029-8832 ; 850. Band). Wien: Verlag der Österreichischen Akademie der Wissenschaften 1st edn.

Edmonds, C. J. 1955. Prepositions and Personal Affixes in Southern Kurdish. *Bulletin of the School of Oriental and African Studies* 17(3). 490–502.

Ekici, Deniz. 2007. *Kurmanji Kurdish Reader*. Dunwoody Press.

Elfenbein, J. 1982. Notes on the Balochi-Brahui Linguistic Commensality. *Transactions of the Philological Society* 80(1). 77–98. doi:10.1111/j.1467-968X.1982.tb01195.x.

Elfenbein, Josef. 1961. A Baluci text, with translation and Notes. *Bulletin of the School of Oriental and African Studies, University of London* 24(1). 86–103.

Elfenbein, Josef. 1989. "Bal\=o\v{c}\={ı}. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, 350–362. Wiesbaden: Reichert Verlag.

Elfenbein, Josef. 1997. Balochi Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 37, 761–776. Winona Lake, Indiana: Eisenbrauns.

Emeneau, M. B. 1964. Linguistic Desiderata in Baluchistan. In *Indo-iranica: mélanges présentés à georg morgenstierne, à l'occasion de son soixante-dixième anniversaire*, 73–77. Wiesbaden: Otto Harrassowitz.

Emmerick, Roland E. 1998. Khotanese ei. In Nicholas Sims-Williams (ed.), *Proceedings of the third european conference of iranian studies: held in cambridge, 11th to 15th, 1995: Part 1 old and middle iranian studies (beiträge zur iranistik 17)*, 93–98. Wiesbaden: Reichert.

Erschler, David & Vitaly Volk. 2011. On Negation, Negative Concord, and Negative Imperitives in Digor Ossetic. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 135–152. Wiesbaden: Reichert.

Farrell, Tim. 2003. Linguistic Influences on the Balochi Spoken in Karachi. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbours: Ethnic and linguistic contact in balochistan in historical and modern times*, 169–210. Weisbaden: Reichert.

Fattah, Ismaïl Kamandâr. 2000. *Les dialectes kurdes méridionaux: étude linguistique et dialectologique (Iranica 37)*. Leuven: Peeters.

Filippone, Ela. 1996. *Spacial Models and Locative Expressions in Baluchi*. Naples: Istituto Universitario Orientale Dipartimento di Studi Asiatici.

Filippone, Ela. 2011. The language of the Qor'\=an-e Qods and its Sistanic dialectal background. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 179–237. Wiesbaden: Reichert.

Franco-Rita, Ludovico & Manzini-Leonardo Savoia. 2012. Some call it article, some ezafe: "Linkers" as "agreement". *Linguistic Review* 2. 1–54.

Friedman, Victor A. 1991. Case in Romani: Old Grammar in New Affixes. *Journal of the Gypsy Lore Society (Fifth Series)* 1(2). 85–102.

Ganjavi, Shadi. 2011. On Direct Objects in Persian: The Case of the Non-r\^a-Marked DOs. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 167–182. Wiesbaden: Reichert.

Gebhart, Lewis. 2008. Classifiers, plural and definiteness in Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 35–56. Newcastle upon Tyne: Cambridge Scholars Publishing.

Geiger, Wilhelm. 1895a. Die Sprache der Balūtschen. In Wilhelm Geiger (ed.), *Grundriss der iranischen philologie*, chap. vi, 231–48. Strassburg: Trübner.

Geiger, Wilhelm. 1895b. *Grundriss der iranischen Philologie*. Strassburg: Trübner.

Geiger, Wilhelm. 1895c. Kleinere Dialekte und Dialektgruppen. In Wilhelm Geiger (ed.), *Grundriss der iranischen philologie*, chap. viii, 287–423. Strassburg: Trübner.

Geiger, Wilhelm. 1901. Kleinere Dialekte und Dialektgruppen: IV. Allgemeine Übersicht über die Dialekte und ihre Gruppirung. In Wilhelm Geiger & Ernst Kuhn (eds.), *Grundriss der iranischen philologie*, I/2:412–423. Strassburg: Trübner.

Gershevitch, Ilya. 1964. Iranian Chronological Adverbs. In *Indo-iranica : mélanges présentés à georg morgenstierne a l'occasion de son soixante-dixième anniversaire*, 78–88. Wiesbaden: Otto Harrassowitz.

Gershevitch, Ilya. 1985a. Ancient Survivals in Ossetic (1952). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 114–126. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985b. Bactrian Inscriptions and Manuscripts (1967). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 83–113. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985c. Dialect Variation in Early New Persian (1965). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 194–222. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985d. Etymological Notes on Persian mih, naxc\={\i}r, b\=eg\=ane, and b\={\i}m\=ar (1964). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 190–193. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985e. Genealogical Descent in Iranian (1973). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 265–277. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985f. Iranian Notes (1949). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 25–32. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985g. Iranian words containing -an- (1971). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 237–261. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985h. Sogdian Compounds (1949). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 6–18. Wiesbaden: Reichert.

Gershevitch, Ilya. 1985i. The Crushing of the Third Singular Present (1970). In Nicholas Sims-Williams (ed.), *Philologia iranica*, 223–236. Wiesbaden: Reichert.

Gholami, Saloumeh. 2011. Definite Articles in Bactrian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 11–22. Wiesbaden: Reichert.

Gholami, Saloumeh. 2018a. *Endangered Iranian Languages*. Wiesbaden: Reichert Verlag.

Gholami, Saloumeh. 2018b. Pronominal clitics in Zoroastrian Dari (Behd\={\i}n\={\i}) of Kerman. In Saloumeh Gholami (ed.), *Endangered iranian languages*, 111–122. Wiesbaden: Reichert.

Ghomeshi, Jila. 1997. Non-Projecting Nouns and the Ezafe: Construction in Persian. *Natural Language & Linguistic Theory* 15. 729–788.

Ghomeshi, Jila. 2008. Markedness and bare nouns in Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 85–112. Newcastle upon Tyne: Cambridge Scholars Publishing.

Ghomeshi, Jila. 2020. The additive particle in Persian: A case of morphological homophony between syntax and pragmatics. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 4, 57–84. Amsterdam; Philadelphia: John Benjamins.

Ghomeshi, Jila, Carina Jahani & Agnes Lenepveu-Hotz. 2016. *Further topics in Iranian linguistics: proceedings of the 5th International Conference on Iranian Linguistics, held in Bamberg on 24-26 August 2013*. Association pour l'avancement des études Iraniennes.

Gilbertson, George Waters 1860 & Ghāno Khān. Haddiānī. 1923. *The Balochi language, a grammar and manual*. Hertford: Printed by S. Austin & Sons, for the author. http://link.library.utoronto.ca/eir/EIRdetail.cfm?Resources{_}{_}ID=538255{&}T=F.

Gippert, Jost. 2007. Albano-Iranica. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 99–108. Wiesbaden: Harrassowitz.

Gippert, Jost. 2009. Zur dialektalen Stellung des Zazaki. *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse* 47(1). 77–107.

Goodrick, Sophie. 1999. The measurement of style in Persian texts: a question of validity. In Charles Melville (ed.), *Proceedings of the third european conference of iranian studies: held in cambridge, 11th to 15th, 1995: Part 2 medieval and modern persian studies (beiträge zur iranistik 17)*, 33–38. Wiesbaden: Reichert.

Gündogdu, Songül, Ergin Opengin, Erik Anonby & Geoffrey Haig. 2019. *Current Issues in Kurdish Linguistics*. Bamberg: University of Bamberg Press.

Gündoğdu, Songül. 2015. Noun-Verb Complex Predicates In Kurmanji Kurdish: A syntactic account. In *Console xxiii*.

Gündoğdu, Songül. 2017a. Variation in the Ergative Pattern of Kurmanji. In Katharina Brizić, Agnes Grond, Christoph Osztovics, Thomas Schmidinger & Maria Six-Hohenbalken (eds.), *Wiener jahrbuch für kurdische studien 5*, 45–62. Wien: Praesens.

Gündoğdu, Songül. 2017b. Variation in the Ergative Pattern of Kurmanji. In Katharina Brizić, Agnes Grond, Christoph Osztovics, Thomas Schmidinger & Maria Six-Hohenbalken (eds.), *Wiener jahrbuch für kurdische studien 5*, 45–62. Wien: Praesens.

Gündoğdu, Songül. 2018. *ARGUMENT-ADJUNCT DISTINCTION IN KURMANJI KURDISH*: BOĞAZİÇİ UNIVERSITY dissertation.

Gündoğdu, Songül. 2019. Asymmetries in Kurmanji morphosyntax. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 93–116. Bamberg: University of Bamberg Press.

Hadank, Karl. 1930. *Mundarten der Gûrân : besonders das Kändûläî, Auramânî und Badschälânî*. Berlin: Verlag der Preussischen Akademie der Wissenschaften.

Hadank, Karl. 1932. *Mundarten der Zâzâ: hauptsächlich aus Siwerek und Kor*. Berlin: Verlag der Preussischen Akademie der Wissenschaften : In Kommission bei W. de Gruyter.

Haghkerdar, Michael. 2009. *Western Iranian Ezafe: A Comparative Syntactic Analysis*: Yale dissertation.

Haider, Hubert & Ronald Zwanziger. 1984. Relatively attributive: The 'ezäfe'-construction from Old Iranian to Modem Persian. In J Fisisk (ed.), *Historical syntax*, 137–172. Berlin, New Yourk: Mouton.

Haig, Geoffrey. 2004. *Alignment in Kurdish: a diachronic perspective*: University of Kiel Habilitation sschrift.

Haig, Geoffrey. 2007. Grammatical borrowing in Kurdish (Northern Group). In Yaron Matras (ed.), *Grammatical borrowing in cross-linguistic perspective*, 165–183. Berlin; New York: Mouton de Gruyter.

Haig, Geoffrey. 2008a. *Alignment change in Iranian languages: a construction grammar approach*. New York: Mouton de Gruyter.

Haig, Geoffrey. 2008b. The emergence of ergativity in Iranian: reanalysis of extension? In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 113–128. Newcastle upon Tyne: Cambridge Scholars Publishing.

Haig, Geoffrey. 2011. Linker, relativizer, nominalizer, tense-particle On the Ezafe in West Iranian. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in asian languages: Diachronic and typological perspectives [typological studies in language 96]*, 363–390. Benjamins.

Haig, Geoffrey. 2013. *The subject/object assymetry in bound person indexing: diachronic and discourse considerations*: University of Bamberg dissertation.

Haig, Geoffrey. 2015. Ergativity in Iranian.

Haig, Geoffrey. 2019a. Debonding of inflectional morphology in Kurdish and beyond. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 117–144. Bamberg: University of Bamberg Press.

Haig, Geoffrey. 2019b. Northern Kurdish (Kurmanji). In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspective*, 106–158. Berlin; Boston: Walter de Gruyter.

Haig, Geoffrey. 2019c. The Iranian languages of northern Iraq. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspectiveanguag*, 267–304. Berlin; Boston: Walter de Gruyter.

Haig, Geoffrey. 2020. The pronoun-to-agreement cycle in Iranian: Subjects do, objects don't. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 5, 85–106. Amsterdam; Philadelphia: John Benjamins.

Haig, Geoffrey & Geoffrey Khan. 2019. *The Languages and Linguistics of Western Asia: An areal perspective*. Berlin; Boston: Walter de Gruyter.

Haig, Geoffrey & Masoud Mohammadirad. 2019. Definiteness in Central Kurdish: sources and outcomes. Presented at the 4th International Conference on Kurdish Linguistics, University of Rouen, Paris-France.

Haig, Geoffrey & Baydaa Mustafa. 2019. Language choice and patterns of usage among Kurdish speakers of Duhok: An empirical intergenerational study. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 145–168. Bamberg: University of Bamberg Press.

Haig, Geoffrey & Ergin Öpengin. 2014. Introduction to special issue Kurdish: A critical research overview. *Kurdish Studies* 2(2). 99–122.

Haig, Geoffrey & Ergin Öpengin. 2018. Kurmanji Kurdish in Turkey: structure, varieties, and status. In Christiane Bulut (ed.), *Linguistic minorities in turkey and turkic speaking minorities of the peripheries*, 157–230. Wiesbaden: Harrassowitz.

Hale, Mark. 2004a. Avestan. In Roger D Woodard (ed.), *The cambridge encyclopedia of the world's ancient languages*, chap. 29, 742–763. Cambridge: Cambridge University Press.

Hale, Mark. 2004b. Pahlavi. In Roger D Woodard (ed.), *The cambridge encyclopedia of the world's ancient languages*, chap. 30, 764–776. Cambridge: Cambridge University Press.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2020. Glottolog 4.2.1. doi:10.5281/zenodo.3754591. https://glottolog.org/.

Hasandust, Muhammad. 2011a. Etymological notes on some Classical New Persian words. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 289–298. Wiesbaden: Reichert.

Hasandust, Muhammad. 2011b. *Farhang-i tatbiqi-mawzu'i-i zubanha u guvishha-i irani-i naw*. Tehran: Nashar Asar.

Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries*, 197–226. Berlin, New York: De Gruyter. doi:https://doi.org/10.1515/9783110331127.

Hazhār. 1991. *Farhang-i Kurdī-Fārsī*. tehran: Surūsh.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge; New York: Cambridge University Press.

Herin, Bruno. 2012. The Domari language of Aleppo, Syria. *Linguistic Discovery* 10(2). 1–52. doi:10.1349/PS1.1537-0852.A.412. http://journals.dartmouth.edu/cgi-bin/Web{%}0AObjects/Journals.woa/1/xmlpage/1/article/412.

Herkenrath, Annette. 2019. Temporal noun squishes in Kurmanji academic writing: From lexicality via NP-level junction to clausal subordination. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 169–204. Bamberg: University of Bamberg Press.

Hock, Hans Henrich. 1991. *Principles of Historical Linguistics*. Berlin; New York: Mouton de Gruyter.

Holmberg, Anders & David Odden. 2008. The Noun Phrase in Hawrami*. In Vida Samiian, Donald Stilo & Simin Karimi (eds.), *Aspects of iranian linguistics*, 129–152. Newcastle upon Tyne: Cambridge Scholars Publishing.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization, vol. i*, 17–36. Amsterdam: John Benjamins.

Horn, Paul. 1895. Neupersische Schriftsprache. In Wilhelm Geiger (ed.), *Grundriss der iranischen philologie*, chap. iv, 1–200. Strassburg: Trübner.

Huyse, Philip. 2017. The dialectology of Iranian. In Jared S Klein, Brian D Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 37, 599–608. Berlin: De Gruyter Mouton.

Igartua, I. 2015. From cumulative to separative exponence in inflection: Reversing the morphological cycle. *Language* 91(3). 676–722. doi:10.1353/lan.2015.0032. http://proxy.lib.ohio-state.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=edselc{&}AN=edselc.2-52.0-84942456166{&}site=eds-live.

Ivanow, Wladimir. 1929. Notes on the dialect of Khūr and Mihrijān. *Acta Orientalia* 8. 45–61.

Ivanow, Wladimir. 1931. The Dialect of Gozärkhon in Alamut. *Acta Orientalia* 9. 352–369.

Izady, Mehrdad R. 1992. *The Kurds: a concise handbook*. Washington: Crane Russak. http://www.gbv.de/dms/hbz/toc/ht004369273.PDF.

Jackson, A V Williams. 1968. *An Avesta grammar in comparison with Sanskrit*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Jacobson, C. M. 1993. *Rastnustena Zonê Ma, Handbuch für die Rechtschreibung der Zaza-Sprache*. Bonn, Germany: Verlag für Kultur und Wissenschaft.

Jahani, Carina. 2003. The Case System in Iranian Balochi in a Contact Linguistic Perspective. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbours: Ethnic and linguistic contact in balochistan in historical and modern times*, 113–132. Wiesbaden: Reichert.

Jahani, Carina. 2008a. Expressions of future in Classical and Modern New Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 153–176. Newcastle upon Tyne: Cambridge Scholars Publishing.

Jahani, Carina. 2008b. Restrictive Relative Clauses in Balochi and the Marking of the Antecedent ‐Linguistic Influence from Persian? In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and others*, 139–167. Wiesbaden: Reichert.

Jahani, Carina. 2015. On the definite marker in Modern Spoken Persian. Paper presented at the Sixth International Conference on Iranian Linguistics (ICIL6).

Jahani, Carina. 2019. *A Grammar of Modern Standard Balochi*. Uppsala: Acta Universitatis Upsaliensis.

Jahani, Carina, Behrooz Barjasteh-Delforooz, Serge Axenov & Maryam Nourzaei. 2010. Impersonal constructions in Balochi. *Orientalia Suecana* 59. 168–181. doi:10.1075/tsl.46.09rob.

Jahani, Carina & Agnes Korn. 2003. *The Baloch and their Neighbors: Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times*. Wiesbaden: Reichert.

Jahani, Carina & Agnes Korn. 2009. Balochi. In Gernot Windfuhr (ed.), *The iranian languages*, 634–692. London: Routledge.

Jahani, Carina, Agnes Korn & Paul Titus. 2008. *The baloch and others: linguistic, historical and socio-political perspectives on pluralism in Balochistan*. Wiesbaden: Reichert Verlag.

Jahāngiri, Nāder. 2003. *Guyeš-e Gilaki-e Lāhijān: vižegihā-ye āvāyi va sāxtvāžei [Gilaki dialect of Lahijan: a study on phonology, morphology and lexicon]. (Studia Culturae Islamicae, 74:1-3.)*. Tokyo: Institute for the Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Jasbi, Masoud. 2020a. The meaning of the Persian object marker rā: What it is not, and what it (probably) is. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 7, 119–136. Amsterdam; Philadelphia: John Benjamins.

Jasbi, Masoud. 2020b. The suffix that makes Persian nouns unique. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 6, 107–118. Amsterdam; Philadelphia: John Benjamins.

Jeremiás, Éva M. 1993. On the Genesis of the Periphrastic Progressive in Iranian Languages. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 99–116. Leuven: Peeters.

Jeremiás, Éva M. 1999. Grammar and linguistic consciousnss in Persian. In Charles Melville (ed.), *Proceedings of the third european conference of iranian studies: held in cambridge, 11th to 15th, 1995: Part 2 medieval and modern persian studies (beiträge zur iranistik 17)*, 19–32. Wiesbaden: Reichert.

Jeremiás, Éva M. 2016. The history of grammatical ideas in Persian: kitābatan-lafzan in Classical Persian sources. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 55–71. Association pour l'avancement des études Iraniennes.

Joseph, Brian. 2009. Greek dialectal evidence for the role of the paradigm in inflectional change. *Morphology* 19(1). 45–57.

Josephson, Judith. 1993. The Preverb be in some Late Book Pahlavi Texts. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 117–126. Leuven: Peeters.

Josephson, Judith. 2011. Definiteness and deixis in Middle Persian. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 23–40. Wiesbaden: Reichert.

Josephson, Judith. 2016. The expression of modality in Late Middle Persian. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 71–90. Association pour l'avancement des études Iraniennes.

Jügel, Thomas. 2009. Ergative Remnants in Sorani Kurdish? *Orientalia Suecana* LVIII.

Jügel, Thomas. 2014. On the linguistic history of Kurdish. *Kurdish Linguistics* 2(May). 123–142. www.kurdishstudies.net.

Jügel, Thomas. 2016. Enclitic Pronouns in Middle Persian and the Placeholder Construction. In Mahmoud Jaafari Dehaghi (ed.), *Quaterly journal of language and inscription 1/1 [1396 h.š.], dedicated to professor mansour shaki*, 41–63. Tehran.

Jügel, Thomas. 2017. The syntax of Iranian. In Jared S Klein, Brian D Joseph & Matthias Andreas Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 35, 549–566. Berlin: de Gruyter Mouton.

Jügel, Thomas, Cemile Çelebi & Diako Nahid. 2017. Deixis and near future expressions in Kurdish. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and proximative in turkic, iranian and beyond*, Wiesbaden: Reichert Verlag.

Jügel, Thomas & Pollet Samvelian. 2020. Topic agreement, experiencer constructions, and the weight of clitics. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 8, 137–154. Amsterdam; Philadelphia: John Benjamins Publishing Company. doi:https://doi.org/10.1075/cilt.351.08jug.

Jurafsky, Daniel. 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72(3). 533–578. doi:10.2307/416278. http://www.jstor.org.proxy.lib.ohio-state.edu/stable/416278.

Kahnemuyipour, Arsalan. 2014. Revisiting the Persian Ezafe construction: A roll-up movement analysis. *Lingua* 150. 1–24. doi:10.1016/j.lingua.2014.07.012.

Kahnemuyipour, Arsalan. 2016. The Ezafe Construction: Persian and Beyond. Paper presented at the second annual Conference on Central Asian Languages and Linguistics (ConCALL2).

Kamioka, Koji., Ataollah. Rahbar & Ali Akbar. Hamidi. 1986. *Comparative basic vocabulary of Khonjī and Lārī : Lārestānī studies 2*. Tokyo: Institute for the study of languages and cultures of Asia and Africa.

Kamioka, Koji. & Minoru. Yamada. 1979. *Lārestāni studies*. Tokyo-.: Institute for the Study of Languages and Cultures of Asia and Africa.

Kanga, Kavasji Edalji. 2003. *A complete dictionary of the Avesta language, in Guzerati and English*. Mumbai: K.R. Cama Oriental Institute.

Kareem, Rebeen. 2016. *the Syntax of Verbal Inflection in Central Kurdish*: University of Newcastle upon Tyne Phd.

Karim, Shuan Osman. 2018. The Kurmancî tense ezafe and its Deep Roots in Western Iranian. Presented at the 8th International Conference on Iranian Linguistics.

Karim, Shuan Osman. 2019. Competition between formatives and the diversity of ezafat. Presented at the 24th International Conference on Historical Linguistics (ICHL24).

Karim, Shuan Osman. 2021a. *From phrases to formatives: the diachrony of Soranî (Central Kurdish) applicative markers*. Columbus, OH: (under review).

Karim, Shuan Osman. 2021b. Rethinking Kurdish imperfective affixes. Tech. rep. The Ohio State University Columbus, OH.

Karim, Shuan Osman. 2021c. The Ezafe and the Article. To appear in the Proceedings of the 2nd North American Conference on Iranian Linguistics.

Karim, Shuan Osman & Ali Salehi. 2020. Soranî Valence Changing Affixes: Teetering on the boundary between morphology and syntax. Presented at the 94th Annual Meeting of the Linguistic Society of America.

Karimi, Simin. 2008. Raising and Control in Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 177–208. Newcastle upon Tyne: Cambridge Scholars Publishing.

Karimi, Simin & Michael Brame. 2012. A Generalization Concerning the EZAFE Construction in Persian. *Linguistic Analysis* 38. 111–143.

Karimi, Simin & Ryan Walter Smith. 2020. Another look at Persian rā: A single formal analysis of a multi-functional morpheme. In Richard K. Larson, Sedigheh Moradi

& Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 9, 155–173. Amsterdam; Philadelphia: John Benjamins.

Karimi, Yadgar. 2007. Kurdish Ezafe construction: Implications for DP structure. *Lingua* 117(12). 2159–2177.

Karimi-Doostan, Gholamhossein. 2008. Event structure of verbl nouns and light verbs. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 209–226. Newcastle upon Tyne: Cambridge Scholars Publishing.

Karvovskaya, Lena. 2016. Comparative constructions in Ishkashimi. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 91–112. Association pour l'avancement des études Iraniennes.

Kaye, Alan S. 1997. *Phonologies of Asia and Africa*. Winona Lake, Indiana: Eisenbrauns.

Kaye, Alan S. 2007. *The Morphologies of Asia and Africa*. Winona Lake, Indiana: Eisenbrauns.

Kent, Roland G. 1944. The Old Persian Relative and Article. *language* 20(1). 1–10.

Kent, Roland G. 1950. *Old Persian: Grammar Texts Lexicon*. New Haven, Connecticut: American Oriental Society.

Key, Gregory. 2008. Differential object marking in a Medieval Persian text. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 227–248. Newcastle upon Tyne: Cambridge Scholars Publishing.

Khan, Geoffrey. 2019. The Neo-Aramaic dialects of northern Iraq. In Geoffrey Khan & Geoffrey Haig (eds.), *The languages and literatures of western asia - an areal perspective*, 305–353. Berlin: De Gruyter.

Khonji, Loṭf-ᶜAli. 2000. *Dastur-e zabān-e lārestāni bar mabnā-ye guyeš-e Xonji.* Shiraz: Dānešnāme Fārs.

Klavans, Judith L. 2017. The Independence of Syntax and Phonology in Cliticization. *Linguistic Society of America* 61(1). 95–120.

Klein, Jared S, Brian D Joseph & Matthias Andreas Fritz. 2017. *Handbook of comparative and historical Indo-European linguistics Band 1.* De Gruyter Mouton.

Koch, Christoph. 1999. Weiterungen des relativen Attributivkonnexes: Die westslavischen Adverbialadjektiva auf -ej's-. *Zeitschrift für Slavistik* 44(4). 455–475.

Korn, Agnes. 2003. Balochi and the Concept of North-Western Iranian. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbours: Ethnic and linguistic contact in balochistan in historical and modern times*, 49–60. Wiesbaden: Reichert.

Korn, Agnes. 2005. *Towards a Historical Grammar of Balochi: studies in Balochi historical phonology and vocabulary.* Wiesbaden: Reichert.

Korn, Agnes. 2008a. A New Locative Case in Turkmenistan Balochi*. *Iran and the Caucasus* 12. 83–100.

Korn, Agnes. 2008b. Marking of Arguments in Balochi Ergative and Mixed Constructions. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of iranian linguistics*, Newcastle upon Tyne: Cambridge Scholars Publishing.

Korn, Agnes. 2008c. The Nominal Systems of Balochi: How Many Grammars? In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and others: Linguistic, historical and socio-political perspectives on pluralism in balochistan*, 167–196. Wiesbaden: Reichert.

Korn, Agnes. 2011. Pronouns as Verbs, Verbs as Pronouns: Demonstratives and the Copula in Iranian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 53–70. Wiesbaden: Reichert.

Korn, Agnes. 2016a. A partial tree of Central Iranian. *Indogermanische Forschungen* 121(1). 401–434. doi:10.1515/if-2016-0021.

Korn, Agnes. 2016b. Western Iranian Pronominal Clitics. *Suecana, Orientalia* 159–171.

Korn, Agnes. 2017a. Notes on the Nominal System of Bashkardi. *Transactions of the Philological Society* 115(1). 79–97.

Korn, Agnes. 2017b. The evolution of Iranian. In Jared S Klein, Brian D Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 38, 608–624. Berlin: De Gruyter Mouton.

Korn, Agnes. 2019a. Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics* 9(2). 239–281.

Korn, Agnes. 2019b. "Digging through Layers of Language Contact: Elements of Diglossia and Multilingualism in Balochi". Presented at the 2nd North American Conference on Iranian Linguistics.

Korn, Agnes & Desmond Durkin-Meisterernst. 2009. Parthian seen from a balochi perspective. *Studia Iranica* 38. 7–23.

Korn, Agnes, Geoffrey Haig, Simin Karimi & Pollet Samvelian. 2011. *Topics in Iranian Linguistics (Beiträge zur iranistik 34)*. Wiesbaden: Reichert.

Kubota, Yusuke. 2010. *(In)flexibility of constituency in Japanese in Multi-Modal Categorial Grammar with Structured Phonology*: Ohio State University Phd.

Kubota, Yusuke. 2014. The logic of complex predicates: A deductive synthesis of "argument sharing" and "verb raising". *Natural Language and Linguistic Theory* 32(4). 1145–1204.

Kubota, Yusuke. 2015. Nonconstituent coordination in Japanese as constituent coordination: An analysis in Hybrid Type-Logical Categorial Grammar. *Linguistic Inquiry* 46(1). 1–42.

Kubota, Yusuke & Robert Levine. 2012. Gapping as like-category coordination. In Denis Béchet & Alexander Dikovsky (eds.), *Logical aspects of computational linguistics 2012*, 135–150. Heidelberg: Springer.

Kubota, Yusuke & Robert Levine. 2013a. Coordination in Hybrid Type-Logical Categorial Grammar. *OSU Working Papers in Linguistics* 60. 21–50.

Kubota, Yusuke & Robert Levine. 2013b. Determiner gapping as higher-order discontinuous constituency. In Glyn Morrill & Mark-Jan Nederhof (eds.), *Proceedings of formal grammar 2012 and 2013*, 225–241. Heidelberg: Springer.

Kubota, Yusuke & Robert Levine. 2014. Pseudogapping as pseudo-VP ellipsis. In Nicholas Asher & Sergei Soloviev (eds.), *Logical aspects of computational linguistics 2014*, 122–137. Heidelberg: Springer.

Kubota, Yusuke & Robert Levine. 2016. Gapping as hypothetical reasoning. *Natural Language and Linguistic Theory* 34(1). 107–156.

Kubota, Yusuke & Robert D. Levine. 2020. *Type-Logical Syntax*. Cambridge, Mass; London, England: The MIT Press.

Kümmel, Martin Joachim. 2014. The development of laryngeals in Indo-Iranian. In *The sound of indo-european*. Opava.

Kuryłowicz, Jerzy. 1947. La nature des procés dits Analogiques. *Acta Linguistica* 5. 17–34.

Kārang, 'Abd-al-'Ali. 1953. *Tāti va Harzani, do lahja az zabān-i bāstān-e Āz̲arbāyjān*. Tabriz: Tabriz University Press.

Lambek, Joachim. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65(3). 154–170.

Lambton, Ann K S 1912-2008. 1938. *Three Persian dialects*. London: Royal Asiatic Society.

Larson, Richard. 2021. Turkish inflyuence on Iranian Relative Clause Syntax. Presented at the Linguistics Association of the University of Kurdistan.

Larson, Richard, Sedigheh Moradi & Vida Samiian. 2020. *Advances in Iranian Linguistics*. Amsterdam; Philadelphia: John Benjamins.

Larson, Richard & Vida Samian. 2020. Ezafe, PP and the Nature of Nominalization. *Natural Language & Linguistic Theory* doi:10.1007/s11049-020-09471-1.

Larson, Richard & Vida Samiian. 2020. The Ezafe Construction Revisited. In Richard K. Larson, Sadigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, 173–236. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Larson, Richard & Hiroko Yamakido. 2008. Ezafe and the deep position of nominal modifiers. In Louise McNally & Kennedy (eds.), *Adjectives and adverbs: syntax, semantics, and discourse*, 43–70. Oxford; New York: Oxford University Press.

Lazard, Gilbert. 1981. Le Dialecte des Juifs de Kerman. In *Monumentum georg morgenstierne i*, 333–346. Leiden: Brill.

Lazard, Gilbert. 1989. Le persan. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 4.1.1, 263–293. Wiesbaden: Reichert Verlag.

Lazard, Gilbert. 1990. Le Dialect Rudbar (Gilan). In *Iranica varia: Papers in honor of professor ehsan yarshater*, 110–123. Leiden: Brill.

Lazard, Gilbert. 1992. Le dialecte laki d'Aleshtar. *Studia Iranica* 21(2). 215–245.

Lazard, Gilbert. 2007. La versification en parthe et son heritage persan. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 161–172. Wiesbaden: Harrassowitz.

Lazard, Gilbert. 2011. Homonymie et polysemie: be\‘eve note \‘a propos des deux enclitiques -i du Persan. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 89–94. Wiesbaden: Reichert.

Le Coq, Pierre. 1989a. Les dialectes du caspians et les dialectes du nord-ouest de l'Iran. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 4.1.2.2, 313–326. Wiesbaden: Reichert Verlag.

Le Coq, Pierre. 1989b. Les dialectes du sud-ouest de l'Iran. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 4.1.2.6, 341–349. Wiesbaden: Reichert Verlag.

Lecoq, Pierre. 2002. *Recherches sur les dialectes kermaniens (Iran central) : grammaire, textes, traductions et glossaires*. Leuven: Peeters.

Leezenberg, Michiel. 1994. *The Shabak and the Kakais: dynamics of ethnicity in Iraqi Kurdistan*. Amsterdam: University of Amsterdam.

Leezenberg, Michiel. 2015. Gorani influence on central Kurdish: Substratum or prestige borrowing.

Leumann, M. 1964. Altpersisch hagmat\=a. In *Indo-iranica: mélanges présentés à georg morgenstierne, à l'occasion de son soixante-dixième anniversaire*, 124–128. Wiesbaden: Otto Harrassowitz.

Lewis, David. 2002. Adverbs of Quantification. In Paul Portner & Barbara H Partee (eds.), *Formal semantics: the essential readings*, 178–188. Oxford: Blackwell.

Lorenz, M. 1982. Die direkte Rede im T\=a\v{g}ischen. In *Monumentum georg morgenstierne ii*, 5–14. Leiden: Brill.

Lorenz, Manfred. 1999. Partikeln in der modernen Tadshikischen Sprache. In Charles Melville (ed.), *Proceedings of the third european conference of iranian studies: held in*

cambridge, 11th to 15th, 1995: Part 2 medieval and modern persian studies (beiträge zur iranistik 17), 39–44. Wiesbaden: Reichert.

MacKenzie, David N. 1954. Gender in Kurdish. *Bulletin of the School of Oriental and African Studies, University of London* 16(3). 528–541.

Mackenzie, David N. 1961. *Kurdish dialect, studies 1-2. studies 1-2.* London; New York: Oxford University Press.

MacKenzie, David N. 1961. The Origins of Kurdish. *Transactions of the Philological Society* 60(1). 69–86. doi:10.1111/j.1467-968X.1961.tb00987.x.

MacKenzie, David N. 1964. The 'Indirect Affectee' in Pahlavi. In *Dr. j.m. unvala memorial volume*, Bombay: Dr. J.M. Unvala Memorial Volume Sub-Commmittee.

MacKenzie, David N. 1966. *The dialect of Awroman (Hawraman-i Luhon): Grammatical sketch, texts, and vocabulary.* Kobenhavn: Kommissionaer : Munksgaard.

MacKenzie, David N. 1982. Matal\=una. In *Monumentum georg morgenstierne ii*, 15–46. Leiden: Brill.

MacKenzie, David N. 1990. Pahlavi compound abstracts. In *Iranica varia: Papers in honor of professor ehsan yarshater*, 124–130. Leiden: Brill.

MacKenzie, David N. 1993. Clitics in Khwarezmian. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 135–146. Leuven: Peeters.

MacKenzie, David N. 1995a. Notes on Southern Zaza (Dimili). In *Proceedings of the second european conference of iranian studies: held in bamberg, 30th september to 4th october 1991, by the societas iranologica europaea*, 401–414. Istituto italiano per il Medio ed Estremo Oriente.

MacKenzie, David N. 1995b. Proceedings of the second European Conference of Iranian Studies. In Bert G. Fragner (ed.), *Proceedings of the second european conference of iranian studies (held in bamberg, 30th september to 4th october 1991, by the societas iranologica europaea)*, 401–414. Istituto italiano per il Medio ed Estremo Oriente.

MacKenzie, David N. 1999a. Bājalānī (1956). In Daniel Nathan Mackenzie, Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa i*, 409–426. Roma: Istituto italiano per l'Africa e l'Oriente. doi:10.1163/1573-3912_islam_SIM_1003LK-https://osu.on.worldcat.org/oclc/5862823479.

MacKenzie, David N. 1999b. Gender in Kurdish (1954). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa ii*, 353–367. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999c. Pahlavi compound abstracts (1990). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa i*, 275–282. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999d. Pseudoprotokurtica (1963). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa ii*, 402–402. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999e. The 'Indirect affectee' in Pahlavi (1964). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa i*, 3–6. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999f. The language of the Medians (1959). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa ii*, 367–368. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999g. The vocabulary of the Lahore Tafsir (1972). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa ii*, 347–352. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 1999h. When is a postposition not a postposition (1990). In Carlo G Cereti & Ludwig Paul (eds.), *Iranica diversa i*, 283–292. Roma: Istituto italiano per l'Africa e l'Oriente.

MacKenzie, David N. 2011. An index to "An Early Jewish-Persian argument". In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 237–246. Wiesbaden: Reichert.

Macuch, Maria, Mauro Maggi & Werner Sundermann. 2007. *Iranian languages and texts from Iran and Turan: Ronald E. Emmerick memorial volume (Iranica 13)*. Wiesbaden: Harrassowitz.

Maggi, Mauro. & Paola. Orsatti. 2011a. *The Persian language in history (Beiträge zur Iranistik 33)*. Wiesbaden: L. Reichert.

Maggi, Mauro & Paola Orsatti. 2011b. Two Syro-Persian Hymns for Palm Sunday and Maundy Thursday. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 247–288. Wiesbaden: Reichert.

Mahalingappa, Laura J. 2009. *The Acquisition of Split-Ergativity in Kurmanji Kurdish*: University of Texas at Austin Phd.

Mahand, Mohammad Rasekh & Zaniar Naghshbandi. 2014. The effect of discourse factors on case system in Hawrami. *Language Related Research* 4(4). 87–109.

Mahmoudveysi, Parvin & Denise Bailey. 2013. *The Gorani language of Zarda, a village of West Iran*. Wiesbaden: Reichert.

Mahmoudveysi, Parvin & Denise Bailey. 2019. Hawrāmī of western Iran. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspective*, 533–568. Berlin; Boston: Walter de Gruyter.

Mahmoudveysi, Parvin., Denise. Bailey, Ludwig. Paul & Geoffrey. Haig. 2012. *The Gorani Language of Gawraju, a village of West Iran: texts, grammar, and lexicon*. Wiesbaden: Reichert.

Mahootian, Shahrzad. 2008. Inversion and topicalization in Farsi discourse: A comparitive study. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 277–288. Newcastle upon Tyne: Cambridge Scholars Publishing.

Maisel, Sebastian. 2018. *The Kurds: An Encyclopedia of Life, Culture, and Society*. Greenwood: ABC-CLIO.

Malmasi, Shervin. 2016. Subdialectal Differences in Sorani Kurdish. In *Proceedings of the vardial workshop*. 89–96. Osaka, Japan.

Mann, Oskar. 1909. *Die Tâjîk-Mundarten der Provinz Fârs*. Berlin: Reimer.

Matras, Yaron, et al. 2016. The dialects of Kurdish. http://kurdish.humanities.manchester.ac.uk/.

Matras, Yaron. 2012. *A Grammar of Domari*. Berlin; Boston: Walter De Gruyter.

Matras, Yaron. 2019. Revisiting Kurdish dialect geography: Findings from the Manchester Database. In Songül Gündogdu, Ergin Opengin, Erik Anonby & Geoffrey Haig (eds.), *Current issuses in kurdish linguistics*, 225–242. Bamberg: University of Bamberg Press.

Mazdeh, Mohsen Mahdavi. 2020. Quantitative meter in Persian folk songs and pop lyrics. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 11, 237–256. Amsterdam; Philadelphia: John Benjamins.

McCarus, Ernest. 1956. *Descriptive analysis of the Kurdish of Sulaimaniya, Iraq*: University of Michigan dissertation.

McCarus, Ernest. 1997. Kurdish Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 34, 691–706. Winona Lake, Indiana: Eisenbrauns.

McCarus, Ernest. 2007. Kurdish Morphology. In Alan S Kaye (ed.), *Morphologies of asia and africa (vol. 2)*, chap. 36, 1021–1050. Winona Lake, Indiana: Eisenbrauns.

McCarus, Ernest. 2009. Kurdish. In Gernot Windfuhr (ed.), *The iranian languages*, 587–633. London and New York: Routledge.

McKinnon, Colin. 2001. GIŌNI. https://iranicaonline.org/articles/gioni-giani.

McKinnon, Colin. 2011. LORI LANGUAGE i. LORI DIALECTS. https://www.iranicaonline.org/articles/lori-dialects.

Melville, Charles. 1999. *Proceedings of the Third European Conference of Iranian Studies: held in Cambridge, 11th to 15th, 1995: Part 2 Medieval and Modern Persian Studies (Beiträge zur Iranistik 17)*. Wiesbaden: Reichert Verlag.

Miller, B. V. 1953. *Tali\v{s}ski Yazlik*. Moskva: Izdatelitstvo.

Miller, Corey. 2016. Theoretical issues in counting Persian words. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 113–134. Association pour l'avancement des études Iraniennes.

Mo3tamadi, 3A. 2001. *Rudān: Behešte Janub*. Bandar Abbas.

Mockler, Edward. 1877. *A Grammar of the Baloochee Language as it is spoken in Makrān (ancient Gedrosia), in the Persi-Arabic and Roman characters*. London: Henry S.~King.

Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*: Université Sorbonne Nouvelle Paris 3 dissertation. https://tel.archives-ouvertes.fr/tel-02988008/document.

Mohebbi Bahmani, Hassan. 2006. *Barresi va tawsif-e zabānšenāxti-ye guyeš-e Minābi [A linguistic study and description of the Minabi dialect]*. Tehrān.

Moortgat, Michael. 1997. Categorial type logics. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of logic and language*, 93–177. Amsterdam: Elsevier.

Morgenstierne, Georg. 1964. *Indo-Iranica; mélanges présentés à Georg Morgenstierne, à l'occasion de son soixante-dixième anniversaire*. Wiesbaden: O. Harrassowitz.

Morgenstierne, Georg. 1982. *Monumentum Georg Morgenstierne II*. Leiden: E.J. Brill.

Morrill, Glyn. 1994. *Type Logical Grammar: Categorial logic of signs*. Dordrecht: Kluwer.

Moyne, J A. 1971. Reflexive and Emphatic. *Language* 47(1). 141–163. doi:10.2307/412192. http://www.jstor.org.proxy.lib.ohio-state.edu/stable/412192.

Moyne, John & Guy Carden. 1974. Subject Reduplication in Persian. *Linguistic Inquiry* 5(2). 205–249. http://www.jstor.org.proxy.lib.ohio-state.edu/stable/4177817.

Naderi, Navid & Marc Van Oostendorp. 2011. Reducing the Number of Farsi Epentheric Consonants. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 153–167. Wiesbaden: Reichert.

Nasīr Xan, Aḥmadza'i. 1984. *The grammar of Balochi language*. Quetta: Balochi Acadamy.

Nawata, T. 1982. The Masal Dialekt of Talyshi. In *Monumentum georg morgenstierne ii*, 93–118. Leiden: Brill.

Neill, Paul O. 2016. Lexicalism, the Principle of Morphology-free Syntax and the Principle of Syntax-free Morphology. In Andrew Hippisley & Gregory Stump (eds.), *The cambridge handbook of morphology*, Cambridge: Cambridge University Press. doi: doi:10.1017/9781139814720.010.

Nilipour, Reza. 2008. Aspects of agrammatic langage in Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 289–308. Newcastle upon Tyne: Cambridge Scholars Publishing.

Noorlander, Paul M & Donald Stilo. 2015. On the Covergence of Verbal Systems of Aramaic and its Neighbours. Part I: Present-Based Paradigms. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-aramaic in its linguistic context*, Piscataway, NJ: Gorgias Press.

Nourzaei, Maryam. 2017. *Participant reference in three Balochi dialects: Male and female narrations of folktales and biographical tales*. Uppsala: Uppsala University dissertation.

Nourzaei, Maryam. 2020. Definiteness Marking from Evaluative Morphology in Balochi: Internal Variation and Diachronic Pathway. *Iranian Studies* doi:10.1080/00210862.2020.1813555.

Nourzaei, Maryam & Carina Jahani. 2013. The Distribution and Role of the Verb Clitic =a/a=in Different Balochi Dialects. *Orientalia Suecana* 61(614). 170–186.

Nourzaei, Maryam, Carina Jahani, Erik Anonby & Abbas Ali Ahangar. 2015. *Koroshi A Corpus-based Grammatical Description*. Uppsala: Uppsala Universitet. doi:10.1017/CBO9781107415324.004.

Oehrle, Richard. 1996. Austinian pluralities. In Jeremy Seligman & Dag Westerståhl (eds.), *Language, logic, and computation*, 433–441. Stanford: CSLI.

Öpengin, Ergin. 2013. Adpositions and Argument Indexing in the Mukri Variety of Central Kurdish : Focus on Ditransitive Constructions. *Orientalia Suecana* 61. 187–198.

Öpengin, Ergin. 2016. *The Mukri Variety of Central Kurdish*. Wiesbaden: Reichert Verlag.

Öpengin, Ergin. 2019. Accounting for the combinations of clitic and person markers in Central Kurdish. In Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig & Erik Anonby (eds.), *Current issuses in kurdish linguistics*, Bamberg: University of Bamberg Press.

Öpengin, Ergin & Geoffrey Haig. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish Studies* 2(2). 143–176.

Orsatti, Paola. 2011. The deictic suffix y\=a-ye e\v{s}\=arat: a hypothesis on the origin of the relative -i in Persian. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 41–88. Wiesbaden: Reichert.

Pakendorf, Brigitte & Ija V Krivoshapkina. 2014. Ėven nominal evaluatives and the marking of definiteness. *Linguistic Typology* 18(2). 289–331.

Parsafar, Parviz. 2010. Syntax, Morphology, and Semantics of Ezafe. *Iranian Studies* 43(5). 637–666.

Partee, Barbara H. 1976. *Montague Grammar*. Amsterdam: Elsevier.

Partee, Barbara H. 2002. Noun Phrase Interpretation and Type-shifting Principles. In Paul Portner & Barbara H Partee (eds.), *Formal semantics: the essential readings*, chap. 15. Oxford: Blackwell. doi:doi.org/10.1002/9780470758335.

Paul, Daniel. 2008a. The individuating function of the Persian 'indefinite suffix'. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 309–328. Newcastle upon Tyne: Cambridge Scholars Publishing.

Paul, Daniel. 2011a. *A comparative dialectal description of Iranian Taleshi*: University of Manchester Phd.

Paul, David. 2011b. A Glance at the Deixis of Nominal Demonstratives in Iranian Taleshi. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 89–102. Wiesbaden: Reichert.

Paul, Ludwig. 1998a. The position of Zazaki among West Iranian languages. *Proceedings of the Third European Conference of Iranian Studies: held in Cambridge, 11th to 15th, 1995: Part 1 Old and Middle Iranian Studies (Beiträge zur Iranistik 17)* 163–177.

Paul, Ludwig. 1998b. *Zazaki : Grammatik und Versuch einer Dialektologie*. Wiesbaden: Reichert.

Paul, Ludwig. 2003. The Position of Balochi Among the Western Iranian Languages: The Verbal System. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The baloch and their neighbors: Ethnic and linguistic contact in balochistan in historical and modern times*, 61–71. Wiesbaden: Reichert.

Paul, Ludwig. 2007. Zur Lage der Gorani-Dialekte im Iran und ihrer Erforschung. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 285–296. Wiesbaden: Harrassowitz.

Paul, Ludwig. 2008b. Some remarks on Persian -ra as a general and historical issue. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 329–338. Newcastle upon Tyne: Cambridge Scholars Publishing.

Paul, Ludwig. 2009. Zazaki. In Gernot. Windfuhr (ed.), *The iranian languages*, 545–586. London: Routledge.

Paul, Ludwig. 2013. *A grammar of early Judaeo-Persian*. Wiesbaden: Reichert Verlag.

Paul, Ludwig. 2019. Persian. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspective*, 569–624. Berlin; Boston: Walter de Gruyter.

Perry, John R. 2007. Persian Morphology. In Alan S Kaye (ed.), *Morphologies of asia and africa*, chap. 35, 975–1020. Winona Lake, Indiana: Eisenbrauns.

Pierce, E. 1875. A Description of the Mekranee-Beloochee Dialect. *Journal of the Bombay branch of the Royal Asiatic Society* XI. 1–98.

Pirooz, Mohammadreza. 2011. Finite Control in Persian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 183–197. Wiesbaden: Reichert.

Plank, Frans. 1999. Split morphology: How agglutination and flexion mix. *Linguistic Typology* 3. 279–340.

Provasi, Elio. 2011. New Persian texts in Manichaean script from Turfan. In Mauro Maggi & Paola Orsatti (eds.), *The persian language in history (beiträge zur iranistik 33)*, 119–178. Wiesbaden: Reichert.

Pufahl, April & Arthur G Samuel. 2014. How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology* 70. 1–30. doi:10.1016/j.cogpsych.2014.01.001.

Qazzaz, Shafiq. 2000. *The Sharezoor Kurdish-English dictionary.* Erbil, Kurdistan: Aras.

Rasekh-Mahand, Mohammad & Raheleh Izadifar. 2016. Compensating ergative alignment loss in Tākestāni. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 135–156. Association pour l'avancement des études Iraniennes.

Rasekh-Mahand, Mohammad & Zaniar Naghshbandi. 2013. Argument Strength and Differential Case Marking: Some Evidence from Hawrami. *Western Iranian Languages and Dialects* 1(1). 1–20.

Rasekhi, Vahideh. 2016. Missing objects in Persian. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 157–174. Association pour l'avancement des études Iraniennes.

Rasekhi, Vahideh. 2020. Stripping structures with negation in Persian. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 12, 257–274. Amsterdam; Philadelphia: John Benjamins.

Rastegar, Nosratollah. 1999. Abschlußbericht über das Neupersische Personennamenbuch (NpPNB) Zur Erstellung der Namenslemmata im NpPNB. In Charles Melville (ed.), *Proceedings of the third european conference of iranian studies: held in cambridge, 11th to 15th, 1995: Part 2 medieval and modern persian studies (beiträge zur iranistik 17)*, 1–18. Wiesbaden: Reichert.

Rastorgueva, V S, Aza Alimovna. Kerimova, Akhmed Kerimovich. Mamedzade, L A Pireĭko, D I Ėdel man & Ronald M Lockwood. 2012. *The Gilaki language*. Uppsala: Uppsala Universitet.

Rastorgueva, Vera S. 1971. *Gilyanskij Jazyk*. Moskva: NAUKA.

Ravel, Lili. 1961. *Khasi, a language of Assam*. Baton Rouge: Louisiana State University Press.

Rawlinson, Major. 1839. Notes on a march from Zoháb, at the foot of Zagros, along the mountains to Khúzistán (Susiana), and from thence through the province of Luristan to Kirmánsháh, in the year 1836. *Journal of the Royal Geographical Society of London* 9. 26–116.

Rießler, Michael. 2016. *Adjective attribution*. Berlin: Languge Science Press.

Rogova, N. V. 1999. *Severo-zapadnye iranskie Yazyki. (Yazyki mira: Iranskie Yazyki, II.)*. Moskva: Indrik.

Rossi, A. 1982. Bal\=o\v{c}\={\i} Miscellanea. In *Monumentum georg morgenstierne ii*, 157–184. Leiden: Brill.

Rzehak, Lutz. 2003. Some Thoughts and Material on Balochi in Afghanistan. In Carina Jahani & Agnes Korn (eds.), *The baloch and their neighbors: Ethnic and linguistic contact in balochistan in historical and modern times*, 259–278. Wiesbaden: Reichert.

Rzehak, Lutz. 2009. Code-Copying in the Balochi Language of Sistan. *Iranian Journal of Applied Language Studies* 1(1). 15–141.

Sadovski, Velizar. 2017. The lexicon of Iranian. In Jared S Klein, Brian D Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 36, 566–599. Berlin: De Gruyter Mouton.

Salehi, Ali. 2020. *Adjunct Incorporation in Soranî*: Bucharest dissertation.

Salemann, Carl. 1895. Mittelpersisch. In *Grundriss der iranischen philologie*, chap. iii, 249–332. Strassburg: Trübner.

Samiian, Vida. 1983. *Structure of phrasal categories in Persian: An X-bar analysis*. University of California, Los Angeles.

Samiian, Vida. 1994. The Ezafe Construction: Some Implications for the Theory of X-bar Syntax. In M Marashi (ed.), *Persian studies in north america*, 17–41. Bethesda, Md.: Iranbooks.

Samvelian, Pollet. 2005. When morphology does better than Syntax: The Ezafe construction in Persian. *(unpublished ms.)* .

Samvelian, Pollet. 2007a. A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics* 43(3). 605–645.

Samvelian, Pollet. 2007b. What Sorani Kurdish absolute prepositions tell us about cliticization. *Texas Linguistic Society IX: The Morphosyntax of Underrepresented Languages* 265.

Samvelian, Pollet. 2008. The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of iranian linguistics*, 339–361. Cambridge Scholars Ltd. https://halshs.archives-ouvertes.fr/halshs-00673182.

Samvelian, Pollet. 2018. Specific Features of Persian Syntax: The Ezâfe Construction, Differential Object Marking and Complex Predicates. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *The oxford handbook of persian linguistics*, 226–269. Oxford: Oxford University Press.

Schenker, Alexander M. 1964. *Polish declension: a descriptive analysis (Slavistic printings and reprintings 39)*. The Hague: Mouton.

Schmitt, Rüdiger. 1989. *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.

Schmitt, Rüdiger. 2002. Grundriss der iranischen Philologie. https://www.iranicaonline.org/articles/grundriss-der-iranischen-philologie.

Schmitt, Rüdiger. 2004. Old Persian. In Roger D Woodard (ed.), *The cambridge encyclopedia of the world's ancient languages*, chap. 28, 717–741. Cambridge: Cambridge University Press.

Schulze, Wolfgang. 2000. *Northern Talysh*. München: Lincom Europa. http://bvbr.bib-bvb.de:8991/F?func=service{&}doc{_}library=BVB01{&}local{_}base=BVB01{&}doc{_}number=009280266{&}line{_}number=0001{&}func{_}code=DB{_}RECORDS{&}service{_}type=MEDIA.

Seddiqi Nezhad, Bakhtiar. 2010. *Bar-resi-ye touṣifi-ye saxt-važe-ye guyeš-e bašakardi [A Descriptive Study of Bashakardi Dialect Morphology]*: Zahedan: Danešgah-e Sistan-va Balučestan Ma.

Seraji, Mojgan. 2015. *Morphosyntactic Corpora and Tools for Persian (ACTA UNIVERSITATIS UPSALIENSIS 16)*. Uppsala: Uppsala universitet.

Shahbakhsh, Azim. 2004. *The Balochi Verb: An Etymological Study*: University of London School of Oriental and African Studies (SOAS) Phd.

Shaked, Sh. 1982. Pahlavi Notes. In *Monumentum georg morgenstierne ii*, 197–206. Leiden: Brill.

Shaked, Sh. 1993. Iranian Elements in Middle Aramaic: Some Particles and Verbs. In Wojciech Skalmowski & Alois van Tongerloo (eds.), *Medioiranica: proceedings of the international colloquium organized by the katholieke universiteit leuven from the 21st to the 23rd of may 1990 (orientalia lovaniensia analecta 48)*, 147–156. Leuven: Peeters.

Sheyholislami, Jaffer. 2018. [The Kurdish] Language. In Sebastian Maisel (ed.), *The kurds: An encyclopedia of life, culture, and society*, 141–162. Greenwood: ABC-CLIO.

Shirtz, Shahar. 2016. Indirect participants as core arguments in Middle Persian. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 175–194. Association pour l'avancement des études Iraniennes.

Shokri, Guiti. 2016. Volition and obligation in four Caspian linguistic varieties. In Jila Ghomeshi, Carina Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in iranian linguistics : proceedings of the 5th international conference on iranian linguistics, held in bamberg on 24-26 august 2013*, 195–208. Association pour l'avancement des études Iraniennes.

Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: why objects don't make it. *Folia Linguistica* 33((1-2)). 225–251.

Sims-Williams, Nicholas. 1982. The Double System of Nominal Inflexion in Sogdian. *Transactions of the Philological Society* 80(1). 67–76.

Sims-Williams, Nicholas. 1998. *Proceedings of the Third European Conference of Iranian Studies: held in Cambridge, 11th to 15th, 1995: Part 1 Old and Middle Iranian Studies (Beiträge zur Iranistik 17)*. Wiesbaden: Reichert Verlag.

Sims-Williams, Nicholas. 2004. The Parthian Abstract Suffix -yft. In John H. W. Penney (ed.), *Indo-european perspectives. studies in honour of anna morpurgo davies*, 539—547. Oxford: Oxford University Press.

Sims-Williams, Nicholas. 2007. The Sogdian potentialis. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 377–386. Wiesbaden: Harrassowitz.

Sims-Williams, Nicholas. 2011. Differential Object Marking in Bactrian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 23–38. Wiesbaden: Reichert.

Sims-Williams, Nicholas (ed.). 1985. *Philologia Iranica*. Wiesbaden: Reichert Verlag.

Skalmowski, Wojciech & Alois van Tongerloo. 1993. *Medioiranica: proceedings of the International Colloquium organized by the Katholieke Universiteit Leuven from the 21st to the 23rd of May 1990 (Orientalia Lovaniensia analecta 48)*. Leuven: Peeters.

Skjærvø, Prods Oktor. 1988. BAŠKARDI.

Skjærvø, Prods Oktor. 1989a. Languages of Southeast Iran: Lārestānī, Kumzārī, Baškardī. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, 363–369. Wiesbaden: Reichert.

Skjærvø, Prods Oktor. 1989b. Modern East Iranian Languages. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, 370–383. Wiesbaden: Reichert.

Skjærvø, Prods Oktor. 2005. Avestica III: Notes on the Avestan Locative Singular. In Dieter Weber (ed.), *Languages of iran: past and present : Iranian studies in memoriam david neil mackenzie (iranica 8)*, 197–206. Wiesbaden: Harrassowitz.

Skjærvø, Prods Oktor. 2007. Avestan and Old Persian Morphology. In Alan S. Kaye (ed.), *Morphologies of asia and africa*, 853–940. Eisenbrauns.

Skjærvø, Prods Oktor. 2009a. Middle West Iranian. In Gernot Windfuhr (ed.), *The iranian languages*, chap. 4, 196–278. London and New York: Routledge.

Skjærvø, Prods Oktor. 2009b. Old Iranian. In Gernot Windfuhr (ed.), *The iranian languages*, chap. 3, 43–195. London; New York: Routledge.

Skjærvø, Prods Oktor. 2017a. The documentation of Iranian. In Jared S Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 32, 471–480. Berlin: De Gruyter Mouton.

Skjærvø, Prods Oktor. 2017b. The morphology of Iranian. In Jared S Klein, Brian D Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical indo-european linguistics band 1*, chap. 34, 503–549. Berlin: De Gruyter Mouton.

Soane, E B. 1913. *Grammar of the Kurmanji or Kurdish Language*. London: Luzac & Co.

Socin, Albert. 1895. Die Sprache der Kurden. In Wilhelm Geiger (ed.), *Grundriss der iranischen philologie*, chap. vii, 249–86. Strassburg: Trübner.

Spooner, Brian. 2006. Review of The Baloch and Their Neighbours: Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times by Carina Jahani and Agnes Korn. *Iranian Studies* 39(1). 145–147.

Stilo, Donald. 2005. Iranian as a Buffer Zone between Turkic and Semitic. In Éva Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic convergence and areal diffusion. case studies from iranian, semitic and turkic*, 35–63. London: Routledge Curzon.

Stilo, Donald. 2007. Isfahan xxi. PROVINCIAL DIALECTS. https://iranicaonline.org/articles/isfahan-xxi-provincial-dialects.

Stilo, Donald. 2008a. Case In Iranian: From Reduction and Loss to Innovation and Renewal. In Andrew Spencer & Andrej L Malchukov (eds.), *The oxford handbook of case*, 700–715. Oxford; New York: Oxford University Press.

Stilo, Donald. 2018a. Dikin Marāqei Tati of Alamut: an undocumented conservative Tati lanuguage. In Saloumeh Gholami (ed.), *Endangered iranian languages*, 41–70. Wiesbaden: Reichert.

Stilo, Donald. 2018b. The Caspian region and South Azerbaijan: Caspian and Tatic. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western asia: An areal perspective*, 659–829. Berlin, Boston: Walter de Gruyter.

Stilo, Donald & Paul M Noorlander. 2015. On the Convergence of Verbal Systems or Aramaic and its Neighbors. Part II: Past Paradigms Derived from Present Equivalents. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-aramaic in its linguistic context*, Piscataway, NJ: Gorgias Press.

Stilo, Donald L. 1971. *A Grammar of Vafsi-Tati: an Application of a Transformational Computer Model*: University of Michigan dissertation.

Stilo, Donald L. 2004. *Vafsi Folk Tales: Twenty Four Folk Tales in the Gurchani Dialect of Vafsi as Narrated by Ghazanfar Mahmudi and Mashdi Mahdi and Collected by Lawrence P. Elwell-Sutton*. Wiesbaden: Reichert Verlag.

Stilo, Donald L. 2008b. Two sets of mobile verbal person markers in the Northern Talyshi language. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 363–390. Newcastle upon Tyne: Cambridge Scholars Publishing.

Stilo, Donald L, Simin Karimi & Vida Samiian. 2008. *Aspects of Iranian linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Stump, Gregory T. 2001. *Inflectional morphology*. Cambridge University Press.

Suleymanov, Murad. 2020a. *A Grammar of Şirvan Tat (Beiträge zur Iranistik 46)*. Wiesbaden: Reichert.

Suleymanov, Murad. 2020b. Oblique marking and adpositional constructions in Tat: A mosaic of dialectal convergence and divergence. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in iranian linguistics*, chap. 13, 275–300. Amsterdam; Philadelphia: John Benjamins.

Sundermann, Werner. 1989a. Mittelpersisch. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 3.1.2, 138–164. Wiesbaden: Reichert Verlag.

Sundermann, Werner. 1989b. Parthisch. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 3.1.1, 114–137. Wiesbaden: Reichert Verlag.

Sundermann, Werner. 1989c. Wesrmitteliranische Sprachen. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 3.1.0, 106–113. Wiesbaden: Reichert Verlag.

Taleghani, Azita. 2008. Mood and modality in Persian. In Donald L Stilo, Simin Karimi & Vida Samiian (eds.), *Aspects of iranian linguistics*, 391–418. Newcastle upon Tyne: Cambridge Scholars Publishing.

Tedesco, Paul. 1921. Dialektologie der westiranischen Turfantexte. *Monde oriental* 15. 184–258.

Tegey, Habibullah & Barbara Robson. 1996. *A Reference Grammar of Pashto*. Washington, DC: Center for Applied Linguistics. http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED399825.

Testen, David. 1997a. Old Persian and Avestan Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 29, 569–600. Winona Lake, Indiana: Eisenbrauns.

Testen, David. 1997b. Ossetic Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 35, 707–732. Winona Lake, Indiana: Eisenbrauns.

Thackston, Wheeler M. 1993. *An introduction to Persian*. Bethesda, Md.: Iranbooks.

Thackston, Wheeler M. 2006a. Kurmanji Kurdish: A Reference Grammar with Selected Readings. https://sites.fas.harvard.edu/{~}iranian/Kurmanji.

Thackston, Wheeler M. 2006b. *Sorani Kurdish: A Reference Grammar with Selected Readings*. Cambridge, MA: Harvard University. https://sites.fas.harvard.edu/{~}Iranian/Sorani.

Thiesen, F. 2005. Eleven Etymologies. In Dieter Weber (ed.), *Languages of iran: past and present : Iranian studies in memoriam david neil mackenzie (iranica 8)*, 213–216. Wiesbaden: Harrassowitz.

Thomason, Sarah G & Terrence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

Thordarson, Fridrik. 1982. Preverbs in Ossetic. In *Monumentum georg morgenstierne ii*, 251–262. Leiden: Brill.

Thordarson, Fridrik. 1989. Ossetic. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 4.2.5. Wiesbaden: Reichert.

Thordarson, Fridrik. 1990. Old Ossetic Accentuation. In *Iranica varia: Papers in honor of professor ehsan yarshater*, 256–266. Leiden: Brill.

Todd, Terry Lynn. 2002. *A Grammar of Dimili Also Known as Zaza*. Stockholm: Iremet Förlag.

Toma, Shivan. 2018. Object and Subject Case Marking in Behdini. *European Scientific Journal* 14(5).

Trubeckoj, Nikolaj Sergeevič. 1937. *Gedanken über die slovakische Deklination (Sborník Matice slovenskej 15)*. Martin, Slovakia: Matica Slovenska.

Utas, Bo. 2013a. A Multiethnic Origin of New Persian. In Carina Jahani & Mehrdad Fallahzadeh (eds.), *From old to new persian: collected essays (beiträge zur iranistik 38)*, 181–192. Wiesbaden: Reichert.

Utas, Bo. 2013b. *From Old to New Persian: collected essays (Beiträge zur Iranistik 38)*. Wiesbaden: Reichert Verlag.

Utas, Bo. 2013c. Old Persian Miscellanea (Orientalia Suecana 14-15 (1965-1966)). In Carina Jahani & Mehrdad Fallahzadeh (eds.), *From old to new persian: collected essays (beiträge zur iranistik 38)*, 1–24. Wiesbaden: Reichert.

Utas, Bo. 2013d. The Grammatical Transition from Middle to New Persian. In Carina Jahani & Mehrdad Fallahzadeh (eds.), *From old to new persian: collected essays (beiträge zur iranistik 38)*, 251–259. Wiesbaden: Reichert.

Utas, Bo. 2013e. Traces of Evidentiality in Early New Persian (2000). In Carina Jahani & Mehrdad Fallahzadeh (eds.), *From old to new persian: collected essays (beiträge zur iranistik 38)*, 167–180. Wiesbaden: Reichert.

Vahman, Fereidun & G.S. Asatrian. 1990. Gleanings from Z\=az\=a vocabulary. In *Iranica varia: Papers in honor of professor ehsan yarshater*, 267–275. Leiden: Brill.

van Emde Boas, Evert, Albert Rijksbaron, Luuk Huitink & Mathieu de Bakker. 2019. *The Cambridge grammar of Classical Greek*. Cambridge: Cambridge University Press.

von Humboldt, Wilhelm. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts*. Berlin: Druckerei der königlichen Akademie der Wissenschaften.

Vydrin, Arseniy. 2011. Counterfactual Mood in Iranian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 71–88. Wiesbaden: Reichert.

Wahby, Taufiq. & Cecil J Edmonds. 1966. *A Kurdish-English dictionary*. Oxford: Clarendon Press.

Weber, Dieter. 1997. Pahlavi Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 30, 601–636. Winona Lake, Indiana: Eisenbrauns.

Weber, Dieter. 2005. *Languages of Iran: past and present : Iranian studies in memoriam David Neil MacKenzie (Iranica 8)*. Wiesbaden: Harrassowitz.

Weber, Dieter (University of Göttingen). 2007. Pahlavi Morphology. In A.S. Kaye (ed.), *Morphologies of asia and africa*, chap. 34, vol. 2, 941–973. Winona Lake, Indiana: Eisenbrauns.

Weir, E. M. H. 1987. Footprints of Yesterday's Syntax: Diachronic Development of Certain Verb Prefixes in an OSV Language (Nadëb). *Lingua* 68. 291–316.

Wendtland, Antje. 2011. The Emergence and Development of the Sogdian Perfect. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics (beiträge zur iranistik 34)*, 39–52. Wiesbaden: Reichert.

Werner, Brigitte. 2018. Forms and Meanings of the Ezafe in Zazaki. In Saloumeh Gholami (ed.), *Endangered iranian languages*, 71–91. Wiesbaden: Reichert Verlag.

Whitney, William Dwight. 1993. *Sanskrit Grammar*. Cambridge, Mass; London: Harvard University Press 2nd edn.

Windfuhr, Gernot. 1979. *Persian Grammar: History and State of its Study*. The Hage, Paris, New York: Mouton.

Windfuhr, Gernot. 1982a. The verbal category of inference in Persian. In *Monumentum georg morgenstierne ii*, 263–287. Leiden: Brill.

Windfuhr, Gernot. 1989a. New West Iranian. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, 251–262. Weisbaden: Reichert Verlag.

Windfuhr, Gernot. 1989b. Western Iranian Dialects. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, chap. 4.1.2, 294–295. Wiesbaden: Reichert Verlag.

Windfuhr, Gernot. 1991. CENTRAL DIALECTS. https://www.iranicaonline.org/articles/central-dialects.

Windfuhr, Gernot. 1995. DIALECTOLOGY. https://www.iranicaonline.org/articles/dialectology.

Windfuhr, Gernot. 1997. Persian Phonology. In Alan S Kaye (ed.), *Phonologies of asia and africa : (including the caucasus) (vol. 2)*, chap. 33, 675–690. Winona Lake, Indiana: Eisenbrauns.

Windfuhr, Gernot. 2009a. Dialectology and Topics. In Gernot Windfuhr (ed.), *The iranian languages*, chap. 2, 5–42. London; New York: Routledge.

Windfuhr, Gernot. 2009b. *The Iranian languages*. London; New York: Routledge.

Windfuhr, Gernot & John R. Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The iranian languages*, chap. 8, 416–544. London; New York: Routledge.

Windfuhr, Gernot L. 1982b. AFTARĪ. http://www.iranicaonline.org/articles/aftari-dialect-of-aftar.

Windfuhr, Gernot L. 2012. FĀRS viii. Dialects. https://iranicaonline.org/articles/fars-viii.

Woodard, Roger D. 2004. *The Cambridge Encyclopedia of the World's Ancient Languages*. Cambridge: Cambridge University Press.

Yarshater, E. 1960. The Tāti Dialect of Kajal. *Bulletin of the School of Oriental and African Studies, University of London* 23(2). 275–286. http://www.jstor.org.proxy.lib.ohio-state.edu/stable/609698.

Yarshater, Ehsan. 1964. The Dialects of Alvir and Vidar. In *Indo-iranica : mélanges présentés à georg morgenstierne a l'occasion de son soixante-dixième anniversaire*, 177. Wiesbaden: Otto Harrassowitz.

Yarshater, Ehsan. 1969. *A Grammar of the Southern Tati Dialects*. The Hague: Mouton.

Yarshater, Ehsan. 1982. *Encyclopaedia Iranica*. London: Routledge & Kegan Paul. https://iranicaonline.org/.

Yarshater, Ehsan. 1989. BĪDGOL. https://www.iranicaonline.org/articles/bidgol-and-bidgoli-dialect.

Yarshater, Ehsan. 1990. ČĀL. https://iranicaonline.org/articles/cal{#}prettyPhoto.

Yarshater, Ehsan. 2005. The Tati Dialect of Kalasur. In Dieter Weber (ed.), *Languages of iran: past and present : Iranian studies in memoriam david neil mackenzie (iranica 8)*, 269–284. Wiesbaden: Harrassowitz.

Yarshater, Ehsan. 2007. The Dialect of Karingan. In Maria Macuch, Mauro Maggi & Werner Sundermann (eds.), *Iranian languages and texts from iran and turan: Ronald e. emmerick memorial volume (iranica 13)*, 443–462. Wiesbaden: Harrassowitz.

Yoshida, Yutaka. 2009. Sogdian. In Gernot L Windfuhr (ed.), *The iranian languages*, 279–335. New York, NY: Routledge.

Yoshie, Satoko. 1996. *Sāri dialect*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Zhang, Guangda & Xinjian Rong. 1998. A Consice History of the Turfan Oasis and its Exploration. *Asia Major* 11(2). 13–36. https://www.jstor.org/stable/41645540.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23(1). 133–148. http://www.jstor.org.proxy.lib.ohio-state.edu/stable/4175870.

Zwicky, Arnold M. 1994. What is a clitic? In Joel A. Nevis, Brian D. Joseph, Dieter Wanner & Arnold M. Zwicky (eds.), *Clitics—a comprehensive bibliography 1892-1991*, xii–xx. Benjamins.